

# Enzyklopädie der Psychologie

# ENZYKLOPÄDIE DER PSYCHOLOGIE

In Verbindung mit der  
Deutschen Gesellschaft für Psychologie

herausgegeben von

Prof. Dr. Niels Birbaumer, Tübingen

Prof. Dr. Dieter Frey, München

Prof. Dr. Julius Kuhl, Osnabrück

Prof. Dr. Wolfgang Prinz, München

Prof. Dr. Franz E. Weinert, München

Themenbereich B

Methodologie und Methoden

Serie I

Forschungsmethoden der Psychologie

Band 1

Methodologische Grundlagen der Psychologie



Hogrefe · Verlag für Psychologie  
Göttingen · Bern · Toronto · Seattle

# Methodologische Grundlagen der Psychologie

Herausgegeben von

Prof. Dr. Theo Herrmann, Mannheim  
und Prof. Dr. Werner H. Tack, Saarbrücken



Hogrefe · Verlag für Psychologie  
Göttingen · Bern · Toronto · Seattle

© by Hogrefe-Verlag GmbH & Co. KG, Göttingen **1994**  
Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten.

ISBN 3-8017-0522-6

# Autorenverzeichnis

*Prof. Dr. Jürgen Bredenkamp*

Psychologisches Institut  
der Universität Bonn

Römerstraße 164  
D-53117 Bonn

*Prof. Dr. Dietrich Dörner*

Lehrstuhl Psychologie II  
der Universität Bamberg

Markusplatz 3  
D-96045 Bamberg

*Dr. Edgar Erdfelder*

Psychologisches Institut  
der Universität Bonn

Römerstraße 164  
D-53117 Bonn

*Prof. Dr. Volker Gadenne*

Fakultät für Pädagogik  
der Universität der Bundeswehr

Werner Heisenberg-Weg 39  
D-85577 Neubiberg

*Dipl.-Psych. Peter Gerjets*

Institut für Psychologie  
der Universität Göttingen

Goßlerstraße 14  
D-37073 Göttingen

*Prof. Dr. Theo Herrmann*

Lehrstuhl für Psychologie III  
der Universität Mannheim

Universität, Schloß  
D-68131 Mannheim

*Dr. Ronald Hitzler*

Institut für Soziologie  
der Universität München

Konradstraße 6  
D-80801 München

*Prof. Dr. Walter Hussy*

Psychologisches Institut  
der Universität zu Köln

Herbert-Lewin-Straße 2  
D-50931 Köln

*Prof. Dr. Rainer Mausfeld*

Institut für Psychologie  
der Universität Kiel

Olshausenstraße 40  
D-24098 Kiel

*Dr. Holger Möller*

Jahnstraße 12  
D-42929 Wermelskirchen

*PD Dr. Klaus Opwis*

Psychologisches Institut  
der Universität Freiburg

Niemensstraße 10  
D-79085 Freiburg

*PD Dr. Axel Ostmann*  
Fachrichtung Psychologie  
der Universität des Saarlandes  
Postfach 151150  
D-66041 Saarbrücken

*Dr. Jürgen Rehm*  
Institut suisse de prophylaxie  
de l'alcoolisme  
Case postale 870  
CH-1001 Lausanne

*Prof. Dr. Hans-Georg Soeffner*  
Fachbereich Erziehungs-, Sozial-  
und Geisteswissenschaften der  
Fernuniversität -  
Gesamthochschule  
Feithstraße 133  
D-58097 Hagen

*Prof. Dr. Hans Spada*  
Psychologisches Institut  
der Universität Freiburg  
Niemensstraße 10  
D-79085 Freiburg

*PD Dr. Rolf Steyer*  
Fachbereich 1 - Psychologie  
der Universität Trier  
Postfach 3825  
D-54286 Trier

*Prof. Dr. Fritz Strack*  
Fachbereich 1 - Psychologie  
der Universität Trier  
Postfach 3825  
D-54286 Trier

*Prof. Dr. Werner H. Tack*  
Fachrichtung Psychologie  
der Universität des Saarlandes  
postfach 151150  
D-66041 Saarbrücken

*Prof. Dr. Rainer Westermann*  
Institut für Psychologie  
der Universität Göttingen  
Goßlerstraße 14  
D-37073 Göttingen

*Dipl.-Psych. Joachim Wutke*  
Fachrichtung Psychologie  
der Universität des Saarlandes  
Postfach 151150  
D-66041 Saarbrücken

# Inhaltsverzeichnis

Teil A: Einführung . . . . .	1
1. Kapitel: Ziele und Aufgaben einer Allgemeinen Methodenlehre der Psychologie. Von Werner H. Tack . . . .	3
1. Versuch einer Abgrenzung . . . . .	3
1.1 Allgemeine vs. spezifische Methodenlehre . . . . .	4
1.2 Psychologiespezifität . . . . .	5
2. Psychologie als empirische Wissenschaft . . . . .	7
2.1 Empirisch, formal, normativ . . . . .	7
2.1.1 Reine und angewandte Disziplin . . . . .	8
2.1.2 Normativität . . . . .	11
2.2 Empirische und rationale Erkenntnis . . . . .	12
2.2.1 Klassisch rationales Vorgehen . . . . .	13
2.2.2 Metaprinzipien . . . . .	15
2.3 Empirische Methoden . . . . .	17
2.3.1 Datenrestriktion . . . . .	19
2.3.2 Manipulationsgrad . . . . .	20
2.3.3 Interne und externe Validität . . . . .	21
2.3.4 Laborbedingungen und Alltagssituationen . . . . .	23
3. Komplexität . . . . .	26
3.1 Komplexität von Systemen . . . . .	26
3.1.1 Komplexitätsvergleich . . . . .	27
3.1.2 Komplexitätsreduktion . . . . .	27
3.2 Wahrscheinlichkeiten, Fehler, Störvariablen . . . . .	30
3.2.1 Störvariablen . . . . .	30
3.2.2 Konfundierung und Effektunterdrückung . . . . .	31
3.2.3 Begründung probabilistischer Konzepte . . . . .	33
4. Themen Allgemeiner Methodenlehre . . . . .	34
4.1 Zugänge . . . . .	34
4.2 Theorien und Programme . . . . .	37
4.3 Forschungsschritte . . . . .	39

Teil B: Zugänge . . . . .	45
2. Kapitel: Erzeugung und Verwendung empirischer Daten. Von Edgar Erdfelder . . . . .	47
1. Beobachtungen und die Erzeugung von Daten . . . . .	50
1.1 <i>Wissenschaftliche Beobachtung</i> . . . . .	51
1.2 <i>Selbstbeobachtung mentaler Prozesse</i> . . . . .	55
1.3 <i>Heuristische Selbst- und Fremdbeobachtung</i> . . . . .	66
1.4 <i>Von wissenschaftlichen Beobachtungen zu Daten</i> . . . . .	68
1.4.1 Klassifikationsprobleme . . . . .	69
1.4.2 Indizierungsprobleme . . . . .	70
1.4.3 Interpretationsprobleme . . . . .	71
1.4.4 Quantifizierungsprobleme . . . . .	72
2. Verwendung von Daten im Entdeckungszusammenhang . . . . .	75
2.1 <i>Kritik exploratorischer Anwendungen konfirmatorischer Analyseverfahren</i> . . . . .	76
2.2 <i>Exploratorische und robuste Datenanalyse</i> . . . . .	78
3. Verwendung von Daten im Begründungszusammenhang . . . . .	81
3.1 <i>Der induktive Zugang</i> . . . . .	82
3.2 <i>Der deduktive Zugang</i> . . . . .	86
4. Konfirmatorische und exploratorische Verwendung empirischer Daten: Ein Antagonismus? . . . . .	89
5. Fazit . . . . .	90
3. Kapitel: Qualitatives Vorgehen - „Interpretation“. Von Hans-Georg Soeffner und Ronald Hitzler . . . . .	98
1. Verstehen in Alltag und Wissenschaft . . . . .	98
1.1 <i>Phänomenologie des Verstehens</i> . . . . .	98
1.2 <i>Geschichte des hermeneutischen Verstehens</i> . . . . .	100
1.3 <i>Besonderheiten sozialwissenschaftlichen Verstehens</i> . . . . .	104
2. Hermeneutik als selbstreflexives Unternehmen: Die Relativität der Deutung und der Verlust des Einmaligen . . . . .	106
3. Probleme methodisch kontrollierten Verstehens . . . . .	111
3.1 <i>Der konkrete Fall und der ‚ideale‘ Typus</i> . . . . .	111
3.2 <i>Die Datenkonstitution</i> . . . . .	113
3.2.1 Divergenzen und Konvergenzen standardisierter und nicht standardisierter Verfahren . . . . .	114



3.3	<i>Der Text und das Vertextete</i> . . . . .	115
3.4	<i>Die Sequenzanalyse</i> . . . . .	117
3.4.1	Die Methodik der sequentiellen Interpretation . . . . .	118
3.1	<i>Prinzipien der sinnschließenden Rekonstruktion</i> . . . . .	121
3.5.1	Rekonstruktion der egologisch-monothetischen Perspektive eines Sprechers . . . . .	122
3.5.2	Übernahme der polythetisch-interaktionsbezogenen Perspektive der Alltagshermeneutik . . . . .	123
3.5.3	Konstitution des einheitlichen Interaktionssinnes . . . . .	124
4.	Hermeneutik als selbstreflexives Unternehmen: Die Unumgänglichkeit der Interpretation in den Sozialwissenschaften	126
5.	Resümee . . . . .	128
4.	Kapitel: Methodologische Grundlagen und Probleme der Psychophysik. Von Rainer Mausfeld . . . . .	137
1.	Der Beginn der Psychophysik . . . . .	139
1.1	<i>Die Entstehung der psychophysikalischen Zugangsweise</i> . . . . .	139
1.2	<i>Fechners Psychophysik</i> . . . . .	140
2.	Vom Status psychophysikalischer Skalen . . . . .	144
2.1	<i>Die Psychophysik als perzeptuelle Physik</i> . . . . .	145
2.2	<i>Was ‚messen‘ eindimensionale psychophysikalische Skalen?</i> . . . . .	146
3.	Die Meßinstrumentkonzeption der Wahrnehmung . . . . .	148
4.	Die Entwicklung von Theorien und Modellen der perzeptuell-kognitiven Informationsverarbeitung . . . . .	150
5.	Von der Meßinstrumentkonzeption zu einer funktionalistischen Perspektive . . . . .	155
5.1	<i>Das Problem der Reizbestimmung</i> . . . . .	155
5.2	<i>Psychophysik aus physikalistischer vs. funktionalistischer Perspektive</i> . . . . .	156
5.3	<i>Internalisierte Regularitäten in der Wahrnehmung</i> . . . . .	157
5.4	<i>Was ist ein Reiz?</i> . . . . .	162
5.5	<i>Die Psychophysik in ihrem Verhältnis zur Neurophysiologie und subjektiven Sinnesphysiologie</i> . . . . .	165
6.	Theorien in der Psychophysik . . . . .	167
6.1	<i>Metaphern, Heuristiken und Metaprinzipien der Theoriebildung</i>	167
6.2	<i>Theorietypen der Psychophysik</i> . . . . .	178
7.	Experiment, Beobachtung und Daten in der Psychophysik . . . . .	183
8.	Implizite Annahmen der experimentellen Psychophysik . . . . .	188

5. Kapitel: Modellierung mit Hilfe wissensbasierter Systeme. Von Klaus Opwis und Hans Spada . . . . .	199
1. Prolog: Von Brücken aus Bausteinen, Ökosystemen in Form von Differentialgleichungen und Problemlöseprozessen auf Rechnern	199
2. Modellierung als Rekonstruktion . . . . .	202
3. Grundlagen wissensbasierter Systeme . . . . .	205
3.1 <i>Konzeptuelle Modelle, formale Kalküle und wissensbasierte Systeme am Beispiel semantischer Netzwerkmodelle</i> . . . . .	207
3.1.1 Konzeptualisierung: Abstraktionshierarchien als psychologische Modellvorstellung . . . . .	207
3.1.2 Formalisierung: Prädikatenlogische Darstellung von Abstraktionshierarchien . . . . .	209
3.1.3 Implementation: Abstraktionshierarchien als wissensbasierte Systeme . . . . .	211
3.2 <i>Algorithmen, Berechenbarkeit und symbolverarbeitende Systeme</i>	213
3.3 <i>Mehrebenenbetrachtung von Theorien der Informationsverarbeitung</i> . . . . .	217
4. Bewertung wissensbasierter Systeme als psychologische Modelle	221
4.1 <i>Empirische Adäquatheit</i> . . . . .	221
4.1.1 Eine theoretische Rahmenvorstellung . . . . .	221
4.1.2 Probleme . . . . .	224
4.1.3 Empirie als Vorbild . . . . .	226
4.1.4 Modellperformanz als Datenquelle . . . . .	227
4.1.5 Vergleich von Modell- und Personendaten . . . . .	229
4.1.6 Modellvergleichende Argumentation . . . . .	231
4.2 <i>Weitere Bewertungskriterien</i> . . . . .	232
5. Wissenschaftstheoretische Anmerkungen . . . . .	234
5.1 <i>Der Turing-Test</i> . . . . .	235
5.2 <i>Der Funktionalismus</i> . . . . .	236
5.3 <i>Das chinesische Zimmer</i> . . . . .	238
6. Epilog . . . . .	240
 Teil C: Theorien und Forschungsprogramme . . .	 249
6. Kapitel: Forschungsprogramme. Von Theo Herrmann	251
1. Was sind Forschungsprogramme? . . . . .	252
1.1 <i>Begriffsexplikation</i> . . . . .	252

1.2 Erläuterungen . . . . .	253
1.2.1 Raum-zeitliche Spezifikation . . . . .	253
1.2.2 Probleme und Annahmenkerne . . . . .	253
1.2.3 Zur Methodenwahl . . . . .	256
1.2.4 Programmspezifische Indisponibilität . . . . .	259
1.2.5 Soziales System und Regulation . . . . .	259
2. Probleme P und strukturalistische Rekonstruktionen . . . . .	261
2.1 Fragestellung . . . . .	261
2.2 Einige Grundannahmen des metatheoretischen Strukturalismus . . . . .	261
2.3 Annahmenkerne und Theorien im Lichte des Strukturalismus . . . . .	263
3. Programmnetze und Austauschbeziehungen . . . . .	265
4. Zur Bearbeitung von Problemen in Forschungsprogrammen . . . . .	266
5. Programmtypen . . . . .	269
5.1 Übersicht . . . . .	269
5.2 Grundlagenwissenschaftliche und technologische Programme . . . . .	270
5.2.1 Grundlagenwissenschaftliche Programme . . . . .	270
5.2.2 Technologische Programme . . . . .	272
5.3 Technologische Techniken-Programme und „Wissensprogramme“ . . . . .	273
5.4 Grundlagenwissenschaftliche Sachproblem- und Theorieprogramme . . . . .	275
6. Einige externe Einflüsse auf Forschungsprogramme . . . . .	279
6.1 Internale und extemale Einflüsse . . . . .	279
6.2 Wissenschaftsparadigmen . . . . .	282
6.3 Moden und Metaphern . . . . .	284
6.4 Weitere externe Einflußquellen . . . . .	286
6.5 Zusammenfassung . . . . .	287
7. Kapitel: Theorien. Von Volker Gadenne . . . . .	295
1. Aufbau und Struktur von Theorien . . . . .	296
1.1 Ein Beispiel: Die Theorie der Leistungsmotivation . . . . .	297
1.2 Theorien als deduktive Systeme von Aussagen . . . . .	302
1.3 Theorie und Gesetz . . . . .	304
1.4 Zur Präzisierung der formalen Struktur einer Theorie: Formalisierung und Mathematisierung . . . . .	305
2. Zur Beziehung zwischen Theorie und Empirie . . . . .	308
2.1 Operationale Definition versus Operationalisierung . . . . .	309
2.2 Zur Abgrenzung zwischen theoretischen und Beobachtungsbegriffen . . . . .	310
2.3 Zum Status von Zuordnungsannahmen . . . . .	311

3. Zur Beziehung zwischen Theorie und Realität . . . . .	313
3.1 <i>Realismus und Instrumentalismus</i> . . . . .	313
3.2 <i>Der Gegenstand psychologischer Theorien</i> . . . . .	318
3.3 <i>Idealisierung und Unvollständigkeit</i> . . . . .	321
4. Theorie und Erklärung . . . . .	324
4.1 <i>Deduktiv-nomologische Erklärung</i> . . . . .	324
4.2 <i>Statistische und unvollständige Erklärung</i> . . . . .	326
4.3 <i>Weitere Erklärungsarten</i> . . . . .	328
4.4 <i>Theorie, Erklärung und Kausalität</i> . . . . .	329
5. Eine Alternative zur Aussagenkonzeption: Der Non-Statement-View . . . . .	332
8. Kapitel: Heuristik der Theorienbildung. Von Dietrich Dörner . . . . .	343
1. Theorienbilden und Theorienprüfen als wissenschaftliche Tätigkeiten . . . . .	343
2. Die Welt der „schwarzen Kästen“ . . . . .	345
3. Wie kommt man in das Innere von schwarzen Kästen? . . . . .	348
3.1 <i>Die Form der Theorie</i> . . . . .	348
3.2 <i>Die Introspektion und das „Verstehen“</i> . . . . .	352
3.3 <i>Der „Positivismus“</i> . . . . .	354
3.3.1 <i>Die „analytische Prozedur“ und Wechselwirkungen</i> . . . . .	354
3.3.2 <i>Positivismus und Statistik</i> . . . . .	360
3.4 <i>Die Heuristik der Zweckmäßigkeit</i> . . . . .	362
3.5 <i>Analogien</i> . . . . .	363
3.6 <i>Modelle und Simulation</i> . . . . .	367
3.6.1 <i>Modelle</i> . . . . .	367
3.6.2 <i>Der Gebrauch von Modellen bei der Konstruktion von             Theorien</i> . . . . .	378
4. <i>Schlußbemerkungen</i> . . . . .	385
9. Kapitel: Theoriebewertung. Von Volker Gadenne . . . . .	389
1. Logische und semantische Bewertungskriterien . . . . .	391
1.1 <i>Logische Konsistenz</i> . . . . .	391
1.2 <i>Semantische Einheitlichkeit</i> . . . . .	394
1.3 <i>Informationsgehalt, Prüfbarkeit, Tiefe</i> . . . . .	396
1.4 <i>Einfachheit</i> . . . . .	403
2. Empirische Bewertungskriterien . . . . .	406

2.1 Die Möglichkeit empirischer Prüfung und Kritik . . . . .	407
2.2 Strategien der empirischen Prüfung . . . . .	413
2.3 Empirische Bestätigung und Wahrheit . . . . .	418

## 10. Kapitel: Induktion.

Von Rainer Westermann und Peter Gerjets . . . . .	428
1. Charakteristika und Probleme induktiver Argumente . . . . .	430
1.1 Enumeration und andere Arten von Induktion . . . . .	430
1.2 Starkes und schwaches Induktionsproblem . . . . .	432
2. Probabilistische Explikationssysteme der Induktion . . . . .	433
2.1 Carnaps induktive Logik . . . . .	434
2.2 Einzelfallbestätigung . . . . .	437
3. Deduktivistische Umgehung der Induktion . . . . .	439
3.1 Poppers deduktivistische Methodologie . . . . .	440
3.2 Induktive Anteile in der deduktivistischen Methodologie . . . . .	441
3.3 Deduktive und induktive Bestätigung . . . . .	443
4. Pragmatische Rechtfertigungen der Induktion . . . . .	445
4.1 Reichenbachs pragmatische Argumentation . . . . .	446
4.2 Anpassung von Prinzipien und Praxis . . . . .	447
4.3 Einbeziehung von Hintergrundwissen . . . . .	448
5. Psychologische Analysen der Induktion . . . . .	450
5.1 Eine allgemeine Induktionskonzeption . . . . .	451
5.2 Induktionspsychologie und -philosophie . . . . .	454
6. Strukturalistische Analyse der Induktion . . . . .	457
6.1 Induktion im Wissenschaftsprozess . . . . .	457
6.2 Berechtigung empirischer Vermutungen . . . . .	459
7. Zusammenfassende Schlußbemerkungen . . . . .	464

## Teil D: Forschungsschritte . . . . . 473

## 11. Kapitel: Hypothesen.

Von Walter Hussy und Holger Möller . . . . .	475
1. Zur Definition, Klassifikation, Generierung und Funktion von Hypothesen . . . . .	475
1.1 Wissenschaftliche und nicht-wissenschaftliche Hypothesen . . . . .	475
1.2 Allgemeine Begriffsbestimmungen . . . . .	476

1.3 Arten von Hypothesen . . . . .	476
1.4 Generierung von Hypothesen . . . . .	479
1.5 Die Hypothese im Forschungsprozeß . . . . .	481
2. Zur Logik der Überprüfung wissenschaftlicher Hypothesen . . .	483
2.1 Widerspruchsfreiheit und Operationalisierbarkeit von Hypothesen	483
2.2 Die Überprüfung wissenschaftlicher Hypothesen . . . . .	485
2.2.1 Der Weg von den Inhalten zur Statistik . . . . .	486
2.2.2 Die Beziehung von inhaltlichen Hypothesen und statistischen Vorhersagen . . . . .	490
2.2.3 Die Fehlerkontrolle von $\alpha$ und $\beta$ . . . . .	496
2.3 Der Weg von der Statistik zurück zu den Inhalten . . . . .	498
2.3.1 Die Ebene des Signifikanztests . . . . .	498
2.3.2 Die Ebene der statistischen Vorhersage . . . . .	498
2.3.3 Die Ebene der inhaltlichen Hypothesen . . . . .	499
3. Zur Validität der Überprüfung wissenschaftlicher Hypothesen . .	501
3.1 Die Zuordnung der empirisch-inhaltlichen zu den theoretisch-inhaltlichen Hypothesen . . . . .	502
3.2 Die Zuordnung der statistischen Vorhersagen zu den empirisch-inhaltlichen Hypothesen . . . . .	503
3.3 Die Zuordnung der Signifikanztests zu den statistischen Vorhersagen . . . . .	504
3.4 Die Validität des statistischen Schlusses . . . . .	505
4. Abschließende Bemerkungen . . . . .	505
12. Kapitel: Kontrolltechniken. Von Jürgen Rehm und Fritz Strack . . . . .	508
1. Beobachtung und Schlußfolgerung . . . . .	509
2. Das psychologische Experiment als Idealtypus kontrollierter Beobachtung . . . . .	510
2.1 Experiment und Kontrolle durch Randomisierung . . . . .	510
2.2 Zur Rolle von systematischen Störfaktoren . . . . .	515
2.3 Kennzeichen eines guten Experiments - experimenteller Realismus	516
2.4 Probleme bei der Realisierung der abhängigen Variablen: Verhaltensmessung und Kontexteinflüsse . . . . .	520
2.1 Zur Frage von intraindividuellen vs. interindividuellen Versuchsplänen . . . . .	525
3. Alternativen zum randomisierten Experiment . . . . .	527
3.1 Allgemeine Grundsätze zur Erstellung von nicht-experimentellen Forschungsplänen . . . . .	527

3.2 Kohortenstudie und Fall-Kontroll-Studie . . . . .	531
3.3 Einzelfallstudie . . . . .	534
4. Forschungsprogramme und ihr Einfluß auf die Planung psychologischer Untersuchungen . . . . .	538
5. Statistische Kontrolltechniken . . . . .	539
6. Anmerkungen zur Ethik psychologischer Untersuchungen . . . .	541
6.1 Vorbemerkung . . . . .	541
6.2 Mögliche Schädigungen . . . . .	542
6.3 Durchführungsregeln für experimentelle Studien . . . . .	547

13. Kapitel: Von Zahlzeichen zu Skalen.

Von Rainer Mausfeld . . . . .	556
-------------------------------	-----

1. Der Zahlgebrauch in der Psychologie . . . . .	556
2. Qualität und Quantität: Ist Psychisches meßbar? . . . . .	560
3. Arten der Skalenkonstruktion . . . . .	566
3.1 Globales, deterministisches Vorgehen I: Die Stevensschen Verfahren . . . . .	568
3.2 Globales, deterministisches Vorgehen II: Die additiv-verbundene Messung . . . . .	571
3.3 Lokales, deterministisches Vorgehen . . . . .	573
3.4 Lokales, probabilistisches Vorgehen . . . . .	575
3.5 Globale, probabilistische Modelle . . . . .	578
3.6 Vergleichende Zusammenfassung . . . . .	580
4. Anmerkungen zum Fehlerbegriff . . . . .	582
5. Zur Rolle von Invarianzkonzepten bei der Konstruktion psychologischer Skalen . . . . .	588
5.1 Skalenniveau, zulässige Transformationen und meaningfulness-Konzepte . . . . .	589
5.2 Die ‚möglichen Formen‘ psychophysikalischer Funktionen und Gesetze . . . . .	593
5.3 Wie sinnvoll sind meaningfulness-Betrachtungen? . . . . .	594

14. Kapitel: Hypothesenprüfung.

Von Edgar Erdfelder und Jürgen Bredenkamp . . . . .	604
---	-----

1. Grundzüge einer deduktivistischen Theorie hypothesentestender Untersuchungen . . . . .	606
1.1 Warum werden psychologische Hypothesen statistisch geprüft? . .	606

1.2 Die Notwendigkeit der simultanen Kontrolle von $\alpha$ und $\beta$ . . . . .	609
1.3 Die Bedeutung von Randomisierungstests . . . . .	610
1.4 Das Theorie-Empirie-Überbrückungsproblem . . . . .	611
1.5 Die Bedeutung der Situationsvalidität . . . . .	612
2. Eine formale Analyse der statistischen Prüfung psychologischer Hypothesen . . . . .	613
2.1 Was ist eine strenge und faire statistische Prüfung psychologischer Hypothesen? . . . . .	613
2.2 Determinanten von Strenge und Fairneß . . . . .	616
2.3 Methodologische Folgerungen aus der Forderung nach Strenge und Fairneß . . . . .	618
2.4 Randomisierte oder nichtrandomisierte Untersuchungen? . . . . .	622
2.5 Sind statistische Aggregathypothesen zulässig? . . . . .	624
2.6 Skalenniveau und Statistik . . . . .	625
3. Empfehlungen zur statistischen Entscheidungsstrategie . . . . .	627
3.1 Zur Festlegung von $\alpha$ und $\beta$ . . . . .	627
3.2 Zur Festlegung der Effektstärke . . . . .	628
3.3 Mehrfache Signifikanztests . . . . .	631
3.4 Nonparametrische Verfahren . . . . .	632
4. Stochastische Modelle mit latenten Variablen als Bestandteile einer deduktivistischen Methodologie . . . . .	634
4.1 Probleme der Operationalisierung . . . . .	634
4.2 Die stochastische Formulierung psychologischer Hypothesen mit theoretischen Größen . . . . .	635
4.3 Eine Auswahl wichtiger stochastischer Rahmenmodelle . . . . .	638
4.4 Probleme der Modellgeltungsprüfung . . . . .	640
4.5 Einwände gegen stochastische Modellbildung zwecks Überprüfung psychologischer Hypothesen . . . . .	643
15. Kapitel: Stochastische Modelle. Von Rolf Steyer . . . . .	649
1. Bestandteile und Gegenstand stochastischer Modelle . . . . .	653
1.1 Zufallsexperiment . . . . .	653
1.2 Wahrscheinlichkeitsraum . . . . .	655
1.3 Zufallsvariablen . . . . .	657
1.4 Zusammenfassende Bemerkungen . . . . .	658
2. Stochastische Abhängigkeitsbegriffe . . . . .	659
2.1 Stochastische Abhängigkeit . . . . .	660
2.1.1 Stochastische Abhängigkeit zwischen Ereignissen . . . . .	660
2.1.2 Stochastische Abhängigkeit zwischen Zufallsvariablen . . . . .	661
2.2 Regressive und korrelative Abhängigkeiten . . . . .	662



2.2.1 Einfache regressive Abhängigkeit . . . . .	664
2.2.2 Einfache lineare quasi-regressive Abhängigkeit . . . . .	664
2.2.3 Korrelative Abhängigkeit . . . . .	665
2.2.4 Partielle lineare regressive Abhängigkeit . . . . .	666
2.2.5 Bedingte regressive Abhängigkeit . . . . .	666
2.2.6 Zusammenfassende Bemerkungen . . . . .	668
2.3 <i>Kausale regressive Abhängigkeiten</i> . . . . .	669
2.3.1 Vorbereitende Definitionen . . . . .	670
2.3.2 Schwache kausale regressive Abhängigkeit . . . . .	674
2.3.3 Starke kausale regressive Abhängigkeit . . . . .	674
2.3.4 Zusammenfassende Bemerkungen . . . . .	675
3. Stochastische Meßmodelle . . . . .	676
3.1 <i>Grundbegriffe der Klassischen Theorie psychometrischer Tests</i> . . . . .	677
3.2 <i>Das Modell essentiell <math>\tau</math>-äquivalenter Variablen</i> . . . . .	679
3.2.1 Existenz . . . . .	679
3.2.2 Eindeutigkeit und Bedeutsamkeit . . . . .	680
3.2.3 Testbarkeit . . . . .	681
3.2.4 Identifizierbarkeit . . . . .	682
3.3 <i>Zusammenfassende Bemerkungen</i> . . . . .	683
4. Zusammenfassung und Ausblick . . . . .	684
4.1 <i>Stochastische Meßmodelle</i> . . . . .	685
4.2 <i>Stochastische Abhängigkeitsbegriffe</i> . . . . .	686
4.3 <i>Ausblick</i> . . . . .	687
16. Kapitel: Statistische Entscheidung.	
Von Axel Ostmann und Joachim Wutke . . . . .	694
1. Einleitung . . . . .	694
2. Statistische Grundbegriffe . . . . .	697
3. Der Signifikanztest nach Fisher . . . . .	709
4. Die Testtheorie von Neyman und Pearson . . . . .	716
5. Eine Variante des Signifikanztests in der Psychologie . . . . .	722
6. Hypothesenprüfung nach Bayes . . . . .	726
7. Schlußbemerkungen . . . . .	730
Autoren-Register . . . . .	739
Sach-Register . . . . .	751

Teil A  
Einführung

# Kapitel 1

## Ziele und Aufgaben einer Allgemeinen Methodenlehre der Psychologie

von Werner H. Tack

Dieser Band der „Enzyklopädie der Psychologie“ beschäftigt sich mit methodologischen Grundlagen psychologischer Forschung. Das einführende Kapitel geht zum einen der Frage nach, was unter „methodologischen Grundlagen“ überhaupt zu verstehen ist, und versucht zum anderen aufzuzeigen, wie Themen und Gegenstände der nachfolgenden Kapitel zusammenhängen und Struktur und Inhalte einer „Allgemeinen Methodenlehre der Psychologie“ exemplifizieren, die sich mit methodologischen Grundlagen befaßt.

**Methodenlehre** ist wissenschaftliche Beschäftigung mit Methoden und nicht lediglich deren sachgerechte Anwendung. **Allgemeine Methodenlehre der Psychologie** bedarf einer doppelten Abgrenzung sowohl gegenüber speziellen Methodenlehren als auch gegenüber einer zwar allgemeinen Methodenlehre, die jedoch keine solche der Psychologie ist. Beide Abgrenzungen werden exemplifiziert, wobei Psychologie-Spezifität am Charakter der Psychologie als einer empirischen Wissenschaft und an der Komplexität der von ihr untersuchten Gegenstände und Phänomene aufgezeigt wird. Empirische Wissenschaft bedarf empirischer Methoden, kommt aber auch nicht ohne analytische (formale und normative) Methodik aus. Rationales Erkenntnis wird dabei vornehmlich als eine Heuristik gesehen, deren Resultate empirisch zu kontrollieren sind. Wahrscheinlichkeiten, Fehler und Störvariablen sind Konzepte, auf denen Methoden zum Umgang mit hoch komplexen Phänomenbereichen basieren.

Neben die Abgrenzung einer Allgemeinen Methodenlehre der Psychologie nach außen tritt eine Skizze möglicher interner Strukturierungen. Kapitel 1 stellt dazu drei Möglichkeiten vor. Unterschiedliche Zugänge zum Gegenstands- und Phänomenbereich der Psychologie eröffnen Fragen nach zugangsspezifischen methodologischen Vorannahmen und deren Vergleichbarkeit. Daneben steht die methodenorientierte Beschäftigung mit zugangsunspezifischen Konzepten wie etwa dem der „Theorie“. Die Orientierung an einzelnen Schritten im Forschungsprozeß bietet eine dritte Möglichkeit. Diese drei Strukturierungsansätze entsprechen der Gliederung der folgenden Kapitel dieses Bandes.

## 1. Kapitel

# Ziele und Aufgaben einer Allgemeinen Methodenlehre der Psychologie

**Werner H. Tack**

Was verbirgt sich hinter der Vorstellung einer „Allgemeinen Methodenlehre“, die unbeschadet ihrer Allgemeinheit doch eine solche der Psychologie ist, und deren Existenz - oder zumindest deren Möglichkeit - der Titel dieses Kapitels voraussetzt? Sicher kann diese Frage hier nicht voll zufriedenstellend beantwortet werden; erste Versuche einer Explikation, Prolegomena also, sind aber möglich und notwendig. Dieser Beitrag eröffnet in einer Enzyklopädie der Psychologie innerhalb einer Serie über Forschungsmethoden einen Band, der sich mit **methodologischen Grundlagen** beschäftigt. Zur gleichen Serie gehören Bände über Klassen von Methoden, die spezifischen Teilaufgaben im Forschungsprozeß dienen (Datenerhebung, Messen und Testen, Strukturierung und Reduzierung von Daten, Hypothesenprüfung). Bezeichnen wir die dort vorherrschenden Beschäftigungen mit mehr oder weniger eng abgegrenzten Gruppen von Methoden und Techniken als „spezielle“ Methodenlehren, so ist die im Titel dieses Bandes angesprochene Behandlung methodologischer Grundlagen das „allgemeine“ Pendant. Die Frage nach Inhalt und Methodik einer Allgemeinen Methodenlehre der Psychologie ist somit gleichbedeutend mit der Frage, was unter „methodologischen Grundlagen“ zu verstehen und mithin in den weiteren Kapiteln dieses Bandes zu erwarten ist.

### **1. Versuch einer Abgrenzung**

Eine erste Annäherung an die Konzeption einer Allgemeinen Methodenlehre der Psychologie liefern Überlegungen zu Fragen oder Themen, die ihr zuzuordnen sind. Damit sei keineswegs unterstellt, Wissenschaften und deren Teildisziplinen bedürften zu ihrer Abgrenzung eines jeweils spezifischen Gegenstandes. Unbeschadet der Frage nach der **Notwendigkeit einer Spezifikation** durch Gegenstandsangabe gibt es aber die **Möglichkeit der Explikation** der

Konzeption einer Wissenschaft oder eines Teilgebietes durch Verweis auf Gegenstände, mit denen sie sich beschäftigt.

Wir gehen also nicht von der Vorstellung aus, „Allgemeine Methodenlehre der Psychologie“ sei abschließend **definierbar**, sondern lediglich davon, sie sei durch Verweis auf Gegenstände, mit denen sie sich befaßt, **explizierbar**. Dabei können die Gegenstände mehr oder weniger typisch sein. Zur Annäherung an den weder hinreichend klar bekannten noch beschreibbaren Prototypen (einen fiktiven „zentralen“ Gegenstand Allgemeiner Methodenlehre bzw. einer „zentralen“ methodologischen Grundlage) wird durch Rückgriff auf möglichst typische Gegenstände vielleicht der Eindruck einer sehr engen Konzeption erweckt. Diese thematische Enge dient lediglich der Verdeutlichung; wie breit das Gesamtkonzept in der Realität ist, das zeigt die Vielfalt der Themen, Fragestellungen und Ansätze in den nachfolgenden Kapiteln dieses Bandes.

Sind es bei Herrmann im Kapitel 6 dieses Bandes „Forschungsprogramme“, mit denen sich eine Wissenschaft - und mithin auch die Psychologie - beschäftigt, so spielen in diesem Kapitel „Gegenstände“ (der Allgemeinen Methodenlehre) eine analoge Rolle. Werden Programme der Psychologie dort durch „Probleme“ angebbbar, so sollen hier Gegenstände einer Allgemeinen Methodenlehre durch „Fragestellungen“ - nämlich durch „methodologische Grundfragen“ beschrieben werden.

„Methodologie“ - das sei vorab festgehalten - ist **wissenschaftliche Beschäftigung mit Methoden**, die hier also nicht lediglich in der Forschung eingesetzt werden, sondern selbst Objekt der Forschung sind. Soll sich wissenschaftliche Tätigkeit nicht mit der Angabe isolierter Befunde bescheiden, sondern darüber hinaus Ergebnisse systematisieren (siehe hierzu Gadene im einführenden Abschnitt zu Kap. 7 dieses Bandes), so muß auch „Methodologie“ nicht lediglich Methoden darstellen, sondern diese analysieren und über sie theoretisieren.

## 1.1 Allgemeine vs. spezifische Methodenlehre

Zur Spezifikation einer Allgemeinen Methodenlehre der Psychologie ist eine doppelte Abgrenzung sinnvoll. Auf der einen Seite ist zwischen „Allgemeiner“ und „Spezifischer“ Methodenlehre zu differenzieren, auf der anderen Seite zwischen Allgemeiner Methodenlehre „der Psychologie“ und der einer beliebigen anderen Wissenschaft oder sogar der Wissenschaft allgemein. Die erste dieser beiden Unterscheidungen ist relativ einfach, eine Methodenlehre ist **spezifisch**, wenn sie sich mit jeweils einer bestimmten Methode (wie etwa dem Experiment, dem statistischen Schließen, der Messung, der Psychodiagnostik, der Faktorenanalyse, dem hierarchischen Clustering oder ähnlichem) beschäftigt.

Diese Abgrenzung hat zwei wichtige Konsequenzen. Da jede Entwicklung und Darstellung einer Methode spezifisch für eben diese Methode ist, gehören Methodenentwicklung und -darstellung **nicht** zum Arbeitsbereich einer Allgemeinen Methodenlehre. Was bleibt ist die Methoden-Systematisierung und -Analyse, also die **Methodologie**. Zum zweiten folgt, daß im Zentrum einer Allgemeinen Methodenlehre Fragen stehen, die sinnvoll für **verschiedene** Methodenarten gestellt werden können. Derartige Fragen gibt es in größerer Zahl, als ein Blick in die meisten einschlägigen Methoden-Lehrbücher der Psychologie vermuten läßt.

Es wird oft gesagt, es bestehe ein enger Zusammenhang zwischen Forschungsfragen und -methoden; wie aber könnte dieser genauer untersucht werden als durch Vergleich eines breiten Spektrums von Fragestellungen mit einem angemessenen Vorrat unterschiedlicher Methoden? Methoden unterscheiden sich bezüglich der (oft impliziten) Vorannahmen, die man unterstellt, um ihren Einsatz rechtfertigen zu können; wie anders als durch Methodenvergleich soll man die Unterschiedlichkeit solcher Vorannahmen analysieren? Die Möglichkeit, Forschungsergebnisse unterschiedlich zu formulieren, ist nicht unabhängig vom methodenbedingten Ausmaß der Ergebnisspezifikation; auch für Arbeiten zur Bedeutsamkeit von Befunden ist ein Ansatz erforderlich, der über Methodenklassen hinausgreift. Dies sind nur beispielhafte Hinweise, die zeigen, daß der Gegenstandsbereich einer Allgemeinen Methodenlehre alles andere als leer ist.

## 1.2 Psychologiespezifität

Was macht eine Allgemeine Methodenlehre zu einer solchen **der Psychologie**? Sicherlich eine vorrangige Berücksichtigung jener Fragen und Fragen-Varianten, die für psychologische Forschung besonders wichtig sind. Was das im einzelnen ist, hängt zum großen Teil davon ab, was man für Psychologie hält. Tseëlon (1991) vergleicht ein Methodenarsenal mit einer Sprache, in der man eine vorgegebene Realität nicht lediglich **beschreiben** kann, sondern in der man aufgrund der gegebenen Möglichkeiten Realität stets auch **konstruiert**. Methoden sind keine neutralen Instrumente, durch die hindurch eine objektive Realität mehr oder weniger so erscheint, wie sie ist. Ist die Verbindung zwischen Methoden und Ergebnissen derart eng, dann bestimmt psychologische Forschungsmethodik weitgehend, was als Psychologie geschieht, erfassbar und darstellbar ist. Es ist damit eine methodologische Aufgabe, sich um methodenbedingte Grenzen dessen zu kümmern, was im Fach als Forschungsergebnis möglich ist.

Wenn im folgenden von „Spezifika der Psychologie“ die Rede ist, dann sei damit keineswegs unterstellt, es gebe Eigenschaftsangaben oder - anders for-

muliert - Aussagen, die ausschließlich auf Psychologie zutreffen. Gemeint ist lediglich, daß bestimmte Merkmale des Faches von besonderer Bedeutung für die Entwicklung seines Methodenkanons sind und (vielleicht) in ihrer Kombination ein „Spezifikum“ der Disziplin im engeren Sinne darstellen.

Zum Selbstverständnis der Psychologie gehört zunächst, daß sie eine **Wissenschaft** ist. Akzeptable Methoden müssen mithin **wissenschaftliche Methoden** sein. Man muß nicht unbedingt auf Feyerabend (1976, 1980) verweisen, um zu sehen, daß aktuelle wissenschaftstheoretische Positionen sehr vielfältig sind. Will man dieser Vielfalt gerecht werden, so bleibt nur - als kleinster gemeinsamer Nenner - die Feststellung, daß Wissenschaft ein soziales Unternehmen ist, das die Kontrolle seiner (inhaltlichen) Aussagen inkorporiert hat.

Agnew und Pyke (1969) verdeutlichen dies, indem sie Wissenschaft mit Journalismus vergleichen. Beide berichten über etwas, das sie herausgefunden haben. Dem Journalisten gestehen wir häufig nicht nur bei Leitartikeln, Theaterkritiken und anderen Arten von Kommentaren einen relativ großen Spielraum für Subjektivität zu. Wird dieser Spielraum überzogen, so sprechen wir zwar von schlechtem Journalismus; es ist aber ungewöhnlich, etwa in einem Zeitungsbericht Angaben über das Vorgehen des Berichterstatters bei der Erkenntnisgewinnung zu erwarten, die so detailliert sind, daß andere Personen die Richtigkeit seiner Aussagen überprüfen oder - bei nicht wiederholbaren Ereignissen - aufgrund einer Analyse des Vorgehens bewerten können. Einem Sportreporter geben wir zumindest gelegentlich auch das Recht, in einem Fußballspiel ein Abseits zu sehen, das Schiedsrichter, Linienrichter und 20000 Zuschauer nicht gesehen haben; er ist nicht verpflichtet, möglichst genau mitzuteilen, wie er zu seiner Wahrnehmung gekommen ist. Genau diese Verpflichtung, die **Kontrolle von Aussagen** durch andere erst ermöglicht, gilt aber für **wissenschaftliche** Aussagen.

Die Konsequenzen für ein Unternehmen, das sich - wie die Psychologie - als Wissenschaft begreift, sind offensichtlich. Zum ersten müssen Ergebnisse in einer mehreren Menschen zugänglichen Form - als Bericht objektiviert - vorliegen. Private Erkenntnis ist noch nicht Wissenschaft; Methoden müssen Handwerkszeug für die Herstellung von Berichten enthalten. Zum zweiten ist es erforderlich, nicht nur Ergebnisse mitzuteilen, sondern auch Informationen über den Weg zu vermitteln, auf dem sie gewonnen wurden. Es muß ein Methodenarsenal geben, auf dessen jeweils benutzte Techniken verwiesen werden kann; die in diesem Arsenal verfügbaren Methoden müssen so präzise dargestellt sein, daß sie bis auf unwesentliche Variation von verschiedenen Personen eingesetzt werden können. Hier ist eines der Aufgabenfelder der spezifischen Methodenlehren, die auch untersuchen, was im jeweiligen Kontext als wesentliche Änderung einer Methode und was als unwesentliche Va-

riante gilt. Ein mögliches Kriterium hierfür ist die Frage, welche Abweichungen den Anwendungsbereich der resultierenden Ergebnisse affizieren.

Soviel zur **Wissenschaftlichkeit** der Psychologie und den daraus resultierenden Minimalforderungen an ihre Methoden. Von den zahlreichen Spezifika des Faches seien nun im weiteren zwei herausgegriffen, deren methodologische Konsequenzen nicht nur an vielen Stellen in diesem Buch, sondern auch in anderen Kontexten immer wieder eine Rolle spielen. Das erste ist die Aussage, Psychologie sei eine **empirische** Wissenschaft, das zweite die Behauptung, ihr Gegenstand sei von **hoher Komplexität**.

## 2. Psychologie als empirische Wissenschaft

Was macht Psychologie zu einer **empirischen** Wissenschaft? Von de Groot (1969, S. 1) stammt die Aussage, eine Wissenschaft sei empirisch, wenn sie Wissen über die Welt bereitstellt, also über die Realität, in der wir leben. Dabei entstehen Theorien, die nach Newell (1990, S. 13) als Wissensbestände angesehen werden können, aus denen man Antworten auf Fragen erhalten kann. Sowohl Vorhersagen als auch Erklärungen oder Handlungsanweisungen sind gleichermaßen mögliche Varianten von Antworten. Die Tatsache, daß damit auch die Wissenschaften selbst als organisierte Wissensbestände interpretierbar sind, bezeichnet bereits Nagel (1961) als etwas, was „zweifelsfrei der Fall ist“ (S. 3, vom Verf. übersetzt). Die Aussagen von de Groot, Newell und Nagel lassen sich problemlos kombinieren, wenn man als Ziel einer **empirischen** Wissenschaft die Bereitstellung von Wissensbeständen ansieht, aus denen man Antworten über die **empirisch erfahrbare Welt** gewinnen kann.

### 2.1 Empirisch, formal, normativ

Fragt man, welches denn Wissenschaften sind, die auch in einem sehr liberalen Empiriekonzept als „nicht empirisch“ zu gelten haben, so gerät man in Schwierigkeiten. Betrachten wir ein **formales axiomatisches System**. Popper (1966) weist darauf hin, daß ein solches System auf zweifache Weise interpretiert werden kann: „Man kann die Axiome [i] als Festsetzungen betrachten oder [ii] als empirisch-wissenschaftliche Hypothesen“ (S. 42). Faßt man sie als Festsetzungen auf, dann legen sie fest, wie die in ihnen auftretenden Begriffe zu benutzen sind; sie sind implizite Definitionen eben dieser Begriffe. Da diese Begriffe als Variablen angesehen werden können, hat das System selbst den Charakter einer Struktur von Aussagenfunktionen. Es sei nun vereinbart, für den Vektor aller Variablen eines solchen Systems nur Wertekombinationen einzusetzen, bei denen die Aussagen des Axiomensystems gelten. Eine solche



Wertekombination ist ein „Modell“ des Systems. Setzt man nur Modelle ein, dann sind die resultierenden Aussagen in jedem Fall wahr, denn sonst wäre das, was man eingesetzt hat, kein Modell. Aus einem als Festsetzung interpretierten Axiomensystem läßt sich also eine Menge von Aussagen gewinnen, die unbeschadet aller Gegebenheiten in der uns umgebenden Welt stets wahr sind (Popper, 1966, S.42-43). Eine Wissenschaft, die lediglich einen Wissensbestand dieser Art bereitstellt, ist sicher nicht empirisch und soll zunächst als „formal“ gekennzeichnet werden.

### 2.1.1 Reine und angewandte Disziplin

Die angesprochene doppelte Interpretationsmöglichkeit axiomatischer Systeme führt dazu, daß es zu jeder **formalen Wissenschaft** ein **empirisches Pendant** gibt, das die gleichen Systeme als idealisierende Aussagen über eine äußere Realität behandelt. Diese beiden Geschwister werden oft als „reine“ und „angewandte“ Disziplin bezeichnet. Den Zusammenhang zwischen beiden diskutiert Nagel (1966, S. 215-233) am Beispiel der reinen und der angewandten Geometrie. Er bedient sich dazu eines einfachen Beispiels: Will man die Spitzen zweier Masten miteinander verbinden, die 30 Meter voneinander entfernt stehen, und von denen der eine 20 und der andere 15 Meter hoch ist, so werden die meisten von uns den zu überbrückenden Abstand zwischen den Mastspitzen nicht direkt ausmessen, sondern - unter Nutzung des Pythagoräischen Lehrsatzes - als  $\sqrt{30^2 + (20 - 15)^2}$  berechnen, was einen Abstand zwischen den Mastspitzen von etwa 30,41 Metern ergibt. Was rechtfertigt die Durchführung dieser Berechnung? Die Gültigkeit des Pythagoräischen Lehrsatzes folgt „rein formal“ aus den Axiomen der Euklidischen Geometrie. Was aber rechtfertigt den Rückgriff auf diese Axiome?

Wechseln wir das Beispiel und betrachten die - zwar schon etwas beharrte, dafür aber vielleicht gut bekannte - Theorie des individuellen Wahlverhaltens von Luce (1959). Sie ist probabilistisch formuliert; es gelten also zunächst alle Grundannahmen der Wahrscheinlichkeitstheorie. Es werden nun als „Wahlen“ bezeichnete Ereignisse eingeführt, die man durch Ausdrücke der Form  $\alpha \dashv \beta$  darstellen kann. Dabei ist für  $\beta$  eine Menge einsetzbar, für  $\alpha$  hingegen eine Teilmenge oder ein Element der für  $\beta$  substituierten Menge. Die Einführung von Mengen impliziert die Grundannahmen der Mengenlehre, die aber bereits durch die Einführung von Wahrscheinlichkeiten, die mit dem Reden über Ereignisse Objekte mit Mengeneigenschaften voraussetzen, ohnehin als gültig unterstellt wird. Die Wahl „ $b \dashv B$ “ (gelesen: „ $b$  aus  $B$ “) mit  $b \in B$  wird interpretiert als Wahl der Alternative  $b$  bei gegebener Optionmenge  $B$ . Entsprechend steht „ $A \dashv B$ “ mit  $A \subset B$  für die Wahl irgendeines Elements aus  $A$  bei gegebenem  $B$ . Die entscheidende Annahme ist nun, daß für  $x \in A \subseteq B$  mit **Prob** ( $x$ )  $\neq 0$  stets gilt:

$$\text{Prob } \{x \vdash B\} = \text{Prob } \{x \vdash A\} \text{Prob } \{A \vdash B\} \quad (2.1)$$

Führt man eine Gesamt-Optionenmenge  $\Omega$  als Vereinigung aller im Kontext einer bestimmten inhaltlichen Fragestellung denkbaren Optionenmenge ein, dann folgt im endlichen Fall formal die Existenz von Funktionen

$$v : \Omega \rightarrow \mathfrak{R} \quad (2.2)$$

derart, daß für alle  $x \in A \subseteq \Omega$  gilt:

$$\text{Prob } \{x \vdash A\} = \frac{v(x)}{\sum_{a \in A} v(a)}. \quad (2.3)$$

Vergleicht man - bei beliebiger Optionenmenge  $A$ , die  $x$  und  $y$  enthält - die Entscheidungswahrscheinlichkeiten für zwei Optionen  $x$  und  $y$ , dann ist

$$\frac{\text{Prob } \{x \vdash A\}}{\text{Prob } \{y \vdash A\}} = \frac{v(x)}{v(y)}. \quad (2.4)$$

Das ist die Grundlage der Bradley-Terry-Luce-Skalierung (Bradley & Terry, 1952; Luce, 1959). Führen wir eine weitere Skala ein, indem wir den Logarithmus eines  $v$ -Wertes mit  $u$  bezeichnen, dann ist

$$\log(\text{Prob } \{x \vdash A\}) - \log[\text{Prob } \{y \vdash A\}] = u(x) - u(y). \quad (2.5)$$

Damit aber sind wir bei einem Ausdruck aus dem Reservoir der log-linearen Modelle angelangt, die in diesem Band von Steyer (Kap. 15) vorgestellt werden, und die sich in der empirischen psychologischen Forschung beachtenswerter Beliebtheit erfreuen.

Nicht nur in der Geometrie, auch in der Psychologie gibt es also Anwendungen formaler Systeme. Dabei sei nicht unterstellt, daß die formale Einführung der „Anwendung“ zeitlich voraus geht. In den meisten Fällen ist es umgekehrt. Die von ägyptischen Landvermessern benutzten Techniken der Neuvermessung des Landes nach Nil-Überschwemmungen sind wesentlich älter als die „Elemente“ des Euklid, probit- und logit-Analysen gab es auch schon vor der Luceschen Konzeption individuellen Wahlverhaltens, Vertrauensbereiche für individuelle Testwerte und Reliabilitäten wurden bereits berechnet, bevor Lord und Novick (1968) das formale Konzept der „propensity distribution“ einführten.

Formale Wissenschaft liefert häufig im nachhinein Begründungen für das Vorgehen empirischer Forschung und alltäglicher Praxis. Sie schlägt Grundannahmen vor und weist nach, daß sich die Richtigkeit bestimmter Vorgehensweisen daraus ableiten läßt. Handelt es sich bei den Vorgehensweisen um Forschungsmethoden, so geht es also um die Analyse der „hidden assumptions“, die

diesen zugrundeliegen. Die Rede von den „hidden assumptions“ ist allerdings leicht mißverständlich. Da die gleiche Methode in durchaus unterschiedlichen formalen Systemen begründbar ist, darf man nicht der Vorstellung verfallen, es gebe zu Forschungsmethoden jeweils **die** verborgenen Grundannahmen, die zu akzeptieren jeder verpflichtet ist, der die entsprechende Methode benutzt, und die zu akzeptieren man ihm mithin auch unterstellen kann.

Die Vorteile der Analyse von Grundannahmen sind offensichtlich. Da formale Annahmensysteme Aussagen enthalten, die für alle Modelle trivial wahr sind, kann man Phänomenbereiche daraufhin untersuchen, ob sie als Modelle des jeweiligen Systems in Frage kommen. Wenn beispielsweise bei bestimmten Arten von Entscheidungen zwischen je zwei Alternativen bereits aufgrund von Plausibilitätsüberlegungen mit intransitiven Präferenzen zu rechnen ist (siehe etwa van Acker, 1977), dann liefern diese Entscheidungsbereiche keine Modelle für Theorien des Wahlverhaltens, deren Annahmen Präferenz-Transitivität implizieren. Bei dieser Art des Folgerns sind oft Methodenanalyse und psychologisch-inhaltliche Fragestellung kaum sinnvoll auseinanderzuhalten. Die Untersuchung der Grenzen eindimensionaler Präferenz-Skalierung wirft die Frage auf, wodurch sich denn ein- und mehrdimensionale Präferenzsysteme unterscheiden; zu deren Beantwortung bedarf es inhaltlicher Hypothesen, etwa über die Wahrnehmung und Analyse von Optionen und/oder über den Prozeß der Bildung von Präferenzurteilen.

Die Abstraktion von einer konkreten Forschungsmethode zu einem formalen Annahmensystem liefert überdies die Möglichkeit, von einem ursprünglich „üblichen“ Anwendungsbereich, der oft sogar die Terminologie des methodischen Apparates geprägt hat, zu einem anderen überzugehen. Ein Beispiel dafür ist die Verwendung der „Signal-Entdeckungs-Theorie“, die in ihrem Namen ihre Herkunft aus der Nachrichtentechnik (Peterson, Birdsall & Fox, 1954; van Meter & Middleton, 1954) nicht verleugnet, in der psychologischen Gedächtnisforschung (Egan, 1958). Die Verschränkung von formaler Analyse und empirischen Befunden verdeutlichen Mausfelds Ausführungen zur Messung im 13. Kapitel dieses Bandes; zur Nutzung formaler Theorien aus anderen Bereichen der Psychologie finden sich Beispiele in seiner Darstellung der Psychophysik im 4. Kapitel.

Empirische Forschung hat also oft einen formalen Partner; das Selbstverständnis der Psychologie als einer **empirischen Wissenschaft** begründet mithin keine **Beschränkung** ihres Methodenarsenals auf **empirische Methoden**, sondern fordert sogar, auch **formale Methoden** zumindest zur Begründung ihres empirischen Vorgehens einzusetzen. Wie steht es nun mit der Abgrenzung von einer Wissenschaft, die man „normativ“ nennen kann?

## 2.1.2 Normativität

Im Umfeld der Psychologie taucht die Rede vom **Normativen** häufig im Zusammenhang mit Rationalitätskonzepten auf, bei denen zwischen einer deskriptiven, einer präskriptiven und einer normativen Nutzung unterschieden wird (Bell, Raiffa & Tversky, 1988). Dabei wird normative Nutzung oft so aufgefaßt, daß hiermit der Anspruch verknüpft sei, jeder Mensch möge sich entsprechend der jeweiligen Konzeption verhalten. So verstandene normative Wissenschaft liefert also Handlungsanweisungen; ihre Systematisierung besteht aus dem Aufbau von Systemen, in denen sich diese Anweisungen ableiten lassen. In der mathematisch orientierten Spiel- und Entscheidungstheorie, die sich ausgiebig mit Rationalitätsvorstellungen beschäftigt, herrscht dagegen eine andere Sichtweise vor. Prämissen (für Systeme von Handlungsanweisungen) werden als Normen im Sinne von **Maßstäben** angesehen; sie definieren implizit eine Klasse von Handlungen als „normgerecht“ im Sinne der jeweiligen Prämissen. Mit formalen Methoden kann man dann untersuchen, welche Eigenschaften derartigen Handlungen und ihren Konsequenzen zukommen und wie sie zu finden sind.

Selten (1989) unterscheidet zwischen „praktischer“ und „idealer“ Normativität. **Praktische Normativität** sieht die durch einen Satz von Prämissen beschriebene Norm als Handlungsanweisung, **ideale Normativität** versucht, unbeschadet der Möglichkeiten und Grenzen realer Menschen zur Klärung des jeweiligen Normkonzeptes beizutragen. Die hierbei eingesetzten Methoden sind die der formalen Analyse; da damit stets Implikationen der Prämissen - also analytisch wahre Aussagen - gewonnen werden, nennt Aumann (1988) auch die zugehörigen Methoden „analytisch“. In verschiedenen Kontexten hat sich offensichtlich ein unterschiedlicher Sprachgebrauch eingebürgert; die „analytischen“ Methoden sind genau die gleichen, die wir zuvor (siehe S. 8) als „formal“ bezeichnet haben. Auch die Systeme selbst sind oft formalisiert; die Prämissen werden als Axiome dargestellt, „formale“ Systeme und „normative“ Systeme sehen gleich aus. **Ein** Unterschied aber bleibt. Hatten wir es zuvor mit Grundannahmen zu tun, die auch als Festsetzungen oder als idealisierende Aussagen über eine **gegebene** äußere Realität angesehen werden konnten, so haben wir es im Bereich des Normativen mit Aussagen zu tun, die als Festsetzungen oder als Aussagen über **mögliches** menschliches Handeln interpretierbar sind.

Natürlich hindert nichts daran, auch Prämissen eines normativen formalen Systems als empirische Hypothesen über real vorfindbares Verhalten zu interpretieren. Dieses Vorgehen wird als „deskriptive Nutzung“ normativer Systeme bezeichnet; diese aber hat sich - zumindest im Bereich der Rationalitätskonzepte - als wenig brauchbar erwiesen. Es gibt jedoch mindestens zwei andere Verwendungsmöglichkeiten. Wir können real beobachtbare **Handlun-**

**gen** durch Vergleich mit verschiedenen Normsystemen beschreiben und klassifizieren. Hier wird „Norm“ als „Vergleichsmaßstab“ aufgefaßt. Für jedes System läßt sich dann untersuchen, unter welchen Bedingungen normgerechtes Handeln (im Sinne eines jeweils spezifischen Satzes von Prämissen) zu erwarten ist, und unter welchen nicht. Zur Erklärung dabei auftretender Befunde kann man dann (beispielsweise) der Frage nachgehen, welche kognitiven oder motivationalen Mechanismen zum normgerechten Handeln führen. Eine zweite Möglichkeit ist, **Situationen** daraufhin zu untersuchen, welche Möglichkeiten normgerechten Handelns sie überhaupt bieten. Es wäre überraschend, wenn Situationsarten, die sich auf diese Weise unterscheiden lassen, nicht auch durch Unterschiede des in ihnen auftretenden Realverhaltens gekennzeichnet wären. Normative Systeme können also die Grundlage für Methoden der Situationsbeschreibung und -klassifikation auf der Basis der Analyse situationsspezifisch erfüllbarer bzw. nicht erfüllbarer Handlungsnormen bieten. Tack (1991) exemplifiziert ein derartiges Vorgehen am Beispiel der Normen kooperativer Rationalität.

Diese Überlegungen zum Verhältnis von empirischer, formaler und normativer Wissenschaft zeigen, daß aus dem Selbstverständnis der Psychologie als einer empirischen Wissenschaft zunächst lediglich folgt, daß sie neben einer empirischen Methodik stets auch der formalen bzw. analytischen Methoden bedarf, um Teile ihres Vorgehens in größeren Kontexten begründen und normative Vorstellungen sowohl zur Handlungsbeschreibung als auch zur Situationsanalyse nutzen zu können. Aus der Empirizität der Psychologie folgt **keine Beschränkung** auf empirische Methoden, sondern bestenfalls, daß diese **unverzichtbar** sind. Wenn also „empirische Methoden“ nicht schlicht als „Methoden einer empirischen Wissenschaft“ umschrieben werden können, dann müssen wir nach anderen Abgrenzungs-Gesichtspunkten suchen.

## 2.2 Empirische und rationale Erkenntnis

Bei der Frage, was Psychologie zu einer empirischen Wissenschaft macht, sind wir - in Anlehnung an de Groot (1969) - davon ausgegangen, sie stelle Wissen über die Welt, in der wir leben, zur Verfügung. Im vorausgegangenen Abschnitt haben wir gezeigt, daß daraus aufgrund der doppelten Interpretierbarkeit von Aussagensystemen kein Verzicht auf formale Methoden, sondern sogar deren Notwendigkeit folgt. Es dürfte aber auch klar geworden sein, daß formale Methoden allein ebenfalls nicht ausreichen, da sie keine Grundlagen für die Behauptung und Kontrolle der **faktischen Richtigkeit** oder - zumindest - **Brauchbarkeit** von Tatbestandsaussagen bereitstellen.

Notwendig sind also Methoden, mit denen man Aussagen **als faktisch gültig rechtfertigen** kann, indem man Beweismaterial für ihre Geltung beibringt. Be-

trachtet man Erkenntnis jeder Art, also auch wissenschaftliche Erkenntnis, aus einer individualistischen Position als Überzeugung von einzelnen, dann führen solche Überlegungen zunächst auf einen Regreß, der nach Chalmers (1989, S. 116) mindestens seit Platon bekannt ist. Zur Rechtfertigung einer Aussage werden andere Aussagen herangezogen; um diese zu rechtfertigen wieder andere und so fort. Um diesen Regreß zu stoppen, braucht man Aussagen, die selbst keiner weiteren Rechtfertigung bedürfen, weil sie „in gewissem Sinne durch sich selbst gerechtfertigt sind“ (Chalmers, 1989, S. 116). Der „klassische Rationalismus“ geht davon aus, daß solche Grundlagen der Erkenntnis durch Nachdenken gewonnen werden können, da sie sich als eindeutig, unabhängig und selbstverständlich erweisen, während der „Empirismus“ die Grundlagen der Erkenntnis durch Sinneswahrnehmung zu sichern bemüht ist.

### **2.2.1 Klassisch rationales Vorgehen**

Descartes (1637) wird oft als der erste klassische Rationalist der Neuzeit bezeichnet. Da es für uns nur schwer vorstellbar ist, daß man durch bloßes Nachdenken zu faktisch wahren Aussagen kommt, sei ein Beispiel skizziert, das gleichzeitig die Tücken einer solchen Methodik veranschaulicht. Damerow, Freudenthal, McLaughlin und Renn (1992, S. 30 ff.) analysieren unter anderem eine Tagebuchnotiz Descartes' aus dem Jahre 1618, in der dieser den Zusammenhang zwischen Zeit und Weg beim freien Fall herauszufinden versucht. Zur Notiz gehört eine Skizze, die in Abbildung 1 dargestellt ist.

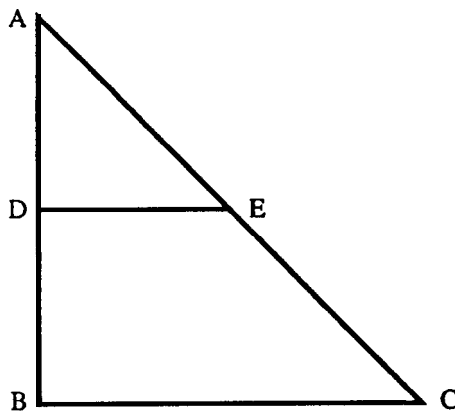


Abb. 1: Skizze zu einer Tagebuchnotiz Descartes' von 1618

Descartes notiert, ein ausnehmend kluger Mann (womit Beekman gemeint ist) habe ihn mit einem Problem konfrontiert, das - stark verkürzt - im Kern

fragt: „Ein Stein fällt in einer Stunde von **A** nach **B** und wird von der Erde mit einer gleichbleibenden Kraft angezogen; in welcher Zeit wird **AD** durchfallen?“ Descartes sagt, er habe das Problem gelöst. Fläche **ABC** steht für die „Bewegung“, der Unterschied zwischen dem Punkt **A** und der Strecke **BC** zeigt die Ungleichheit der Bewegung. Somit wird Strecke **AD** in der durch Fläche **ADE** gegebenen Zeit durchlaufen, und **DB** in der durch **DEBC** repräsentierten Zeit. Dabei müsse man allerdings beachten, daß die kleinere Fläche für langsamere Bewegung steht. Nun ist **ADE** ein Drittel von **DEBC**, und **AD** wird somit dreimal so langsam durchfallen wie **AB**.

Diese Argumentation wirkt heute etwas chaotisch. Flächen werden als Kehrwerte des Zeitbedarfs interpretiert. Ist die zugehörige Fallstrecke in mehreren Fällen gleich (wie bei **AD** und **DB**), so können wir diese Strecke als Einheit wählen; die Flächen sind dann proportional den mittleren Geschwindigkeiten. Dabei erhalten die Längen der basisparallelen Strecken zwischen **AB** und **AC** den Charakter von Geschwindigkeiten, die somit proportional der jeweils zurückgelegten Fallstrecke sind (und nicht - wie man heute bereits in der Schule lernt - proportional zur Fallzeit). Damerow et al. (1992) untersuchen in ihrem Buch die Möglichkeiten einer konsistenten Interpretation der damals benutzten Sprache, und sie analysieren, wann und wie Galileo, Descartes, Beekman und andere erfaßt haben, daß die Geschwindigkeit eines frei fallenden Körpers nicht proportional zur durchfallenen Strecke, sondern zur verflossenen Zeit ist. Immerhin: Dieser Wechsel fand statt. Eine Fehlannahme über die uns umgebende Welt wurde durch Nachdenken korrigiert und nicht durch Beobachtung und/oder Messung.

Zwei Fragen dazu. Die Erste: Warum versucht man im 17. Jahrhundert die Frage nach einem faktisch geltenden Gesetz durch Nachdenken zu lösen, wo doch empirische Forschung, die nachschaut und registriert, zu dieser Zeit großen Aufschwung nimmt. Olaus Römer mißt die Lichtgeschwindigkeit, Evangelista Toricelli und Otto von Guericke experimentieren mit dem Luftdruck, Robert Boyle legt empirische Daten zur Nicht-Existenz der klassischen vier Elemente vor, Antony van Leeuwenhoek sieht zum ersten Mal unter dem Mikroskop Bakterien und Thomas Hobbes propagiert eine „soziale Physik“, deren Erkenntnisgrundlage Selbstbeobachtung durch Introspektion ist (Störig, 1970, S. 299-360).

**Eine** mögliche Antwort ist rein pragmatisch. Praktisch realisierbare Fallhöhen sind so klein und Meßtechniken so ungenau, daß man allein mit technisch unterstützter Beobachtung keine Daten erhält, aus denen sich das Fallgesetz ergibt. Die uns umgebende Luft bremst statt zu beschleunigen; empirische Daten werden nicht einmal symmetrisch um ideale Werte streuen. Die in der Psychologie so beliebte Mittelwerts-Strategie hilft also auch nicht. Könnte man extrem genau messen und gäbe es extrem lange Fallstrecken, dann ließe

sich das Gesetz des freien Falls in empirischen Daten auch nicht wiederfinden, da während eines jeden Falles eine Annäherung an den Erdmittelpunkt stattfindet, womit die (beim Fallgesetz als konstant unterstellte) Erdanziehung zunimmt.

Damit sind wir bei einer **zweiten** Antwort. Der freie Fall ist eine **Idealisierung**, eine kontrafaktische Aussage darüber, was geschehen würde, wenn bestimmte Annahmen erfüllt wären, deren Erfüllung praktisch nicht zu sichern ist. Andererseits kann man über die „Störfaktoren“ und deren Effekte ebenfalls Aussagen machen, so daß zunehmende Berücksichtigung von immer mehr „sonstigen“ Einflüssen eine zunehmend bessere Anpassung empirischer Daten ermöglicht. Gegenstand der rationalen Erkenntnissuche ist also etwas, das als Idealisierung in der uns umgebenden Welt durchaus vorhanden ist, sich aber wegen der Überlagerung von Effekten einer Bedingungsvariation, die im Idealfall nicht vorgesehen ist, der direkten Beobachtung und Registrierung entzieht. Die Brücke zur Empirie besteht darin, daß man unter expliziter Berücksichtigung zusätzlicher Bedingungsfaktoren etwas darüber aussagen kann, wie sich die Diskrepanz zwischen Realdaten und Idealphänomen ändert, wenn man eben diese zusätzlichen Faktoren kontrolliert.

### 2.2.2 Metaprinzipien

Nun zu der Frage: Was ist es, das in rationalen Analysen letztendlich als so eindeutig, unabhängig und selbstverständlich angesehen wird, daß man es als Erkenntnisgrundlage nutzt, ohne es weiter in Frage zu stellen, und das wir ein „Metaprinzip“ nennen wollen? Im Beispiel des freien Falls ist dies ein **Einfachheits-Prinzip**. Wenn die Geschwindigkeit beim freien Fall nicht konstant ist, sondern offensichtlich zunimmt, dann ist die einfachste Möglichkeit eine Proportionalität der Geschwindigkeit zu einer Größe, deren Zunahme während des Falles trivial ist. Als solche Größen stehen Fallstrecke und Fallzeit zur Verfügung. Descartes' Pech war, daß er zunächst auf die falsche Karte setzte, nämlich auf die Fallstrecke.

Ein anderes Prinzip ist **Optimalität**. Schoemaker (1991) bezeichnet es als sehr erfolgreiches und flexibles wissenschaftliches Metaprinzip und verweist auf Nutzenmaximierung in der Ökonomie, Aufwandsminimierung in der Physik, Entropie in der Chemie und das Überleben der am besten angepaßten genetischen Ausstattung in der Biologie. **Optimalitätsprinzip** ist allgemein die Annahme, ein Verhalten oder ein anderes empirisch faßbares Phänomen verlaufe so, daß eine geeignete Zielfunktion unter Berücksichtigung gegebener Randbedingungen maximiert oder minimiert wird. Verschiedene Zielfunktionen entsprechen unterschiedlichen **Optimalitätsstandards**. Er weist darauf hin, daß bei Coombs, Dawes und Tversky (1970) Beispiele dafür zu finden sind, daß



sich auch in der Psychologie beispielsweise im Kontext der Signalentdeckungstheorie, der Erforschung des Lernens und insbesondere der operanten Konditionierung sowie in der Entscheidungsforschung Formalisierungen finden, denen Optimalisierungsannahmen zugrundeliegen (Schoemaker, 1991, S. 207).

Am Beispiel der Optimalität läßt sich zeigen, wie Wissenschaft heute mit Meta-Prinzipien umzugehen pflegt. Der klassische Rationalismus ging davon aus, daß letzte Prinzipien, die als eindeutig, unabhängig und selbstverständlich jeder gegen sie gerichteten Argumentation standhalten, stets richtig sind, und daß mithin aus ihnen abgeleiteten Aussagen über die äußere Welt faktische Richtigkeit zukommt. Schoemaker (1991) bezeichnet hingegen Meta-Prinzipien als „Heuristiken“. Eine Heuristik führt zwar oft und einfach, aber **niemals immer** zur richtigen Lösung eines Problems. Da ist die Gefahr einer Fehlattribution. Kitcher (1985) weist darauf hin, daß allein die Erklärbarkeit von Verhalten in einem Optimierungsmodell noch kein hinreichender Grund dafür ist, Optimierung als einzig mögliches Erklärungsprinzip und damit als in der äußeren Welt gegeben zu betrachten. Eine überzogene Bestätigungstendenz kann dazu führen, daß für eine Optimalitätsannahme Daten aus Bereichen angehäuft werden, die mit hoher Wahrscheinlichkeit die jeweilige Annahme stützen, also nicht „kritisch“ sind. Schoemaker (1991, S.212) führt diese Fehlermöglichkeit explizit an; sie dürfte allerdings kaum spezifisch für den Umgang mit Meta-Prinzipien sein, sondern jeden plausibel erscheinenden Erklärungsansatz treffen. Überziehung des (angenommenen) Geltungsbereichs und zu schwache empirische Kontrolle wegen allzu eindrucksvoller Plausibilität sind weitere Gefahren. Meta-Prinzipien liefern somit lediglich Aussagen mit hypothetischem Charakter, die der Kontrolle durch Validierung bedürfen. Das Methodenarsenal des klassischen Rationalismus wird hier als Werkzeug zur Generierung und Elaboration (hypothetischer) Theorien benutzt. Über Heuristiken der Theorie- und Hypothesenkonstruktion findet sich mehr in einem Beitrag von Dörner (Kap. 8 dieses Bandes).

Was sind nun die Kriterien der Validierung rational begründeter Theorien und Hypothesen? Schoemaker (1991, S. 238) führt fünf Kriterien an, von denen das erste die Genauigkeit ist, mit der existierende Daten erklärt oder vorhergesagt werden können, und das zweite die Generierung neuer Hypothesen und deren Validierbarkeit. Damit ist offensichtlich, daß - im Gegensatz zum klassischen Rationalismus - die rationale Begründung nicht als hinreichend gesehen wird, sondern als möglicherweise zu Fehlschlüssen führende Heuristik, die der Kontrolle durch Anwendung empirischer Methoden bedarf. Ansonsten ist der Kriterienkatalog eine von vielen Varianten der Zusammenstellung unterschiedlicher Aspekte der Theoriebewertung, mit denen sich in diesem Buch Gadenne (Kap. 9 dieses Bandes) ausführlicher beschäftigt.

Beispiele für die Bedeutung derartiger Überlegungen im Rahmen psychologischer Forschung bieten in breiter Auswahl die Kommentare zu Schoemakers (1991) Artikel, von denen mehr als die Hälfte von Psychologen stammen. Dabei zeigt sich die Spannweite von einer eng empirizistischen Position, die allein Daten gelten läßt und rational begründete Prinzipien für überflüssig hält, über eine Einstellung, derzufolge rationale Systeme die Bedeutung globaler „Nullhypothesen“ haben, bis hin zu Beispielen dafür, daß die Nutzung von Meta-Hypothesen neue Einsichten und Erklärungsmöglichkeiten eröffnet. Die „Nullhypothesen-Position“ wird oft bezüglich des Modells vom rationalen Entscheider vertreten. Danach sind **es** die **Abweichungen** vom rationalen Entscheidungsverhalten, die psychologisch relevant und erklärungsbedürftig sind. Eine andere Einstellung vertreten etwa Baron (1991) und Tack (1991), die die **Realisierung** rationalen Verhaltens für genauso erklärungsbedürftig halten. Die Tatsache, daß es sowohl reales rationales Verhalten als auch Abweichungen davon gibt, wird - zunächst wieder rein rational und unter Plausibilitäts Gesichtspunkten - durch die Existenz von Heuristiken erklärt, die unter bestimmten und im Alltag häufigen Bedingungen erfolgreich sind, während sie unter anderen Bedingungen Abweichungen liefern. Die Identifikation solcher Heuristiken und die Überprüfung der aus ihnen ableitbaren Hypothesen über die Situationsabhängigkeit realisierter Rationalität ist es, die dann mit empirischer Methodik zu untersuchen ist. Diese Sichtweise entspricht jener, die Anderson (1991) bezüglich der Optimalität des menschlichen Gedächtnisses vertritt.

Rationale Theorie- und Hypothesenbildung mit der zugehörigen Methodik fällt - sofern die jeweiligen Prinzipien und deren Varianten hinreichend formalisiert sind - mit der zuvor angesprochenen formalen Analyse zusammen. Wie in den Bemerkungen zur Validierung erwähnt, wird - damit von empirischer Wissenschaft die Rede sein kann - darüber hinaus noch eine Methodenart benötigt, die Aussagen mit Beobachtungen verknüpft. Diese Methodenart ist gemeint, wenn wir im folgenden von **empirischen Methoden** reden.

## 2.3 Empirische Methoden

Empirische Methoden liefern Aussagen, die Beobachtungsergebnisse möglichst direkt wiedergeben, und deren faktische Richtigkeit sich ohne weiteren Begründungszwang aus dem jeweils beobachteten Sachverhalt ergibt. Man kann zweifeln, ob es überhaupt Vorgehensweisen gibt, die dies leisten können. Vorausgesetzt wird sowohl die Veridikalität von Beobachtungen als auch die Möglichkeit einer Sprache, die Beobachtungsergebnisse unverzerrt zu repräsentieren vermag. Eine solche Sprache ist in der Wissenschaftstheorie unter dem Namen „Beobachtungssprache“ (siehe Stegmüller, 1970, S. 293-399) be-

kannt. Es gibt viele Argumente, mit denen man die Möglichkeit einer derartigen Sprache in Zweifel ziehen kann; nahezu jede wissenschaftstheoretische Position hat hierzu eine eigene Meinung geäußert und begründet (siehe die entsprechenden Absätze bei Chalmers, 1989). Das gleiche gilt für die Frage nach der Möglichkeit unbelasteter und wahrheitsgetreuer Beobachtung. Hierzu findet man gelegentlich den Hinweis, am ehesten fehlerfrei feststellbar seien Raum-Zeit-Koinzidenzen; die zuverlässigsten Beobachtungsaussagen seien also solche, in denen festgestellt wird, daß sich Objekt **A** zum Zeitpunkt  $t$  am Punkt mit den Ortskoordinaten  $\langle x, y, z \rangle$  befindet bzw. befunden hat. Da Veränderungen von Raum-Zeit-Koinzidenzen Bewegungen sind, sind auch diese problemlos feststellbar.

In der Psychologie hat eine derart enge Konzeption empirischer Letztbegründung wissenschaftlicher Aussagen außer in extremen Spielarten des Behaviorismus kaum eine Rolle gespielt. Soweit sich Psychologie mit Verhalten beschäftigt, hat sie es mit beobachtbaren Phänomenen zu tun, die bestenfalls in manipulativ extrem kontrollierten Situationen durch hinreichend einfach beschreibbare Klassen von Raum-Zeit-Koinzidenzen oder deren Veränderungen erfaßt werden können. Hinzu kommt, daß nach einer ebenfalls zur Psychologie gehörenden Tradition, für die etwa William James (1890) steht, Gedanken und Gefühle zu den grundlegenden Daten des Faches gehören, für die man entweder eine besondere Art von Beobachtung - „Introspektion“ genannt - postulieren muß, oder eine Zugänglichkeit durch Interpretation beobachtbarer Sachverhalte.

Die Charakterisierung empirischer Methoden bedarf daher der Abschwächung. Verlangt werde nun lediglich, daß empirische Methoden Aussagen liefern, von denen allgemein **unterstellt** wird, sie gäben Beobachtungsergebnisse möglichst direkt wieder, und für deren Richtigkeit keine weiteren Begründungen **gefordert** werden. Das gibt dem Ganzen einen konventionalistischen Anstrich, der durchaus Sinn macht. Wenn Psychologie - wie jede Wissenschaft - ein soziales Unternehmen ist, dann muß eine Aussage, zu der keine weiteren Begründungen verlangt werden, für andere so verständlich sein, daß die zugrundeliegende Feststellung (Beobachtung) nachvollzogen werden kann. Dies setzt sozialen Konsens darüber voraus, was als „in sich selbst begründet“ gelten möge. Andererseits ist die abgeschwächte Charakterisierung aber auch nicht ausschließlich konventionalistisch. Der allgemeine Verzicht auf weitere Begründung bei bestimmten Aussagenarten kann jederzeit in Frage gestellt werden. Diskussionen über Erhebungsartefakte oder Kontrollen der Inter-Rater-Reliabilität von Beobachtungsdaten sind Beispiele dafür.

Zur weiteren Explikation empirischer Methodik sollen nun zwei Aspekte betrachtet werden, in denen sich empirische Methoden unterscheiden, nämlich das Ausmaß der „Datenrestriktion“ und der „Manipulationsgrad“.

### 2.3.1 Datenrestriktion

Mit „**Datenrestriktion**“ ist gemeint, wie stark und auf welche Weise der im Prozeß einer wie auch immer gearteten Beobachtung bzw. Registrierung stattfindende Übergang von einem Phänomen zu einem objektiviert vorliegenden Datum die Menge der erfaßten Phänomenaspekte einschränkt. Das eine Extrem - mit minimaler Restriktion - wäre beispielsweise die Video-Aufzeichnung eines Verhaltens mit möglichst vielen Kameras aus verschiedenen Perspektiven, das andere Extrem die rechnergesteuerte Erhebung eines Testwertes, bei der lediglich eine Gesamtzahl richtiger Antworten ausgewiesen wird und nicht einmal mehr das Antwortmuster erkennbar ist.

Beide Extreme - wie auch die dazwischen liegenden Möglichkeiten - haben ihre spezifischen Anwendungsbereiche. Eine möglichst geringe Datenrestriktion ist das Ideal einer **interpretierenden Wissenschaft**. Soll Interpretation, wie von Soeffner und Hitzler im 3. Kapitel dieses Bandes ausgeführt, die Vielfalt der Deutungsmöglichkeiten eines Sachverhaltes erschließen, dann ist zunächst unbekannt, welche Möglichkeiten es gibt und welche Aspekte des Sachverhalts zur Begründung und Kontrolle benötigt werden. Die ideale Interpretationsbasis kommt mithin einer vollständigen Repräsentation der zu deutenden Gegebenheit, die diese als eine Art Kopie über die Zeit hinweg bewahrt, möglichst nahe. Dem dient ein empirischer Zugang durch möglichst freie und umfassende Beobachtung.

Nun zum anderen Extrem. Eine starke Datenrestriktion setzt voraus, daß im vorhinein bekannt ist, zu welchen Aspekten beobachtbarer Gegebenheiten welche Informationen benötigt werden. Das ist beim streng **theorie- und hypothesesengeleiteten Vorgehen** oft der Fall, wenn die benutzten Hypothesen derart spezifiziert sind, daß sie die Art der zu ihrer Kontrolle benötigten Informationen festlegen. Soll eine Hypothese zur ‚perceptual defense‘ geprüft werden, die unterschiedliche mittlere Erkennungszeiten für Wörter verschiedener emotionaler Tönung bei Personen mit unterschiedlicher Neuroseform behauptet, dann benötigt man als Daten Worterkennungszeiten, eine Kennzeichnung der emotionalen Tönung der benutzten Wörter, klinisch-psychologische Diagnosen der Versuchspersonen und sonst nichts.

Es gibt noch einen anderen guten Grund für Datenrestriktion, der in der alltäglichen Forschungspraxis oft eine Rolle spielt. Als sogenannte „Theorie“ oder „Hypothese“ liegt lediglich eine unscharfe Vermutung über einen nur wenig spezifizierten Zusammenhang vor (wie bei der Katharsis-Theorie, der Frustrations-Aggressions-Theorie und ähnlichem mehr). Weitere Elaboration führt auf Annahmen über Zusammenhänge, die als Korrelationen ungleich Null verstanden werden, über Mittelwertsunterschiede Verschiedenster Art, die man als varianzanalytische Haupt- und/oder Wechselwirkungseffekte deu-

tet, über Ähnlichkeitsbeziehungen, die faktoren- oder clusteranalytischen Konzepten entsprechen, und so fort. Diesen und ähnlichen Elaborationen ist gemeinsam, daß sie eine bestimmte **Strategie der Datenauswertung** nahelegen. Jede Entscheidung für eine bestimmte Technik der Auswertung ist aber zugleich die Festlegung auf Daten, deren Eigenschaften die gewünschte Auswertung überhaupt ermöglichen.

Damit die in Erhebungen generierten Daten für bestimmte Auswertungen brauchbar sind, benötigt man einen konzeptuellen Rahmen, in dem sich entsprechende Datenformate darstellen und analysieren lassen. Dabei muß gezeigt werden, aufgrund welcher Regeln bei der Datengenerierung Formate mit Eigenschaften resultieren, die Voraussetzung für unterschiedliche Möglichkeiten der Weiterverarbeitung sind. Eine Basis hierzu bietet die **Datentheorie** von Coombs (1964), deren weitere Entwicklung etwa bei Roskam (1983) dargestellt wird. Allgemein spricht man dann, wenn Restriktionen zu Daten in einem vorher festgelegten Format führen, auch von einer „**datenförmigen Empirie**“.

### 2.3.2 Manipulationsgrad

Der **Manipulationsgrad** kann unabhängig von der Datenrestriktion variieren. Dies wird oft übersehen, da in der einschlägigen Literatur häufig **Felduntersuchungen und Laboruntersuchungen** als Extrema eines Kontinuums unterschiedlicher Lebensnähe betrachtet werden (siehe Bortz, 1984, S. 33), wobei manipulative Kontrolle und Datenrestriktion als Reduktionen dieser „Lebensnähe“ nicht immer klar auseinanderzuhalten sind.

Betrachten wir eine verkehrspsychologische Untersuchung, bei der an verschiedenen gestalteten Ampelkreuzungen durch Lichtschranken registriert wird, wie lange nach einem Wechsel von Grün auf Rot wieviele Autos noch in den Kreuzungsbereich einfahren. Manipulative Kontrolle ist so gut wie nicht vorhanden; die Autofahrer sind in der gleichen alltäglichen Situation, in der sie auch wären, wenn keine Registrierung stattfinden würde. Die Datenrestriktion ist allerdings massiv; die Erhebungstechnik liefert Zeitverteilungen und sonst nichts. Die Untersuchungssituation ist lebensnah, die Daten sind hingegen ausgesprochen realitätsfern abstrahierend. Übrigens wäre es sinnvoll, sich in diesem Kontext den Ausdruck „Lebensnähe“ abzugewöhnen und besser von „Alltagsnähe“ zu sprechen, denn auch in der Laborsituation ist die Versuchsperson selbst unter extremer manipulativer Kontrolle immer noch ein lebender Mensch. Das Labor, in dem Versuchspersonen nur noch durch präparierte Teile vertreten sind, gibt es in der Psychologie nicht. Die Konzeption einer „lebensfernen“ Laborsituation sollten wir den Wissenschaften überlassen, in denen diese wirklich vorkommt.

Zeigt das verkehrspsychologische Beispiel die Vereinbarkeit eines geringen Manipulationsgrades mit einer starken Datenrestriktion, so wäre ein Gegenstück die Untersuchung einer Versuchsperson in einer perfekt abgedunkelten camera silens, bei der mit Tonband und Infrarot-Kamera registriert wird, wie sich im Laufe der Zeit Äußerungen zu Halluzinationen und Indikatoren eines Verlusts des emotionalen Gleichgewichts einstellen. Die Situation ist stark manipuliert und entsprechend alltagsfern, die Daten sind kaum restringiert und entsprechend realitätsnah und anschaulich.

### **2.3.3 Interne und externe Validität**

Eng verknüpft mit dem Manipulationsgrad ist die Konzeption einer „**internen**“ und einer „**externen Validität**“, die alles andere als klar ist. Die Angabe, eine Untersuchung sei intern valide, „... wenn ihre Ergebnisse eindeutig interpretierbar sind (Bortz, 1984, S.29)“, besagt solange wenig, wie nicht klar gestellt ist, was als Ergebnis und was als Interpretation gelten soll; die Gleichsetzung externer Validität mit Generalisierbarkeit über die Bedingungskonstellation einer Untersuchungssituation und den Kreis der Versuchspersonen hinaus ist ohne Präzisierung des Generalisierungsbegriffs ebenso unklar.

Zur Klärung beschränken wir uns auf empirische Methoden, die Angaben darüber liefern, daß bei bestimmten Bedingungskonstellationen bestimmte Phänomene auftreten. Die Bedingungskonstellationen seien durch Werte endlich vieler Variablen charakterisiert, die „**unabhängige Variablen**\*“ heißen, während die Beschreibung der resultierenden Phänomene durch Wertangabe endlich vieler anderer Variablen erfolgt, die „**abhängig**“ genannt werden.

Sind unabhängige und abhängige Variablen getrennt, dann ist zu fragen, ob eine **neue** Beobachtung mit einer bereits früher realisierten Kombination von Werten der **unabhängigen** Variablen mit der gleichen Wertekombination **abhängiger** Variablen verknüpft ist wie zuvor. Wenn ja, dann ist die durchgeführte Untersuchung - und insbesondere die benutzte empirische Methode - „valide“. Dabei läßt sich das Prinzip der vollständigen Übereinstimmung von Wertekombinationen nach Einführung einer Unterschiedlichkeits-Metrik auf verschiedene Weise sinnvoll probabilistisch abschwächen.

In der Psychologie stammt die Bezeichnung „Validität“ aus der Theorie psychodiagnostischer Verfahren und deren Anwendung; sie kennzeichnet dort die Brauchbarkeit eines Diagnostikums für Vorhersagen in neuen Situationen. Die Vorstellung, daß Validität stets eine Art **prognostischer Brauchbarkeit** ist, bleibt beim hier skizzierten Präzisierungsversuch erhalten. Liegen valide empirische Daten vor, dann können wir - nach üblichem Sprachgebrauch - auch davon sprechen, die durch unabhängige Variablen beschriebene Bedingungs-

kombination „bewirke“ die durch abhängige Variablen erfaßten Phänomenaspekte. Daß **A** ein **B** bewirkt, besagt, daß **A** - wie immer es auch zustandekommen möge - stets mit **B** einhergeht, daß also der Befund, **A** trete mit **B** kombiniert auf, valide ist.

Ein zusätzliches Problem soll nicht unerwähnt bleiben: Wenn Validität eine Art prognostischer Eignung ist, dann wird dazu nicht nur eine empirische Erhebungsmethode benötigt, die Einzelaussagen liefert, sondern zusätzlich eine Methode zur Generalisierung dieser Aussagen, um aus deren Ergebnis eine Individualaussage über einen neuen Fall ableiten zu können. Der Weg von Einzelbeobachtungen zu generalisierenden Aussagen ist aber keineswegs trivial; hierzu bedarf es einer Methode, die durch vorliegende Einzeldaten „vernünftig begründete“ allgemeine Aussagen ermöglicht und damit eine Lösung des „schwachen Induktionsproblems“ (siehe Westermann & Gerjets in Kap. 10 dieses Bandes) anbietet.

Dieser Versuch einer Präzisierung des Validitätsbegriffes ist keineswegs der einzig vertretbare. Eine Alternative stellen Erdfelder und Bredenkamp im 14. Kapitel dieses Bandes vor. Sie gehen ebenfalls davon aus, daß Validität mit der Beziehung zwischen einem empirisch gewonnenen Datensatz und einer allgemeineren Aussage zu tun hat, die als Hypothese aufgefaßt wird. Funktion des Datensatzes ist die Prüfung dieser Hypothese, die „streng“ und „fair“ sein soll. Validität ist bei Erdfelder und Bredenkamp eine Funktion von „Strenge“ und „Fairneß“. Diese beiden Konzepte werden in Kapitel 14 dieses Bandes genauer spezifiziert.

Wie steht es nun im hier angedeuteten Ansatz mit dem Unterschied zwischen interner und externer Validität? **Interne Validität** geht davon aus, daß die neuen Vergleichsfälle „unter Untersuchungsbedingungen“ realisiert werden, das heißt, daß alle nicht kontrollierten Variablen genau so variieren wie bei den Ausgangsbeobachtungen. Somit ist interne Validität in einer streng manipulierten Laborsituation, die nicht kontrollierte Variablen auf einen möglichst engen Variationsbereich einschränkt, sicher größer als bei alltagsnahen Beobachtungen in Situationen, in denen sehr viele Variablen breit variieren.

Faßt man die eingangs erwähnte Umschreibung interner Validität als möglichst eindeutiger Ergebnisinterpretation so auf, daß damit auch eine geringe Wahrscheinlichkeit unkontrollierter konfundierter Variablen gemeint ist, so ist dieser Aspekt durch den hier vorgestellten Präzisierungsentwurf nicht abgedeckt. Es ist wenig einsichtig, daß manipulative Kontrolle die Konfundierungswahrscheinlichkeit verringern sollte, sofern man nicht eine sehr spezifische Kontrollvariante unterstellt, die alle nicht registrierten Bedingungsaspekte auf geheimnisvolle Weise konstant zu halten versteht und so eine ideale Begründung für ceteris-paribus-Klauseln liefert. Wir werden bei Überlegungen

zur Komplexität des Phänomenbereichs der Psychologie auf diesen Punkt zurückkommen (siehe S. 26-30).

### **2.3.4 Laborbedingungen und Alltagssituationen**

Wird zur Prüfung **interner** Validität ein Vergleichsfall unter Untersuchungsbedingungen herangezogen, **so** geht es bei **externer Validität** um den Vergleich mit Daten aus anders gearteten Situationen, wobei im allgemeinen an Alltagssituationen gedacht ist. Dabei tauchen zunächst mindestens zwei Schwierigkeiten auf: Die unter Untersuchungsbedingungen realisierte Bedingungskonstellation kann in Alltagssituationen überhaupt nicht vorkommen, oder sie kommt vor, das Ergebnis entspricht **nicht** dem Validitätskriterium, ist aber dennoch mit den jeweiligen Untersuchungsbefunden vereinbar.

Der erstgenannte Fall liegt vor, wenn Versuchspersonen in einer Untersuchungssituation etwas tun, was sie im Alltag nie tun. Tachistoskopisch dargebotene Wörter zu erkennen ist im Alltag genau so unüblich wie ein vollständiger Paarvergleich oder eine weitgehend leere Landkarte, in die man mehr oder weniger gut bekannte Städte einzeichnet. Bedeutet dies, daß Befunde aus solchen Experimenten für praktische Entscheidungshilfen wertlos sind? Offensichtlich nicht. Man kann die Beschreibungsebene wechseln, also Darbietungszeiten im Tachistoskopversuch als Indikatoren erschwerter Wahrnehmbarkeit, Paarvergleiche als spezifische Konkretisierung von Entscheidungen bei wechselnden und überlappenden Optionenmengen, und leere Landkarten als Komponenten einer Indikatorsituation für subjektive Repräsentationen räumlicher Beziehungen auffassen. Solche Übertragungen von Aussagen von einem auf einen anderen Gegenstandsbereich unter Nutzung eines geeigneten Abstraktionsniveaus sind Analogieschlüsse. Zu einer Methodologie, die sich um bessere Aufklärung der Zusammenhänge zwischen **Forschungs- und Anwendungsmethodik** bemüht, gehört also sicher auch die Analyse von **Analogisierungen**.

Im zweiten Fall sind sowohl Verhaltensweisen der Versuchspersonen als auch Bedingungskonstellationen aus der Untersuchungssituation im Alltag wiederzufinden; dennoch sind die Kriterien externer Validität nicht erfüllt. Beim assoziativen Lernen weitgehend sinnlosen Materials im Labor findet man oft Daten, die den Angaben eines einfachen stochastischen Modells mit linearen Operatoren, dem sogenannten „a-Modell“ (Bush & Mosteller, 1951), gut entsprechen. Betrachtet man ähnliche Lernprozesse im Alltag, etwa das Lernen von Vokabeln einer noch weitgehend unbekanntes Sprache, dann sehen die Daten völlig anders aus. Hier gibt es Plateaus in den Lernkurven, die im Labor registrierte negativ beschleunigte Exponentialfunktionen nicht kennen. Dennoch wäre es aberwitzig zu behaupten, die in den Labordaten zum Aus-



druck kommenden psychischen Mechanismen seien im Alltag schlicht abwesend; stattdessen wird man Erklärungen suchen, denen zufolge andere Prozesse (im Beispielfall Superierung und verwandte Formen der kognitiven Reorganisation von Lernmaterial) zu Überlagerungen einer basalen Einprägungsmechanik führen, aus denen sich die beobachtbaren Abweichungen ergeben. Das auf Dingler (1928, 1955) zurückgehende und in der Psychologie vor allem von Holzkamp (1964) propagierte Verfahren des Exhaurierens, das ursprünglich eine Theorie durch Einführung „störender Bedingungen“ gegen unpassende Daten verteidigt, ist entsprechend verwendbar, um Daten aus einer manipulativ kontrollierten Laborsituation gegen Alltagserfahrungen zu verteidigen. Dabei darf man allerdings nicht übersehen, daß auch die Behauptung störender Bedingungen wiederum der empirischen Kontrolle bedarf.

Externe Validität ist somit kein Gütekriterium an sich; sie sagt lediglich etwas darüber, aus, wie weit von Befunden **direkt** auf Aussagen über Verhalten in Alltagssituationen geschlossen werden kann. **Anwendung von Wissenschaft auf Alltagsfragen** setzt voraus, daß man die jeweils zu problematisierende Situation auf Komponenten untersucht, zu denen wissenschaftliche Aussagen verfügbar sind, und daß man die so gewonnenen Aussagen derart in Beziehung setzt, daß eine partielle Repräsentation der Alltagssituation entsteht. Diese Repräsentation kann dann als Aussagensystem genutzt werden, aus dem Entscheidungshilfen ableitbar sind. Die derzeit übliche Psychologie hat oft einen überstark vereinfachend praxeologischen Zug, der wissenschaftliche Aussagen ohne Anpassung an die Eigenheiten der jeweiligen Alltagssituation direkt anzuwenden bemüht ist. Ein Physiker würde kaum untersuchen, wie weit er Ablösungszeitpunkt und Fallbahn eines Ahornblattes im Herbstwind vorherzusagen vermag, und auch in der Biologie umfaßt die Untersuchung des Verhaltens in freier Wildbahn nur einen recht geringen Teil der Forschungslandschaft. Viele Psychologen streben dagegen nach einer Forschung, deren wissenschaftliche Ergebnisse im Alltag nach Art einer Ahornblatt-Physik **direkt** anwendbar sind.

In diesem Zusammenhang ist auch die altherwürdige Kritik am **Experiment** zu erwähnen. Das Experiment setzt voraus, daß man recht genau weiß, welche Bedingungsvariation erforderlich ist. Das ist notwendig zur Realisierbarkeit manipulativer Kontrolle. Ist die Fragestellung einer Untersuchung derart vorstrukturiert, dann folgt meistens auch, welche Verhaltensaspekte zu registrieren sind. Das aber bietet die Möglichkeit zur Datenrestriktion. Hoher Manipulationsgrad und starke Datenrestriktion sind im psychologischen Experiment kombiniert; dies mag einer der Gründe dafür sein, daß sie in der Literatur nicht immer klar getrennt werden.

Hermann (1990) setzt sich mit Argumenten auseinander, die immer wieder gegen das Experiment in der Psychologie ins Feld geführt werden. Dabei zeigt

sich, daß solche Argumente oft erst dadurch möglich werden, daß man an das Experiment Forderungen stellt, die seinem eigenen (oft wesentlich bescheideneren) Anspruch nicht angemessen sind. Auch die Rede von der externen Validität ist von dieser Art; sie unterstellt eine direkte Anwendbarkeit von Befunden im Alltag, ohne nach den Methoden einer angemessenen Analyse von Alltagssituationen und einer entsprechenden Synthese wissenschaftlicher Befunde zu einem praktisch nutzbaren Aussagensystem zu fragen.

Wir lassen offen, wie man überhaupt von einer Substanztheorie ausgehend zu Aussagen kommen kann, die einer direkten empirischen Kontrolle zugänglich sind, wie man also empirische Daten in einem größeren wissenschaftlichen Zusammenhang verwenden kann. Hiermit beschäftigt sich Erdfelder im 2. Kapitel dieses Buches. Auch in der Psychologie zur Zeit eher ungewöhnliche Arten des Umgangs mit empirischen Daten sollen hier nicht genauer vorgestellt werden. Dazu gehören etwa Formen einer „mittleren Datenrestriktion“, die nicht die volle Spielbreite einer freien Interpretierbarkeit eröffnen, wohl aber unterschiedliche Auswertungsmöglichkeiten im Zusammenhang mit verschiedenen Fragestellungen. Die Konsequenz ist, solche Daten in einem „Corpus“ verfügbar zu machen, der für wechselnde Zwecke genutzt werden kann. Ein Beispiel hierzu liefert die als „Ulmer Textbank“ (Mergenthaler, 1985; Kächele, 1986) bekannte Sammlung psychotherapeutischer Gespräche. Ebenfalls (noch) ungewöhnlich ist eine Form des empirischen Zugangs, der in kognitionswissenschaftlich orientierten Arbeiten an Bedeutung gewinnt. Kaum restringierte verbale Daten (etwa aus lautem Denken oder aufgabenbegleitenden Befragungen) werden als Äußerungen interpretiert, die bestimmte Wissensinhalte voraussetzen. Die resultierenden Interpretate werden in einem stark restringierten Repräsentationsformat niedergelegt und als Daten weiterer Verarbeitung zugeführt, indem sie beispielsweise in die Datenbasis eines wissensbasiert agierenden Modells eingebaut werden. Opwis und Spada zeigen im 5. Kapitel dieses Bandes Möglichkeiten der Modellierung kognitiver Prozesse durch wissensbasierte Systeme, bei denen neben Experimenten herkömmlicher Art auch empirische Zugänge eine Rolle spielen, die interpretierende und datenförmige Ansätze miteinander verknüpfen.

Überlegungen zu den methodischen Konsequenzen der Auffassung von Psychologie als einer **empirischen Wissenschaft** haben wir mit der Unterstellung eingeleitet, **es** müsse **empirische Methoden** geben, deren sie sich bedient. Die Betrachtung formaler und normativer methodischer Ansätze erwies diese allerdings ebenfalls als unverzichtbar; aus der Empirizität der Psychologie folgt keine Beschränkung auf den ausschließlichen Einsatz empirischer Methoden. Der mit analytischen Methoden möglichen rationalen Erkenntnis haben wir dann allerdings im Kontext empirischer Wissenschaft einen **heuristischen** Status zugesprochen, bei dem Ergebnisse nicht als letztbegründet akzeptiert werden und daher einer Überprüfung bedürfen, für die **empirische Methoden** un-

verzichtbar sind. Zusammengefaßt: In der Psychologie als empirischer Wissenschaft ist der Einsatz empirischer Methoden nicht hinreichend, aber **unverzichtbar**. Die Vielfalt der Fragen nach Kriterien akzeptabler empirischer Methoden, die nicht zuletzt wegen der skizzierten zentralen Stellung der Empirie in der Psychologie besonderer Beachtung bedürfen, werden in diesem Kapitel nicht weiter verfolgt, da sie ohnehin Gegenstand der weitaus meisten nachfolgenden Beiträge dieses Bandes sind.

### **3. Komplexität**

Die für Psychologie relevanten methodologischen Fragestellungen haben oft damit zu tun, daß Psychologie als eine empirische Wissenschaft - in einem sehr breiten und für zahlreiche Varianten offenen Sinne des Wortes - aufgefaßt wird. Ihre besondere Tönung erhalten sie durch zusätzliche Spezifika, in denen sich Psychologie von anderen, ebenfalls empirischen Wissenschaften unterscheidet. Hierzu gehört unter anderem die oft beschworene große Komplexität des Phänomenenbereiches.

Außerhalb der Komplexitätstheorie, bei der es um den zur Berechnung berechenbarer Funktionen bzw. zur Entscheidung entscheidbarer Aussagen erforderlichen Aufwand geht (siehe Paul, 1987), ist die Bedeutung von „Komplexität“ vielfältig und unklar. De Groot (1969, S. 327-342) analysiert die Komplexität der zu untersuchenden **Phänomene** wie auch der **Instrumente** (Methoden, Theorien, Modelle), derer sich die Psychologie bedient. Breuer (1977, S. 20) spricht von der „Komplexität der Systembedingungen“ des **Objekts** der Psychologie, die diese unangemessen vereinfache, und nennt Beispiele für derartige Vereinfachungen ohne zu explizieren, was denn Komplexität nun sein möge. Die Endlichkeit menschlichen Wissens - und der für wissenschaftliche Publikationen verfügbaren Seitenzahlen - bringt stets Produkte hervor, die nicht alle Aspekte und Möglichkeiten berücksichtigen, die zur Lösung eines Problems oder zur Analyse eines Sachverhaltes hätten herangezogen werden können. Mangelnde Berücksichtigung von Komplexität ist so ein stets wohlfeiles Argument gegen den jeweils anderen.

#### 3.1 Komplexität von Systemen

Simon (1977) und auch Nelson (1977) bemühen sich um eine Präzisierung der Rede von der Komplexität eines **Systems**. Danach wächst Komplexität mit der Anzahl der Systemkomponenten, mit der Anzahl und dem Ausmaß der Verschiedenheit unterscheidbarer Komponentenklassen und mit der Zahl der Relationen zwischen ihnen. Aus dieser Sicht kommt Komplexität nur einem

System zu, das Ergebnis der Analyse einer Gegebenheit auf einem bestimmten Auflösungsgrad ist; es geht also in der Psychologie weniger um die Komplexität des Menschen als vielmehr um die eines angemessenen Menschenbildes.

### **3.1.1 Komplexitätsvergleich**

Diese Konzeption von Komplexität ist zunächst komparativ; man kann lediglich unter bestimmten Bedingungen bei zwei Systemen ausmachen, welches **komplexer als** das andere ist. Nelson (1989, S.61-73) verdeutlicht dies am Beispiel von „Peter the sleeper“. Peter ist ein relativ einfach funktionierendes Wesen. Es gibt für ihn nur zwei Zustände: Entweder schläft er, oder er ist wach. Es gibt zwei Reize, auf die er reagiert: das Klingeln eines Weckers und ein Stoß zwischen die Rippen. Die Kombinationen ergeben vier für ihn relevante Situationen: „Klingeln mit Stoß“, „Klingeln ohne Stoß“, „Stoß ohne Klingeln“, „weder Klingeln noch Stoß“. Wenn er schläft, kann ihn nur ein Stoß zwischen die Rippen in den Wachzustand überführen. In diesem Fall reagiert er mit Gähnen, sofern er nur einen Rippenstoß erhalten hat, während er beim gleichzeitigen Weckerklingeln zuschlägt. Ist er wach, dann gähnt er und schläft prompt wieder ein, sofern er weder gestoßen noch durch Weckerrasseln wachgehalten wird. Unter allen anderen Bedingungen bleibt er wach und reagiert mit irgendwelchen streng festgelegten Kombinationen seiner Möglichkeiten „Zuschlagen“ und „Gähnen“.

„Peter the sleeper“ ist ein deterministischer endlicher Automat. Damit zumindest einige von Peters Verhaltensmöglichkeiten zum Tragen kommen können, muß er einen Wecker erhalten, der klingeln kann, und einen Zimmergenossen, der ihm einen Rippenstoß versetzen kann. Zwischen den drei Komponenten des so entstehenden Superautomaten müssen bestimmte Beziehungen gegeben sein. man könnte etwa daran denken, daß der Schlag, den Peter manchmal ausführt, den Zimmergenossen trifft, sofern dieser überhaupt im Zimmer ist, und ähnliches mehr. Die Kombination aus Peter, dem Wecker und dem Zimmergenossen ist immer noch ein endlicher deterministischer Automat, aber von höherer Komplexität als „Peter the sleeper“, da neue Komponenten anderer Art und vorher überhaupt nicht mögliche Relationen zwischen den drei kombinierten Komponenten hinzugekommen sind.

### **3.1.2 Komplexitätsreduktion**

Ein solcher Aufbau immer komplexerer Systeme durch Kombination vorhandener Komponenten dient hier lediglich der beispielhaften Erläuterung des Komplexitätskonzepts. Für das Vorgehen einer empirischen Wissenschaft ist

der umgekehrte Weg interessanter. Empirisch vorfindbare Phänomene werden erklärt, indem man das zugrundeliegende System als ein Gefüge aus Komponenten auffaßt. So wird der menschliche Geist zu einer Struktur aus sensorischen Puffern, diversen Lang- und Kurzzeitgedächtnissen, motorischen Programmen, Motivsystemen, Zielstrukturen, Bewertungsfunktionen, Entscheidungseinheiten, und ähnlichem mehr. Erst in der Darstellung als Kombination von weniger komplexen Teilsystemen erhält er im Vergleich mit eben diesen Teilsystemen seine oft beschworene „Komplexität“.

Palmer und Kimchi (1986) betrachten eine solche „Komplexitätsreduktion“ als charakteristisch für jeden Ansatz, der den Menschen als informationsverarbeitendes System auffaßt. Hier wird eine Operation, beispielsweise die Beantwortung einer Frage, als ein Gefüge aus Teiloperationen (syntaktische Analyse, Worterkennung, semantische Analyse, verschiedene Formen des Gedächtniszugriffs, Operationen der Zielverwaltung, Bildung einer Antwortabsicht, Umsetzung in Sprache, Realisierung eines motorischen Programms und so weiter) dargestellt; die Relationen zwischen diesen stehen für Parallelität und Sequentialität, Unterbrechungen, Reaktivierungen und ähnliches mehr. Selbstverständlich kann man auch Teiloperationen wieder in Komponenten zerlegen und so weiter und so fort. Auf diese Weise entsteht eine Kette von Komplexitätsreduktionen. Sie beginnt mit dem zu analysierenden Phänomenbereich, über dessen Komplexität wir zunächst nicht aussagen, und führt zur Reduktion auf immer mehr und immer einfachere Komponenten, deren größere Einfachheit nichts weiter widerspiegelt als das triviale Vergleichsergebnis, daß sie gegenüber dem Ausgangsphänomen als weniger komplex auszeichnet.

Eine solche Sicht der Dinge legt die Frage nahe, ob es „einfachste“ Operationen gebe, bei denen die Komplexitätsreduktion endet. Ein hemmungsloser Reduktionist wird vielleicht darauf hinweisen, daß alle geistigen Prozesse einer neurophysiologischen Basis bedürfen, daß somit jeder Fortschritt der Neurophysiologie ein Schritt näher zum Ende der Psychologie sei, daß Neurophysiologie ihrerseits Vorgänge analysiere, die sich biochemisch erklären lassen, daß Biochemie nur ein Teilbereich der Chemie schlechthin sei, daß deren Aussagen im Endeffekt auf Gesetze der Teilchenphysik reduzierbar seien und daß jede konsequent fortgesetzte Komplexitätsreduktion schließlich hier anlangen müsse. Ernsthaft vertreten wird ein solcher rigoroser Reduktionismus kaum; das reduktionistische Programm ist unrealistisch, weil es nicht nur einen sehr hohen Stand der jeweiligen Basiswissenschaft voraussetzt, sondern auch die Beschäftigung mit den Gefügen von Relationen zwischen Komponenten, die zum Gegenstandsbereich der Basiswissenschaft gehören. Psychologie könnte erst dann durch Neurophysiologie ersetzt werden, wenn diese bestimmte Klassen neuronaler Aktivitätsmuster (von denen die eine vielleicht die Klasse der möglichen physiologischen Implimentationen dessen ist, was wir „Fru-

stration“ nennen) daraufhin untersucht, welche Effekte die Anwesenheit eines Elements einer solchen Klasse auf die Realisierung eines Musters hat, das zu einer anderen Klasse gehört (deren Elemente beispielsweise für individuelle Varianten von „Aggression“ stehen). Dies aber wurde im Endeffekt bedeuten, daß eine derartige Neurophysiologie auch ursprünglich psychologische Konzepte untersuchen wurde; es läge keine Reduktion, sondern lediglich eine neuartige institutionelle Verknüpfung vor.

Diese Argumentation basiert auf der Vorstellung, daß bei (gegenüber ihren Komponenten) komplexeren Systemen Regelhaftigkeiten auftreten können, bei denen Relationengefüge auf Komponenten-Teilmengen relevant sind. Ein solches Relationengefüge ist dann eine „emergente Eigenschaft“ (Nagel, 1961, S. 366-380), auf die nicht verzichtet werden kann, wenn die in Frage stehende Regelhaftigkeit darstellbar sein soll. Damit aber mißlingt das radikale reduktionistische Programm.

Die Frage nach dem möglichen Ende einer Komplexitätsreduktion ist so zu modifizieren, daß die Möglichkeit kleinster Einheiten bei Systemdarstellungen zu erkunden ist, die im Rahmen einer Wissenschaft oder einer Teildisziplin bleiben. Für das Gemischtunternehmen „Psychologie“ liegt hierzu kein überzeugender Antwortentwurf vor. Nun macht zur Zeit innerhalb des Faches der „kognitive“ Ansatz immer mehr von sich reden, dessen fachübergreifende Komponente einer allgemeinen „Kognitionswissenschaft“ (siehe etwa Posner, 1989) zuzurechnen ist. Kognitionswissenschaft stellt den Anspruch, intelligentes Verhalten mit den dazu erforderlichen Strukturen und Prozessen so darzustellen und zu analysieren, daß die über verschiedene konkrete Realisierungen hinweg invarianten Aspekte erfaßt werden. Dabei ist bei den Realisierungen vorwiegend an intelligentes Verhalten von Menschen einerseits und das ebenfalls als „intelligent“ bezeichnete Verhalten von Rechnern andererseits gedacht. Die übergreifend abstrahierende Struktur ist die **kognitive Architektur**, die menschen- und/oder maschinenspezifische Realisierung deren **Implementation**. Diese Differenzierung zwischen kognitiver Architektur und Implementation macht es möglich, nach „kleinsten Einheiten“ zu fragen, die noch nicht Implementationskomponenten sind. Dabei läßt sich diese Frage verschieden wenden. Betrachtet man intelligente Systeme als wissensbasierte Agenten (wie von Opwis und Spada in Kap.5 dieses Bandes genauer dargestellt), so geht es um Eigenschaften **elementarer Wissenseinheiten**. Zerlegt man (wie Palmer und Kimchi, 1986) informationsverarbeitende Operationen in Komponenten-Operationen, **so** geht **es** um **elementary information processes (eips)**’ sensu Newell und Simon (1972).

## 3.2 Wahrscheinlichkeiten, Fehler, Störvariablen

Im alltäglichen Wissenschaftsbetrieb wird man bei einer systemorientierten Betrachtung die Analyse durch Einführung von Komponentengefügen irgendwann beenden, wohl wissend, daß man weitermachen könnte. Breite Bereiche psychologischer Forschung denken überhaupt nicht an eine systemorientierte Komplexitätsreduktion, sondern betrachten Abhängigkeiten und Zusammenhänge zwischen empirisch manipulierbaren und/oder registrierbaren Variablen, ebenfalls wohl wissend, daß eine feinere Analyse möglich wäre. Dies zeigt eine andere Variante des Komplexitätsproblems. Was auch immer in einer Untersuchung oder einem Forschungsprogramm erfaßt und berücksichtigt wird, es besteht stets Grund zu der Annahme, daß die beobachteten und zu erklärenden Phänomene auch von anderen nicht-kontrollierten Gegebenheiten abhängen oder modifiziert werden, die erst auf einem feineren Auflösungsgrad faßbar wurden. Es gilt also, die Möglichkeit unbekannter relevanter Variablen zu berücksichtigen. Hier kommen Konzeptionen von Wahrscheinlichkeit, Zufall und Fehler ins Spiel.

### 3.2.1 Störvariablen

Betrachten wir - ohne Verlust der Allgemeinheit und lediglich zur Vereinfachung - eine Untersuchungssituation mit nur **einer** reellwertigen abhängigen Variablen. Wir untersuchen, welche Werte diese Variable annimmt, wenn die Werte einer bestimmten Kombination unabhängiger Variablen systematisch variiert werden. Ohne allzu überrascht zu sein, stellen wir fest, daß bei wiederholter Realisierung der gleichen Wertekombination der unabhängigen Variablen verschiedene Werte der abhängigen Variablen resultieren. Wir führen dies auf den Einfluß **nicht-kontrollierter Variablen** zurück, deren Existenz angesichts der Komplexität, die sich bei weiterer komplexitätsreduzierender Analyse des untersuchten Systems ergeben würde, plausibel ist. Dazu nehmen wir an, daß die aus der skizzierten Variation resultierende Varianz, die wir gern „Fehlervarianz“ nennen, durch Kontrolle zusätzlicher relevanter Variablen verringert werden kann. Letztendlich nutzen wir diese Überlegung methodisch zur Entwicklung von Versuchsplan-Techniken und zur Einführung korrespondierender statistischer Verfahren der Datenauswertung.

Bei einer solchen Art von Begründung statistischer und versuchsplan-technischer Methoden, die von Rehm und Strack im Kapitel 12 dieses Bandes ausführlicher vorgestellt werden, ist zu berücksichtigen, daß die oft als **Störvariablen** bezeichneten unbekannt relevanten Variablen sowohl bei den Beschreibungsgrößen des menschlichen Geistes zu suchen sind wie auch im System der uns umgebenden äußeren Welt. Erstere versucht man gern durch

Versuchspersonenauswahl oder Erhebung als Kovariate zu kontrollieren, letztgenannte durch Manipulation.

Existieren Störvariablen, dann ist bei gleicher Wertekombination der kontrollierten unabhängigen Variablen mit unterschiedlichen Werten der abhängigen Variablen zu rechnen. Gleiche Werte der abhängigen Variablen bei gleicher Wertekonstellation der unabhängigen Variablen kommen kaum noch vor und können für die Einführung eines Validitätskriteriums nicht mehr verwandt werden. An ihre Stelle tritt die **Unterschiedlichkeit** zwischen den Werten abhängiger Variablen bei Realisierung **verschiedener** Werte der unabhängigen Variablen. Könnten wir Paare von Untersuchungssituationen herstellen, bei denen mit Sicherheit alle nicht kontrollierten Bedingungsaspekte - also auch die Werte aller Störvariablen - gleich wären, während sich die Werte der kontrollierten unabhängigen Variablen in bekannter Weise unterscheiden, so wäre zu fordern, daß die resultierenden **Unterschiede** auf der abhängigen Variablen - also die „**Effekte**“ kontrollierter Variation - stets gleich sind. Dies ist eine sehr strenge Version der Vorstellung, daß alle Aussagen über empirisch registrierbare Effekte stets unter dem Vorbehalt stehen, alles nicht kontrollierte Sonstige (cetera omnia) sei gleich (pars). Diese - oft stillschweigende - Zusatzannahme ist die ‚ceteris paribus‘-Klausel in ihrer wohl strengsten Variante.

Nun leidet auch diese Vorstellung darunter, daß sie in vielen Fällen weder realisierbar noch überzeugend approximierbar ist. Es bietet sich daher an, den Begriff des „Sonstigen“ etwas anders zu fassen, und an die Stelle der Gleichheit nicht-kontrollierter Bedingungen die schwächere Forderung nach deren gleicher Verteilung zu setzen. An dieser Stelle braucht man Zusatzannahmen, aus denen prüfbare Aussagen über die als „Fehler“ bezeichneten Abweichungen registrierter Werte vom Mittelwert bei der jeweiligen kontrollierten Bedingungskombination ableitbar sind. Die strengste Möglichkeit wäre, den kombinierten Effekt aller Störvariablen auf die abhängige Variable als unabhängig vom und additiv zum kombinierten Effekt der kontrollierten Bedingungsaspekte anzunehmen. Dann wären die Fehler über alle kontrollierten Kombinationen hinweg stets gleich verteilt. Selbstverständlich sind auch schwächere - und in vielen Fällen realistischere - Annahmen möglich. Weitere Ausführungen zu sinnvoll probabilisierten ceteris paribus-Varianten und daraus resultierenden Fehlermodellen findet man bei Steyer im Kapitel 15 dieses Bandes.

### **3.2.2 Konfundierung und Effektunterdrückung**

Zwei Fragen, die in diesem Zusammenhang bislang offen geblieben sind, seien kurz angesprochen. Erstens: Es ist nie auszuschließen, daß auch schwache Forderungen nach Unabhängigkeit zwischen Störvariablen und kontrollierter



Bedingungsvariation nicht erfüllt sind. Die Differentielle Psychologie ist voll von Nachweisen über Korrelationen zwischen menschlichen Eigenschaften. Wird also bei einer psychologischen Untersuchung Intelligenz, Ängstlichkeit, Neurotizismus, Ich-Stärke oder was auch immer durch entsprechende Probandenauswahl kontrolliert, so müssen wir damit rechnen, daß entsprechend den Korrelationen anderer Persönlichkeitsvariablen mit den kontrollierten Größen abhängig von der jeweils realisierten Bedingungsvariation diese anderen Variablen unterschiedlich verteilt sind. Sind sie - was zunächst unbekannt ist - als Störvariablen wirksam, kann jeder gefundene Effekt statt auf die kontrollierten Größen auf eine Kombination solcher **konfundierter Variablen** zurückgeführt werden. Es wurde bereits erwähnt (S.22), daß auch die Abwesenheit konfundierter Störvariablen als eine der Forderungen interner Validität angesehen wird. Die Erfüllung dieser Forderung kann allerdings nicht effektiv festgestellt werden, weil dazu die Nicht-Existenz einer konfundierten Störvariablen in einer offenen Population definierbarer Bedingungskomponenten nachzuweisen wäre.

Die zweite Frage: Die gesamte Konzeption der „Fehler“ und „Störvariablen“ basiert auf der Annahme, daß nicht kontrollierte Aspekte einer Bedingungsvariation - sofern sie einen Einfluß auf die jeweils abhängigen Variablen haben - stets zu einer **Varianzerhöhung** führen, wobei die Störvariablen mit **Varianzquellen** identifiziert werden. Es ist aber durchaus denkbar, daß nicht-kontrollierte Bedingungsaspekte beispielsweise zu einem Korrekturmechanismus führen, der unter anderen Bedingungen manifeste Effekte überdeckt. So könnte etwa die oft gescholtene Frustrations-Aggressions-Hypothese durchaus richtig sein; in allen Untersuchungen, in denen Daten gegen sie sprechen, führen irgendwelche nicht identifizierten Aspekte, die etwa mit Angst vor den Folgen einer aggressiven Handlung, mit Vorstellungen vom sozial erwarteten Verhalten, mit der Überlagerung durch andere Motivsysteme oder mit ähnlichem zu tun haben, dazu, daß eine durchaus existente frustrationsbedingte Aggressionstendenz in den registrierten Daten nicht manifest wird. Um Mißverständnisse zu vermeiden: Ich behaupte nicht, daß dies so sei, sondern lediglich, daß wir eine solche Möglichkeit im Kontext der Fehler/Störvariablen-Methodik überhaupt nicht berücksichtigen.

Es wäre daher auf den zweiten Blick gar nicht so abwegig, eine statistische Methodik zu entwickeln, deren Verfahren nicht - wie bei der üblichen Signifikanzstatistik - der Entscheidung über eine Hypothese dienen, sondern vielmehr der Entscheidung darüber, ob **fehlerhafte Daten** vorliegen, bei denen störende Variablen nicht zur Varianzerhöhung, sondern zur **Effektüberdeckung** geführt haben. Tradition hätte ein solches Vorgehen. Frühe Anwendungen statistischer Techniken in der Astronomie dienten nicht der Entscheidung über Theorien oder Hypothesen, sondern der Kontrolle der Brauchbarkeit einzelner Daten (Swijtink, 1987). War ein Datum aufgrund der

Erwartung extrem unwahrscheinlich, wurde nicht Erwartung, sondern das Datum als fehlerhaft verworfen.

### **3.2.3 Begründung probabilistischer Konzepte**

Man kann davon ausgehen, daß probabilistische Methoden genutzt werden, um mit mangelndem Wissen über Systemkomponenten fertig zu werden, die bei gegebenem Auflösungsgrad der Analyse nicht bekannt sind, mit deren Existenz aber zu rechnen ist. Man kann Wahrscheinlichkeiten als einen Ausdruck partieller Informiertheit benutzen. Wenn wir die Wahrscheinlichkeit der Entnahme einer roten Kugel aus einer Urne mit 50 roten und 50 schwarzen Kugeln als  $\frac{1}{2}$  bezeichnen, setzen wir voraus, daß die Person, für die diese Wahrscheinlichkeit gilt, nicht in die Urne hineinschaut und sieht, daß alle schwarzen Kugeln unten und alle roten Kugeln oben liegen, denn dann wäre - bei Entnahme von oben - die Wahrscheinlichkeit einer roten Kugel wohl besser bei 1 anzusiedeln. Wahrscheinlichkeiten werden also als Relationen zwischen Sachverhalten und Informiertheiten interpretiert.

Diese Sichtweise bezeichnen Gigerenzer und Murray (1987, S.xii) als die **epistemische Interpretation** probabilistischer Ideen. Zufalls- und Wahrscheinlichkeitsannahmen bilden Grenzen unserer Erfahrung ab. Diese Sicht sollte aber keinesfalls dahingehend mißverstanden werden, als seien alle Zufallsvorstellungen in Theorien und Methoden der Psychologie ausschließlich aus der Diskrepanz zwischen Komplexität und eingeschränkter Erfahrung abzuleiten. Man kann auch die Meinung vertreten, es gebe Prozesse, die als nicht weiter reduzierbare natürliche Phänomene inhärent probabilistisch sind, also auch ohne unsere bekannten Wissenslücken nicht anders als durch Wahrscheinlichkeitsangaben beschrieben werden können. Dies ist die **ontische Interpretation** probabilistischer Konzepte.

Schließlich und endlich gibt es auch noch Wahrscheinlichkeitskonzepte, die Relationen zwischen empirischen Aussagen über endlich viele Fälle und allgemeinen Hypothesen, Annahmen und Gesetzen beschreiben, die die Form von Allaussagen über offenen Populationen haben. Solche Wahrscheinlichkeiten haben wenig mit der Beschreibung von Phänomenen oder einem begrenzten Wissen darüber zu tun, um so mehr aber mit Formen und Möglichkeiten induktiver Folgerungen (siehe Westermann und Gerjets in Kap. 10 dieses Bandes). Hierher gehört auch die Verwendung probabilistischer Konzepte bei der statistischen Entscheidung, deren Varianten in diesem Band von Ostmann und Wutke (Kap. 16) dargestellt und analysiert werden.

Komplexität mit den methodischen Konsequenzen der systemorientierten Komplexitätsreduktion und der Verwendung probabilistisch begründbarer Methoden sind nicht die einzigen Psychologie-Spezifika, die eine fachspezi-

fische Allgemeine Methodenlehre berücksichtigen muß. Da gibt es etwa noch die Frage nach den Kriterien der Angemessenheit einer Methode an eine Fragestellung, die ihre Tönung durch die Eigenschaften des Bereichs der Psychologie-üblichen Forschungsprogramme erhält, oder die Frage nach der Eindeutigkeit von Befundaussagen, die analog dem Eindeutigkeitskonzept der Meßtheorie behandelt werden kann, mit dem sich in diesem Buch Mausfeld (Kap. 13) auseinandersetzt. Aus spezifischen Teilbereichen des Faches ergeben sich weitere methodologische Fragestellungen. So hat die Entwicklungspsychologie mit der Abgrenzbarkeit und empirischen Identifizierbarkeit verschiedener Arten von Entwicklungsverläufen zu tun und die Differentielle Psychologie mit der oft nicht unbeträchtlichen Schwierigkeit der Entscheidung darüber, ob ein Befund (nur) Aussagen über Populationen oder (auch) solche über Individuen erlaubt. Sind Faktoren der Intelligenz Beschreibungsgrößen einer Kovarianzstruktur in einer Population oder sind es auch Variablen, die für irgendetwas bei jedem Individuum stehen? Beschreibt die Differenzierungshypothese der Intelligenz eine altersabhängige Veränderung einer Variabilitätsstruktur in einer Population oder sagt sie etwas über irgendwelche Prozesse aus, die in jedem einzelnen ablaufen?

Solche und ähnliche Fragen liefern weitere Beispiele für Themen einer Allgemeinen Methodenlehre der Psychologie. Derartige Überlegungen sollen hier nicht weiter verfolgt werden; stattdessen zeigen wir am Beispiel der Themen dieses Enzyklopädie-Bandes Zusammenhänge auf und skizzieren eine erste Ordnung.

## ***4. Themen Allgemeiner Methodenlehre***

In den bisherigen Überlegungen und Ausführungen wurde versucht, Allgemeine Methodenlehre der Psychologie nicht durch eine - kaum sinnvoll mögliche - strenge Definition, sondern vielmehr durch Verweis auf möglichst typische Themen und Fragestellungen nach außen abzugrenzen, und zwar einerseits gegenüber einer Spezifischen Methodenlehre des Faches und andererseits gegenüber einer zwar Allgemeinen Methodenlehre, die aber keine solche der Psychologie ist. Es stellt sich nun die Frage nach möglichen Binnengliederungen einer Psychologie-Methodologie, die zumindest Arbeits- und Aufgabenbereiche aufzeigt.

### **4.1 Zugänge**

Ein Blick auf die aktuelle Psychologie zeigt, daß es eine Vielzahl von Zugängen zum Gegenstands- und Phänomenbereich, von Forschungsmöglichkeiten und -Varianten gibt. Eine Suche nach der „richtigen Art, Psychologie zu be-

treiben“ (Grawe, Hänni, Semmer & Tschan, 1991) ist vergebliches Bemühen. Wir finden eine Vielzahl von Varianten psychologischer Forschung, die sich sowohl in den jeweils untersuchten Fragestellungen und Phänomenbereichen als auch in ihren Methodenarsenalen mehr oder weniger deutlich unterscheiden.

Allgemeine Methodenlehre kann solche Ansätze psychologischer Forschung darstellen und untersuchen, welche Themen und Fragestellungen jeweils mit welchen Methoden auf der Basis welcher Vorannahmen bearbeitet werden. Analog zur Abgrenzung einer Allgemeinen Methodenlehre der Psychologie nach außen erhält man so gewissermaßen eine Menge gegeneinander abgrenzbarer Allgemeiner Methodenlehren von XY-Psychologien. Hierauf können Arbeiten aufbauen, die etwa über die XY-Psychologien hinweg Korrespondenzen zwischen Fragestellungen, Rahmenkonzepten und Methodenvorräten analysieren. Teil B dieses Bandes mit den Kapiteln 2 bis 5 zeigt Spezifika einiger Zugänge der Psychologie zu ihren Gegenständen, die hier als Beispiele ohne jeden Anspruch auf Vollständigkeit dienen.

Erdfelder (Kap. 2) skizziert die Grundlagen eines Ansatzes, der **empirische Daten** generiert und nutzt. Dabei versteht er als ein „Datum“ nicht - wie bislang in diesem Kapitel - jede Aussage, von der allgemein unterstellt wird, sie gebe ein Beobachtungsergebnis wieder und für die keine weitere Begründung gefordert wird. Er unterscheidet vielmehr zwischen „Beobachtungsprotokollen“ und „Daten“ in einem engeren Wortsinne, die das Ergebnis der Abbildung von Beobachtungen in einen **Datenraum** sind. Dieser hinwiederum ist eine Menge möglicher Daten, die durch einen Satz von Variablen beschrieben werden kann, von denen jede bei jedem Einzeldatum mit einem Wert belegt ist. Welche Variablen dabei berücksichtigt werden, hängt von der jeweiligen Fragestellung ab, zu deren Beantwortung Daten erhoben werden. Daten sind bei Erdfelder stets soweit restringiert, daß sie möglichst genau die zur Beantwortung einer Fragestellung notwendige Information enthalten,

Von einem solchen „datenförmigen“ Vorgehen deutlich unterschieden sind **interpretierende Ansätze**, mit denen sich Soeffner und Hitzler im 3. Kapitel dieses Bandes beschäftigen. Das empirische Material besteht hier aus wenig restringierten Daten, die oft als (sprachliche wie auch nicht-sprachliche) „Dokumente“ bezeichnet werden. Der Vergleich zwischen dem 2. und dem 3. Kapitel macht deutlich, daß die Entscheidung für unterschiedliche Möglichkeiten der Datenrestriktion eng zusammenhängt mit verschiedenen Zielen wissenschaftlicher Forschung. Datenförmige Empirie, die Daten im Begründungszusammenhang verwendet, setzt „... ein psychologisches Gesetz, eine Hypothese, eine technologische Prognose oder eine psychologische Theorie ...“ (Erdfelder, Kap. 2 in diesem Band) voraus, die etwas über bestimmte Aspekte möglicher Daten aussagt. Damit kann der Datenraum auf eben diese Aspekte

restringiert werden. Bei Verwendung von Daten im Entdeckungszusammenhang gibt es „offene Fragestellungen“, in denen ebenfalls bestimmte Qualitäten beobachtbarer Sachverhalte vorkommen, auf deren Berücksichtigung der Datenraum restringiert werden kann. Sinnvolle Datenrestriktion wird also dadurch möglich, daß spezifizierte offene oder geschlossene Fragen Ausgangspunkt von Forschungen sind. Interpretation zielt dagegen auf Verstehen ab, auf die Entfaltung einer möglichst breiten Vielfalt von Sinn-Potentialen. Ausgangspunkt ist keine vorformulierbare Fragestellung, die einen Variablenrahmen vorgibt, sondern ein Dokument, für das Deutungsmöglichkeiten erschlossen werden sollen. Eine solche Forschungsaufgabe bietet nicht den geringsten Ansatzpunkt für eine wie auch immer geartete Datenrestriktion.

Zu den Spezifika der Psychologie gehört sicherlich, daß sie sich mit Beziehungen zwischen (äußeren) wahrnehmbaren Gegebenheiten und (inneren) Wahrnehmungsinhalten sowie zwischen (innerem) Erleben und (äußerm) Verhalten beschäftigt, also in der einen oder anderen Weise mit dem Leib-Seele-Problem befaßt ist. Weder datenförmiges noch qualitativ-interpretierendes Vorgehen sind Forschungsansätze, die spezifisch auf Fragen des Verhältnisses zwischen einem Innen und einem Außen abgestellt sind; dies gilt aber sehr wohl für die **Psychophysik**, die Mausfeld im 4.Kapitel skizziert. Zwei Fragenkomplexe charakterisieren Psychophysik: die Erforschung von Mechanismen der Wahrnehmung und die Frage nach der Meßbarkeit des Psychischen. Zum Methodenarsenal gehören Techniken der Darstellung und Analyse von Mechanismen der Wahrnehmung als Systeme, auf denen informationsverarbeitende Prozesse laufen. Eine solche Konzeption setzt voraus, daß eine zwischen einem Innen und einem Außen differenzierende Vorstellung gegeben ist, von der ausgehend die in physikalisch beschreibbaren (Reiz-)Gegebenheiten repräsentierte Information in ein kognitiv beschreibbares Format transformierbar wird. Die Frage nach der Meßbarkeit des Psychischen kann dann als Frage nach einer bestimmten Form der Repräsentierbarkeit mentaler Gegebenheiten im kognitiven Modell interpretiert werden.

Psychophysik liefert zum einen ein Beispiel für die Beschäftigung mit deutlich Psychologie-spezifischen Fragestellungen und daraus resultierende methodischen Konsequenzen, zum anderen schafft sie mit ihrem Anspruch nach Repräsentierbarkeit mentaler Gegebenheiten einen Übergang zum Kapitel 5 von Opwis und Spada über Möglichkeiten der Modellierung mit Hilfe **wissensbasierter Systeme**. Auch hier wird die Innen-Außen-Differenzierung thematisiert. Menschen verfügen über eine „interne, subjektive Repräsentation von Ausschnitten ihrer tatsächlichen oder einer vorgestellten Umgebung“ (Opwis und Spada, Kap. 5 dieses Bandes, Abschnitt 2). Die Frage ist, wie eine solche interne Repräsentation durch eine Symbolstruktur (als Konstruktion zweiter Ordnung) repräsentiert werden kann, auf der ein formales kognitives System operiert. Dabei soll zwischen dem Verhalten dieses artifiziellen Systems und

menschlichem Verhalten eine Analogiebeziehung bestehen, die es gestattet, das formale wissenschaftliche System als Modell des (natürlichen) mentalen Systems aufzufassen.

Diese vier Beispiele für Zugänge der Forschung zum Gegenstandsbereich der Psychologie mit unterschiedlichen Methodeninventaren lassen zumindest einige Zusammenhänge zwischen Forschungsansätzen und Methoden deutlich werden. Die Annahme, am Anfang eines Forschungsprozesses stehe eine bestimmte Art von Fragestellung (Kap. 2), legt eine angemessene Datenrestriktion nahe, während die Offenheit gegenüber einer Vielfalt von Deutungsmöglichkeiten (Kap. 3) sie geradezu ausschließt. Bestimmte Spezifika der Psychologie wie (in diesem Fall) ihre Beschäftigung mit einer Innen/Außen-Differenzierung spielen bei bestimmten Ansätzen keine Rolle (Kap.2), werden bei anderen teilweise relevant (Kap. 3) und rücken schließlich bei wieder anderen ins Zentrum der behandelten Fragestellungen (Kap.4 und 5). Die jeweils interessierenden Aspekte des (inneren) mentalen Systems werden als Variablen-Systeme (Kap. 4) oder als Symbolstrukturen in einem Wissensrepräsentations-Format (Kap. 5) erfaßt. Teil B dieses Bandes deckt so zumindest einen Teil der Variation ab, die Varianten des forschenden Zugangs zu psychologisch relevanten Phänomenen aufspannen.

## **4.2 Theorien und Programme**

Die Darstellung und methodologische Analyse einzelner Forschungsansätze verstellt allzu leicht den Blick auf Aspekte, die den verschiedenen Zugängen - oder zumindest einer großen Mehrheit davon - gemeinsam sind. Allgemeine Methodenlehre erfordert mithin einen zweiten Themenbereich, der solche Gemeinsamkeiten behandelt und in diesem Band im Teil C exemplifiziert wird.

Hierher gehört zumindest das, womit sich wissenschaftliche Forschung in doppeltem Sinne beschäftigt: Einmal, indem sie es zum Ausgangspunkt ihrer Bemühungen nimmt, und zum anderen, indem sie es durch ihre Bemühungen bewertet, verändert oder gar neu konzipiert. Einer weit verbreiteten Vorstellung zufolge ist dieses „Etwas“ die Theorie. Forschung dient oft der Bewertung von Theorien; Forschung verändert häufig vorhandene Theorien oder führt gar zu einer Datenlage, die Neukonzeptionen nahelegt oder zumindest anregt. Liest man allerdings Artikel, in denen einzelne Untersuchungen dargestellt werden, so findet man oft einleitend statt eines Verweises auf eine spezifische zu prüfende Theorie Hinweise auf eine Art Thema und auf einige andere Untersuchungen, die ebenfalls diesem Thema zugeordnet sind und dabei ähnliche Konzeptionen benutzen wie der Verfasser der jeweils vorliegenden Arbeit. Untersuchungen werden damit eingeordnet in ein jeweils von mehreren Wissenschaftlern benutztes System von Annahmen über einen Ge-

genstandsbereich, in dem spezifizierbare und noch nicht beantwortete Fragen möglich sind. In Kapitel 6 dieses Bandes werden solche Systeme von Herrmann genauer abgegrenzt, als „**Forschungsprogramme**“ bezeichnet, vorgestellt und analysiert.

Bei der Identifikation konkreter spezifischer Forschungsprogramme geht es nicht ohne die Festlegung einer Korngröße. Bezeichnet man etwa die Forschungen zur Modularität des menschlichen Gedächtnisses (Engelkamp, 1990) als ein Forschungsprogramm, so ließe sich - bei größerer Granularität - die gesamte kognitiv orientierte Gedächtnispsychologie sicherlich auch als ein solches auffassen. Prinzipiell spricht nichts dagegen, jeden der im Teil B dieses Bandes dargestellten Ansätze oder sogar die gesamte Psychologie ein „Forschungsprogramm“ zu nennen. Von Bedeutung ist, daß eine solche Sicht bestimmte Darstellungs- und Analysemöglichkeiten eröffnet, die Herrmann (Kap.6 dieses Bandes) vorstellt und erläutert.

**Theorien** sind nun keine „kleinsten Forschungsprogramme“, sondern - wie Herrmann im 6. Kapitel dieses Bandes ausführt - „Mittel zur Lösung der in Forschungsprogrammen behandelten Probleme“. Im Detail gibt es eine Vielzahl von Versuchen, das zunächst recht unscharfe Konzept einer Theorie zu präzisieren. Gadenne zeigt im 7. Kapitel dieses Bandes solche Spezifikationsvarianten auf, problematisiert die Beziehungen zwischen Theorie und Empirie sowie zwischen Theorie und Realität, untersucht die Rolle der Theorie bei verschiedenen Arten von Erklärungen und weist darauf hin, daß es neben der Ansicht, Theorien seien stets als Aussagensysteme darstellbar, auch andere Theoriekonzeptionen - wie etwa den Strukturalismus (Sneed, 1971) - gibt.

Zu Theorien gibt es zwei methodologische Problemkreise: einen zur Frage, wie Theorien zustandekommen, und einen zweiten zu der Aufgabe, Theorien zu bewerten. Mit diesen beiden Komplexen befassen sich in diesem Band zwei Kapitel von Dörner (Kap. 8 zur Theoriebildung) und von Gadenne (Kap.9 zur Theoriebewertung). Dabei wird klar, daß keine Regeln die **Konstruktion von Theorien** ermöglichen und dabei die Brauchbarkeit ihrer Produkte garantieren. Es gibt aber sehr wohl Regeln zur Theoriekonstruktion, deren Produkte **möglicherweise** brauchbare Theorien sind. Regelsysteme, die angewandt auf eine Fragestellung ein konkretes Ergebnis produzieren, heißen auch „Produktionensysteme“; Regelsammlungen, deren Produkte lediglich möglichst oft (aber nicht immer) den Ausgangsanforderungen genügen, nennt man oft „Heuristiken“, wobei die Regeln selbst die „Heurismen“ sind. Dörner (Kap. 8 dieses Bandes) geht davon aus, daß wir lediglich über Heurismen der Theoriekonstruktion verfügen. Konsequenz ist, daß wir Methoden der Bewertung von Theorien (Gadenne in Kap.9) benötigen.

Zur **Theoriebewertung** gibt es eine Vielzahl von Kriterien. Dabei kann das Bewertungsergebnis je nach Kriterium von unterschiedlicher Art sein. Fragen

wir etwa nach der logischen Konsistenz einer Theorie, so ist diese gegeben oder nicht, kümmern wir uns um ihren Informationsgehalt, so ist dieser mehr oder weniger reichhaltig, und fragen wir schließlich gar danach, ob sie empirisch gültig ist oder nicht, so können wir unter günstigen Umständen vielleicht behaupten, sie sei mit höherer Wahrscheinlichkeit gültig als eine konkurrierende Theorie. Hinzu kommt das Problem, verschiedene Kriterien gegeneinander abzuwägen und eine multivariate Entscheidung zu treffen.

Da Psychologie - wie in Abschnitt 2 dieses Kapitels näher erläutert - in ihrem Selbstverständnis eine empirische Wissenschaft ist, spielt bei der Theoriebewertung die aufgrund von Daten mögliche empirisch basierte Bewertungskomponente eine entscheidende Rolle. Dies führt auf die Frage, ob und inwieweit überhaupt Schlüsse von endlichen Datensätzen auf Systeme von Allgemeinaussagen begründbar sind. Diesem Problem der **Induktion** ist das 10. Kapitel dieses Bandes von Westermann und Gerjets gewidmet. Alle Übergänge von Aussagen über einen limitierten Gegenstandsbereich zu Aussagen über einen größeren Bereich (die sogenannten „ampliativen Inferenzen“) sind insoweit unsicher, als sie die Richtigkeit der produzierten neuen Aussagen unter der Bedingung der Richtigkeit der Ausgangsbehauptungen nicht garantieren können. Damit gibt es keine (sicheren) induktiven Schlüsse, sondern lediglich (unsichere) induktive Argumente, für die allerdings Qualitätskriterien und rationale Begründungen existieren. Westermann und Gerjets untersuchen, welche Begründungen man zur Stützung induktiver Argumente heranziehen kann und welche Voraussetzungen dabei zu beachten sind.

Insgesamt ergibt sich aus den Ausführungen in Teil C dieses Bandes die Unsicherheit und Vorläufigkeit aller wissenschaftlichen Befunde und Ergebnisse. Nun ist es sicher trivial darauf hinzuweisen, daß Resultate wissenschaftlicher Forschung keine unumstößlichen Wahrheiten sind. Die Autoren der Kapitel aus Teil C untersuchen, aus welchen Bedingungen Unsicherheit und Vorläufigkeit resultieren, und nutzen die Ergebnisse dieser Analysen, um Möglichkeiten methodischer Kontrolle aufzuzeigen, die zu solchen Forschungsergebnissen führen, die mit vernünftigen Argumenten als akzeptabel und brauchbar (nach jeweils zu spezifizierenden Kriterien) anzusehen sind.

### 4.3 Forschungsschritte

Sowohl die Betrachtung verschiedener Zugänge psychologischer Forschung zu ihrem Gegenstandsbereich als auch die Analyse von Forschungsprogrammen, Theorien und damit zusammenhängenden Problemkreisen liefern relativ große Analyseeinheiten. Stärker mit methodischen Details befaßte Themen findet man, wenn einzelne Schritte des Forschungsprozesses und die dabei eingesetzten Methoden analysiert werden. Einen großen Teil aktueller psy-



chologischer Forschungen machen Arbeiten aus, in denen irgendwelche Hypothesen geprüft werden sollen. Es ist daher nicht verwunderlich, daß zur hypothesengeleiteten Forschung auch die meisten methodischen und methodologischen Arbeiten vorliegen. Dieser Band versucht dem gerecht zu werden, indem er einen Teil D enthält, in dem Forschungsschritte beim hypothesengeleiteten Vorgehen untersucht werden.

Die Orientierung der Aufeinanderfolge der einzelnen Kapitel von Teil D am Forschungsablauf ist offensichtlich. Es beginnt mit einer Hypothese (Hussy & Möller, Kap. 11). Bei der Planung von Untersuchungen zur Prüfung dieser Hypothese geht es um den Einsatz von Kontrolltechniken (Rehm & Strack, Kap. 12). Dabei wird häufig unterstellt, daß relevante Variablen meßbar sind und daß entsprechend sinnvoll interpretierbare Skalen zur Verfügung stehen (Mausfeld, Kap. 13). Auf der Basis erhobener Daten wird die jeweilige Ausgangshypothese geprüft (Erdfelder & Bredenkamp, Kap. 14). Dabei operiert man in den weitaus meisten Fällen mit Wahrscheinlichkeits- und Zufallskonzepten, benutzt also stochastische Modelle (Steyer, Kap. 15). Der Prozeß endet mit einer statistischen Entscheidung (Ostmann & Wutke, Kap. 16).

Da alle Beiträge aus Teil D dieses Bandes Teile eines Forschungsprozesses behandeln, die auf das gleiche Ziel ausgerichtet sind, gibt es notwendig zahlreiche Überschneidungen. Man kann nun einmal - um nur ein Beispiel zu erwähnen - Kontrolltechniken bei der Versuchsplanung nur einsetzen, wenn man die Prüflöge kennt und weiß, welche Kontrollen sie ermöglicht. Auf der anderen Seite macht es wenig Sinn, Verfahren der Hypothesenprüfung zu analysieren ohne Verweis auf Kontrollmöglichkeiten, die bereits bei der Versuchsplanung zu berücksichtigen waren.

Das in diesem Zusammenhang zentrale Konzept der **Hypothese** wird in den einzelnen Beiträgen von Teil D zunehmend differenziert. Hussy und Möller (Kap. 11) beginnen mit der alltagssprachlichen Rede von „Vermutungen“ und zeichnen wissenschaftliche Hypothesen dadurch aus, daß ihren Inhalten ein Wahrheitswert zukommt und daß sie prinzipiell überprüfbar sind. Sie führen mehrere Differenzierungen ein, darunter jene zwischen einer theoretisch-inhaltlichen Hypothese, einer möglicherweise durch Operationalisierungen, operationale Reduktionen oder möglicherweise Vorgehensweisen abgeleiteten empirisch-inhaltlichen Hypothese, einer durch Präzisierung von Häufigkeits- oder Intensitätsangaben formulierbaren statistischen Hypothese und einer für statistische Entscheidungen benötigten Prüfhypothese. Während bei Hussy und Möller (Kap. 11) auch Vorhersagen über individuelle Ereignisse durchaus als „wissenschaftliche Hypothesen“ angesprochen werden, gehen Rehm und Strack (Kap. 12, Abschnitt 1) davon aus, daß es sich im allgemeinen um Vermutungen handelt, die sich auf raum-zeitlich begrenzte „offene Populationen“ beziehen. Sie betrachten bei ihren weiteren Überlegungen auch nur noch diese

Art von Hypothesen, und so tun es auch die Verfasser aller noch folgenden Kapitel. Bei Erdfelder und Bredenkamp (Kap. 14) spielt dann wiederum die Unterscheidung zwischen einer **psychologischen**, einer **empirischen** und einer **statistischen** Hypothese eine Rolle, wobei diese bereits bei Hussy und Möller (Kap. 11) vorgestellte Differenzierung hier ausführlicher spezifiziert wird. Dabei geht es insbesondere um die Frage nach der Relation zwischen psychologischen und jeweils zugeordneten statistischen Hypothesen. Wie kann man es begründen, statistische Entscheidungen über statistische Hypothesen als Grundlage für Bewährungsurteile über psychologische Hypothesen zu benutzen? Sowohl Steyer (Kap. 15) als auch Ostmann und Wutke (Kap. 16) gehen - der Themenstellung ihrer Beiträge entsprechend - vom Umgang mit probabilistisch formulierten Hypothesen aus, wobei Aussagen über Verteilungen und deren Parameter ein häufiger Spezialfall sind.

Zur Kennzeichnung der Qualität einer hypothesenorientierten Untersuchung wird oft der Ausdruck „**Validität**“ benutzt. Im Umgang mit diesem Ausdruck tauchen teils beträchtliche Bedeutungsunterschiede auf. Unter 2.3.3 wurde in diesem Kapitel bereits auf ein recht unscharfes Konzept der Validität einer Untersuchung bei Bortz (1984) hingewiesen, das dann in einer Weise präzisiert wurde - nämlich als Güte der Prognostizierbarkeit eines neuen Datums -, bei der Validität zum Merkmal eines Befundes wird. Hussy und Möller (Kap. 11) sprechen von verschiedenen Validitätskriterien, die teilweise unterschiedlichen Komponenten eines hypothesenorientierten Forschungsprozesses zugeordnet sind. So ist bei ihnen etwa die Rede von der „Ableitungsvalidität“ einer statistischen Hypothese (Kap. 11, Abschnitt 2.3.3), von einer „Populationsvalidität“, einer „ökologischen Validität“, einer „Situationsvalidität“ und einer „internen Validität“, die auch als „Ceteris-Paribus-Validität“ angesprochen wird (Kap. 11, Abschnitt 3). Erdfelder und Bredenkamp (Kap. 14) stellen ein auf Untersuchungen bezogenes komparatives Validitätskonzept vor. Danach ist Untersuchung **A** valider als Untersuchung **B**, wenn bei **A** sowohl die Wahrscheinlichkeit eines falschen Bewährungsurteils für die (in Wirklichkeit falsche) in Frage stehende psychologische Hypothese als auch die Wahrscheinlichkeit für ein falsches Nichtbewährungsurteil bei in Wirklichkeit richtiger Hypothese kleiner ist als bei **B**. Man sieht, daß derartige Vergleiche nur beschränkt möglich sind; in vielen Fällen kann man von zwei Untersuchungen nicht aussagen, welche von ihnen valider ist.

Selbstverständlich gibt es auch eine große Zahl von Konzepten, die in den Kapiteln zu Teil D dieses Bandes einheitlich behandelt werden. So wird von fast allen Autoren unterstellt, daß bei den zur statistischen Prüfung anstehenden Hypothesenformulierungen problemlos Verteilungsparameter benutzt werden können, die quantitative Variablen voraussetzen. Werte und Verteilungen quantitativer Variablen sind aber nur dann sinnvoll interpretierbar, wenn diese Variablen Skalen sind, die ausgewählte Aspekte bestimmter Ob-

jekte und Phänomene im jeweils betrachteten Realitätsbereich repräsentieren. Damit zusammenhängende Fragen analysiert Mausfeld im Kapitel 13. Einheitlichkeit herrscht auch bezüglich der Auffassung, Hypothesenprüfung müsse die Wahrscheinlichkeiten beider möglicher Arten von Fehlurteilen (falsches Bewährungsurteil bei falscher Hypothese, falsches Nichtbewährungsurteil bei richtiger Hypothese) berücksichtigen. Ostmann und Wutke (Kap. 16) stellen verschiedene Strategien der Hypothesentestung vergleichend nebeneinander. Sie analysieren sowohl die Vorteile einer Neyman-Pearson-Testung, die beide Fehlermöglichkeiten berücksichtigt, gegenüber der simpleren Fisher-Testung als auch die Möglichkeiten einer an Bayes orientierten Entscheidungsstrategie.

Insgesamt zeigt Teil D dieses Bandes beispielhaft auf, wie die Identifikation von Forschungsschritten zu Themen methodologischer Untersuchungen führen kann. Zusammen mit der Identifikation unterschiedlicher Zugänge psychologischer Forschung zu ihren Gegenständen und Phänomenen (Teil B) und der - orthogonal dazu konzipierten - Zuwendung zu gemeinsamen Gegebenheiten und Problemen (wie der Existenz von Forschungsprogrammen, der Arbeit mit und an Theorien und der Auseinandersetzung mit den Varianten des Induktionsproblems in Teil C) wird so in diesem Band ein erstes und sicher vorläufiges System von Themen aufgespannt, das die Reichhaltigkeit der Einzelfragen innerhalb einer Allgemeinen Methodenlehre der Psychologie erkennen läßt.

## **Literatur**

- Agnew, N.M. & Pyke, S.W.(1969). **The science game. An introduction to research in the behavioral sciences.** Englewood Cliffs, N.J.: Prentice Hall.
- Anderson, J. R. (1991). Optimality and human memory. **Behavioral and Brain Sciences**, **14**, 215-216.
- Baron, J. (1991). Optimality as an evaluative Standard in the study of decision-making. **Behavioral and Brain Sciences**, **14**, 216.
- Bell, D. E., Raiffa, H. & Tversky, A. (1988). **Decision making: Descriptive, normative, and prescriptive interactions.** Cambridge: Cambridge University Press.
- Bortz, J. (1984). **Lehrbuch der empirischen Forschung für Sozialwissenschaftler.** Berlin: Springer.
- Bradley, R.A. & Terry, M.E. (1952). Rankanalysis of incomplete block designs. 1. The method of paired comparisons. **Biometrika**, **39**, 324-345.
- Bush, R. R. & Mosteller, F. (1951). A mathematical model for simple learning. **Psychological Review**, **58**, 313-323.
- Carnap, R. (1936/1937). Testability and meaning. **Philosophy of Science**, **314**.

- Chalmers, A.F. (1989). **Wege der Wissenschaft** (zweite, durchgesehene Auflage; Original erschienen 1976: What is this thing called science?). Berlin: Springer.
- Coombs, C. H. (1964). A theory of **data**. New York: Wiley.
- Coombs, C. H., Dawes, R. M. & Tversky, A. (1970). **Mathematical psychology: An introduction**. Englewood Cliffs, N.J.: Prentice Hall.
- Damerow, P., Freudenthal, G., McLaughlin, P. & Renn, J. (1992). **Exploring the limits of pre-classical mechanics. A study of conceptual development in early modern science: Descartes, Galileo, Beekman**. New York: Springer.
- Descartes, R. (1637/1984). **Abhandlung über die Methode des richtigen Vernunftgebrauchs** (französisches Original 1637). Stuttgart: Reclam.
- de Groot, A.D. (1969). **Methodology. Foundations of inference and research in the behavioral sciences**. Den Haag: Mouton.
- Dingler, H. (1928). **Das Experiment: Sein Wesen und seine Geschichte**. München: Reinhardt.
- Dingler, H. (1955). **Die Ergreifung des Wirklichen**. München: Eidos.
- Egart, J.P. (1958). **Recognition memory and the operating characteristic** (Technical Note AFCRC-TN-58-51). Indiana: University Hearing and Communication Laboratory.
- Engelkamp, J. (1990). **Das menschliche Gedächtnis**. Göttingen: Hogrefe.
- Feyerabend, P. (1976). **Wider den Methodenzwang. Skizzen einer anarchistischen Erkenntnistheorie**. Frankfurt: Suhrkamp.
- Feyerabend, P. (1980). **Erkenntnis für freie Menschen** (veränderte Ausgabe). Frankfurt: Suhrkamp.
- Gigerenzer, G. & Murray, D.J. (1987). **Cognition as intuitive statistics**. Hillsdale, N.J.: Erlbaum.
- Grawe, K., Hänni, R., Semmer, N. & Tschan, F. (Hrsg.) (1991). **über die richtige Art, Psychologie zu betreiben**. Göttingen: Hogrefe.
- Herrmann, T. (1990). Die Experimentiermethodik in der Defensive? **Spruche und Kognition, 9**, 1-11.
- Holzkamp, K. (1964). **Theorie und Experiment in der Psychologie**. Berlin: de Gruyter.
- James, W. (1890). **Principles of psychology** (Volume I and II). New York: Holt.
- Kächele, H. (1986). **Maschinelle Inhaltsanalyse in der psychoanalytischen Prozeßforschung**. Ulm: PSZ-Verlag.
- Lord, F. & Novick, M. R. (1968). **Statistical theories of mental test scores**. Reading, Mass.: Addison-Wesley.
- Luce, R.D. (1959). **Individual choice behavior**. New York: Wiley.
- Mergenthaler, E. (1985). **Textbank systems: Computer science applied in the field of psychoanalysis**. Berlin: Springer.
- Nagel, E. (1961). **The structure of science. Problems in the logic of scientific explanation**. New York: Harcourt, Brace & World.
- Nelson, R. J. (1977). Structure of complex systems. In F. Suppe & P.D. Asquith (Eds.), **PSA 1976, Vol. 2: Symposia**. East Lansing, Mich.: Philosophy of Science Association.

- Nelson, R. J. (1989). **The logic of mind** (2nd ed.). Dordrecht: Kluwer.
- Newell, A. (1990). *Unified theories of cognition***. Cambridge, Mass.: Harvard University Press.
- Newell, A. & Simon, H. A. (1972). **Human problem solving**. Englewood Cliffs, N.J.: Prentice Hall.
- Palmer, S. E. (1986). The information processing approach to cognition. In T. J. Knapp & L.C. Robertson (Eds.), **Approaches to cognition: Contrasts and Controversies** (S. 37-77). Hillsdale, N.J.: Erlbaum.
- Paul, W. J. (1978). **Komplexitätstheorie**. Stuttgart: Teubner.
- Peterson, W. W., Birdsall, T. G. & Fox, W.C. (1954). The theory of signal detectability. **Transactions of the Professional Group on Information Theory** [Institute of Radio Engineers], 4, 171-212.
- Popper, K.R. (1966). **Logik der Forschung** (zweite, erweiterte Auflage). Tübingen: Mohr.
- Posner, M. I. (Ed.) (1989). **Foundations of cognitive science**. Cambridge, Mass.: MIT Press.
- Roskam, E. E. (1983). Allgemeine Datentheorie. In H. Feger & J. Bredenkamp (Hrsg.), **Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie I: Forschungsmethoden der Psychologie, Band 3: Messen und Testen (S. 1-135)**. Göttingen: Hogrefe.
- Schoemaker, P. J. H. (1991). The quest for optimality: A positive heuristic of science? **Behavioral and Brain Sciences, 14, 205-245**.
- Selten, R. (1989). Evolution, **learning and economic behavior** (1989 Nancy L. Schwartz Memorial Lecture. Sonderforschungsbereich 303, discussion paper B-132). Bonn: Rheinische Friedrich-Wilhelms-Universität.
- Simon, H. (1977). How complex are complex systems? In F. Suppe & P.D. Asquith (Eds.), **PSA 1976, Vol. 2: Symposia**. East Lansing, Mich.: Philosophy of Science Association.
- Sneed, J. D. (1971). **The logical structure of mathematical physics**. Dordrecht: Reidel.
- Störig, H. J. (1970). **Kleine Weltgeschichte der Wissenschaft** (zwei Bände; Originalausgabe 1954, Stuttgart: Kohlhammer). Frankfurt: Fischer.
- Swijtink, Z.G. (1987). The objectification of observation: Measurement and statistical methods in the nineteenth century. In L. Krüger, L.J. Daston & M. Heidelberger (Eds.), **The probabilistic revolution. Vol. 1: Ideas in history** (Pp.261-285). Cambridge, Ma.: MIT Press.
- Tack, W. H. (1991). Rationales Handeln in sozialen Situationen. **Zeitschrift für Sozialpsychologie, 22, 151-165**.
- Tseëlon, E. (1991). The method is the message: On the meaning of methods as ideologies. **Theory and Psychology, 1, 299-316**.
- van Acker, P.O.F.C. (1977). **Models for intransitive choice**. Doctoral Dissertation, Katholieke Universiteit te Nijmegen, Nijmegen.
- van Meter, D. & Middleton, D. (1954). Modern statistical approaches to reception in communication theory. **Transactions of the Professional Group on Information Theory** [Institute of Radio Engineers], 4, 119-145.

**Teil B**  
**Zugänge**

## **Kapitel 2**

### **Erzeugung und Verwendung empirischer Daten**

von Edgar Erdfelder

## **Kapitel 3**

### **Qualitatives Vorgehen - „Interpretation“**

von Hans-Georg Soeffner und Ronald Hitzler

## **Kapitel 4**

### **Methodologische Grundlagen und Probleme der Psychophysik**

von Rainer Mausfeld

## **Kapitel 5**

### **Modellierung durch wissensbasierte Systeme**

von Klaus Opwis und Hans Spada

Teil B dieses Bandes stellt verschiedene Zugänge zum Gegenstands- und Phänomenbereich der Psychologie mit ihren methodischen Implikationen vor. Die Auswahl der dargestellten Forschungsansätze versucht, einige wichtige Unterschiede deutlich werden zu lassen; sie erhebt keinerlei Anspruch auf eine wie auch immer geartete Vollständigkeit oder Relevanzbeurteilung.

In Kapitel 2 geht es um eine Auffassung von der Erzeugung und Verwendung empirischer Daten, bei der eine bestimmte Art von Fragestellung zur Erhebung und Nutzung angemessen restringierter Daten führt. Kapitel 3 betrachtet dagegen bei der „Interpretation“ die Erschließung einer Vielfalt von Deutungsmöglichkeiten als Forschungsziel; methodische Konsequenz ist die Verwendung möglichst wenig restringierter Daten, um für unterschiedliche Deutungsmöglichkeiten vielfältige Datenqualitäten und -aspekte nutzen zu können. Die beiden weiteren Kapitel befassen sich mit Forschungsansätzen, für die eine in der Psychologie mögliche Innen/Außen-Differenzierung konstitutiv ist. Psychophysik (Kapitel 4) erfaßt die jeweils interessierenden Gegebenheiten eines (inneren) mentalen Systems als Variablensysteme, während sie bei der Modellierung durch wissensbasierte Systeme (Kapitel 5) als Symbolstrukturen dargestellt und analysiert werden.

Diese Aspekte lassen zumindest einen bedeutsamen Teil der methodologisch relevanten Variation unterschiedlicher Forschungszugänge deutlich werden.

## 2. Kapitel

# Erzeugung und Verwendung empirischer Daten

***Edgar Erdfelder***

In dem Maße, in dem sich die Psychologie als Erfahrungswissenschaft etabliert hat, ist die Erzeugung und Verwendung empirischer Daten für Lehrende, Studierende und Anwender des Faches zu einer Selbstverständlichkeit geworden. Fragen zu den Charakteristika, zum Stellenwert und zur Funktion empirischer Daten im Forschungsprozeß geraten deshalb leicht aus dem Blickfeld. Was sind überhaupt empirische Daten? Handelt es sich lediglich um eine bestimmte Klasse von Beobachtungsaussagen? Ist ihre Verwendung auf Erfahrungswissenschaften beschränkt oder können sie auch in anderen Wissenschaften eine Rolle spielen? Weiterhin: Wie gewinnt man empirische Daten, die zur Beantwortung einer bestimmten psychologischen Fragestellung geeignet sind? Gibt es eine verbindliche Menge von Regeln für die „korrekte“ Abbildung empirischer Sachverhalte in empirische Daten, welche empirisch arbeitende Wissenschaftlerinnen und Wissenschaftler lediglich zu erlernen haben? Wie sind Daten im Hinblick auf eine forschungsleitende Fragestellung angemessen zu analysieren? Und schließlich: Begründen empirische Daten psychologisches Wissen? Falls sie dies nicht leisten: Welche andere Funktion im Forschungsprozeß kann ihnen zugesprochen werden? Derartige Fragen umreißen den Problemkreis, mit dem sich der vorliegende Beitrag auseinandersetzt.

Abbildung 1 veranschaulicht die Teilprozesse, die bei der Erzeugung und Verwendung empirischer Daten eine Rolle spielen. Das vorliegende Kapitel folgt im wesentlichen den möglichen Sequenzen dieser Teilprozesse. Im ersten Abschnitt wird die Beziehung zwischen Beobachtungen und empirischen Daten erörtert. In Übereinstimmung mit Coombs (1964) und Roskam (1983, im Druck) werden empirische Daten als Resultat der Abbildung von registrierten Beobachtungen in einen Datenraum aufgefaßt. Der Datenraum - die Menge möglicher Daten - sowie die Regeln für die Zuordnung von Daten zu registrierten Beobachtungen müssen fragestellungsabhängig *vor* jeder Datenerhebung festgelegt werden. Die Beobachtungen, welche den Ausgangspunkt der Datenerhebung bilden, determinieren weder die Struktur des Datenraums noch die Art der Zuordnung von Daten zu Beobachtungen.



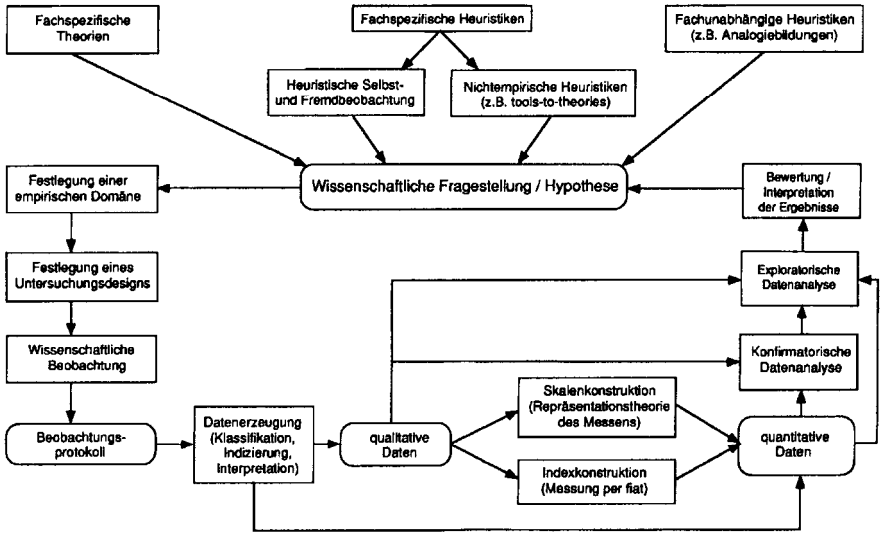


Abb. 1: Mögliche Sequenzen von Forschungsaktivitäten (in eckigen Kästen) und Forschungsprodukten (in abgerundeten Kästen) bei der Erzeugung und Verwendung empirischer Daten.

So verstanden sind empirische Daten notwendig an Erfahrungswissenschaften gebunden. Andere Wissenschaften können Beobachtungen im Forschungsprozeß durchaus verwenden oder gar ohne bestimmte Klassen registrierter Beobachtungen völlig undenkbar sein; allerdings verfahren diese Disziplinen mit Beobachtungen in grundsätzlich anderer Weise und mit anderen Zielsetzungen als Erfahrungswissenschaften. Für letztere ist ein „datenerzeugender“ und „datenganalisierender“ Zugang charakteristisch (Tack spricht in Kapitel 1 dieses Bandes kurz von „datenförmig“), wobei die Daten - anders als beispielsweise beim hermeneutischen Zugang - nicht einfach schriftliche Dokumentationen der ursprünglichen Beobachtungen darstellen, sondern zumeist durch erhebliche Informationsreduktion aus diesen hervorgehen. Diese Auffassung schließt die Möglichkeit ein, daß eine bestimmte Menge registrierter Beobachtungen - etwa literarische Texte oder Gesprächsprotokolle - sowohl Gegenstand „datenförmiger“ empirischer Wissenschaft als auch Gegenstand hermeneutischer Wissenschaft sein kann (vgl. z.B. Groeben, 1972). Nur der datenerzeugende und datenganalisierende Zugang zu Beobachtungen soll in diesem Kapitel untersucht werden; der interpretativ-hermeneutische Zugang wird von Soeffner und Hitzler in Kapitel 3 dieses Bandes behandelt.

Die Hauptthese des ersten Abschnitts lautet, daß sowohl die Abbildung empirischer Sachverhalte in Beobachtungen als auch die Transformation von Beobachtungen in empirische Daten viele Freiheitsgrade bieten, auch wenn diese

nicht immer offensichtlich sind. Datenerzeugung verläuft somit nicht geradlinig nach einem vorgegebenen Kodierungsschema für empirische Sachverhalte, das allgemein verbindlich ist. Empirische Wissenschaftlerinnen und Wissenschaftler spielen keine passiv-registrierende, sondern eine aktiv-gestaltende Rolle bei der Erzeugung psychologierelevanter empirischer Daten. Daten werden nicht „gefunden“, sondern stets „kreiert“. Dies ist *unvermeidlich*, d.h. die im Forschungsprozeß involvierten Personen befinden sich in einer gestaltenden Rolle auch dann, wenn sie sich dessen nicht bewußt sind oder wenn sie diese Rolle nicht wünschen. Deshalb liegt es nahe, Datenerzeugung generell als Prozeß zu betrachten, der keineswegs annahmefrei und neutral, sondern zwangsläufig „theoriedurchtränkt“ ist (vgl. Gigerenzer, 1980; Roskam, 1983).

Wenn man im Prozeß der Datenerhebung ebenso wie bei der anschließenden Datenanalyse notwendigerweise massiv Einfluß nimmt, liegt die Frage nahe, wie diese Einflußnahme optimalerweise aussehen sollte. Die Antwort auf diese Frage kann nicht undifferenziert für alle denkbaren Forschungskontexte gegeben werden, sondern hängt davon ab, mit welcher Zielsetzung und unter welchen substanzwissenschaftlichen Voraussetzungen empirische Forschung im konkreten Fall betrieben wird. In Anlehnung an Reichenbach (1938) kann man zunächst grob und vereinfachend zwischen der Verwendung empirischer Daten im Entdeckungszusammenhang (*context of discovery*) und der Verwendung empirischer Daten im Begründungszusammenhang (*context of justification*) unterscheiden. Dieser Dichotomie entspricht im wesentlichen die Unterscheidung von **offenen** und geschlossenen psychologischen Fragestellungen. Offene Fragestellungen sind beispielsweise von der Form „Wovon hängt das Auftreten von X ab?“, „Von welcher Art ist der Zusammenhang oder die Beziehung zwischen X und Y?“, „Welche Auswirkungen hat X?“, „Welche potentiellen Determinanten von Y kommen in Frage?“, „Wie entsteht Y?“, „Wie läßt sich Y beeinflussen?“, „Welche Typen von X gibt es?“, es handelt sich um Fragen zu empirischen Sachverhalten, für die a priori keine festen Antwortkategorien vorgegeben sind. Daten werden erzeugt und analysiert, um mögliche Antworten auf derartige Fragen - und damit potentielle Hypothesen oder Theorien zur interessierenden empirischen Domäne - **explorieren** zu können.

Stehen offene Fragen im Vordergrund, so werden empirische Daten **im Entdeckungszusammenhang** verwendet. Konsequenzen für Prinzipien der Datenanalyse werden im zweiten Abschnitt erörtert. Es wird die These formuliert und begründet, daß die üblichen hypothestetestenden Analysemethoden im Entdeckungszusammenhang nur sehr selten und nur unter besonderen Vorichtsmaßnahmen geeignete Forschungsinstrumente darstellen.

Geschlossene psychologische Fragestellungen haben demgegenüber immer die Struktur „Trifft es zu, daß X?“, wobei X ein psychologisches Gesetz, eine

Hypothese, eine technologische Prognose (Brocke, 1979) oder eine Theorie (mit primär wissenschaftlicher oder technologischer Schwerpunktsetzung, vgl. Herrmann, 1979) bezeichnet. Daten werden erhoben und analysiert, um eine der beiden möglichen Antworten „ja“ oder „nein“ empirisch stützen zu können. In diesem Fall liegt eine Verwendung empirischer Daten **im Begründungszusammenhang** vor, was in Abschnitt 3 diskutiert wird. Zwei Verwendungszwecke - die „induktive“ und die „deduktive“ - werden unterschieden und kurz skizziert.

Abschnitt 4 macht deutlich, daß der Entdeckungs- und der Begründungszusammenhang nicht als disjunkte Kategorien von Forschungsaktivitäten aufgefaßt werden sollten. Es wird die These vertreten, daß exploratorische Datenanalyseverfahren auch im Kontext der strengen Hypothesenprüfung eine wichtige Rolle spielen, da sie Theorierevisionen nahelegen können, wenn hypotheseentestende Verfahren noch „Theorieverträglichkeit“ von Daten signalisieren.

Die Arbeit schließt mit einer Zusammenfassung der wichtigsten Ergebnisse und Thesen (Abschnitt 5). Als Hauptresultat wird sich ergeben, daß die Funktion empirischer Daten weniger darin besteht, daß mit ihrer Hilfe psychologisches Wissen erzeugt und begründet werden kann; die wichtigere Funktion ist die **Ermöglichung der Kritik** psychologischer Annahmen und (synthetischer) Aussagen. Dies gilt unabhängig davon, ob Daten im Entdeckungs- oder im Begründungszusammenhang Verwendung finden.

## **1. Beobachtungen und die Erzeugung von Daten**

Daten implizieren zunächst Beobachtungen: Ohne protokollierte Beobachtungen ist die Erzeugung empirischer Daten unmöglich. Doch welche Eigenschaften müssen protokollierte Beobachtungen aufweisen, um als Grundlage für die Erzeugung empirischer Daten geeignet zu sein? Im folgenden wird zunächst der Begriff der **wissenschaftlichen Beobachtung** eingeführt, und es wird begründet, warum empirische Daten aus wissenschaftlichen Beobachtungen erwachsen sollten (Abschnitt 1.1). Anschließend werden Selbstbeobachtungen mentaler Prozesse (Abschnitt 1.2) und heuristische Beobachtungen (Abschnitt 1.3) von wissenschaftlichen Beobachtungen abgegrenzt. Es wird erörtert, welcher Stellenwert und welche Funktion diesen Beobachtungsformen in der empirischen Psychologie zukommt. Eine Diskussion von Problemen bei der Erzeugung empirischer Daten auf der Grundlage wissenschaftlicher Beobachtungen (Abschnitt 1.4) schließt den ersten Abschnitt ab.

## 1.1 Wissenschaftliche Beobachtung

Von wissenschaftlicher Beobachtung soll gesprochen werden, wenn abgrenzbare empirische Sachverhalte in **objektiver** und **prinzipiell wiederholbarer** Weise unter Zugrundelegung eines vorgegebenen Vokabulars in Beobachtungsprotokolle überführt werden (vgl. auch Glück, 1971; Graumann, 1966, S. 87; Greve & Wentura, 1991, S.5; Herrmann, 1972, S.40). Die Objektivitätsforderung impliziert, daß die empirischen Sachverhalte „öffentlich“ in dem Sinne sind, daß **potentiell** verschiedene Personen Beobachtungsprotokolle zu dem gleichen Sachverhalt anfertigen können und daß diese (weitgehend) in ihren Beobachtungsprotokollen übereinstimmen müssen. Prinzipielle Wiederholbarkeit impliziert, daß einmalige, flüchtige, nicht systematisch replizierbare empirische Sachverhalte auch nicht wissenschaftlich beobachtet werden können. **Nicht** zwangsläufig impliziert ist die Identität von Beobachtungsprotokollen im Falle wiederholter Beobachtungen eines Individuums: Entwicklungs- und Lernvorgänge sowie veränderte Kontextbedingungen bei der Beobachtung können (müssen aber selbstverständlich nicht) veränderte Protokolle zur Folge haben (Herrmann, 1972, S. 39/40). Replizierbarkeit von Beobachtungen muß nicht notwendig anhand wiederholter Beobachtungen eines Individuums zu verschiedenen Zeitpunkten geprüft werden; sie kann sich auch auf Beobachtungen einer Verhaltensweise an unterschiedlichen Individuen oder verschiedener (theoretisch miteinander verbundener) Verhaltensweisen bei einem Individuum beziehen (Feger, 1983). Welche Methode der Replizierbarkeitsprüfung angemessen ist, kann erst auf dem Hintergrund der wissenschaftlichen Fragestellung entschieden werden, deren Beantwortung intendiert ist. Wenn die forschungsleitende Hypothese Veränderungen über die Zeit als Vernachlässigbar ausweist, mag man sich intraindividuelle Replikationen bedienen; werden dagegen interindividuelle Unterschiede weitgehend ausgeschlossen, so kann man Replikationen über verschiedene Individuen heranziehen, usw. Feger (1983, S. 3) betont deshalb mit Recht, daß zur wissenschaftlichen Beobachtung immer auch die **Absicht** gehört, eine bestimmte wissenschaftliche Fragestellung oder Hypothese zu untersuchen (vgl. auch Glück, 1971). Ohne einen solchen Hintergrund kann nicht wissenschaftlich beobachtet werden, da die Kriterien für eine wissenschaftliche Beobachtung erst auf dem Hintergrund der zu beantwortenden Fragestellung eindeutig spezifiziert werden können.

Schließlich ist auch ein **Beobachtungsprotokoll** - die Abbildung auf ein fest vereinbartes System von Zeichen - notwendiger Bestandteil wissenschaftlicher Beobachtung (Glück, 1971). Beim vereinbarten Zeichensystem kann es sich um die Umgangssprache handeln (einen Überblick hierzu bietet z.B. Faßnacht, 1979). Oft wird aber schon ein einfacheres Vokabular (basierend z.B.

auf Zahlen, die Antwortzeiten repräsentieren, oder Ketten der Zeichen „0“ und „1“ für falsche und richtige Antworten) völlig ausreichen.

Es soll hier darauf verzichtet werden, wissenschaftliche Beobachtung von anderen empirischen Methoden definitorisch abzugrenzen. Eine solche Grenzziehung kann im Detail nur willkürlich sein und hätte auch keinerlei methodologische Relevanz. Ähnlich wie z.B. bei Coombs (1964) wird „wissenschaftliche Beobachtung“ stattdessen als Sammelbezeichnung für alle empirischen Methoden aufgefaßt, die den Bedingungen Objektivität und Replizierbarkeit genügen. Ebenfalls soll darauf verzichtet werden, wissenschaftliche Beobachtung an das psychologische Experiment zu binden (wie es z.B. W. Wundt getan hat, vgl. Abschnitt 1.2) oder aber strikt davon abzugrenzen. Wissenschaftliche Beobachtung kann sowohl unter strenger als auch unter reduzierter versuchsplanerischer Kontrolle sinnvoll sein. Ob ein Experiment durchgeführt wird oder nicht, hängt in erster Linie von der Präzision der zugrundeliegenden psychologischen Fragestellung, von ethischen Erwägungen und von der Manipulierbarkeit der darin erwähnten Variablen ab (vgl. auch Rehm & Strack, Kapitel 12 dieses Bandes).

Bilden wissenschaftliche Beobachtungen im o. g. Sinne einen geeigneten Ausgangspunkt für die Erzeugung psychologischer Daten? Eine ganze Reihe der in der Psychologie üblichen empirischen Methoden lassen sich sicher problemlos unter diesen Beobachtungsbegriff subsumieren: die Erhebung psychophysiologischer Indikatoren wie Hautwiderstand, Pulsfrequenz und das evozierte Potential, ferner Reaktionszeitmessungen, automatisierte Augenbewegungsregistrierungen, Größenschätzungen, Kategorialurteile, intermodale Zuordnungen, Fragebogen-, Rating-, Sortiermethoden und vieles andere mehr. Charakteristisch für alle genannten Beispiele ist eine starke Selektivität der Beobachtungen: Nur ein sehr kleiner Ausschnitt des „Universums potentieller Beobachtungen“ (Coombs, 1964, S. 4) wird als relevant erachtet. Entsprechend klein ist jeweils die Anzahl überhaupt möglicher distinkter Beobachtungsprotokolle bzw. die Dimensionalität der Menge möglicher Beobachtungsprotokolle. Dies erleichtert die Erfüllung der Objektivitätsforderung, die für den wissenschaftlichen Beobachtungsbegriff zentral ist. Darüber hinaus trägt die Benutzung automatisierter Registriertechniken zur Objektivität der Beobachtungen bei. In einigen der genannten Fälle dominiert die Registriertechnik so sehr, daß man vielleicht gar nicht mehr bereit sein mag, den Begriff „Beobachtung“ im Sinne einer menschlichen Wahrnehmungsaktivität zu benutzen (Greve & Wentura, 1991, S.17f.). Dennoch ist es nützlich und richtig, die genannten Beispiele ausnahmslos als Instanzen wissenschaftlicher Beobachtung aufzufassen: Letztendlich sind es immer empirische Wissenschaftlerinnen oder Wissenschaftler, welche die Ausgabe des Registrierinstruments akzeptieren (und damit ein Beobachtungsprotokollfestsetzen) oder auch zurückweisen

(und damit ein Protokoll verweigern), weil sie z.B. der Meinung sind, daß ein bestimmtes Instrument nicht korrekt funktioniert hat.

So gut der wissenschaftliche Beobachtungsbegriff auf die genannten empirischen Methoden anwendbar ist, so problematisch erscheint er im Hinblick auf andere Methoden, die in der Psychologie ebenfalls häufig Verwendung finden. Verhaltensbeobachtungen, projektive Tests und Interviews mit offenen Antwortalternativen zählen beispielsweise hierzu. Charakteristisch ist in diesen Fällen ein breiter, komplexer und weitgehend unstrukturierter Bereich von als relevant erachteten empirischen Sachverhalten (im Extremfall: der Tagesablauf eines Kindes, vgl. Barker & Wright, 1951) sowie Unklarheit über das geeignete Beobachtungsvokabular. Beides erschwert die Erfüllung der Objektivitätsforderung bzw. macht ihre Umsetzung sogar unmöglich. Soweit Objektivität nicht erfüllt ist, ist man in der Tat gezwungen, die o. g. Methoden aus dem Bereich wissenschaftlicher Beobachtungen (und damit aus der empirischen Wissenschaft) auszugrenzen (Herrmann, 1972, S. 39). Allerdings ist die „Objektivierung“ o. g. Methoden mit Hilfe bestimmter Beobachtungsschemata, Kategoriensysteme und Kodierungsschlüssel in Verbindung mit einem entsprechenden Beobachtungs- und Kodiertraining oftmals möglich, was für den Bereich der Verhaltensbeobachtung z.B. die Übersicht bei Manns, Schultze, Herrmann und Westmeyer (1987) belegt. Somit kann Methodenklassen nicht immer global das Prädikat „wissenschaftliche Beobachtung“ zu- oder abgesprochen werden. Es kommt im Detail auf die Art der Verwendung an. Verhaltensbeobachtungen können sicherlich als wissenschaftliche Beobachtungen klassifiziert werden, wenn standardisierte Beobachtungsschemata und darin geschulte Beobachtende herangezogen werden. Ähnliches gilt aber nicht für unsystematische und unstandardisierte Verhaltensbeobachtungen z.B. am eigenen Kind, die dann in Form von Tagebuchaufzeichnungen festgehalten werden.

Wenn Objektivität von Beobachtungsprotokollen und Wiederholbarkeit des Beobachtungsvorgangs gesichert sind, ist insofern eine solide Grundlage für die Erzeugung empirischer Daten gegeben, als intersubjektive Übereinkunft über Daten potentiell herstellbar ist. Allerdings darf Objektivität nicht mit „Gültigkeit“ oder „Wahrheit“ einer Beobachtung gleichgesetzt werden, d.h. mit der Annahme, daß das Beobachtungsprotokoll die schlechthin „richtige“, einzig korrekte und unanzweifelbare Darstellung der empirischen Realität liefert. Daß Objektivität und Gültigkeit verschiedene Dinge sind, ist unmittelbar offensichtlich, wenn - wie im Falle der Verwendung von bestimmten Systemen der Verhaltensbeobachtung (vgl. z.B. Faßnacht, 1979; Manns et al., 1987) - der Beobachtungsvorgang eine explizite Theorie darüber voraussetzt, **welche** empirischen Sachverhalte **wie** in Beobachtungsprotokolle zu überführen sind. Eine andere Theorie (ein anderes Kategoriensystem) wurde zu anderen Be-

obachtungsprotokollen führen. Dies allein relativiert den möglichen Geltungsanspruch, auch wenn die Beobachtungsobjektivität perfekt sein sollte.

Bei einfacheren Beobachtungsleistungen ist die Fraglichkeit eines uneingeschränkten Geltungsanspruchs weniger offensichtlich. Wenn z.B. sehr viele Beobachtende darin übereinstimmen, daß eine Versuchsperson (Vp) das Gesicht einer jungen Frau im Profil gezeichnet hat, neigt man dazu, nicht mehr daran zu zweifeln, daß sie wirklich das Gesicht einer jungen Frau gezeichnet hat. Doch diese Auffassung ist naiv. Leeper (1935) hat z.B. empirisch nachgewiesen, daß ein- und derselbe Reiz (eine mehrdeutige Strichzeichnung) in zwei verschiedenen Kontexten in intersubjektiv übereinstimmender Weise zu zwei verschiedenen figuralen Wahrnehmungsurteilen führen kann (im Beispiel: „junge Frau“ vs. „alte Frau“). Die Objektivität eines Wahrnehmungsurteils schließt also nicht aus, daß identische empirische Sachverhalte in verschiedenen zeitlichen oder räumlichen Kontexten zu unterschiedlichen Beobachtungsprotokollen führen können. Wahrnehmung im allgemeinen und Beobachtung im besonderen ist keine passive Spiegelung von „Realität“, sondern ein Vorgang, dessen Ergebnis nur z.T. durch den wahrgenommenen empirischen Sachverhalt, z.T. aber auch durch dessen raum-zeitlichen Kontext bzw. die dadurch ausgelösten Informationsverarbeitungs- und Gedächtnisprozesse auf seiten der beobachtenden Person mitbestimmt wird (Epstein & Rock, 1960; Wallach, 1949). Man kann zwar versuchen, Beobachtungsbedingungen so zu konstruieren und Beobachtende so zu schulen, daß Kontexteinflüsse minimiert werden (vgl. Greve & Wentura, 1991, Kap.3). Wenn verschiedene Beobachtungsmethoden zur Beantwortung einer wissenschaftlichen Fragestellung simultan eingesetzt werden, gibt es darüber hinaus effiziente Methoden der Validitätsprüfung (vgl. Feger, 1983, Abschnitt 6). Allerdings ändert all dies nichts an dem grundsätzlichen Problem, daß Beobachtung nicht einfach „Spiegel der Realität“ sein kann.

Wissenschaftliche Beobachtungsprotokolle bilden also trotz ihrer Objektivität kein prinzipiell unanfechtbares, sicheres Fundament empirischer Wissenschaft. Popper (1935/1982, S. 75f.) formuliert treffend:

So ist die empirische Basis der objektiven Wissenschaft nichts ‚Absolutes‘; die Wissenschaft baut nicht auf Felsen Grund. Es ist eher ein Sumpfland, über dem sich die kühne Konstruktion ihrer Theorien erhebt; sie ist ein Pfeilerbau, dessen Pfeiler sich von oben her in den Sumpf senken - aber nicht bis zu einem natürlichen, ‚gegebenen‘ Grund. Denn nicht deshalb hört man auf, die Pfeiler tiefer hineinzutreiben, weil man auf eine feste Schicht gestoßen ist: wenn man hofft, daß sie das Gebäude tragen werden, beschließt man, sich vorläufig mit der Festigkeit der Pfeiler zu begnügen.

Treten Widersprüche zwischen einer Theorie und wissenschaftlichen Beobachtungsprotokollen auf, so ist die Verwerfung der Theorie also nicht notwendig die zwingende Konsequenz. Grundsätzlich denkbar und vertretbar ist immer eine konventionalistische Position (Dingler, 1931; in der Psychologie:

Holzcamp, 1964), bei der die theoretischen Aussagen „exhaustiert“ und die widersprechenden Beobachtungsaussagen verworfen werden. Derartige „konventionalistische Wendungen“ sind logisch einwandfrei; dies wird auch von entschiedenen Gegnern des Konventionalismus zugestanden (vgl. Popper, 1935/1982). Man kann ihnen allein durch den **methodologischen Beschluß** begegnen, sie nicht zur Widerspruchsbeseitigung heranzuziehen (Gadenne, 1984; Popper, 1935/1982). Dies bedeutet umgekehrt, daß auch wissenschaftliche Beobachtungsprotokolle letztlich nur **per Beschluß** Wirksamkeit erlangen können. Vom Charakter her ist ein solcher Beschluß immer vorläufig und prinzipiell revidierbar. Er kann auf keine Weise - weder logisch noch empirisch - erzwungen werden (Andersson, 1988).

## 1.2 Selbstbeobachtung mentaler Prozesse

Eine Teilklasse von Beobachtungen, für die immer wieder behauptet worden ist, daß ihr im Rahmen der Psychologie eine besondere Bedeutung zukommt, blieb bislang unerwähnt: Selbstbeobachtungen spontan auftretender oder experimentell induzierter mentaler Vorgänge und Zustände wie z.B. Gedanken, Vorstellungen, Empfindungen, Gefühle usw., die dann zumeist in Form verbaler Protokolle festgehalten werden. Für Bühler (1927/1978, S. 17) war Psychologie „(...) seit Descartes und Locke gedacht als die Wissenschaft von den **Erlebnissen**, als eine Theorie dessen, was der sogenannten inneren Wahrnehmung, der Selbstbeobachtung, zugänglich ist“. Folglich ist es „(...) eitel Selbsttäuschung, wenn man glaubt, irgendwo in der Psychologie ohne die Selbstbeobachtung auskommen zu können“ (Bühler, 1908, S. 101). Diese Auffassung hat in der Wissenschaftsgeschichte eine bis auf Aristoteles zurückgehende Tradition (vgl. Ericsson, 1990; Eysenck, 1990). In der Psychologie haben sich Persönlichkeiten wie z.B. Wilhelm Wundt, Edward B. Titchener, William James und Hubert Rohracher ganz ähnlich geäußert (zusammenfassend: Feger & Graumann, 1983, S. 80ff; Lyons, 1986, Kap. 1). Die Koppelung der Psychologie an die Methode der Selbstbeobachtung erschien u.a. Kant (z.B. 1804/1977a, S.648; 1798/1977b, S.401, 415, 417, 428) und Comte (1830-1842/1974, S. 11f.) so zwingend, daß sie glaubten, über die Kritik der Selbstbeobachtung bzw. die Reflexion ihrer Grenzen zugleich zeigen zu können, daß Psychologie als Wissenschaft zur Erfolglosigkeit verurteilt oder gar unmöglich ist (vgl. auch Bühler, 1908; Feger & Graumann, 1983, S. 82; Gadenne & Oswald, 1991, S. 61f; Lyons, 1986, S. 9ff.; Pongratz, 1984, S. 71f.). Ebenso wie Kant und Comte hielten auch Watson und Skinner eine mentalistisch orientierte wissenschaftliche Psychologie vor allem aufgrund der ihrer Meinung nach mit den Prinzipien einer objektiven Wissenschaft nicht verträglichen Selbstbeobachtungsmethode für unmöglich. Ihre radikale Konsequenz - der Behaviorismus (für Bühler, 1927/1978, S. 17, der Auslöser der „Krise der



Psychologie“) - beinhaltet im Kern eine Neudefinition von Psychologie als „Wissenschaft vom menschlichen Verhalten“ bei gleichzeitiger Verbannung aller mentalistischen Konzepte.

Vor allem zwei Fragen stellen sich in diesem Zusammenhang. Zunächst ist zu klären, ob Selbstbeobachtungen für eine Psychologie, die mentalistische Konzepte zuläßt, wirklich unverzichtbar sind. Ist eine selbstbeobachtungsfreie empirische Psychologie nur auf der Grundlage des Behaviorismus oder des Operationalismus (Bridgman, 1927) möglich? Gibt es einen dritten Weg zwischen Selbstbeobachtungsmethodik und klassischem Behaviorismus bzw. Operationalismus? Darüber hinaus muß untersucht werden, ob die Vorbehalte, die gegen die Methode der Selbstbeobachtung immer wieder vorgebracht wurden, auch im Lichte der Weiterentwicklungen, die diese Methode inzwischen erfahren hat, noch berechtigt sind.

Verschiedene Formen der Selbstbeobachtung sind zunächst zu unterscheiden. Soweit Selbstbeobachtung sich auf **eigenes Verhalten** bezieht (Groeben & Scheele, 1977, S.212ff; Teegen, Grundmann & Röhrs, 1975, Kap.4), ergeben sich keine prinzipiellen Probleme. Selbstverständlich sind Probleme der Reaktivität (Veränderung des Verhaltens durch den Prozeß der Selbstbeobachtung) und der Validität (Verfälschungen aufgrund der Tendenz, sich positiv darzustellen) a priori nicht auszuschließen. Diese Probleme können aber auch bei Verhaltensbeobachtungen durch eine andere Person auftreten und sind zudem mit empirischen Mitteln kontrollierbar. Methodologisch gibt es somit keine prinzipiellen Unterschiede zwischen Selbstbeobachtungen und Fremdbeobachtungen von Verhalten. Beide Beobachtungsformen können hinsichtlich Objektivität und Replizierbarkeit bewertet werden und erfüllen somit die Voraussetzungen wissenschaftlicher Beobachtung.

Komplizierter wird es, wenn sich Selbstbeobachtungen auf **mentale Vorgänge und Zustände** beziehen. Dies kann einerseits in Form introspektiver Betrachtungen von Bewußtseinszuständen und -Vorgängen, andererseits in Form protokollartiger Erlebnisdeskriptionen geschehen (vgl. auch Feger & Graumann, 1983). Die Unterscheidung von Introspektionen und Erlebnisdeskriptionen ist nützlich, da mit ihrer Hilfe einige Positionen zur Selbstbeobachtungsproblematik - etwa die von Wilhelm Wundt - besser zu verstehen sind, und andererseits die Unterschiede zwischen der Introspektionsmethode der Bewußtseinspsychologie und der Methode des lauten Denkens in der modernen Denkpsychologie besser herausgearbeitet werden können.

Introspektive Betrachtungen von Bewußtseinsvorgängen beschränken sich nicht auf die Artikulation von Bewußtseinsinhalten, sondern erläutern diese zugleich im Sinne eines Reflektierens über (abgeschlossene) Bewußtseinsvorgänge. Wundt (1907, 1908), der - wie schon erwähnt - bestimmte Formen der Selbstbeobachtung für unverzichtbar hielt (vgl. etwa Wundt, 1888/1921),

gehörte zugleich zu den schärfsten Kritikern der Introspektionsmethode, wie sie beispielsweise im Rahmen der Würzburger Schule verwendet wurde. Wundts (1907) Kritikpunkte an der „Ausfragemethode“ - wie er sie nannte - decken sich z.T. mit denen, die bereits Kant (1798/1977b, S. 417) und vor allem Comte (1830-1842/1974) zuvor in ähnlicher Weise formuliert hatten: (1) Spezifika der „Ausfragemethode“ (überraschendes Eintreten des zu beobachtenden Ereignisses, Anwesenheit eines Versuchsleiters, Aufmerksamkeitsablenkung) implizieren, daß die interessierenden Bewußtseinsinhalte nicht unverfälscht beobachtet werden können; (2) angestregtes Denken und gleichzeitige Selbstbeobachtung schließen einander gegenseitig aus; (3) Wiederholbarkeit und planmäßige Variierbarkeit der Selbstbeobachtungen sind im Rahmen der „Ausfragemethode“ nicht gegeben. Damit handelt es sich nach Wundt (1907) um „Scheinexperimente“, die den Kriterien eines psychologischen Experiments - für Wundt die maßgebliche Form wissenschaftlicher Beobachtung (vgl. Feger & Graumann, 1983) - nicht genügen. Bühler (1908) hat hierauf mit Klarstellungen geantwortet, so etwa der, daß Interferenzen zwischen Denk- und Selbstbeobachtungsvorgängen deshalb ausgeschlossen seien, weil Bewußtseinstatbestände immer **im Nachhinein** betrachtet wurden (Retrospektion)<sup>1</sup>. Doch schon die Replik von Wundt (1908) macht deutlich, daß die Einwände damit nicht hinfällig werden: Man könne nichts erinnern, was man zuvor nicht beobachtet habe (a.a.O., S.450); im übrigen seien „(...) Erinnerungstäuschungen Tür und Tor geöffnet“ (a.a.O., S.452). Siebzig Jahre später haben Nisbett und Wilson (1977) mit empirischen Mitteln gezeigt, daß diese kritische Haltung gerechtfertigt ist. Verbale Berichte über mentale Vorgänge sind z.T. inkonsistent mit Verhaltensbeobachtungen. Unabhängige Variablen (UVn) zeigen Effekte auf das Verhalten, die in der Selbstbeobachtung bestritten werden; umgekehrt werden in der Selbstbeobachtung Effekte von UVn behauptet, die sich anhand von Verhaltensindikatoren nicht verifizieren lassen.

Wie sehen die Formen der Selbstbeobachtung aus, die in Wundts Augen methodisch vertretbar sind? Für Wundt war die **experimentelle Kontrolle** der Selbstbeobachtung entscheidend. Er verstand hierunter, daß Selbstbeobachter (1) den Eintritt des zu beobachtenden Vorganges selbst bestimmen, (2) die Erlebnisse im Zustand „gespannter Aufmerksamkeit“ verfolgen, (3) die Beobachtungen wiederholen und schließlich (4) das Eintreten der Ereignisse durch planmäßige Variation der begleitenden Umstände herbeiführen, verstärken, abschwächen oder eliminieren können (Wundt, 1907, S. 308). Auf diesem Hintergrund sind z. B. Bestimmungen der Unterschiedsschwelle für verschiedene Reizintensitäten oder Untersuchungen zum Erleben der Müller-Lyer-

<sup>1</sup> Ähnlich gingen auch William James und Titchener in den USA vor, gelangten auf diese Weise jedoch zu Auffassungen über das Bewußtsein, die mit den Thesen der Würzburger Schule nicht harmonieren (vgl. Lyons, 1986, Kap. 1).

sehen Täuschung unter verschiedenen Kontextbedingungen ideale Experimente (vgl. Wundt, 1907). Anhand dieser Beispiele wird deutlich, daß Wundts experimentelle Selbstbeobachtungen eigentlich eher **Erlebnis- bzw. Empfindungsdeskriptionen** unter kontrollierten Randbedingungen entsprechen. Sie weisen damit zu den auch noch in der modernen Psychophysik gebräuchlichen Methoden der Erhebung von Wahrnehmungsurteilen eine größere Nähe auf als zu Selbstbeobachtungen im introspektionistischen Sinne des Wortes. Letztere wurden von Wundt immer abgelehnt (vgl. auch Danziger, 1980; Feger & Graumann, 1983, S. 82).

Können Erlebnisdeskriptionen, die Wundts Kriterien psychologischen Experimentierens genügen, als wissenschaftliche Beobachtungen im o. g. Sinne qualifiziert werden? Zweifellos fallen einige der Kritikpunkte, die gegen die Intro- und Retrospektionsmethode der Würzburger Schule sprechen, hier fort. So scheinen etwa Schmerzen oder der Geschmack eines Weines durch den Prozeß der Selbstbeobachtung nicht maßgeblich beeinflußt zu werden (Gadenne & Oswald, 1991, S. 62). Auf der anderen Seite kann aber nicht bestritten werden, daß Erlebnisdeskriptionen insofern immer subjektiv sein müssen, als sie sich auf nichtöffentliche und nichtempirische Tatbestände beziehen. Da die Objektivitätsforderung für den wissenschaftlichen Beobachtungsbegriff zentral ist, kann man also nicht von einer wissenschaftlichen Beobachtung mentaler Vorgänge oder Erlebnisse sprechen, jedenfalls nicht in dem Sinne, der in Abschnitt 1.1 spezifiziert wurde (vgl. auch Bunge & Ardila, 1987, S. 68).

Heißt das, daß eine Ausklammerung des Mentalen - wie von den Behavioristen gefordert - im Rahmen einer mit wissenschaftlichen Beobachtungsmethoden arbeitenden empirischen Psychologie letztendlich doch unvermeidlich ist? Das ist nicht der Fall. Man kann sich nämlich entschließen, die **verbalisierten Erlebnisse selbst** (also z. B. Wahrnehmungsurteile) als empirische und prinzipiell öffentliche Tatbestände zu betrachten, die dann einer wissenschaftlichen Beobachtung zugänglich sind<sup>2</sup>. Ausgangspunkt für die Erzeugung psychologischer Daten wären demnach nicht Erlebnisse, sondern artikulierte Erlebnisdeskriptionen. Zunächst mag es so scheinen, als würde man sich mit diesem „Trick“ an eine operationalistische Position binden, so daß mentale Vorgänge und Zustände zwar als psychologische Konzepte pro forma zugelassen sind, aber eben nur um den Preis, daß man bereit ist, sie letztlich **per definitionem** mit wissenschaftlich Beobachtbarem gleichzusetzen. Doch diese Konsequenz ist nicht zwingend, wie - um nur ein Beispiel von vielen möglichen zu nennen - die Signalentdeckungstheorie (SDT, vgl. Green & Swets, 1974) lehrt. Im Rahmen der SDT wird z.B. eine Geräuschempfindung nicht auf das Wahrnehmungsurteil „Ich höre etwas“ reduziert. Umgekehrt wird aus

2 Soeffner und Hitzler vertreten aus hermeneutischer Perspektive einen ganz ähnlichen Standpunkt (vgl. Kapitel 3 dieses Bandes).

einem solchen Urteil auch nicht abgeleitet, daß bei der Vp eine Geräuschempfindung vorliegt. Es ist sogar explizit vorgesehen, daß nichtsensorische Faktoren (Antworttendenzen) Einfluß auf das Wahrnehmungsurteil nehmen und daß die Stärke dieser Antworttendenzen zwischen verschiedenen Versuchsbedingungen in unbekanntem Ausmaß variieren kann. Obwohl man also keinerlei Annahmen über direkte Korrespondenzen zwischen Empfindungen und Wahrnehmungsurteilen macht, gelingt es mit Hilfe der SDT, Aussagen über die reine, von Antworttendenzen nicht überlagerte **sensorische Leistungsfähigkeit (Sensitivität)** der Vp abzuleiten, ein theoretisches Konzept, das sicherlich dem Bereich des Mentalen zuzurechnen ist<sup>3</sup>. Natürlich kann eine solche Aussage nur dann als fundiert gelten, wenn die Modellannahmen der SDT im konkreten Anwendungsfall erfüllt sind und Grund zu der Annahme besteht, daß die psychologische Interpretation der Modellparameter berechtigt ist. Ob diese Annahmen haltbar sind, kann im Prinzip immer mit empirischen Mitteln anhand der Häufigkeiten verschiedener Wahrnehmungsurteile unter variierenden experimentellen Bedingungen geprüft werden. Für die Frage der Modellgültigkeit ist es dagegen ohne Bedeutung, welche Beziehung zwischen dem SDT-Konstrukt „Sensitivität“ und der bewußten Lautstärkeempfindung der Vp besteht. Die SDT ist eine Theorie über sensorische und nichtsensorische Determinanten von Wahrnehmungsurteilen im Ja-Nein-Paradigma; „bewußte Lautstärkeempfindungen“ kommen in dieser Theorie nicht vor und können daher weder zur Bestätigung noch zur Widerlegung signalentdeckungstheoretischer Aussagen herangezogen werden. Prinzipiell denkbar ist also auch eine Dissoziation zwischen signalentdeckungstheoretischer und selbsteingeschätzter sensorischer Leistungsfähigkeit, was sich konkret darin ausdrücken würde, daß die Vp überrascht ist, in signalentdeckungstheoretischer Hinsicht besonders gut (oder besonders schlecht) abgeschnitten zu haben. Tatsächlich gibt es derartige Dissoziationen zwischen theoretischen Größen und bewußten Einschätzungen dieser Größen durch eine Vp recht häufig in der Psychologie. Die Diskrepanz zwischen registrierter und eingeschätzter Distanz bei der Mondtäuschung (Kaufman, 1979, S. 331) ist vielleicht eines der bekanntesten Beispiele.

Ich fasse zusammen. Das Problem, daß innere Vorgänge nicht wissenschaftlich beobachtbar sind, kann dadurch gelöst werden, daß (geschriebene oder gesprochene) **Erlebnisdiskriptionen** als zu beobachtende empirische Sachverhalte aufgefaßt werden. Dies impliziert keine antimentalistische Psychologie, sondern ist - wie am Beispiel der SDT erläutert wurde - mit der Verwendung mentalistischer Konzepte durchaus vereinbar. Die Beziehung zwischen diesen Konzepten und dem Bewußtsein bleibt hierbei offen. Insbesondere wird also

---

<sup>3</sup> Oswald (1980) würde hier von einer „kognitiven Struktur“ sprechen. Im Gegensatz zu Kognitionen sind kognitive Strukturen sensu Oswald (1980) prinzipiell nicht bewußtseinsfähig und damit der Selbstbeobachtung auch nicht zugänglich.

keine Gleichsetzung von Mentalem und Bewußtem vorgenommen<sup>4</sup>; das Bewußtsein läßt sich prinzipiell ganz ausklammern oder aber als Teil des Mentalen auffassen (vgl. dazu Bunge & Ardila, 1987; Gadenne & Oswald, 1991; Oswald, 1980). Auf diesem Hintergrund sind Selbstbeobachtungen von Bewußtseinsprozessen prinzipiell Verzichtbar; wenn sie durchgeführt werden, so geschieht dies nicht, weil ohne sie z.B. eine Wahrnehmungspsychologie nicht möglich wäre. Es geschieht im allgemeinen deshalb, um die Frage studieren zu können, welche Beziehung zwischen den Konstrukten psychologischer Theorien und bewußten Vorgängen besteht. So interessant die Antwort auf diese Frage sein mag: Im Hinblick auf die empirische Gültigkeit der Theorien, in die die betreffenden psychologischen Konstrukte eingebettet sind, ist sie völlig irrelevant<sup>5</sup>.

Ein naheliegender Einwand könnte nun lauten, daß das Beispiel der SDT nicht repräsentativ für das Gesamtgebiet der Psychologie ist. Vielleicht kann man Wahrnehmungspsychologie auf die skizzierte Art und Weise „selbstbeobachtungsfrei“ betreiben, weil das Konzept des Bewußtseins hier nicht von zentraler Bedeutung ist. Wie ist es aber z.B. in der Denkpsychologie? Für Wundt (z.B. 1905, S.28f.) waren die „höheren psychischen Vorgänge“ einerseits untrennbar mit dem Bewußtsein verknüpft, andererseits aber den von ihm favorisierten experimentellen Selbstbeobachtungen nicht zugänglich, jedenfalls nicht in methodisch vertretbarer Weise. Deshalb hielt er eine **experimentelle** Denkpsychologie für unmöglich<sup>6</sup>. Lediglich mit völkerpsychologischen Mitteln, d.h. mit der psychologischen Untersuchung der Geisteserzeugnisse einer Kultur, erschien ihm eine indirekte Untersuchung von Denktätigkeit auf kollektiver Ebene möglich: „Demnach verfügt die Psychologie, ähnlich der Naturwissenschaft, über **zwei** exakte Methoden: die erste, die experimentelle Methode, dient der Analyse der einfacheren psychischen Vorgänge; die zweite,

4 Auf die gleiche Position stößt man bereits in der Descartes-Kritik von Leibniz (vgl. Danziger, 1980, S. 242; Pongratz, 1984, S.36f.).

5 Die hier vertretene Auffassung fügt sich sehr gut in eine Bewußtseinskonzeption ein, die jüngst Gopnik (im Druck) vertreten hat. Gopnik stellt auf der Grundlage entwicklungspsychologischer Daten die Auffassung in Frage, daß das Wissen einer Person um den eigenen psychologischen Zustand auf direkter Erfahrung beruht. Denkbar ist - so Gopnik -, daß Wissen um eigene psychologische Zustände ebenso wie das Wissen um psychologische Zustände anderer Personen auf Inferenzen beruht. Die „illusion of direct perception“ (Gopnik, im Druck) ist dann lediglich ein Epiphänomen der Expertise. die die „erste Person“ beim Erschließen eigener psychologischer-Zustände im Verlauf der Ontogenese erwirbt. Illusionen direkter Wahrnehmung sind in vielen Bereichen mit Expertise korreliert. Auf dem Hintergrund dieser Konzeption werden „Bewußtseinstatbestände“ als Kriterien psychologischer Theorieevaluation fragwürdig. Bewußtseinstatbestände können falsch in dem Sinne sein, daß ihnen Inferenzfehler zugrunde liegen bzw. der Inferenzcharakter von Bewußtseinstatbeständen gezeugnet wird (vgl. auch Oswald, 1980).

6 Auf eine ähnliche Position - wenn auch mit einer etwas anderen Begründung - trifft man u.a. in der Philosophie Wittgensteins (vgl. dazu Chapman, 1987).

die Beobachtung der allgemeinen Geisteserzeugnisse, dient der Untersuchung der höheren psychischen Vorgänge und Entwicklungen“ (Wundt, 1905, S. 29).

Wundts Skepsis in bezug auf die Möglichkeit experimenteller Denkpsychologie wurde nur von wenigen experimentell orientierten Psychologen geteilt. Vielen schwebte eine Lösung der methodischen Probleme durch eine Optimierung der Introspektionsmethodik vor, und zwar in die Richtung, die Claparede (1932/1965) mit seiner „Technik lauten Denkens“ beschritten hat (für einen Überblick vgl. Kluwe, 1988). „Lautes Denken“ beschränkt sich im Gegensatz zu Introspektion auf die Artikulation von Gedanken ohne jede Erläuterung ihres Zusammenhanges oder Hintergrundes. Im Grunde genommen handelt es sich hierbei ebenfalls um eine Form der Erlebnisdeskription (s. o.), wobei die Erlebnisse hier der Natur nach nicht Empfindungen, sondern Gedanken sind. Claparede (1932/1965) selbst sprach sogar von einer „(...) behavioristischen Methode, die den Ablauf des inneren Verhaltens zu bestimmen versucht“ (deutsche Übersetzung, S. 110).

Die moderne und theoretisch fundierte Version dieser Technik stammt von Ericsson und Simon (1980, 1984). Bemerkenswert ist an diesem Ansatz zunächst, daß nur solche Anwendungen lauten Denkens zugelassen werden, für die sich kognitionspsychologisch begründet vermuten läßt, daß sie nichtreaktiv und valide sind. Lediglich Artikulationen phonologisch kodierter Inhalte des Arbeitsgedächtnisses erfüllen nach Ericsson und Simon diese Bedingung. Retrospektionen, die zwangsläufig auf Inhalte des episodischen Langzeitgedächtnisses Bezug nehmen müssen, werden explizit verworfen, da einerseits nur ein Teil der Inhalte des Arbeitsgedächtnisses langfristig gespeichert wird und andererseits neben Reproduktions- auch Rekonstruktionsprozesse eine (validitätssenkende) Rolle spielen können. Ferner werden Aufgaben, die die Umkodierung eines nichtverbalen (z. B. imaginalen) Repräsentationsformates verlangen, ebenso ausgeschlossen wie Aufgaben, deren Bearbeitung (z.B. aufgrund weitgehend automatisierter kognitiver Prozesse) keine oder nur eine partielle Beanspruchung des Arbeitsgedächtnisses erfordert.

Eine zweite Besonderheit des Ansatzes von Ericsson und Simon, die ihn von älteren Ansätzen in markanter Weise abhebt, ist der **empirische Gehalt** der zugrundeliegenden Annahmen: Der Ansatz basiert auf einer kognitionspsychologischen Theorie, die empirisch prüfbare Konsequenzen hat. Man muß also nicht an die Annahmen von Ericsson und Simon glauben, sondern kann sie mit empirischen Mitteln kritisch hinterfragen. Diesen Weg haben jüngst u.a. Russo, Johnson und Stephens (1989) beschritten. In ihrer Untersuchung ergaben sich einerseits Ergebnisse, die nach Ericsson und Simon (1980, 1984) zu erwarten sind, beispielsweise erhebliche Diskrepanzen zwischen parallel zur Aufgabebearbeitung erhobenen Laut-Denken-Protokollen und retrospektiven Protokollen. Dies unterstreicht einmal mehr die fragwürdige Vali-

dität retrospektiver Protokolle. Andererseits ergaben sich aber auch Ergebnisse, die mit den Annahmen von Ericsson und Simon nicht harmonieren. So erwies sich lautes Denken überraschend als nichtreaktiv bei Aufgaben, die mutmaßlich eine Umkodierung des Repräsentationsformates verlangen (Matrizen-Test nach Raven), während Aufgaben, deren Bearbeitung eigentlich nicht substantiell mit simultaner Protokollerzeugung interferieren sollte, tatsächlich erheblich hierdurch beeinflußt wurden. Deffner (1989) gelangte auf ähnlichem Wege teilweise zu bestätigenden, teilweise zu unerwarteten Befunden auf dem Hintergrund der Ericsson-Simon-Theorie. Das Fazit kann daher vorerst nur lauten, daß eine empirisch gut abgesicherte Theorie der Erzeugung von Laut-Denken-Protokollen trotz des innovativen und vielversprechenden Ansatzes von Ericsson und Simon (1980, 1984) gegenwärtig noch nicht vorliegt. Dies heißt aber nichts anderes, als daß nach wie vor unklar ist, welche gültigen psychologischen Schlüsse aus Laut-Denken-Protokollen gezogen werden können. Ericsson (1990, S. 195) hält es immerhin für möglich, daß mit ihrer Hilfe „(...) the experimenter can reject incorrect hypotheses about the thought processes“. Dabei setzt er offenkundig voraus, daß die Protokolle die „wirklichen“, vom Prozeß des lauten Denkens unbeeinflußten kognitiven Prozesse reflektieren. Diese Auffassung muß im Lichte der o.g. Ergebnisse mit Skepsis betrachtet werden. Dies schließt jedoch nicht aus, daß Weiterentwicklungen des Ericsson-Simon-Ansatzes sowie die Einhaltung gewisser Vorsichtsmaßnahmen, wie sie z.B. Deffner (1989, S. 109f.) vorgeschlagen hat, zu einer sehr fruchtbaren Verwendung der Methode des lauten Denkens führen können. Absehbar ist allerdings, daß man sich im Kontext der Prüfung kognitionspsychologischer Theorien und Hypothesen auch in Zukunft nicht allein auf Laut-Denken-Protokolle stützen kann.

Festzuhalten bleibt, daß Vorbehalte gegen die Selbstbeobachtungsmethodik auch dann nicht vollständig aus dem Wege geräumt werden können, wenn man die neuesten methodischen Entwicklungen in diesem Bereich berücksichtigt. Die Lücke, die hieraus resultiert, kann allerdings im Rahmen der Denkpsychologie in ähnlicher Weise geschlossen werden wie das für die Wahrnehmungspsychologie schon am Beispiel der SDT erläutert wurde. Voraussetzung ist auch hier, daß man bereit ist, die Gleichsetzung von kognitiven Prozessen mit Bewußtseinsprozessen aufzugeben. Tatsächlich spricht für eine solche Position einiges (vgl. Eysenck, 1990; Gadenne & Oswald, 1991; Oswald, 1980). Die Existenz von verhaltensdeterminierenden mentalen Prozessen, denen kein Bewußtseinsinhalt korrespondiert - z.B. „Blindsight“ (Weiskrantz, 1990), implizite Gedächtnisleistungen, implizites Lernen, unterschwellige Wahrnehmungen und automatische Verarbeitungsprozesse - kann schwerlich bestritten werden. Darüber hinaus gibt es - wie schon die erwähnte Arbeit von Nisbett und Wilson (1977) zeigt - klare empirische Evidenz dafür, daß verbalisierte Bewußtseinsinhalte verhaltensleitende mentale Prozesse falsch oder zumindest

irreführend darstellen können (vgl. auch Oswald, 1980). Bredenkamp (1990) demonstrierte beispielsweise eine Inkonsistenz zwischen den verbalisierten Selbstbeobachtungen eines Rechenkünstlers (G. M.) hinsichtlich dessen Vorgehen bei der Bearbeitung bestimmter Aufgaben und den Ergebnissen der Analyse von G.M.s Antwortzeiten für verschiedene Rechenaufgaben. Interessanterweise änderte G. M. im Lichte der Antwortzeit-Ergebnisse seine Meinung über sein eigenes Vorgehen (Bredenkamp, 1990, S.66). Dies macht eindringlich deutlich, daß die kognitionspsychologische Validität von Selbstbeobachtungen unter bestimmten Umständen sogar von den Personen post hoc in Frage gestellt wird, die sie vorgenommen haben.

Für den Bereich der Konzeptidentifikationsforschung - die Vpn müssen hierbei unter Rückmeldebedingungen eine Reihe von Stimuli als positive oder negative Exemplare eines unbekanntes Begriffs klassifizieren - hat Wilson (1975) in ähnlicher Weise gezeigt, daß verbale Vpn-Berichte Schlüsse auf die Konzept-Hypothesen, die das Klassifikationsverhalten der Vpn steuern, nicht zulassen. Die Dissoziation zwischen Verhalten und verbalen Berichten ist in diesem Bereich besonders verblüffend, da einerseits schwer vorstellbar ist, daß das Testen von Hypothesen über den gesuchten Begriff außerhalb des Bewußtseins abläuft, und andererseits die Bedingungen für nichtreaktive und valide Laut-Denken-Protokolle nach Ericsson und Simon (1980, 1984) optimal sein müßten.

Wie sich trotz dieser Probleme unter Ausklammerung der Selbstbeobachtung und des lauten Denkens ein fundierter Schluß auf die verhaltensleitende „mentale Hypothese“ einer Vp bewerkstelligen läßt, hat Levine (1966) in einem einfachen, aber sehr einfallreichen Experiment gezeigt. Er baute in sein Experiment sog. „blank trials“ ein, d.h. Folgen von jeweils vier Durchgängen, in denen Vpn Stimuli als Exemplare oder Nichtexemplare des gesuchten Begriffs klassifizieren mußten, ohne Rückmeldung über die Korrektheit der Klassifikationen zu bekommen. Da die Vpn in jedem Durchgang zwei Antwortmöglichkeiten hatten („Exemplar“ oder „kein Exemplar“), gab es insgesamt  $2^4 = 16$  mögliche Antwortmuster über die vier Durchgänge. Die vier Stimuli waren zuvor so ausgewählt worden, daß ein bestimmtes der 16 möglichen Antwortmuster resultieren mußte, falls die betreffende Vp ihre Klassifikationen jeweils auf der Grundlage ein und derselben mentalen Hypothese vornahm. Da acht Hypothesen den Vpn gegenüber als möglich ausgewiesen worden waren, durften bei Gültigkeit der Grundannahme von Levine (1966) auch nur acht der 16 Antwortmuster vorkommen. Dies wurde von Levine (1966) überprüft und - mit unerheblichen Abstrichen - auch bestätigt. Somit konnte Levine begründet von den Antwortmustern der Vpn auf die zugrundeliegenden mentalen Hypothesen schließen, die - wenn seine Grundannahme, daß in den „blank trials“ konsistent eine mentale Hypothese verwendet wird, wirklich zutrifft - die Antwortmuster generiert haben **müssen**. Natürlich



ist die Möglichkeit eines Irrtums hierbei nicht ausgeschlossen, weil nur eine notwendige, nicht aber eine hinreichende Bedingung für die Gültigkeit der Grundannahme geprüft wurde. Aber dieses Problem ist in den empirischen Wissenschaften häufig nur schwer zu umgehen (vgl. auch Gadenne, Kapitel 8 dieses Bandes). Verglichen mit den Validitätsproblemen, die Selbstbeobachtungen aufwerfen, erscheint es vertretbar, diese Unsicherheit in Kauf zu nehmen.

Am Beispiel des Vorgehens von Levine kann sehr schön dargestellt werden, wie das generelle Prinzip der selbstbeobachtungsfreien Erschließung mentaler Zustände und Prozesse aus wissenschaftlich beobachtbaren Sachverhalten aussieht. Voraussetzung ist zunächst immer eine **Theorie**, mittels der die interessierenden mentalen Zustände und Prozesse als theoretische Konstrukte eingeführt werden. Diese Theorie muß Aussagen darüber enthalten, in welcher Beziehung die verschiedenen mentalen Zustände oder Prozesse zu beobachtbaren Sachverhalten stehen. Es muß also eine **Abbildung  $f$**  einer Menge theoretischer Sachverhalte (z. B. eine Menge möglicher „mentaler Hypothesen“) in eine Menge wissenschaftlich beobachtbarer Sachverhalte (t. B. eine Menge möglicher Antwortmuster) konstruiert werden, und diese Abbildung muß **injektiv** sein, d.h. verschiedenen theoretischen Zuständen müssen auch unterschiedliche beobachtbare Zustände zugeordnet werden. Darüber hinaus sollte die Abbildung möglichst **nicht surjektiv** sein, so daß bestimmte Elemente aus der Menge möglicher beobachtbarer Sachverhalte ausgeschlossen werden. Das macht die gesamte Theorie zu einer empirisch testbaren, so daß über ihre Annahme oder Verwerfung mit empirischen Argumenten entschieden werden kann. Erweist sich die Theorie als adäquat, so kann die zu  **$f$**  inverse Abbildung der zulässigen empirischen Sachverhalte auf die theoretischen Sachverhalte betrachtet werden, die dann die Erschließung des mentalen Zustands einer beobachteten Person erlaubt.

Wenn eine wissenschaftliche Fragestellung vorliegt, die auf mentale Zustände oder Prozesse abzielt, sollte bei der Auswahl geeigneter empirischer Indikatoren also nicht die mutmaßliche Nähe dieser Indikatoren zu Bewußtseinsinhalten (Gedanken, Gefühlen etc.) ausschlaggebend sein, sondern die Verfügbarkeit einer empirisch testbaren Theorie, die angibt, welche gültigen Schlüsse aus empirischen Sachverhalten im Hinblick auf mentale Vorgänge und Zustände zu ziehen sind. So kann es durchaus sein, daß dem Eindruck nach „bewußtseinsnahe“ empirische Materialien - Laut-Denken-Protokolle etwa - kaum brauchbar sind, weil keine bewährte Theorie vorliegt, die angibt, wie diese Materialien gültig zu interpretieren sind. Auf der anderen Seite können scheinbar „bewußtseinsferne“ Sachverhalte - z.B. Antwortzeiten oder Augenbewegungen (vgl. Lürer, 1988) - hochinformativ bezüglich mentaler Vorgänge sein, wenn nämlich empirisch gehaltvolle und gut bestätigte Theorien existieren, die Aussagen darüber erlauben, welche kognitiven Prozesse die Antwortzeiten oder Blickbewegungsfolgen im einzelnen generiert haben.

Ich komme zum Fazit dieses Abschnitts. Zunächst ist festzuhalten, daß die Selbstbeobachtungsmethodik seit Beginn des Jahrhunderts zweifellos erheblich verbessert wurde. Das Attribut „Methode“, das Bunge und Ardila (1987, S.63) der Introspektion absprechen, kann mit Bezug auf die Technik lauten Denkens inzwischen durchaus verwendet werden. Hervorzuheben ist vor allem, daß eine theoretische Fundierung der Selbstbeobachtungsmethodik auf dem Hintergrund kognitionspsychologischer Erkenntnisse inzwischen immerhin möglich erscheint. Dennoch sind Vorbehalte nach wie vor angebracht, da bis dato keine empirisch gut bewährte Theorie der Erzeugung von Selbstbeobachtungsprotokollen vorliegt. Der Ansatz von Ericsson und Simon (1980, 1984) kommt dem Ziel gegenwärtig am nächsten, aber - wie insbesondere die Arbeit von Russo et al. (1989) zeigt - nicht nah genug.

Auf der anderen Seite muß betont werden, daß Selbstbeobachtung nicht den zentralen Stellenwert für die Psychologie hat, der ihr von unterschiedlichen Seiten immer wieder zugesprochen wurde. Eine selbstbeobachtungsfreie empirische Psychologie ist prinzipiell möglich, ohne daß damit eine antimentalistische Position erzwungen wird. Dies gilt auch für die Denkpsychologie, den Bereich also, der auf das engste mit Bewußtseinsvorgängen verknüpft zu sein scheint. Das besprochene Experiment von Levine (1966) zeigt exemplarisch, daß fundierte Aussagen über kognitive Zustände und Prozesse möglich sind, ohne dabei Selbstbeobachtungsprotokolle heranzuziehen.

Dies muß nun nicht bedeuten, daß verbale Berichte über mentale Zustände und Vorgänge solange aus der empirischen Psychologie fernzuhalten sind, bis eine bewährte Theorie zur Erzeugung von Selbstbeobachtungsprotokollen vorliegt. Verbale Berichte sind als empirische Tatbestände prinzipiell der wissenschaftlichen Beobachtung zugänglich und können somit den Ausgangspunkt für die Erzeugung von Daten bilden, auf die sich dann psychologische Theorien beziehen können. Dies ist so lange unproblematisch, wie die verbalen Protokolle als das betrachtet werden, was sie sind: Beobachtungsprotokolle über **Selbstbeschreibungen** innerer Vorgänge. Diese Perspektive wird z. B. in einigen Forschungsprogrammen eingenommen, in denen die subjektive Sicht der Vp selbst Forschungsgegenstand ist (vgl. z.B. Groeben & Scheele, 1977). Probleme entstehen erst dann, wenn Selbstbeschreibungen mit den beschriebenen inneren Vorgängen explizit oder implizit gleichgesetzt werden. Eine solche Gleichsetzung entbehrt nicht nur einer methodologischen Rechtfertigung, sondern ist zudem mit einer Reihe von empirischen Befunden nicht zu vereinbaren. Selbstbeschreibungen mentaler Vorgänge und die mentalen Vorgänge selbst sind unterschiedliche Sachverhalte, die strikt auseinanderzuhalten sind. Es gibt keinen zwingenden Grund, von einer kognitionspsychologischen Theorie zu verlangen, daß sie Selbstbeschreibungen kognitiver Prozesse erklärt. Umgekehrt muß eine Theorie über Selbstbeschreibungen kognitiver Prozesse - eine Theorie über „subjektive Theorien“ also - nicht not-

wendig Bezüge zu Konstrukten der Kognitionspsychologie aufweisen. Bezüge und Korrespondenzen zwischen beiden Theorieklassen können bestehen, müssen es aber nicht. Häufig sind es gerade die Dissoziationen zwischen kognitionspsychologischen Theorien und subjektiven Theorien, die zu den bedeutenderen - weil überraschenden - Erkenntnissen gehören (Eysenck, 1990).

### 1.3 Heuristische Selbst- und Fremdbeobachtung

Eine denkbare Funktion von Selbstbeobachtungen in der Psychologie blieb im letzten Abschnitt unerwähnt: die heuristische Funktion. Selbstbeobachtungen können im „context of discovery“ (Reichenbach, 1938) insofern eine wichtige Rolle spielen, als sie - z. B. über interpretative Zugänge (vgl. Soeffner & Hitzler, Kapitel 3 dieses Bandes) - psychologische Hypothesen nahelegen, die dann anhand empirischer Daten, die aus wissenschaftlichen Beobachtungen im Sinne von Abschnitt 1.1 gewonnen wurden, überprüfbar sind (vgl. auch Dörner, Kapitel 8 dieses Bandes). So können etwa subjektiv erlebte optische Täuschungen oder Unterschiedsempfindungen zwischen verschiedenen Formen einer Täuschung der Anlaß sein, eine neue Hypothese zu den Determinanten und Mechanismen optischer Täuschungen zu formulieren. Selbstverständlich kann das subjektive Erleben der Person, die die neue Hypothese erfunden hat, im Rahmen einer empirischen Wissenschaft kein zulässiges Argument für die neue Hypothese sein. Daran ändert sich grundsätzlich nichts, wenn ihm oder ihr andere Personen bestätigen, daß sie die Täuschungen genauso erleben. Globale, komplexe Empfindungsurteile sind keine geeigneten Primitiva, auf denen eine empirische Fundierung neuer Hypothesen aufbauen könnte; dies gilt um so mehr, wenn die Urteile sich auf nur wenige und nicht systematisch variierte Stimuli beziehen. Bekannte Artefaktquellen bleiben hierbei unkontrolliert: Versuchsleitereffekte, möglicherweise unterschiedlicher Sprachgebrauch von Vp und VI, Unklarheit über die Replizierbarkeit unter Doppelblindbedingungen, „demand characteristics“ usw. (vgl. hierzu Barber, 1976). Die subjektive Sicht des Hypothesenerfinders oder der Hypothesenerfinderin sowie informelle Kommentare anderer Personen sind bestenfalls als Plausibilitätsargumente im „context of discovery“ zu werten, die mit triftigen empirischen Argumenten im „context of justification“ nicht verwechselt werden dürfen. Dies gilt auch für Demonstrationen von optischen Täuschungen oder anderen Wahrnehmungseffekten in Lehrbüchern, Lehrveranstaltungen usw.; es handelt sich um **Illustrationen**, die ausschließlich didaktische Ziele verfolgen, keineswegs aber um stichhaltige empirische Argumente für oder gegen eine bestimmte Hypothese oder Theorie. Solange keine konsequente empirische Evaluation einer Hypothese erfolgt ist - basierend auf akzeptierten Methoden der Erzeugung empirischer Daten (im Beispiel etwa Paarvergleichs- oder Matchingurteile über systematisch variierte und replizierte Stimuli) -

kann über die empirische Adäquatheit oder Inadäquatheit der neuen Hypothese nichts ausgesagt werden.

Daß Fremdbeobachtungen ebenso wie Selbstbeobachtungen eine heuristische Funktion haben können und eine Reihe von psychologischen Hypothesen und Theorien vermutlich auch auf diese Weise entstanden sind, kann kaum bestritten werden. Eine andere Frage ist, inwieweit gelungene heuristische Beobachtungen - und in der Folge: gegenstandsadäquate Beschreibungen im Detail - sozusagen den Charakter eines Nadelöhrs haben, durch das man hindurch muß, wenn man zu einer adäquaten und erfolgreichen psychologischen Theorie gelangen will. Ist Beobachtung in diesem Sinne eine verbindliche Heuristik? Müssen psychologische Theorien aus langwierigen und detaillierten Beobachtungen des interessierenden Gegenstandes erwachsen? Diese Position wird gerade in jüngerer Zeit wieder häufiger vertreten (siehe z.B. Bischof, 1989; Dörner, 1989; Dörner & Lantermann, 1991; Foppa & von Cranach, 1991).

Man kann die Position, daß heuristische Beobachtung eine **Voraussetzung** adäquater Theorienentwicklung ist, auf zweierlei Weise explizieren: als **normative** Position - dann wäre zu fragen, welche methodologischen Argumente für diese Position sprechen - oder als **deskriptive** Aussage, die sich auf die konkrete Forschungspraxis in der Psychologie bezieht. Als deskriptive Behauptung dürfte die Position allerdings kaum haltbar sein, da sich viele erfolgreiche Hypothesen und Theorien der modernen Psychologie nicht angemessen als aus heuristischen Beobachtungen gewonnene Ideen verstehen lassen. Gigerenzer (1988, 1991) hat an vielen Beispielen aus der kognitiven Psychologie, der Wahrnehmungspsychologie und der Sozialpsychologie deutlich gemacht, daß häufig die **Werkzeuge** empirisch arbeitender Psychologinnen und Psychologen - insbesondere die Statistik und der Computer - die Heuristik bilden, aus der neue Theorien erwachsen'. Gigerenzer (1991) nennt dies die „tools-to-theories“-Heuristik. Grundlegend ist offenbar in vielen Fällen die Idee, daß Menschen im Alltag intuitiv genauso vorgehen wie Wissenschaftler, die bestimmte Fragen beantworten oder Hypothesen testen wollen und geeignete statistische Methoden sowie anderweitige methodische Hilfsmittel hierzu heranziehen. Es handelt sich also um eine Form der Analogiebildung, die sehr häufig mit kreativen Leistungen - speziell mit wissenschaftlicher Kreativität - gekoppelt zu sein scheint (Eysenck & Keane, 1990, p.399-403; vgl. auch Dörner, Kapitel 8 dieses Bandes). Die Forschungsprogramme, die durch die „tools-to-theories“-Heuristik geschaffen wurden, können im allgemeinen als erfolgreich, d.h. als „progressiv“ im Sinne von Lakatos (1970) charakterisiert werden. Da ein direkter Bezug dieser Forschungsprogramme zu

---

7 Die bereits erwähnte Signalentdeckungstheorie, welche den mit Reizentdeckungs- oder Rekognitionsproblemen konfrontierten Menschen letztlich als intuitiven Neyman-Pearson-Statistiker konzipiert, ist ein besonders prägnantes Beispiel.

heuristischen Beobachtungen nicht erkennbar ist, scheint damit der Nachweis erbracht, daß erfolgreiche Theoriebildung auch unter Umgehung heuristischer Beobachtungen möglich ist. Die Psychologie zeichnet sich also nicht dadurch aus, daß Theorien und Hypothesen aus der Anwendung empirischer Methoden - insbesondere der heuristischen Beobachtung - erwachsen, d.h. „empirisch gefunden“ werden. Sie zeichnet sich vielmehr dadurch aus, daß **wie auch immer erfundene** Theorien mit empirischen Mitteln **geprüft** werden. Um es einmal mehr in der Terminologie Reichenbachs (1938) auszudrücken: **Verbindlich** sind empirische Methoden allein im Begründungskontext, nicht im Entdeckungskontext.

Dies bedeutet nun nicht, daß die empirische Psychologie von einer intensiveren Nutzung heuristischer Beobachtungen nicht profitieren könnte. Es kann nicht von Nachteil sein, möglichst viele verschiedene Quellen konsequent heuristisch zu nutzen. Dies gilt insbesondere für **alle** Phasen des Forschungsprozesses. Hierzu zählen nicht nur die statistischen Werkzeuge der Forschenden und Selbst- sowie Fremdbeobachtungen; der Prozeß der Datenerzeugung, die Daten selbst sowie Formen der Datenanalyse können und sollten auch aus heuristischer Perspektive betrachtet werden. Was im einzelnen hierunter zu verstehen ist, wird in Abschnitt 2 noch deutlicher werden.

Eine liberale Grundhaltung gegenüber denkbaren Vorgehensweisen im „context of discovery“ garantiert, daß potentiell fruchtbare Ansätze der Theoriegenerierung immer eine Chance haben. Restriktive normative Prinzipien - z.B. das Postulat, daß Theorien aus heuristischen Beobachtungen erwachsen **müssen** - können dagegen dazu führen, daß fruchtbare Ansätze ungenutzt bleiben. Daß eine liberale Methodologie des Entdeckungskontextes auch mancherlei Unbrauchbares forcieren wird, sei zugestanden. Aber man kann es empirischen und nichtempirischen Methoden der Theorieevaluation überlassen, post hoc die Spreu vom Weizen zu trennen (vgl. dazu Kapitel 9 und 14 dieses Bandes). Solange Heuristiken gewählt werden, die alternative fruchtbare Zugänge nicht unmöglich machen (vgl. dazu Abschnitt 2.1), gibt es keinen Grund, die Heuristiken in „bessere“ und „schlechtere“ oder gar „verbindliche“, „mögliche“ und „verbotene“ einzuteilen. „Anything goes“ (Feyerabend, 1975/1983, Kap. 1) ist im Entdeckungszusammenhang ein methodologisch vertretbarer und substanzwissenschaftlich fruchtbarer Grundsatz.

## 1.4 Von wissenschaftlichen Beobachtungen zu Daten

Protokollierte Beobachtungen werden oftmals mit empirischen Daten gleichgesetzt. Diese Auffassung ist als definatorische Setzung natürlich prinzipiell durchsetzbar, schafft aber insofern Probleme, als sie fälschlich suggeriert, daß Datenanalyseverfahren sich unmittelbar auf Beobachtungsprotokolle als „in-

put“ beziehen. Wer Daten und Beobachtungsprotokolle gleichsetzt, wird häufig Schwierigkeiten mit der korrekten Interpretation von Datenanalyseresultaten haben. Außerdem werden die vielen alternativen Möglichkeiten, aus Beobachtungsprotokollen „input“ für Datenanalyseverfahren zu gewinnen, nicht erkannt; sie bleiben demzufolge heuristisch ungenutzt. Darüber hinaus kann es leicht passieren, daß ein Weg der Datenerzeugung gewählt wird, der relativ zur forschungsleitenden Fragestellung inadäquat ist, weil er die wichtigsten empirischen Aspekte ausklammert. Welchen „Blick“ auf die Beobachtungen Datenanalytiker erhalten, hängt bei Nichtreflexion der Beziehung zwischen Datenanalyseeingabe und Beobachtungsprotokollen vom Zufall bzw. von den „default options“ des Computerprogramms ab, das zur Datenanalyse herangezogen wird.

Vor allem die Datentheorie (Coombs, 1964; Coombs, Dawes & Tversky, 1970/1975; Roskam, 1983, im Druck) beschäftigt sich mit den unterschiedlichen Möglichkeiten, aus protokollierten Beobachtungen Daten für weitere Analyse Zwecke zu gewinnen. Wichtig ist zunächst die Einsicht, daß die Beobachtungen selbst nicht festlegen, was im folgenden zu analysieren ist. Anders als manche Handbücher zu Datenanalyseprogrammen es suggerieren, gehört zu einem bestimmten Beobachtungs- oder Untersuchungsparadigma kein festes System der Datenerzeugung und erst recht kein bestimmtes Analyseverfahren. Man kann im wesentlichen vier Klassen von Problemen unterscheiden, die beim Übergang von Beobachtungen zu Daten zu lösen sind. Sie werden hier als **Klassifikations-, Indizierungs-, Interpretations- und Quantifizierungsprobleme** bezeichnet. Erst wenn zu jeder Frage, die sich in diesen Zusammenhängen stellt, eine befriedigende Antwort vorliegt, kann eine aussagekräftige Datenanalyse erfolgen. Die vier Problemklassen werden im folgenden der Reihe nach besprochen.

### 1.4.1 **Klassifikationsprobleme**

**Klassifikationsprobleme** stellen sich immer dann, wenn die protokollierten Beobachtungen zunächst Beobachtungskategorien zuzuweisen sind, bevor eine weitergehende Datenanalyse erfolgen kann. Dies ist z.B. sehr häufig bei Verwendung von Interviews, Fragebögen mit offenen Antwortalternativen, projektiven Tests usw. der Fall. Unmittelbar beobachtet werden hierbei konkrete (gesprochene oder geschriebene) Sätze, Wörter oder Texte. Von Interesse ist aber lediglich, in welche theoretisch vorgegebene Kategorie verbalen Materials sie fallen. Ein Beispiel ist das von Heckhausen (1963) vorgeschlagene Kategoriensystem zur Auswertung der Antworten auf einige Bilder des Thematischen-Apperzeptions-Tests (TAT). Für Datenanalysen zum Leistungsmotivationskonstrukt „Hoffnung auf Erfolg“ ist es demnach allein relevant, wieviele

„Erfolgsthemen“ eine Person erwähnt hat, wobei „Erfolgsthemen“ in einem ausführlichen Anhang des Buches von Heckhausen (1963) extensional charakterisiert werden. Das ursprüngliche Beobachtungsprotokoll - insbesondere also der genaue Wortlaut der „Erfolgsthemen“ - ist dagegen für die weitergehende Datenanalyse ohne Bedeutung.

Coombs (1964) und Roskam (1983) behandeln Klassifikationsprobleme im Rahmen der Datentheorie vermutlich deshalb nicht explizit, weil sie diesen Problemkreis eher den Beobachtungsmethoden zurechnen. Wo man genau die Grenze zwischen Beobachtungsmethoden und Datenerzeugung aus Beobachtungen zieht, ist letztendlich willkürlich. Entscheidend ist, daß sich Klassifikationsprobleme irgendwo auf dem Wege von empirischen Sachverhalten zu empirischen Daten stellen können. Sie lassen sich prinzipiell immer in unterschiedlicher Weise beantworten und verlangen deshalb eine **explizite Formulierung** und **Begründung** des favorisierten Lösungsvorschlags.

### 1.4.2 Indizierungsprobleme

Klassifikationsprobleme werden selten übersehen, weil ihre Lösung - sei sie nun gut oder schlecht - eine Voraussetzung dafür ist, daß man überhaupt zu analysierbaren Daten gelangt. Anders sieht es mit Interpretations- und Indizierungsproblemen aus. Sie werden oftmals als solche gar nicht erkannt, weil die verbreiteten Datenanalyseprogramme „default options“ für ihre Lösung anbieten, die häufig unkritisch akzeptiert werden. Das **Indizierungsproblem** besteht darin, den Datenraum, d.h. die Menge möglicher empirischer Daten, sowie die Zuordnung von Beobachtungsprotokollen (bzw. Beobachtungskategorien) zu empirischen Daten genau zu definieren. Wenn im Beobachtungsprotokoll steht, daß eine Vp ein Wort in einem Gedächtnistest korrekt rekogniziert hat, ist nicht ohne weiteres klar, welche **Struktur** das empirische Datum haben muß, das diese wissenschaftliche Beobachtung abbildet. Man kann z.B. schlicht feststellen, daß ein Wort rekogniziert wurde. Dann hat man sich implizit auf einen Datenraum mit nur zwei Elementen „Wort wurde rekogniziert“ versus „Wort wurde nicht rekogniziert“ festgelegt. Man kann aber auch feststellen, daß Person **i** ein Wort rekogniziert hat oder Person **i** das Wort **j** rekogniziert hat, wobei **i** und **j** Elemente aus bestimmten Mengen darstellen. Falls **n** Personen und **m** Items untersucht wurden, hat der Datenraum im letzteren Fall  $n \cdot m \cdot 2$  Elemente, wobei jedes Element ein Tripel (**i**, **j**, rekogniziert) bzw. (**i**, **j**, nicht rekogniziert) ist. Weitere Indizierungsmöglichkeiten bestehen darin, einen Index für die Zeit bzw. den Lerndurchgang, den Lernkontext, den Abrufkontext und vieles andere mehr einzuführen. Das hier nicht näher zu behandelnde **Facetten-Design** von Guttman (vgl. Borg, 1981; Feger, 1983; Roskam, im Druck) ist ein geeigneter Rahmen, um das

Indizierungsproblem zu konkretisieren bzw. Vorschläge zur Lösung des Problems anschaulich zu machen.

Von der Art der gewählten Indizierung hängt es ab, welche Modelle der Datenanalyse überhaupt anwendbar sind. Insbesondere zieht die Lösung des Indizierungsproblems Konsequenzen hinsichtlich der Strukturierung der Beobachtungen in verschiedene Variablen- und verschiedene Realisationen einer Variablen nach sich. So impliziert etwa die Wahl des zweielementigen Datenraums [Wort rekonstruiert, Wort nicht rekonstruiert] eine einzige dichotome Variable; Rekonstruktionsleistungen verschiedener Vpn bezüglich verschiedener Wörter werden - soweit erhoben - als Realisationen derselben Variablen aufgefaßt. Bestimmte Fragestellungen - z.B. ob sich konkrete und abstrakte Wörter oder ältere und jüngere Personen unterscheiden - lassen sich dann natürlich nicht mehr untersuchen. Außerdem sind in Verbindung mit der anschließenden Datenanalyse oftmals bestimmte (problematische) Homogenitätsannahmen unvermeidlich, z.B. die, daß die Rekonstruktionsleistungen verschiedener Vpn - als Zufallsvariablen interpretiert - unabhängig und identisch verteilt sind.

Ebenso wie die Klassifikation verlangt auch die Indizierung der Beobachtungen eine Begründung, die immer nur unter Rückgriff auf die forschungsleitende (offene oder geschlossene) Fragestellung erfolgen kann. Das Indizierungsproblem läßt sich nicht dadurch umgehen, daß man es ignoriert. Wer unter Nichtreflexion dieses Problems beispielsweise eine Rohdatenmatrix anlegt und deren Struktur über eine Datenformatangabe einem Computerprogramm „mitteilt“, hat sich implizit auf eine Antwort zur Indizierungsfrage festgelegt, u. U. in fragestellungsunangemessener Weise. Werte in unterschiedlichen Spalten einer Rohdatenmatrix werden beispielsweise im Regelfall als Werte auf unterschiedlichen Variablen interpretiert, so daß es z.B. ausgeschlossen ist, die Rekonstruktionsurteile auf zwei Wörter, die in unterschiedlichen Spalten abgetragen sind, als Realisationen einer Variablen zu analysieren. Gerade dies kann aber unter bestimmten Umständen viel sinnvoller sein, als etwa die Urteile zweier Vpn auf ein Wort als Realisationen derselben Variablen aufzufassen.

### 1.4.3 Interpretationsprobleme

Ein weiteres zu lösendes Problem ist das der Interpretation einer wissenschaftlichen Beobachtung hinsichtlich der empirischen Relation, die sich darin ausdrückt. So ist beispielsweise ohne Bezugnahme auf die Instruktion oder eine nachträgliche Befragung der Vp nicht ohne weiteres klar, welche empirische Relation sich in der Aussage eines Tennisspielers manifestiert, daß er am liebsten mit dem Mitglied a seines Tennisklubs spielt (vgl. Coombs et al.,



deutsche Übersetzung 1975, S.49). Es kann sich um eine Näherrelation („proximity“) handeln, falls b damit ausdrücken will, welchem Mitspieler er eine ungefähr gleiche Spielstärke zuspricht; es kann sich aber auch um eine Dominanzrelation handeln, wenn nämlich a ein Spieler ist, gegen den sich b ein leichtes Spiel verspricht.

Die genaue Festlegung des Typs der empirischen Relation entscheidet zusammen mit der gewählten Indizierung der Beobachtungen darüber, welche Formen der Datenanalyse sinnvoll sind. Coombs (1964) hat eine Taxonomie ausgearbeitet, die angibt, welche Datenanalyseverfahren für welchen Datentyp angemessen sind. Daten werden dabei danach differenziert, ob sie Dominanz- oder Näherrelationen behaupten, ob sich diese Relationen auf Elemente einer Menge (z.B. Reize) oder auf Elemente zweier verschiedener Mengen (z.B. Personen und Reize) beziehen und ob diese Mengen aus einfachen Elementen oder aber aus Dyaden (z.B. Rechtecke, die hinsichtlich Höhe und Breite variieren) bestehen. Diesen drei Dichotomien entsprechen  $2^3 = 8$  mögliche Formen von Daten, die datenanalytisch unterschiedlich zu behandeln sind. Tatsächlich lassen sich sehr viele bekannte Skalierungsverfahren in dieses System einordnen (Coombs, 1964), auch wenn sich dieses System letztlich doch als unvollständig erweist (vgl. Roskam, 1983, S. 6).

#### 1.4.4 Quantifizierungsprobleme

Das Quantifizierungsproblem kann als ein bestimmter Aspekt des Interpretationsproblems aufgefaßt werden, ist jedoch von so fundamentaler Bedeutung für die empirische Psychologie, daß es in einem eigenen Unterpunkt angesprochen werden soll. Typischerweise sind empirische Daten in der Psychologie zunächst qualitativer oder ordinaler Natur, d.h. sie drücken qualitative empirische Relationen („a ist schwerer als b“, „x löst Aufgabe y“ usw.) zwischen Elementen einer Menge oder verschiedener Mengen aus<sup>8</sup>. Sehr viele gängige Datenanalyseverfahren setzen aber quantitative, d.h. reellwertige (Zufalls-) Variablen voraus, so z.B. die ganze Vielfalt der Verfahren im Rahmen des Allgemeinen Linearen Modells (ALM), insbesondere Regressions- und Varianzanalyse. Häufig werden deshalb protokollierte Beobachtungen direkt quantitativ interpretiert (z. B. Ratings) - man spricht dann auch von „direkter Skalierung“ - oder es werden quantitative Indizes unter Rückgriff auf vorliegende qualitative Beobachtungen ad hoc definiert, beispielsweise Summensco-

8 Die Bezeichnungen „qualitative Daten“ bzw. „qualitative Datenanalyse“ führen gelegentlich zu Mißverständnissen, da auch im hermeneutischen Kontext gelegentlich von „qualitativen Daten“, „qualitativen Methoden“, „qualitativem Vorgehen“ usw. gesprochen wird (vgl. dazu auch Soeffner & Hitzler, Kapitel 3, in diesem Band). Das ist hier natürlich nicht gemeint. Statt „qualitativ“ könnte man im gegebenen Kontext auch die Bezeichnungen „kategorial“ oder „diskret“ verwenden.

res (Anzahl gelöster Items, Anzahl Ja-Antworten) oder komplexere Leistungsindizes, in die verschiedene Beobachtungsaspekte eingehen, welche durch unterschiedliche arithmetische Operationen verknüpft werden. Man spricht in diesem Zusammenhang auch von **Indexmessung**, **Messung per Definition** oder **Messung per fiat** (vgl. Gigerenzer, 1980, S. 128ff; Tack, 1977, S. 46). Auf diese Weise entstehen Zahlen (Mausfeld spricht in Kapitel 13 dieses Bandes von „Zahlzeichen“), die man wie reelle Zahlen behandeln, d.h. addieren, subtrahieren, multiplizieren, mitteln usw. kann. Die Frage ist allerdings, welche Bedeutung den Ergebnissen solcher Analysen im Hinblick auf die forschungsleitende Fragestellung zugesprochen werden kann.

Die quantitative Interpretation von Beobachtungen **kann** zweifellos unproblematisch sein. Antwortzeiten - als **physikalische** Daten aufgefaßt - enthalten beispielsweise bereits numerische Informationen; es kann also z.B. sinnvoll sein zu sagen, daß die Zeitdifferenz für die Beantwortung der Fragen **a** und **b** genauso groß ist wie die für die Beantwortung der Aufgaben **c** und **d**. Dieser Befund kann unmittelbar psychologische Relevanz haben, etwa im Hinblick auf die Bewertung einer Informationsverarbeitungstheorie. Auf der anderen Seite ist beispielsweise der Status von Größenschätzungen („magnitude estimations“) nicht ohne weiteres klar, auch wenn Größenschätzungen - oberflächlich betrachtet - wie numerische Daten aussehen. Die Aussage einer Vp **i**, daß sie „Reiz **k** für fünfmal so groß wie den Standardreiz **s**“ hält, ist jedoch zunächst mal ein Satz, keine Zahl. Wenn hieraus die reelle Zahl „5“ bzw. das Quadrupel (**i**, **k**, **s**, 5) als empirisches Datum abgeleitet und als Maß für eine Relation von Empfindungsgrößen interpretiert wird, sind eine Reihe von ungeprüften und in diesem Rahmen auch unprüfbar Annahmen implizit vorzusetzen.

Ähnliches gilt für Ratings und komplexere Leistungsindizes, etwa den Prozentsatz Ersparnis beim Wiederlernen eines Lernstoffs, Metakognitionsindikatoren, die Güte der Steuerung eines komplexen, dynamischen Systems usw. Natürlich ist gegen numerische Analysen dieser Variablen nichts einzuwenden, solange sich die Interpretation der Ergebnisse wieder nur auf diese numerischen Variablen bezieht. Faktisch interessieren sie aber immer als Indikatoren **latenter** Größen, z.B. der Gedächtnisleistung oder der Problemlösefähigkeit (Gigerenzer, 1980, S. 128). Welche Aussagen numerische Analysen der Indikatorvariablen bezüglich der eigentlich interessierenden Größen zulassen, bleibt aber völlig unklar. Deutlich wird dies insbesondere dann, wenn man sich klar macht, daß unterschiedliche Möglichkeiten der Definition von Indizes existieren, die sich oftmals gleich gut argumentativ begründen lassen, aber zu unterschiedlichen Resultaten führen können. Tritt dieser Fall ein, so ist offenkundig, daß Aussagen, welche auf die numerischen Indikatorvariablen Bezug nehmen, hinsichtlich der forschungsleitenden Fragestellung nicht bedeutsam sind, weil ihr Wahrheitswert nicht nur von den vorliegenden quali-

tativen Daten, sondern auch von der **willkürlichen** Auswahl eines Indexmaßes aus einer Menge potentieller Kandidaten abhängig ist.

Dieser Zustand ist nicht tolerabel. Deshalb sollten Indexmessungen so weit wie möglich aus dem Datenerzeugungsprozeß eliminiert werden. Dies gilt auch dann, wenn ein Indexmaß (scheinbar) konkurrenzlos ist oder wenn es in eine empirische Untersuchung mit primär exploratorischem Charakter eingebettet ist. Die Konkurrenzlosigkeit des Maßes kann auf einer versteckten Konvention beruhen, die unkritisch tradiert wird, auch wenn sie einer stichhaltigen meß- oder skalierungstheoretischen Begründung entbehrt. Exploratorische Datenanalysen (vgl. Abschnitt 2.2) können in substanzwissenschaftlicher Hinsicht nur dann weiterführen, wenn die analysierten Variablen sinnvoll interpretierbar sind. Insbesondere können sie keine differenzierten Strukturen in Rohdaten aufdecken, wenn diese durch Verwendung komplexer Indexmaße bereits verschleiert sind.

Eine Alternative besteht z.B. darin, auf die Ableitung von quantitativen Variablen ganz zu verzichten und ein Datenanalysemodell zu wählen, das der qualitativen (bzw. ordinalen) Natur der Daten entspricht. Wenn der theoretische und methodische Hintergrund der empirischen Untersuchung schon weitgehend ausgereift ist, sollte auch daran gedacht werden, einen anderen Weg zu beschreiten, der in der Psychophysik schon eine lange Tradition hat (vgl. Tack, 1983; Mausfeld, Kapitel 4 dieses Bandes). Konkret besteht dieser Weg darin, Meß- und Skalierungsmodelle zu formulieren, die die Ableitbarkeit von numerischen Skalen aus qualitativen Daten nicht generell garantieren, sondern von bestimmten empirisch testbaren Bedingungen abhängig machen. Erweisen sich diese Modelle in einem konkreten Anwendungsfall als empirisch adäquat, lassen sich darüber hinaus die Eindeutigkeitseigenschaften der numerischen Zuordnungen angeben, so daß klar ist, welchen numerischen Aussagen empirische Bedeutsamkeit zugesprochen werden kann und welchen nicht. Mausfeld (Kapitel 13 dieses Bandes) behandelt diesen Problemkreis ausführlich.

Allerdings sollte der meßtheoretische Zugang kein Selbstzweck sein. Grundsätzlich ist die Art der Datenverwendung der Fragestellung nachgeordnet, die die betreffende empirische Untersuchung angeregt hat (vgl. Tack, 1977). Die Nichtanwendbarkeit von Meß- oder Skalierungsmodellen rechtfertigt also nicht die Elimination einer Fragestellung bzw. die Ersetzung dieser Fragestellung durch eine andere. Läßt sich zur Ausgangsfrage kein angemessenes Meß- oder Skalierungsmodell finden, sollte man sich auf deskriptive Formen der Datenanalyse beschränken, die die Daten **in der vorliegenden Form** aufgreifen, insbesondere also ohne „Ad-hoc-Quantifizierung“. Eine heuristische Betrachtung derartiger Befunde kann ergiebiger als eine theoretisch nicht motivierbare Anwendung der Repräsentationstheorie des Messens sein. Wenn ein Homomorphismus zwischen einem empirischen und einem numerischen relationalen

System formal gesichert ist, so impliziert dies nicht automatisch theoretische Fruchtbarkeit oder praktische Nützlichkeit (vgl. Schwager, 1991).

## **2. Verwendung von Daten im Entdeckungszusammenhang**

Reichenbach (1938) war der Meinung, daß nur der „context of justification“ - die Begründung und Rechtfertigung von Wissen - legitimer Gegenstand der Epistemologie sein könne, nicht aber der „context of discovery“. Letzterer ist nach Reichenbach von der kognitiven Psychologie zu untersuchen, weil allein sie Aussagen über die wirklichen Denkvorgänge kreativer und in theoretischer Hinsicht erfolgreicher Wissenschaftlerinnen und Wissenschaftler machen kann (a.a.O., S.382). Mit dieser Auffassung steht Reichenbach nicht allein. Viele Wissenschaftstheoretiker - auch solche, die Reichenbach ansonsten sehr kritisch gegenüberstehen - lassen sich in diesem Punkt in ähnlicher Weise charakterisieren (so z.B. Popper, 1935/1982).

In diesem Abschnitt wird die These vertreten, daß es Gründe gibt, den Entdeckungszusammenhang nicht ausschließlich psychologisch-deskriptiv, sondern wenigstens in Teilen auch **methodologisch-normativ** zu betrachten (vgl. Dörner, Kapitel 8 dieses Bandes). Dem liegt die Auffassung zugrunde, daß gute theoretische Einfälle im Regelfall keine unprädictierbaren Zufallsprodukte, keine „lightning strokes“ oder „happy guesses“ besonders begabter Persönlichkeiten sind. Vielmehr können gute Einfälle u.a. durch die Art und Weise der Verwendung von Daten wahrscheinlicher oder unwahrscheinlicher gemacht werden. Trifft diese Auffassung zu - was prinzipiell empirisch entscheidbar ist -, **so** läßt sich eine **methodologische Leitregel der Verwendung von Daten im Entdeckungszusammenhang** rechtfertigen. Sie lautet schlicht: Verwende im Entdeckungszusammenhang solche Datenanalyseverfahren, die gute Einfälle wahrscheinlicher machen; meide gleichzeitig solche Verfahren, die gute Einfälle unwahrscheinlicher machen. Unter einem „guten Einfall“ sind dabei Hypothesen, Theorien oder Modelle zu verstehen, die einerseits den üblichen Gütekriterien genügen (u. a. logische Konsistenz, empirischer Gehalt und Einfachheit, vgl. Gadenne, Kapitel 9 dieses Bandes), gleichzeitig aber in besonderer Weise innovativ und kreativ genannt werden können. Insbesondere bedeutet dies, daß sie sehr viele neue empirische Überprüfungsansätze implizieren, die den Rahmen bekannter empirischer Zugänge zur forschungsleitenden Fragestellung sprengen. Ein „guter Einfall“ drängt also unmittelbar aus dem Entdeckungszusammenhang heraus in den Begründungszusammenhang hinein.

Es kann kaum ein Zweifel daran bestehen, daß in der psychologischen Methodenlehre Verfahren dominieren, die zunächst einmal auf eine Verwendung im Begründungszusammenhang hin angelegt sind. So unverzichtbar beispiels-

weise das psychologische Experiment bei der Theorienüberprüfung ist (vgl. dazu Kapitel 12 und Kapitel 14 dieses Bandes), so wenig garantiert dessen Verwendung im Entdeckungszusammenhang automatisch gute Einfälle im o.g. Sinne (Herrmann, 1990, S.9). Dörner (1989) hat die mangelnde Fruchtbarkeit theorieleeren Experimentierens treffend am Beispiel der „Erforschung“ fiktiver außerirdischer Schildkröten karriert, deren Verhalten formal durch einen Mealy-Automaten beschrieben werden kann (einführend hierzu z.B. Buchner & Funke, 1993). Dörner zeigt, daß das Grundprinzip dieses Automaten kaum erkannt werden kann, wenn man schematisch experimentelle Bedingungsvariation betreibt und die durchschnittlichen oder modalen Effekte dieser Behandlungen betrachtet. Dies gilt auch dann, wenn alle „Schildkröten“ durch denselben finiten Automaten beschrieben werden können.

Natürlich wäre es irreführend, diese Kritik auf das psychologische Experiment per se zu beziehen (Herrmann, 1990, S.9). Zu kritisieren ist allein die Art seiner Verwendung im Entdeckungskontext, d.h. die schematische, unreflektierte Übertragung eines Vorgehens, das im Begründungskontext nicht nur sinnvoll, sondern sogar zwingend erforderlich sein kann (vgl. Erdfelder & Bredenkamp, Kapitel 14 dieses Bandes, Abschnitt 2.4). Im übrigen wirkt sich im von Dörner (1989) erörterten Falle auch nicht das experimentelle Vorgehen an sich unfruchtbar aus - also das Prinzip der Bedingungskontrolle und der Randomisierung -, sondern in erster Linie die ausschließliche Betrachtung von Statistiken, die über Beobachtungsobjekte hinweg aggregiert wurden (Mittelwerte, Modalwerte, relative Häufigkeiten). Die gleichen Probleme, die Dörner (1989) am Beispiel experimenteller Untersuchungen zu den Verhaltensgesetzen „kleiner grüner Schildkröten“ anschaulich macht, lassen sich also im Prinzip auch anhand nichtexperimenteller korrelationsanalytischer Studien z.B. zur kognitiven Struktur der „kleinen grünen Schildkröten“ aufzeigen. Um nur ein Beispiel von vielen möglichen zu nennen: Ist die Regression zwischen zwei Leistungsindikatoren  $Y_1$  und  $Y_2$  für die Hälfte einer Schildkrötenstichprobe linear fallend, für die andere Hälfte dagegen linear steigend, so daß das  $Y_1$ - $Y_2$ -Scattergramm wie ein „X“ aussieht, dann ist die Kovarianz zwischen  $Y_1$  und  $Y_2$  in der Gesamtstichprobe Null. Die artifizialen Konsequenzen, die sich für darauf aufbauende Regressions-, Korrelations-, Diskriminanz-, Faktoren- und Kovarianzstrukturanalysen ergeben, sind offensichtlich.

## 2.1 Kritik exploratorischer Anwendungen konfirmatorischer Analyseverfahren

Das Beispiel der „kleinen grünen Schildkröten“ zeigt, daß die Anwendung sog. konfirmatorischer Analyseverfahren - Verfahren, die die Prüfung von Hypothesen und die Schätzung von Parametern in bestimmten statistischen

Modellen zum Ziel haben - im Entdeckungszusammenhang generell wenig fruchtbar ist. Ursächlich hierfür ist die durch konfirmatorische Verfahren erzwungene Komprimierung der Daten zu Aggregatstatistiken, die den Blick auf heuristisch wertvolle Aspekte der Rohdaten versperren. Stimmt man dieser Auffassung zu, so ist unmittelbar klar, daß sich die Kritik auch auf jene Verfahren beziehen muß, die zwar oft „exploratorisch“ genannt werden, letztlich aber nur Optimierungsverfahren auf dem Hintergrund konfirmatorischer Verfahren darstellen. Beispiele hierfür sind schrittweise Regressionsanalysen, exploratorische Korrelations- und Faktorenanalysen, sukzessive hierarchische log-lineare Modelltests, multiple Effektttests in hochfaktoriellen ANOVA-Modellen sowie „nicht-statische“ Anwendungen des quadratischen Assignierens (Hubert & Schultz, 1976). Ausgangspunkt ist in allen diesen Fällen ein statistisches Rahmenmodell, das unhinterfragt **vorausgesetzt** wird. „Exploriert“ wird lediglich, welcher **Spezialfall** dieses Rahmenmodells für einen gegebenen Datensatz gewissen Optimalitätskriterien genügt. Im Falle der schrittweisen Regressionsanalyse ist das Rahmenmodell beispielsweise das ALM für eine Kriteriumsvariable und eine bestimmte Menge von Prädiktorvariablen. Alle implizierten Modellannahmen (insbesondere Additivität und Linearität der Effekte sowie Varianzhomogenität der Residuen) werden als gültig unterstellt. Die Optimierung besteht hier darin, für einen gegebenen Datensatz denjenigen Spezialfall des Rahmenmodells ausfindig zu machen, in dem nur die Prädiktoren enthalten sind, deren Regressionsgewicht bedeutsam von Null verschieden ist<sup>9</sup>. Die verwendeten Algorithmen garantieren in jedem Fall Konvergenz, so daß es immer zu einem „Ergebnis“ kommt. Dies gilt auch dann, wenn die Annahmen des zugrundeliegenden Rahmenmodells in gravierender Weise verletzt sind. Man sieht, daß das Verfahren nur scheinbar geeignet ist, auf die im Entdeckungszusammenhang typischen offenen Fragestellungen wie z.B. „Welche Variablen haben einen Effekt auf Y?“ oder „Was ist die optimale Vorhersagegleichung für Y?“ eine Antwort zu geben. Beantwortet wird vielmehr die geschlossene Frage nach dem sparsamsten Prädiktorensatz, gegeben **die Gültigkeit der vorausgesetzten Modellannahmen**. Eine solche Frage kann - wenn überhaupt - nur dann sinnvoll aufgeworfen werden, wenn man bereits über eine starke, gut bewährte Theorie verfügt, die die Gültigkeit der voraussetzenden Annahmen impliziert. Im theoretischen Vakuum können schrittweise Regressionsanalysen keine „guten Ideen“ nahelegen, da sie zwangsläufig auf Modellannahmen beruhen, die u.U. überwunden werden müssen, um zu guten Ideen zu gelangen. Die oben erwähnte Abhängigkeitsstruktur der Variablen  $Y_1$  und  $Y_2$ , die im Scattergramm ein „X“ bilden, bleibt bei schrittweisen Regressionsanalysen z.B. zwangsläufig unentdeckt, obwohl kein Zweifel

<sup>9</sup> Die üblichen schrittweisen Regressionsverfahren - „forward inclusion“ und „backward elimination“ - garantieren übrigens nicht die Lösung dieses Optimierungsproblems. Dieses eher technische Problem läßt sich jedoch durch aufwendigere Verfahren, bei denen alle möglichen Teilmengen von Prädiktoren betrachtet werden, prinzipiell in den Griff bekommen.

daran bestehen kann, daß der Zusammenhang von  $Y_1$  und  $Y_2$  ein extrem starker ist, wenn man die Ausprägung einer (noch) unbekanntem Moderatorvariablen konstant hält. Schrittweise Regressionsanalysen machen gute Ideen zur Lösung dieses komplizierten Abhängigkeitsproblems nicht wahrscheinlich; im Gegenteil: Sie versperren den Blick auf das interessante Phänomen und **hindern** damit gute Ideen.

Analoge Vorbehalte lassen sich gegen „exploratorische“ Varianten anderer konfirmatorischer Verfahren benennen. Hinzu kommt, daß viele dieser Verfahren nicht robust sind, d.h. sensibel auf Ausreißer in den Daten reagieren, deren Beachtung hinsichtlich der forschungsleitenden Fragestellung in die Irre führt (einen Überblick hierzu gibt Oldenbürger, im Druck). Exploratorische Anwendungen konfirmatorischer Analyseverfahren reagieren also durchaus auf bestimmte Modellverletzungen, diskriminieren aber nicht zwischen wichtigen und unwichtigen Fällen; sie vermengen beides vielmehr so, daß nicht mehr festgestellt werden kann, was für die zugrundeliegende Datenstruktur eigentlich wesentlich ist. Damit wird Datenanalyse im eigentlichen Sinne des Wortes unmöglich gemacht; analysierbar sind lediglich **Projektionen** der Daten, die durch ein unkritisch und unbegründet akzeptiertes Projektionsinstrument - das betreffende statistische Verfahren - erzeugt wurden. Die Verwendung von Aggregatstatistiken hat hierbei eine Informationsreduktion zur Folge, die einen Rückschluß auf die ursprüngliche Datenstruktur unmöglich macht.

## 2.2 Exploratorische und robuste Datenanalyse

Abschnitt 2.1 zeigte, daß die im Begründungskontext wichtigen konfirmatorischen Verfahren ebenso wie ihre „exploratorischen“ Varianten keine geeigneten Instrumente darstellen, wenn man sich im Entdeckungskontext bewegt. Nunmehr ist konstruktiv zu fragen, wie denn dann die Vorgehensweise zu charakterisieren ist, die „gute Einfälle wahrscheinlicher macht“.

In vielen Arbeiten wird der Weg der Einzelfallbeobachtung favorisiert. So liest man etwa in der bereits erwähnten Arbeit von Dörner (1989, S. 97): „Im Falle unserer Schildkröte ist es möglich, **allein durch Einzelfallbeobachtung** die richtigen Hypothesen über die innere Struktur der Schildkröte zu ermitteln und zu prüfen“. Da die von Dörner erörterten Schildkröten Realisationen desselben deterministischen Mealy-Automaten sind, trifft es in der Tat zu, daß die relevanten empirischen Informationen am Einzelfall erhoben werden können. Auf der anderen Seite ist aber auch klar, daß systematische Einzelfallbeobachtungen **allein** kaum hinreichend sind, wenn man zu einer adäquaten Theorie der Schildkröten gelangen will. Inzwischen liegt recht umfangreiche Literatur zum Erwerb von Wissen über diverse Modelle finiter Auto-

maten vor, die den Schluß rechtfertigt, daß durch Beobachtung und Intervention wohl implizites Wissen i.S. eines „Handhaben-Könnens“, nicht aber explizites Wissen über die Theorie des beobachteten Automaten erworben werden kann (Reber, 1989; Buchner & Funke, 1993). Bedenkt man darüber hinaus, daß das Verhalten der Schildkröten sicher eine grobe Vereinfachung psychologischer Forschungsprobleme ist, so wird klar, daß erfolgversprechende empirische Zugänge im Entdeckungskontext nicht ausschließlich und undifferenziert als Einzelfallbeobachtungen charakterisiert werden können. Vielmehr muß etwas über das „Wie“ der Analyse der erhobenen Daten ausgesagt werden.

Zu diesem Zwecke ist es hilfreich, wenn zunächst einmal genauer und differenzierter herausgearbeitet wird, warum herkömmliche konfirmatorische Analyseverfahren im Entdeckungskontext gute Ideen eher verhindern als fördern. Das bereits besprochene Beispiel einer Nullkovarianz zweier Variablen, hinter der sich ein X-förmiges Scattergramm verbirgt, eignet sich hierfür als Ausgangspunkt. Betrachtet man den bivariaten Datensatz im Rahmen des linearen Modells, so kommt nur eine Form von Abhängigkeit als Option in Frage: die der **linearen** Abhängigkeit. Die Datenanalyse ist also nicht offen für andere denkbare Abhängigkeitsstrukturen; sie ist **voreingenommen**. Doch dies ist nur ein Teil des Problems. Noch schädlicher wirkt sich aus, daß diese Voreingenommenheit nirgendwo im Prozeß der konfirmatorischen Datenanalyse korrigierbar ist. Es bleibt unbeachtet, wie weit die empirische Datenstruktur von den Vorannahmen abweicht. Residuen werden - wenn überhaupt - nur in globaler Form reflektiert; man komprimiert sie zu einem Fehlervarianzterm, der über die Frage der Angemessenheit der Vorannahmen keine Auskunft gibt. Die **lokalen Residuen**, die im Falle des X-förmigen Scattergramms die Inadäquatheit der Varianzhomogenitätsannahme deutlich machen wurden, bleiben unbeachtet. Damit sind zwei zentrale Punkte aufgezeigt, in denen sich ein für den Entdeckungskontext adäquates Datenanalyseverfahren von konfirmatorischen Vorgehensweisen unterscheiden muß: Die Datenanalyse muß so weit wie möglich offen für eine Vielzahl von Zusammenhangs- und Abhängigkeitsstrukturen sein, und sie muß vor allem die Beachtung der lokalen Residuen erzwingen.

Seit den sechziger Jahren hat sich in der angewandten Statistik und Biometrie unter der Bezeichnung „exploratorische Datenanalyse“ (EDA) eine Schule entwickelt, die sich diese beiden Prinzipien konsequent zu eigen macht (Hartwig & Dearing, 1981; Hoaglin, Mosteller & Tukey, 1983; Mosteller & Tukey, 1977; Tukey, 1977; zusammenfassend: Oldenbürger, im Druck). Im Gegensatz zur konfirmatorischen Statistik ist EDA nicht angemessen als Sammlung von Methoden charakterisierbar, die man sich durch bloßes Erlernen aneignen kann: „Exploratory data analysis is an attitude, a flexibility, and a reliance on display. NOT a bundle of techniques, and should be so taught“ (Tukey, 1980,



S.23). Sie ist „(...) a state of mind, a way of thinking about data analysis“ (Hartwig & Dearing, 1981, S. 9), eine Denkweise, die durch **Skeptizismus** bezüglich der Adäquatheit der erfolgten Datenbeschreibung und durch **Offenheit** für alternative Beschreibungsmöglichkeiten gekennzeichnet ist. EDA ist prinzipiell nicht algorithmisierbar und nicht an Computer delegierbar, wenngleich grafikfähige Computersoftware immer häufiger hilfreiche Dienste bei der Datenanalyse erledigt. Letztlich ist auch bezweifelbar, ob EDA überhaupt angemessen als wissenschaftliche Disziplin verstanden werden kann; vielleicht ist die Bezeichnung „Kunst“ angemessener (Good, 1983).

Ausgangspunkt exploratorischer Datenanalysen ist fast immer eine informationshaltige, aber dennoch leicht lesbare **graphische Darstellung** der u.U. mehrdimensionalen Rohdaten. Die Zielsetzung derartiger Graphiken hat Tukey (1977, S. VI) treffend formuliert: „The greatest value of a picture is when it **forces** us to notice what we never expected to see.“ Zu diesem Zweck sind Techniken entwickelt worden, die herkömmlichen Histogrammen und Scattergrammen weit überlegen sind. Dies gilt für uni- und bivariate Datensätze, vor allem aber für multivariate Datenstrukturen (Kleiner & Hartigan, 1981; Riedwyl & Schafroth, 1976; Tukey, 1977; Wainer & Thissen, 1981). Ausgehend von graphischen Darstellungen lassen sich erste Vorstellungen über „wesentliche“ und „unwesentliche“ Aspekte der Daten gewinnen. „Wesentlich“ sind dabei jene Aspekte, die potentiell durch eine Gesetzmäßigkeit beschreibbar erscheinen; „unwesentlich“ sind solche Aspekte, die vorläufig unerklärbar erscheinen (Good, 1983). Diese ersten Vorstellungen münden dann u.U. in Vermutungen darüber ein, welche **Variablentransformationen („reexpressions“)** oder Substichprobenbetrachtungen die Datenstruktur vereinfachen können, so daß kurvilineare Zusammenhangsmuster linear, heterogene Varianzen homogen oder asymmetrische Verteilungen symmetrisch werden<sup>10</sup>. Diese Vermutungen führen dann zu erneuten graphischen Analysen der transformierten Variablen, wobei nunmehr der Residuenbetrachtung eine besondere Bedeutung zukommt. Unter einem Residuum wird in diesem Zusammenhang ganz allgemein die Differenz zwischen einem empirischen Datum und dem zugehörigen vorläufigen Modelldatum („fit“ oder „smooth“) verstanden (Hartwig & Dearing, 1981, S. 10; Tukey, 1977, S. 125). Zeigt die Residuenbetrachtung Irregularitäten wie z.B. Varianzheterogenität oder Korrelationen zwischen Residuum und „smooth“, so ist das ein Hinweis darauf, daß die Datenstruktur noch nicht angemessen beschrieben ist. Es werden dann neue Vermutungen über bessere Beschreibungsmöglichkeiten der Daten entwickelt, die ebenfalls

<sup>10</sup> Man beachte, daß Variablentransformationen nicht generell positiv oder negativ bewertet werden können. Ob Variablentransformationen sinnvoll sind oder nicht, hängt ganz entscheidend vom Forschungskontext ab, in dem man sich bewegt. Während „reexpressions“ im Entdeckungskontext als Teil der EDA hilfreiche Dienste leisten können, sind sie im Begründungskontext oftmals sehr problematisch, weil sie implizit die Hypothese verändern, die zu prüfen ist (Garnes, 1983).

anhand von Residuenbetrachtungen zu evaluieren sind. Der EDA-Prozeß schreitet auf diese Weise zyklisch fort, bis Optimierungsmöglichkeiten der Datenbeschreibung nicht mehr erkennbar sind.

Die Beschreibung der Datenstruktur erfolgt weitgehend in Form robuster Statistiken; Standardschätzfehler und Konfidenzintervalle werden mit Hilfe verteilungsfreier „resampling“-Verfahren wie „jackknife“ oder „bootstrap“ ermittelt (Efron, 1979). Dieses Vorgehen garantiert, daß in die Beschreibung nicht doch wieder inadäquate Verteilungsannahmen oder Ausreißer maßgeblich einfließen. Gelegentlich kann auf diese Vorsichtsmaßnahmen auch verzichtet werden; das ist dann der Fall, wenn es durch Variablentransformationen gelingt, die Annahmen eines herkömmlichen konfirmatorischen Verfahrens einzuhalten.

EDA ist ein Hilfsmittel, auf das man im Entdeckungskontext nicht verzichten sollte. Allerdings ersetzt es andere Heuristiken der Theorienentwicklung nicht; es ist lediglich eine Ergänzung hierzu, Beispielsweise sagt EDA wenig darüber aus, wie die Abbildung von Beobachtungen in Daten erfolgen sollte. EDA setzt erst dort an, wo Daten bereits vorliegen. EDA mündet auch nicht zwangsläufig in „gute Ideen“ ein, es macht sie lediglich wahrscheinlicher. **Garantieren** kann EDA lediglich die Gewinnung angemessener Beschreibungen, die zu empirischen Generalisierungen werden, wenn man eine allgemeine „Gesetzmäßigkeit“ vermutet. Von dort ist es aber u.U. noch ein weiter Weg zu einer psychologischen Hypothese oder Theorie. Andere Heuristiken - vor allem nichtempirische - sind erforderlich, wenn man diesen Weg erfolgreich beschreiten will (vgl. Dörner, Kapitel 8 dieses Bandes).

### **3. Verwendung von Daten im Begründungszusammenhang**

Dieser Abschnitt bezieht sich auf Strategien der Datenerhebung und Datenanalyse, die im Begründungskontext Verwendung finden können. Vorauszusetzen ist also, daß eine hinreichend präzise formulierte psychologische Theorie oder Hypothese vorliegt, die empirisch evaluiert werden soll. Grob kann man zwei Schulen in diesem Zusammenhang unterscheiden: eine eher induktive, die wissenschaftstheoretisch vor allem mit Namen wie Carnap und Reichenbach zu verknüpfen ist, und eine eher deduktive, die sich eher am kritischen Rationalismus in der von Popper und Lakatos geprägten Form orientiert. In der induktiven Schule steht das Bemühen im Vordergrund, psychologische Theorien und Hypothesen aus vorliegenden empirischen Daten heraus empirisch zu stützen (vgl. Westermann & Gerjets, Kapitel 10). Empirische Daten haben in diesem Zusammenhang also in erster Linie eine Bestätigungsfunktion. In der deduktiven Schule dominiert dagegen eher das Bemühen, strenge empirische Prüfungen psychologischer Theorien und Hypothesen vorzunehmen, die falsche Hypothesen mit hoher Wahrscheinlichkeit als em-

pirisch inadäquat ausweisen, zutreffende Hypothesen dagegen nicht (vgl. Erdfelder & Bredenkamp, Kapitel 14 dieses Bandes; Hussy & Möller, Kapitel 11 dieses Bandes). Die Kritikfunktion empirischer Daten steht hierbei im Vordergrund. In der psychologischen Methodenlehre hat man selten explizit versucht, die fachspezifische Methodologie an die Lehren von Carnap oder Reichenbach auf der einen Seite bzw. Popper oder Lakatos auf der anderen Seite anzubinden. Dennoch zeigt eine nähere Analyse psychologiespezifischer Methodologien, daß sie einem der beiden Pole auf dem Induktions-Deduktions-Kontinuum immer mehr oder minder stark zuzuordnen sind. Ein ausführlicher Überblick hierzu ist im folgenden nicht beabsichtigt. Es geht vielmehr darum, die Erscheinungsformen und die spezifischen Probleme induktiver bzw. deduktiver Grundhaltungen in der psychologischen Methodenlehre exemplarisch deutlich zu machen.

### 3.1 Der induktive Zugang

Die wohl bekannteste abgeschlossene Methodologie innerhalb der Psychologie ist die auf Campbell und Stanley (1963/1970) zurückgehende Theorie der internen und externen Validität (vgl. auch Cook & Campbell, 1979). Sie bildet den Ausgangs- und Bezugspunkt für eine Reihe anderer Methodologien (z.B. Brinberg & McGrath, 1985). Die Campbell-Stanley-Theorie basiert im wesentlichen auf zwei Metaregeln, nämlich dem Prinzip der internen und dem Prinzip der externen Validität. Grundsätzlich wird dabei von dem Prinzip der externen Validität ausgegangen, daß die zu prüfende Hypothese die Struktur einer Wenn-dann-Aussage hat, bei der ein kausaler Effekt einer oder mehrerer unabhängiger Variablen (UVn) auf eine oder mehrere abhängige Variablen (AVn) behauptet wird, also z.B. so etwas wie „Die Technik der progressiven Muskelrelaxation reduziert den Juckreiz bei Neurodermitis“. Das Prinzip der internen Validität (PiV) besagt nun, daß Untersuchungsleiter systematische Fehler (oder störende Bedingungen erster Art, vgl. Gadenne, 1976) völlig auszuschließen haben. Systematische Fehler beruhen auf einer Konfundierung der interessierenden W(n) mit (bekannten oder unbekanntem) Störvariablen, die einen Effekt der **W** vortäuschen oder verschleiern können. Die Sicherung der internen Validität verlangt also die Kontrolle aller potentiellen Störvariablen, was nach Campbell und Stanley (1963/1970) durch Konstanthaltung oder durch Sicherung der Unkonfundiertheit (der stochastischen Unabhängigkeit) von interessierenden UVn und allen potentiellen Störvariablen erreicht werden kann (vgl. Steyer, Kapitel 15 dieses Bandes).

Interne Validität (d.h. die vollständige Realisation von PiV) impliziert nach Campbell und Stanley, daß der beobachtete Effekt der konkreten **W** auch

als kausaler<sup>11</sup> Effekt dieser **W unter den konkreten Bedingungen der betreffenden empirischen Untersuchung** interpretiert werden kann. Dies besagt zunächst nichts über den Effekt der **W** bei anderen Vpn der gleichen Grundgesamtheit, beigeringfügig modifizierten Untersuchungsbedingungen, anderen Indikatoren als AV usw. Da die zu prüfenden Hypothesen praktisch immer universellen Charakter haben, insbesondere also nicht auf bestimmte Vpn, Versuchsleiter, Randbedingungen usw. eingeschränkt sind, muß das PiV nach Campbell und Stanley (1963/1970) um das Prinzip der externen Validität (PeV) ergänzt werden. Dieses Prinzip ist erfüllt, wenn eine Untersuchung anhand von Vpn, Randbedingungen, UVn, AVn usw. durchgeführt wurde, die für die in der betreffenden Hypothese angegebenen Grundgesamtheiten **repräsentativ** sind (vgl. Gadenne, 1976, S. 19). Repräsentativität kann im wesentlichen durch **Zufallsauswahl** der Vpn, Randbedingungen, Indikatorvariablen etc. aus den jeweils interessierenden Grundgesamtheiten erreicht werden.

Sind die Prinzipien PiV und PeV erfüllt, so darf ein beobachteter W-Effekt nach Campbell und Stanley (1963/1970) kausal interpretiert und auf die in der Hypothese genannten Grundgesamtheiten von Personen, Situationen und Maßen „generalisiert“ werden. Wörtlich schreiben sie: „Obwohl **innere Gültigkeit die unbedingte Voraussetzung ist** und obwohl die Frage der äußeren **Gültigkeit**, so wie die Frage der induktiven Schlüsse, nie völlig beantwortet werden kann, ist die Wahl von Anordnungen, die hinsichtlich beider Formen der Gültigkeit zufriedenstellend sind, ganz sicher das erstrebenswerte Ideal“ (deutsche Übersetzung, 1970, S. 459f.).

Ob PiV und PeV erfüllt sind, hängt nach Campbell und Stanley (1963/1970) maßgeblich vom gewählten Untersuchungsdesign ab. Da das Problem der adäquaten Versuchsanordnung von den Autoren nahezu ausschließlich behandelt wird, muß der Eindruck entstehen, daß interne und externe Validität in erster Linie **Eigenschaften von Designs** sind. über Vorliegen oder Nichtvorliegen von (interner oder externer) Validität kann demnach im Prinzip ohne Bezugnahme auf die zu prüfende Hypothese und ohne Erörterung anderer Elemente des Überprüfungsprozesses entschieden werden. Folgerichtig steht eine Evaluation von Untersuchungsdesigns anhand von Faktoren, die die interne und externe Validität gefährden können, im Mittelpunkt des Kapitels von Campbell und Stanley (1963/1970). Die Autoren warnen am Ende ihrer Arbeit zwar davor, die Wahl von Untersuchungsdesigns schematisch anhand ihrer Bewertungstabellen vorzunehmen, geben aber kaum Hinweise, welche anderen Faktoren im Überprüfungsprozeß zu berücksichtigen sind.

11 Was „kausal“ in diesem Zusammenhang bedeutet, wurde von Campbell und Stanley (1963) sowie Cook und Campbell (1979) nicht klar expliziert, kann aber im Prinzip präzise angegeben werden (vgl. dazu Steyer, Kapitel 15 dieses Bandes).

Vor allem in diesem Punkt stellt die Arbeit von Cook und Campbell (1979) eine Weiterentwicklung dar. Es wird dort explizit festgestellt, daß Validität eine Eigenschaft von Aussagen ist, nämlich die „beste verfügbare Approximation an die Wahrheit“ einer Aussage (Cook & Campbell, 1979, S.37). Eine valide Aussage ist demnach so etwas wie eine „vermutlich wahre“ oder „am ehesten vertretbare“ Aussage. Validität in diesem Sinne hängt von vielen Aspekten des Überprüfungsprozesses ab, nicht nur vom Untersuchungsdesign. Die Campbell-Stanley-Konzepte der internen und externen Validität werden deshalb um die Konzepte „Validität des statistischen Schlusses“ sowie „Konstruktvalidität der UV<sub>n</sub> und AV<sub>n</sub>“ ergänzt. Interner Validität wird eine zentrale Bedeutung sowohl für die Grundlagen- als auch für die Anwendungsforschung zugesprochen. In der Grundlagenforschung kommt darüber hinaus der Konstruktvalidität und der Validität des statistischen Schlusses besondere Bedeutung zu, während externe Validität eine untergeordnete Rolle spielt. Externe Validität ist nach Cook und Campbell (1979) in erster Linie in der Anwendungsforschung zu fordern.

Die Arbeit von Cook und Campbell (1979) zeichnet sich ferner durch den Versuch aus, die eigene Methodologie in einen größeren wissenschaftstheoretischen Rahmen einzuordnen. Die Autoren tendieren zu einer am kritischen Rationalismus orientierten Position, die mit anderen „postpositivistischen“ Elementen angereichert wird. Der Kausalitätsbegriff wird ausführlich erörtert, ohne allerdings auf die Verbindung zum Konzept der internen Validität näher einzugehen. Eine wissenschaftstheoretische Evaluation der eigenen Methodologie bleibt ebenso wie in der älteren Arbeit von Campbell und Stanley (1963/1970) ausgespart.

Gadenne (1976) hat den Versuch unternommen, diese Lücke zu schließen. Im Mittelpunkt seiner Arbeit steht die Frage, ob und inwieweit die Methodologie vom Campbell und Stanley mit dem Induktionsproblem - dem Problem gehaltenweiternder Schlüsse - behaftet ist. In der Wissenschaftslehre herrscht seit langem Einigkeit darüber, daß mittels induktiver Schlüsse die Wahrheit von Konklusionen nicht begründet werden kann (sog. starkes Induktionsproblem, vgl. Westermann & Gerjets, Kapitel 10, Abschnitt 1.2, in diesem Band). Dies hat erhebliche Konsequenzen für eine Methodologie der Hypothesenprüfung: Eine Methodologie ist - vorsichtig ausgedrückt - problematisch, wenn sie eine Lösung des starken Induktionsproblems zur Gänze oder in Teilen voraussetzt. Sie fordert dann offensichtlich etwas, was man (seit Hume, 1739-1740/1978) **nachweislich** nicht leisten kann.

Gadennes (1976) Hauptthese besagt, daß die Prinzipien PiV und PeV jeweils **zweifach** mit Induktionsproblemen behaftet sind. Zunächst zeigt er, daß die Erfüllung von PiV die Lösung eines starken Induktionsproblems voraussetzt. Die potentiellen Störvariablen in einer bestimmten Untersuchung sind weder sämtlich bekannt noch notwendig von endlicher Anzahl. Demzufolge impliziert die zur Erfüllung von PiV erforderliche Kontrolle aller Störvariablen

einen induktiven Schluß. Dieser läßt sich auch mit Hilfe der Randomisierung nicht rechtfertigen. Randomisierung garantiert lediglich asymptotisch (für  $N \rightarrow \infty$ ) Unabhängigkeit zwischen Treatmentbedingungen und Störvariablen. über **endliche** Stichproben - mit denen man es in empirischen Untersuchungen immer zu tun hat - ist damit nichts ausgesagt. Die Begründung der Randomisierung kann also nicht mit den von Campbell und Stanley (1963/1970), Cook und Campbell (1979) und vielen anderen Autoren favorisierten Argumenten erfolgen. Randomisierung führt nicht zwangsläufig zur Elimination von Störfaktoren in endlichen Stichproben; wäre dem so, müßte man sich auch fragen, warum überhaupt noch andere Kontrollverfahren wie z.B. Parallelisierung eingesetzt werden<sup>12</sup>.

Das Argument „PiV erlaubt kausale Interpretationen des W-Effektes“, das die Forderung nach interner Validität letztlich begründet, verlangt die Lösung eines weiteren Induktionsproblems. Kausalaussagen - so Gadenne (1976) - sind grundsätzlich universeller Natur, da sie implizieren, daß ein Ereignis in gesetzmäßiger Weise regelmäßig ein anderes Ereignis zur Folge hat. Dem Übergang von singulären Beobachtungsaussagen auf Kausalaussagen kann also nur ein gehaltserweiternder Schluß zugrunde liegen. Hierbei ist es letztlich unerheblich, auf welchem Wege man zu den singulären Beobachtungsaussagen gelangt ist.

In ganz ähnlicher Weise läßt sich das Prinzip PeV problematisieren. Einerseits ist dieses Prinzip - zumindest für die psychologische Grundlagenforschung - nicht realisierbar, da Hypothesen hier typischerweise universeller Natur sind und sich auf **offene** Grundgesamtheiten beziehen, z.B. auf die Menge aller bereits gestorbenen, aller lebenden und aller noch nicht geborenen Menschen. Aus derartigen Grundgesamtheiten lassen sich keine repräsentativen Stichproben ziehen. Ähnliches gilt für Grundgesamtheiten möglicher **UVn**, **AVn**, experimenteller Randbedingungen usw. Entschließt man sich, ersatzweise aus einer großen finiten Grundgesamtheit eine repräsentative Stichprobe zu ziehen, so ist dies nur bedingt von Nutzen, da der Status dieser Stichprobe relativ zur interessierenden Grundgesamtheit unklar ist.

Angenommen, PeV ließe sich verwirklichen. Dann stellt sich die weitergehende Frage, wie sich das Argument „PeV erlaubt Generalisierungen des beobachteten W-Effekts“ begründen läßt. Schon Campbell und Stanley (1963/1970) deuteten an, daß dies nur auf induktivem Wege erfolgen kann. Parameter der Grundgesamtheit lassen sich aber nicht induktiv „erschließen“, sondern nur schätzen. Aussagen über einzelne Elemente der Grundgesamtheit, die in der Allgemeinen Psychologie im Zentrum des Interesses stehen, sind

12 Die Frage, welche andere Funktion der Randomisierung im Überprüfungsprozeß zukommen kann, soll in diesem Kapitel ausgespart bleiben. Erdfelder und Bredenkamp (Kapitel 14, Abschnitt 2.4) sowie Steyer (Kapitel 15, Abschnitt 2.3) gehen an anderen Stellen dieses Bandes ausführlich auf dieses Problem ein.

aus Stichprobendaten überhaupt nicht ableitbar. Gadenne (1976) betont, daß Zufallsstichproben nicht die Funktion haben, die Anwendung induktiver Schlüsse auf Grundgesamtheiten zu legitimieren. Für einen solchen „Schluß“ gibt es weder eine logische noch eine statistische Rechtfertigung. „Inferenz-Statistik“ oder „schließende Statistik“ sind deshalb auch sehr unglückliche Wendungen. In der Statistik wird nicht „geschlossen“, sondern „entschieden“; Zufallsstichproben - soweit sie erhoben werden - haben allein die Funktion, den Annahmen eines Stichprobenmodells zu genügen und somit Fehlerrisiken bei statistischen Entscheidungen kontrollierbar zu machen (vgl. Ostmann & Wutke, Kapitel 16 dieses Bandes).

Das Fazit der Arbeit von Gadenne (1976) ist, daß die Theorie der internen und externen Validität das starke Induktionsproblem an zentralen Stellen mehrfach aufwirft. Dies ist trotz anderslautender verbaler Bekundungen von Campbell und Stanley (1963/1970) sowie Cook und Campbell (1979) Anlaß genug, diese Methodologie als „induktiv“ zu klassifizieren. Cook und Campbell (1979, S. 85ff.) sind auf diesen Kritikpunkt - wie auch auf andere, hier nicht aufgeführte Kritikpunkte - kurz eingegangen. Die Frage, inwieweit PiV und die Absicherung von Kausalaussagen induktive Schlüsse erforderlich machen, klammern sie allerdings aus. Dem Einwand, daß PeV induktive Schlüsse impliziert, wird prinzipiell nicht widersprochen. In der Tat sei - so Cook und Campbell (1979, S. 86) - externe Validität „(...) inherently more problematic than even internal validity whose bases are more obviously deductive“. Sie schlagen vor, einige Teilprobleme - z.B. die Frage der Übertragbarkeit von Befunden von einem Teil auf einen anderen Teil einer Grundgesamtheit (z.B. von Psychologiestudenten auf Nicht-Studierende) durch deduktive Hypothesentests innerhalb beider Teilpopulationen zu lösen. Da damit das Prinzip PeV nicht generell hinfällig wird, ist allerdings die Kritik von Gadenne nicht grundsätzlich entkräftet (vgl. auch Gadenne, 1984, Kap. 10).

### **3.2 Der deduktive Zugang**

Gadenne (1976, Kap. 3; 1984, Kap. 10) hat zugleich diskutiert, welche Eigenschaften eine Methodologie der Prüfung psychologischer Hypothesen aufweisen muß, die induktive Elemente weitgehend vermeidet und somit „eine deduktivistische Theorie psychologischer Untersuchungen“ genannt werden kann. Andere Autoren haben - z.T. unabhängig von Gadenne - in mehr oder minder enger Anlehnung an die Positionen von Popper (1935/1982) und Lakatos (1970) weitere wichtige Komponenten deduktiven Vorgehens bei der psychologischen Hypothesenprüfung ausgearbeitet bzw. einzelne Vorschläge Gadennes aufgegriffen, modifiziert und erweitert (vgl. dazu im einzelnen Erd-

felder & Bredenkamp, Kapitel 14 dieses Bandes; Hager, 1992; Hussy & Möller, Kapitel 11 dieses Bandes).

Der Grundgedanke des deduktiven Vorgehens besteht darin, die Frage, welche Kausalaussagen und Generalisierungen aus vorliegenden Befunden abgeleitet werden können, quasi „umzudrehen“. Gefragt wird also zunächst, was alles aus der zu prüfenden psychologischen Hypothese in bezug auf empirische Daten gefolgert werden kann (vgl. bereits Meehl, 1967, und Bredenkamp, 1969, sowie für die Entwicklungspsychologie Kendler, 1979). Insbesondere nichttriviale empirische Implikationen sind von Interesse, d.h. solche Folgerungen, die aus anderen, bereits bekannten und bewährten Hypothesen oder aus dem Hintergrundwissen **nicht** ableitbar sind bzw. alternativen Hypothesen sogar direkt widersprechen. Anhand empirischer Daten wird dann **entschieden**, ob die nichttrivialen Implikationen der Hypothese empirisch erfüllt sind (Bewährungsurteil) bzw. nicht erfüllt sind (Nichtbewährungsurteil). Einzelne Bewährungsurteile bedeuten natürlich keinen „Beweis“ der Hypothese, ebenso wie Nichtbewährungsurteile keine „Falsifikation“ implizieren. Erst im Rahmen ganzer Forschungsprogramme kann eine Hypothese oder Theorie letztendlich bewertet werden (vgl. Herrmann, Kapitel 6 dieses Bandes; Erdfelder & Bredenkamp, Kapitel 14 dieses Bandes).

Validität ist in diesem Kontext ein **Attribut von Hypothesenprüfungen**, nicht ein Attribut von Aussagen. Auf Generalisierungen jeder Art wird verzichtet. Nur für die Vpn und die konkret realisierten Bedingungen einer bestimmten Untersuchung soll jeweils entschieden werden, ob die empirischen Implikationen einer psychologischen Hypothese haltbar sind oder nicht. Die Untersuchung ist dabei so zu planen und durchzuführen, daß Fehlentscheidungen über die interessierende psychologische Hypothese möglichst selten vorkommen. Genauer: Das Risiko fälschlicher Bewährungsurteile und das Risiko fälschlicher Nichtbewährungsurteile bezüglich der zu prüfenden Hypothese soll möglichst gering sein.

Allein an diesem Maßstab der „Strenge und Fairneß einer Hypothesenprüfung“ ist aus deduktiver Perspektive die Validität einer Hypothesenprüfung zu messen. Damit liegt ein Kritikpunkt nahe, dem Campbell und Koautoren - andere Autorinnen und Autoren (z.B. Hultsch & Hickey, 1978) teilweise mit noch mehr Nachdruck - durch die Betonung der Wichtigkeit von „externer Validität“<sup>13</sup> begegnen wollten: der Vorwurf, daß „valide“ Hypothesenprüfungen aufgrund der zur Validitätssicherung erforderlichen Künstlichkeit der Untersuchungssituation ohne jeden Praxis- und Alltagsbezug sind und somit psychologischen Theorien und Hypothesen Vorschub leisten, die nur in einer stark eingeschränkten Laborwelt Gültigkeit beanspruchen können,

13 Was Hultsch und Hickey (1978) unter „externer Validität“ genau verstehen, bleibt unklar. Für sie „(...) external validity is defined as the adequate dimensionalization of the organized complexity and its reciprocal interactions“ (a.a.O., S. 78).



nicht aber den Menschen in seinen natürlichen Lebenszusammenhängen zum Gegenstand haben. Holzkamp (1972, S.64) bringt die Vorbehalte gegenüber dieser Art von Forschung treffend auf den Punkt, indem er feststellt, daß das auf diese Weise akkumulierte „(...) überall gleich Gültige letztlich das Gleichgültige sein könnte“.

Was ist von dieser Kritik zu halten? Das Problem der Übertragbarkeit von Befunden auf den Alltag bzw. auf Praxisfelder der Psychologie wird in der deduktiven Schule in der Tat vom Validitätsproblem völlig getrennt (vgl. Gadenne, 1984, Kap. 10.4). Es wäre aber falsch, hieraus zu folgern, daß das Praxis- und Anwendungsproblem ignoriert wird. Vielmehr wird „Praxisbezug“, „Alltagsrelevanz“ usw. nicht als Problem von **Befunden** hypothesenprüfender Untersuchungen aufgefaßt, sondern als **Problem psychologischer Theorien**, deren möglichst strenger Überprüfung empirische Untersuchungen allein dienen. Anwendungsbezug wird also nicht durch Generalisierungen der Ergebnisse hypothesentestender Untersuchungen hergestellt, sondern dadurch, daß man unter Heranziehung von Hintergrundwissen aus möglichst gut bewährten (empirisch streng geprüften) Theorien **technologische Prognosen** gewinnt, die dann erneut in Anwendungskontexten deduktiv zu überprüfen sind. Natürlich fließen heuristische (wenn man so will: induktive) Elemente in den Prozeß der Gewinnung technologischer Prognosen ein; deshalb wäre es auch falsch, von einer „Ableitung“ technologischer Prognosen (im deduktiven Sinne) zu sprechen. Aber die Annahme oder Verwerfung einer technologischen Prognose wird durch die in ihre Entstehung eingehenden heuristischen Elemente nicht determiniert; grundsätzlich wird immer ein deduktiver Hypothesenprüfungsprozeß „nachgeschoben“, der das Aufdecken von Irrtümern ermöglicht.

Insgesamt kann somit Hypothesenprüfung deduktiv konzipiert werden, ohne auf diese Weise Forschungsfelder der Grundlagenforschung oder der angewandten Forschung auszugrenzen. Anzumerken ist allerdings, daß sich bei der Umsetzung des deduktiven Programms viele nichttriviale Detailprobleme stellen. Da Erdfelder und Bredenkamp (Kapitel 14 dieses Bandes) diesen Problemkreis ausführlich behandeln, kann eine Erörterung hier unterbleiben. Ebenfalls anzumerken ist, daß „deduktive Hypothesenprüfung“ unterschiedlich ausgelegt werden kann. So geben Proponenten deduktiven Vorgehens z.B. unterschiedliche Antworten auf die Frage, wie bei der statistischen Prüfung psychologischer Hypothesen die psychologische und die statistische Hypothese miteinander zu verknüpfen sind. Einige Autoren sehen eine Implikationsbeziehung zwischen psychologischer und statistischer Hypothese letztlich als unverzichtbar an (vgl. Erdfelder & Bredenkamp, Kapitel 14 dieses Bandes), während andere eine statistische Bewertung psychologischer Hypothesen auch dann für möglich halten, wenn die psychologische und die statistische Hypothese nur locker miteinander assoziiert sind (vgl. Hager, 1992, sowie Hussy & Möller, Kapitel 11 dieses Bandes).

#### **4. Konfirmatorische und exploratorische Verwendung empirischer Daten: Ein Antagonismus?**

In Abschnitt 2 wurde gezeigt, daß konfirmatorische Datenanalyseverfahren nicht sinnvoll für exploratorische Zwecke eingesetzt werden können. Exploratorische Datenanalyse verlangt ein grundsätzlich anderes Methodeninventar, das sich vor allem durch Annahmearmut, besondere Beachtung von Residuen, Robustheit gegenüber Ausreißern und graphische Veranschaulichung von Datenstrukturen auszeichnet. Exploratorische Varianten konfirmatorischer Analyseverfahren besitzen diese Eigenschaften nicht.

Auf diesem Hintergrund mag es naheliegend erscheinen, die möglichen Verwendungskontexte empirischer Daten in der Psychologie als zwei disjunkte und exhaustive Bereiche - exploratorische und konfirmatorische Datenanalyse - aufzufassen. Hat man sich für einen der beiden Bereiche entschieden, so wäre demzufolge genau das (und nur das) Werkzeug zu benutzen, das für diesen Bereich entwickelt wurde. Konfirmatorische Analyseverfahren wären somit im exploratorischen Kontext ebenso „verboten“ wie exploratorische Analyseverfahren im konfirmatorischen Kontext.

Zweck dieses Abschnitts ist es, diese Auffassung zu kritisieren, auch wenn die Gliederung des vorliegenden Kapitels ihr auf den ersten Blick ZU entsprechen scheint. Richtig ist zwar, daß konfirmatorische Verfahren im exploratorischen Kontext nur bedingt von Nutzen sind bzw. sogar eher schaden als nutzen, da sie den Blick auf die Rohdatenstrukturen versperren. Umgekehrt gilt jedoch **nicht**, daß exploratorische Verfahren im konfirmatorischen Kontext nur bedingt von Nutzen sind. Im Gegenteil: Kein empirischer Datensatz sollte im Anschluß an eine Hypothesenprüfung ohne Blick auf die Residuen - seien sie auch noch so klein - ad acta gelegt werden. Diese Sichtweise ergibt sich unmittelbar aus der in Abschnitt 1.3 dargelegten methodologischen Grundhaltung, **keine** Phase des empirischen Forschungsprozesses - also auch nicht die Phase der Hypothesenprüfung - heuristisch ungenutzt zu lassen. Durch viele konkrete Beispiele ließe sich belegen, daß Residuen auch dann neue und bessere Hypothesen heuristisch anregen können, wenn sie nicht ausreichen, um eine alte Hypothese (z.B. via Signifikanztestergebnis) zu „belasten“. Aus Platzgründen soll nur ein Beispiel besprochen werden, nämlich die Arbeit von Thomas (1981). Weitere, auf die Psychophysik bezogene Beispiele nennt Mausfeld (Kapitel 4 dieses Bandes).

Thomas (1981) hat entwicklungspsychologische Daten einmal ganz anders betrachtet als das üblicherweise geschieht. Sowohl in längs- als auch in querschnittlichen Analysen analysiert man bevorzugt die bedingte Erwartung von Verhaltens- und Leistungsvariablen für gegebenes Alter - die Lebensspannen-Mittelwertskurve also - und beachtet die Residuen nur insoweit, als sie die

Anwendbarkeit bestimmter statistischer Verfahren aus dem Bereich des ALM möglicherweise in Frage stellen können (Varianzhomogenitätsforderung). Typischerweise ist das Ergebnis, daß die Varianzhomogenität zwar verletzt ist, aber nicht so stark, daß man nicht trotzdem (unter Hinweis auf die „Robustheit“ von ALM-Analysen) die üblichen Mittelwertanalysen vornehmen könnte. Thomas (1981) hat die Residuen (genauer: die Varianz der Residuen) nun von einer Nebenrolle in eine Hauptrolle gerückt. Die sich bei graphischer Veranschaulichung geradezu aufdrängende Feststellung, daß Mittelwerte und Varianzen von Verhaltensvariablen über verschiedene Altersgruppen hinweg positiv korrelieren, hat Thomas (1981) heuristisch genutzt. Er hat sich gefragt, welche Eigenschaften ein Wachstums- und Entwicklungsmodell aufweisen muß, das diesem Befund gerecht wird. Sein Vorschlag läuft auf ein Modell hinaus, auf das man durch Hypothesentests im Rahmen des ALM niemals kommen kann, weil es impliziert, daß eine Modellannahme - eben die Varianzhomogenitätsannahme - für entwicklungspsychologische Datensätze nicht erfüllt sein kann. Offensichtlich handelt es sich um eine „gute Idee“ in dem Sinne, der in Abschnitt 2 herausgearbeitet wurde: Die üblichen Untersuchungsparadigmen der Entwicklungspsychologie werden überwunden, neue empirische Zugangs- und Analysemöglichkeiten werden aufgezeigt. Ferner werden Integrationsmöglichkeiten von Forschungsfeldern deutlich (mittelwerts-, varianz- und kovarianzzentrierte Entwicklungsforschung), die bislang noch nebeneinander existieren und (scheinbar) unabhängige Fragestellungen bearbeiten.

Dieses Beispiel zeigt wie viele andere, daß die Gewinnung neuer Hypothesen und Theorien durch exploratorische Residuenbetrachtungen im Anschluß an konfirmatorische Analysen konstruktiv angeregt werden kann. Dies ist der Grund, weshalb in Abbildung 1 exploratorische Datenanalysen (im Gegensatz zu konfirmatorischen) bei der Verwendung empirischer Daten als **obligatorisch** ausgewiesen wurden. Die Bewertung und Interpretation empirischer Daten im Hinblick auf eine wissenschaftliche Fragestellung oder Hypothese sollte also nicht ohne eine exploratorische Betrachtung der Daten vorgenommen werden. Dies gilt auch dann, wenn der Hauptzweck einer Untersuchung die strenge Prüfung einer psychologischen Hypothese ist.

## 5. Fazit

In den vorstehenden Abschnitten wurde die **Erzeugung** empirischer Daten auf der Grundlage wissenschaftlicher Beobachtungen sowie die **Verwendung** dieser Daten im Entdeckungs- und Begründungszusammenhang untersucht. Empirische Daten wurden dabei als Ergebnis der Abbildung von wissenschaftlichen Beobachtungen in einen Datenraum aufgefaßt. Der Datenraum selbst sowie die Zuordnung von empirischen Daten zu Beobachtungen sind dabei

nicht vorgegeben bzw. durch die Beobachtungsprotokolle automatisch determiniert. Erst eine fragestellungsabhängig zu treffende **Entscheidung** auf seiten des Wissenschaftlers bzw. der Wissenschaftlerin legt die Struktur des Datenraums und die Zuordnungsregel fest.

Man kann grob zwischen der Verwendung von empirischen Daten im Entdeckungszusammenhang (bei eher vagen „offenen wissenschaftlichen Fragestellungen“) und der Verwendung empirischer Daten im Begründungszusammenhang (bei präzisen „geschlossenen wissenschaftlichen Fragestellungen“) unterscheiden. Je nach Forschungskontext erweisen sich unterschiedliche Datenanalyseverfahren als hilfreich bzw. wenig hilfreich. Dennoch sollten beide Forschungskontexte nicht als disjunkte Kategorien von Forschungsaktivitäten aufgefaßt werden. Die zur Anregung von Innovationen erforderliche Metaregel, daß alle Phasen des Forschungsprozesses konsequent heuristisch genutzt werden sollten, verleiht exploratorischen Datenanalysen auch im Begründungskontext eine erhebliche Bedeutung. Umgekehrt sind jedoch konfirmatorische Verfahren im Entdeckungskontext nicht nur wenig nützlich, sondern sogar schädlich.

Insgesamt erscheint die Funktion empirischer Daten in der Psychologie nicht angemessen charakterisiert, wenn vor allem betont wird, daß empirische Daten psychologisches Wissen erzeugen und begründen. In den Prozeß der Gewinnung psychologischer Hypothesen und Theorien gehen wissenschaftliche Beobachtungen und empirische Daten nur neben anderen, nichtempirischen Heuristiken ein, sie spielen hier möglicherweise sogar eine Nebenrolle (vgl. Dörner, Kapitel 8 dieses Bandes). Die Verwendung empirischer Daten ist deshalb im Entdeckungskontext auch nicht obligatorisch. Umgekehrt ist die Verwendung empirischer Daten im Begründungskontext zwar obligatorisch, jedoch auch hier nicht in der Weise, daß sie - im induktiven Sinne - eine Bestätigung von Hypothesen oder Theorien erlauben.

Angemessener erscheint es deshalb, die **Kritikfunktion empirischer Daten** in den Vordergrund zu rücken. Die Diskussion hat gezeigt, daß der Hauptzweck von Datenanalysen - konfirmatorischen wie konsequent angewandten exploratorischen - eigentlich immer darin besteht, explizite oder implizite Vermutungen, Hypothesen oder Modellvorstellungen, die in **beliebiger** Weise gewonnen wurden, als inadäquat ausweisen zu können, wenn diese wirklich inadäquat sind. Empirische Daten sind somit in erster Linie ein Korrektiv. Sie stellen sicher, „(...) daß sich Psychologen **nachweislich** irren können“ (Herrmann, 1979, S.22). Dies gilt unabhängig davon, ob Daten im Begründungs- oder im Entdeckungskontext Verwendung finden.

## Literatur

- Andersson, G. (1988). **Kritik und Wissenschaftsgeschichte. Kuhn, Lakatos' und Feyerabends Kritik des Kritischen Rationalismus.** Tübingen: Mohr.
- Barber, T.X. (1976). **Pitfalls in human research.** New York: Pergamon.
- Barker, R. G. & Wright, H.F. (1951). **One boy's day. A specimen record of behavior.** New York: Harper & Brothers.
- Bischof, N. (1989). Emotionale Verwirrungen. Oder: Von den Schwierigkeiten im Umgang mit der Biologie. **Psychologische Rundschau, 40**, 188-205.
- Borg, I. (1981). **Anwendungsorientierte Multidimensionale Skalierung.** Berlin: Springer.
- Bredenkamp, J. (1969). über die Anwendung von Signifikanztests bei theoretisierenden Experimenten. **Psychologische Beiträge, 11**, 275-285.
- Bredenkamp, J. (1990). Kognitionspsychologische Untersuchungen eines Rechenkünstlers. In H. Feger (Hrsg.), **Wissenschaft und Verantwortung. Festschrift zum 60. Geburtstag von K.J. Klauer** (S.47-70). Göttingen: Hogrefe.
- Bridgeman, P. W. (1927). **The logic of modern physics.** New York: Macmillan.
- Brinberg, D. & McGrath, J.E. (1985). **Validity and the research process.** Beverly Hills: Sage.
- Brocke, B. (1979). Aspekte einer Methodologie der angewandten Sozial- und Verhaltenswissenschaften. **Zeitschrift für Sozialpsychologie, 10**, 2-29.
- Buchner, A. & Funke, J. (1993). Finite state automata: Dynamic task environments in problem solving research. **Quarterly Journal of Experimental Psychology, 46A**, 83-118.
- Bühler, K. (1908). Antwort auf die von W. Wundt erhobenen Einwände gegen die Methode der Selbstbeobachtung an experimentell erzeugten Erlebnissen. **Archiv für die gesamte Psychologie, 12**, 93-122.
- Bühler, K. (1978). **Die Krise der Psychologie.** Frankfurt: Ullstein (Originalausgabe: 1927).
- Bunge, M. & Ardila, R. (1987). **Philosophy of psychology.** New York: Springer.
- Campbell, D.T. & Stanley, J. C. (1970). Experimentelle und quasiexperimentelle Anordnungen in der Unterrichtsforschung. In K. Ingenkamp (Hrsg.), **Handbuch der Unterrichtsforschung, Teil I** (S. 448-631). Weinheim: Beltz. (Originalausgabe: Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.)
- Chapman, M. (1987). Inner processes and outward criteria: Wittgenstein's importance for psychology. In M. Chapman & R. A. Dixon (Eds.), **Meaning and the growth of understanding. Wittgenstein's significance for developmental psychology** (pp. 103-127). Berlin: Springer.
- Claparede, E. (1965). Die Entdeckung der Hypothese. In C.F. Graumann (Hrsg.), **Denken** (S. 109-115). Köln: Kiepenheuer & Witsch. (Originalausgabe: La decouverte de l'hypothese. Journal de Psychologie Normale et Pathologique, 29, 1932, 648-565.)

- Comte, A. (1974). **Die Soziologie. Die positive Philosophie im** Auszug (zweite Auflage). Stuttgart: Kröner. (Originalausgabe: Cours de philosophie positive, Vol. 1-6. Paris, 1830-1842.)
- Cook, T. D. & Campbell, D. T. (1979). **Quasi-experimentation. Design & analysis issues for field settings.** Boston: Houghton Mifflin.
- Coombs, C. H. (1964). **A theory of data.** New York: Wiley.
- Coombs, C. H., Dawes, R. M & Tversky, A. (1975). **Mathematische Psychologie. Eine Einführung.** Weinheim: Beltz. (Originalausgabe: Mathematical Psychology. Englewood Cliffs: Prentice Hall, 1970.)
- Danziger, K. (1980). The history of introspection reconsidered. **Journal of the History of the Behavioral Sciences, 16,** 241-262.
- Deffner, G. (1989). Interaktion zwischen Lautem Denken, Bearbeitungsstrategien und Aufgabenmerkmalen? Eine experimentelle Prüfung des Modells von Ericsson und Simon. **Sprache & Kognition, 8,** 98-111.
- Dingler, H. (1931). **Der Zusammenbruch der Wissenschaft und der Primat der Philosophie** (zweite Auflage). München: Reinhardt.
- Dörner, D. (1989). Die kleinen grünen Schildkröten und die Methoden der experimentellen Psychologie. **Sprache & Kognition, 8, 86-97.**
- Dörner, D. & Lantermann, E. D. (1991). Experiment und Empirie in der Psychologie. In K. Grawe, R. Hänni, N. Semmer & F. Tschan (Hrsg.), **Über die richtige Art, Psychologie zu betreiben** (S. 37-57). Göttingen: Hogrefe.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. **Annals of Statistics, 7,** 1-26.
- Epstein, W. & Rock, I. (1960). Perceptual set as an artifact of recency. **American Journal of Psychology, 73,** 214-228.
- Ericsson, K. A. (1990). Introspection. In M. W. Eysenck (Ed.), **The Blackwell dictionary of cognitive psychology** (pp. 193-195). Oxford: Blackwell.
- Ericsson, K.A. & Simon, H.A. (1980). Verbal reports as data. **Psychological Review, 87,** 215-251.
- Ericsson, K. A. & Simon, H. A. (1984). **Protocol analysis: Verbal reports as data.** Cambridge: MIT Press.
- Eysenck, M. W. (1990). Consciousness. In M. W. Eysenck (Ed.), **The Blackwell dictionary of cognitive psychology** (pp. 84-85). Oxford: Blackwell.
- Eysenck, M. W. & Keane, M.T. (1990). **Cognitive psychology. A student's handbook.** Hove and London: Erlbaum.
- Faßnacht, G. (1979). **Systematische Verhaltensbeobachtung.** München: Reinhardt.
- Feger, H. (1983). Planung und Bewertung von wissenschaftlichen Beobachtungen. In H. Feger & J. Bredenkamp (Hrsg.), **Datenerhebung** (= Enzyklopädie der Psychologie, Themenbereich B, Serie I, Band 2, S. 1-75). Göttingen: Hogrefe.
- Feger, H. & Graumann, C. F. (1983). Beobachtung und Beschreibung von Erleben und Verhalten. In H. Feger & J. Bredenkamp (Hrsg.), **Datenerhebung** (= Enzyklopädie der Psychologie, Themenbereich B, Serie I, Band 2, S.76-134). Göttingen: Hogrefe.

- Feyerabend, P. (1983). **Wider den Methodenzwang** (zweite Auflage). Frankfurt: Suhrkamp. (Originalausgabe: Against method, London: New Left Books, 1975).
- Foppa, K. & Cranach, M. von (1991). De modo ad explorandarum animarum artem colendam maxime idoneo. (Über die richtige Art, Psychologie zu betreiben). In K. Grawe, R. Hänni, N. Semmer & F. Tschan (Hrsg.), **Über die richtige Art, Psychologie zu betreiben** (S. 59-75). Göttingen: Hogrefe.
- Gadenne, V. (1976). **Die Gültigkeit psychologischer Untersuchungen**. Stuttgart: Kohlhammer.
- Gadenne, V. (1984). **Theorie und Erfahrung in der psychologischen** Forschung. Tübingen: Mohr.
- Gadenne, V. & Oswald, M.E. (1991). **Kognition und Bewußtsein**. Berlin: Springer.
- Garnes, P.A. (1983). Curvilinear transformations of the dependent variable. **Psychological Bulletin**, 93, 382-387.
- Gigerenzer, G. (1980). **Messung und Modellbildung in der Psychologie**. München: Reinhardt.
- Gigerenzer, G. (1988). Woher kommen Theorien über kognitive Prozesse? **Psychologische Rundschau**, 39, 91-100.
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. **Psychological Review**, 98, 254-267.
- Glück, G. (1971). Methoden der Beobachtung. In G. Dohmen (Hrsg.), **Forschungstechniken für die Hochschuldidaktik** (S. 57-62). München: Beck.
- Good, I.J. (1983). The philosophy of exploratory data analysis. **Philosophy of Science**, 50, 283-295.
- Gopnik, A. (im Druck). How we know our minds: The illusion of first-person knowledge of intentionality. **Behavioral and Brain Sciences**.
- Graumann, C. F. (1966). Grundzüge der Verhaltensbeobachtung. In E. Meyer (Hrsg.) **Fernsehen in der Lehrerausbildung** (S. 86-107). München: Manz.
- Green, D.M. & Swets, J. A. (1974). **Signal detection theory and psychophysics**. New York: Krieger.
- Greve, W. & Wentura, D. (1991). **Wissenschaftliche Beobachtung in der Psychologie. Eine Einführung**. München: Quintessenz.
- Groeben, N. (1972). **Literaturpsychologie. Literaturwissenschaft zwischen Hermeneutik und Psychologie**. Stuttgart: Kohlhammer.
- Groeben, N. & Scheele, B. (1977). **Argumente für eine Psychologie des reflexiven Subjekts**. Darmstadt: Steinkopff.
- Hager, W. (1992). **Jenseits von Experiment und Quasi-Experiment. zur Struktur psychologischer Versuche und zur Ableitung von Vorhersagen**. Göttingen: Hogrefe.
- Hartwig, E & Dearing, B.E. (1981). **Exploratory data analysis**. Beverly Hills: Sage.
- Heckhausen, H. (1963). **Hoffnung und Furcht in der Leistungsmotivation**. Meisenheim: Glan.
- Herrmann, T. (1972). **Lehrbuch der empirischen Persönlichkeitsforschung (2., unveränderte Aufl.)**. Göttingen: Hogrefe.
- Herrmann, T. (1979). **Psychologie als Problem**. Stuttgart: Klett.

- Herrmann, T. (1990). Die Experimentiermethodik in der Defensive? ***Sprache & Kognition***, 9, 1-11.
- Hoaglin, D.C., Mosteller, F. & Tukey, W. (1983). ***Understanding robust and exploratory data*** analysis. New York: Wiley.
- Holzkamp, K. (1964). ***Theorie und Experiment in der Psychologie. Eine grundlegende kritische Untersuchung***. Berlin: de Gruyter.
- Holzkamp, K. (1972). Verborgene anthropologische Voraussetzungen der allgemeinen Psychologie. In K. Holzkamp, ***Kritische Psychologie. Vorbereitende Arbeiten*** (S. 35-73). Frankfurt: Fischer.
- Hubert, L. & Schultz, J. (1976). Quadratic assignment as a general data analysis strategy. ***British Journal of Mathematical and Statistical Psychology***, 29, 190-241.
- Hultsch, D.F. & Hickey, T. (1978). External validity in the study of human development. ***Human Development***, 21, 76-91.
- Hume, D. (1978). ***Treatise on human nature*** (2nd edition). Oxford: Clarendon. (Originalausgabe: 1739-1740.)
- Kant, I. (1977a). über die von der königl. Akademie der Wissenschaften zu Berlin für das Jahr 1791 ausgesetzte Preisfrage: Welches sind die wirklichen Fortschritte, die die Metaphysik seit Leibnizens und Wolffs Zeiten in Deutschland gemacht hat? In W. Weischedel (Hrsg.), ***Immanuel Kant. Schriften zur Metaphysik und Logik 2. Werkausgabe Band VI*** (S. 583-676). Frankfurt: Suhrkamp. (Originalausgabe: Königsberg: Goebbels und Unzer, 1804.)
- Kant, I. (1977b). Anthropologie in pragmatischer Hinsicht. In W. Weischedel (Hrsg.), ***Immanuel Kant. Schriften zur Anthropologie, Geschichtsphilosophie, Politik und Pädagogik 2. Werkausgabe Band XII*** (S. 399-690). Frankfurt: Suhrkamp. (Originalausgabe: Königsberg: Nicolovius, 1798.)
- Kaufman, L. (1979). ***Perception. The world transformed***. New York: Oxford University Press.
- Kendler, T. S. (1979). Cross-sectional research, longitudinal theory, and a discriminative transfer ontogeny. ***Human Development***, 22, 235-254.
- Kleiner, B. & Hartigan, J. A. (1981). Representing points in many dimensions by trees and castles. ***Journal of the American Statistical Association***, 76, 260-269.
- Kluwe, R.H. (1988). Methoden der Psychologie zur Gewinnung von Daten über menschliches Wissen. In H. Mandl & H. Spada (Hrsg.), ***Wissenspsychologie*** (S. 359-385). München: Psychologie Verlags Union.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), ***Criticism and the growth of knowledge*** (pp.91-195). Cambridge: Cambridge University Press.
- Leeper, R. (1935). A study of a neglected portion of the field of learning: The development of sensory organization. ***Journal of Genetic Psychology***, 46, 41-75.
- Levine, M. (1966). Hypothesis behavior by humans during discrimination learning. ***Journal of Experimental Psychology***, 71, 331-338.
- Lüer, G. (1988). Kognitive Prozesse und Augenbewegungen. In H. Mandl & H. Spada (Hrsg.), ***Wissenspsychologie*** (S. 386-399). München: Psychologie Verlags Union.
- Lyons, W. (1986). ***The disappearance of introspection***. Cambridge: MIT Press.



- Manns, M., Schultze, J., Herrmann, C. & Westmeyer, H. (1987). **Beobachtungsverfahren in der Verhaltensdiagnostik. Eine systematische Darstellung** ausgewählter Beobachtungsverfahren. Salzburg: Müller.
- Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, **34**, 103-115.
- Mosteller, F. & Tukey, J. W. (1977). **Data analysis and regression. A second course in statistics**. Reading: Addison Wesley.
- Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, **84**, 231-259.
- Oldenbürger, H.-A. (im Druck). Exploratorische, graphische und robuste Datenanalyse. In E. Erdfelder, R. Mausfeld & G. Rudinger (Hrsg.), **Methodenlehre. Ein Handbuch in Schlüsselbegriffen**. Weinheim: Psychologie Verlags Union.
- Oswald, M. (1980). **Jenseits der Methodologie des Behaviorismus. Eine Analyse der methodologischen Grundlagen behavioristischer und kognitiver Theorien**. Unveröff. Diss., Universität Mannheim.
- Pongratz, L.J. (1984). **Problemgeschichte der Psychologie** (zweite Auflage). München: Francke.
- Popper, K. R. (1935/1982). **Logik der Forschung** (siebente, verbesserte und durch sechs Anhänge vermehrte Auflage). Tübingen: Mohr. (Originalausgabe: Wien: Springer, 1935).
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, **118**, 219-235.
- Reichenbach, H. (1938). **Experience and prediction. An analysis of the foundations and the structure of knowledge**. Chicago: University of Chicago Press.
- Riedwyl, H. & Schafroth, M. (1976). Grafische Darstellung mehrdimensionaler Beobachtungen. *EDV in Medizin und Biologie*, **1**, 21-24.
- Roskam, E. E. (1983). Allgemeine Datentheorie. In H. Feger & J. Bredenkamp (Hrsg.), **Messen und Testen** (= Enzyklopädie der Psychologie, Themenbereich B, Serie I, Band 3, S. 1-135). Göttingen: Hogrefe.
- Roskam, E.E. (im Druck). Beobachtung und Daten. In E. Erdfelder, R. Mausfeld & G. Rudinger (Hrsg.), **Methodenlehre. Ein Handbuch in Schlüsselbegriffen**. Weinheim: Psychologie Verlags Union.
- Russo, J. E., Johnson, E. J. & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, **17**, 759-769.
- Schwager, K. W. (1991). The representational theory of measurement: An assessment. *Psychological Bulletin*, **110**, 618-626.
- Stachowiak, H. (1973). **Allgemeine Modelltheorie**. Wien: Springer.
- Tack, W. H. (1977). Probleme des Messens im Bereich des Psychischen. In G. Strube (Hrsg.), **Binet und die Folgen** (= Die Psychologie des 20. Jahrhunderts, Band 5, S. 13-64). Zürich: Kindler.
- Tack, W. H. (1983). Psychophysische Methoden. In H. Feger & J. Bredenkamp (Hrsg.), **Messen und Testen** (= Enzyklopädie der Psychologie, Themenbereich B, Serie I, Band 3, S.346-426). Göttingen: Hogrefe.

- Teegen, F., Grundmann, A. & Röhrs, A. (1975). **Sich ändern lernen. Anleitung zur Selbsterfahrung und Verhaltensmodifikation.** Reinbek: Rowohlt.
- Thomas, H. (1981). Eine Wachstumstheorie der Entwicklung. In R.H. Kluwe & H. Spada (Hrsg.), **Studien zur Denkentwicklung** (S. 93-144). Bern: Huber.
- Tukey, J. W. (1977). **Exploratory data analysis.** Reading, Mass.: Addison-Wesley.
- Tukey, J.W. (1980). We need both exploratory and confirmatory. **The American Statistician, 34, 23-25.**
- Wainer, H. & Thissen, D. (1981). Graphical data analysis. **Annual Review of Psychology, 32, 191-241.**
- Wallach, H. (1949). Some considerations concerning the relation between perception and cognition. **Journal of Personality, 18, 6-13.**
- Weiskrantz, L. (1990). Blindsight. In M. W. Eysenck (Ed.), **The Blackwell dictionary of cognitive psychology** (pp. 44-46). Oxford: Blackwell.
- Wilson, A. (1975). The inference of covert hypotheses by verbal reports in concept-learning research. **Quarterly Journal of Experimental Psychology, 27, 313-322.**
- Wundt, W. (1921). Selbstbeobachtung und innere Wahrnehmung. In W. Wundt, **Kleine Schriften. Dritter Band** (S. 423-440). Stuttgart: Kröner. (Originalausgabe: Philosophische Studien, 4, 1888, 292-309.)
- Wundt, W. (1905). **Grundriss der Psychologie** (siebente, verbesserte Auflage). Leipzig: Engelmann.
- Wundt, W. (1907). über Ausfrageexperimente und über die Methoden zur Psychologie des Denkens. **Psychologische Studien, 3, 301-360.**
- Wundt, W. (1908). Kritische Nachlese zur Ausfragemethode. **Archiv für die gesamte Psychologie, 11, 445-459.**

## **Autorenhinweis**

Für aus- und ergiebige Diskussionen zum Thema des vorliegenden Kapitels sowie kritische Anmerkungen zu Vorformen der Arbeit möchte ich Prof. Dr. J. Bredenkamp, Dr. A. Bucher, PD Dr. J. Funke, Prof. Dr. V. Gadenne, Dr. W. Greve, Lisa Irmen, Jörg Mutke, Dr. H.-A. Oldenbürger und Prof. Dr. W.H. Tack herzlich danken.

### 3. Kapitel

## Qualitatives Vorgehen - „Interpretation“

*Hans-Georg Soeffner **und** Ronald Hitzler*

Die Lehre vom **interpretativen Vorgehen**, die Hermeneutik, hat im Verlauf ihrer Geschichte Veränderungen durchgemacht, Erweiterungen erfahren und - als selbstreflexiver Erkenntnisstil - durch Kritik maßgeblich zur eigenen Veränderung beigetragen. In historisch-systematischer Hinsicht ist sie gebunden an das Prinzip der Schriftlichkeit, zunächst an die Aufzeichnung von Sprache (an Texte), im weiteren aber - folgerichtig - an die Qualität des Aufgezeichnet-Seins von ‚Daten‘ (im weitesten Sinne) überhaupt, also an die Fixiertheit und damit tendenziell unendlich wiederholbare Abrufbarkeit (d.h. Diskursivität) von sowohl sprachlichen **als** auch nicht-sprachlichen Dokumenten: von menschlichen (Ent-)Äußerungen, Handlungen und Produktionen, von menschlichen Objektivationen jeder Art.

Ergänzt und verändert worden ist aber auch die generelle Fragestellung der Hermeneutik. Geht es traditionell ausschließlich um das ‚Was‘ des Verstehens, so geht es nun - gerade auch in den Verfahren, die sich unter dem Etikett einer **sozialwissenschaftlichen** Hermeneutik versammeln lassen - mehr und mehr um das ‚Wie‘: um das **Verstehen des Verstehens** selbst, um Verfahren, ‚Regeln‘, ‚Muster‘, implizite Prämissen, sozialisatorisch vermittelte Aneignungs-, Unterweisungs- und Überlieferungsweisen des Deutens und Verstehens.

### **1. Verstehen in Alltag **und** Wissenschaft**

#### 1.1 Phänomenologie des Verstehens

**Verstehen** können wir jenen Vorgang nennen, der einer Erfahrung Sinn verleiht. **Fremdverstehen** können wir jenen Vorgang nennen, bei dem wir einer Erfahrung den Sinn verleihen, daß sie sich auf ein Ereignis in der Welt bezieht, dem alter ego bereits einen Sinn verliehen hat.

Verstehen als Vorgang, der sich auf meine eigenen Bewußtseinsleistungen bezieht, liegt einerseits **logisch** dem Verstehen des alter ego zugrunde, andererseits aber ist **empirisch** Selbstverstehen ein Produkt dessen, was in der Tradition des Symbolischen Interaktionismus seit Charles H. Cooley (1902) als ‚Spiegelungsprozeß‘ bezeichnet wird, also sozusagen eine Übertragung des Verstehens anderer auf mein Bewußtsein. Selbstverstehen ist prinzipiell kontinuierlich und vollständig möglich, denn Erlebnisse und Erfahrungen tragen zunächst **keinen** Sinn in sich. Vielmehr konstituiert das subjektive Bewußtsein Sinn dadurch, daß es die Erfahrung auf anderes bezieht. Dieser Akt der Sinnerschöpfung beinhaltet im wesentlichen das, was Verstehen als Selbstverstehen meint. Fremdverstehen hingegen, und das ist das für das Problem des Verstehens überhaupt Entscheidende, geschieht in **Auffassungsperspektiven**. D. h., Fremdverstehen ist nur diskontinuierlich und partiell möglich. Selbstverstehen ist ein prinzipiell unzweifelhafter Akt; Fremdverstehen ist ein prinzipiell zweifelhafter Akt.

Fremdverstehen basiert auf **meinen** Erlebnissen und Erfahrungen von alter ego. Jeder Sinn, den ich ihm unterstelle, kann abweichen von dem Sinn, den alter ego selber seinen Erfahrungen verleiht. Ich erfasse stets nur Fragmente seines tatsächlichen Erlebens. Und ich verstehe stets nur **möglicherweise** den von ihm subjektiv tatsächlich gemeinten Sinn. Das Bewußtsein von alter ego appräsentiert sich mir über Anzeichen und über Zeichen. Diese weisen drei Sinnschichten auf (dazu auch Eberle, 1984): 1. einen objektivierten, intersubjektiv gültigen Sinn (d.h., das Zeichen ist dem Bezeichneten einsinnig zuordenbar, unabhängig vom Zeichensetzenden und vom Zeichendeutenden, es ist seinem Sinn nach invariant und immer wieder anwendbar; z.B.: Dies ist die Melodie von ‚Hänschen klein‘); 2. einen subjektiven Sinn (d.h., das Zeichen hat dazuhin für den Zeichensetzenden und/oder den Zeichendeutenden eine individuelle Zusatzbedeutung; z. B.: ‚Hänschen klein‘ ist meine Lieblingsmelodie); und 3. einen okkasionellen Sinn (d.h., das Zeichen hat eine spezielle, kontextabhängige Bedeutung, die sich aus dem situativen Gesamtzusammenhang erschließt; z.B.: ein kleines Kind übt die Melodie von ‚Hänschen klein‘ auf der Blockflöte).

Wenn ich alter ego verstehen will, dann muß ich seine subjektiven Motive auslegen und den objektiven, subjektiven und okkasionellen Sinn seiner ‚Bezeichnungen‘ rekonstruieren. Damit dürfte plausibel werden, daß Verstehen fremden Sinnes nur näherungsweise gelingen kann. Der tatsächlich gemeinte Sinn eines Handelnden und das, was von einem anderen als ‚gemeinter Sinn‘ gedeutet wird, ist prinzipiell nicht identisch. Letzteres ist allenfalls ein **typischer** Näherungswert zum ersteren.

Ganz fraglos verfügt ego im Alltag **nicht** über ein Monopol zur Interpretation der Welt, sondern befindet sich empirisch immer schon in einer von anderen

vorinterpretierten Welt. Alltäglich gilt, daß die Standpunkte des einen und des anderen im großen und ganzen austauschbar sind, daß das, was für den einen relevant ist, auch relevant für den anderen wäre, wäre er an dessen Stelle, und daß es immer wieder und bis auf weiteres weiterhin so ist und sein wird. Das heißt, alltäglich ist Fremdverstehen **nicht** an sich problematisch, sondern fraglose Routine. Während erkenntnistheoretisch das Problem des Fremdverstehens also darin besteht, zu erklären, wie letzteres überhaupt möglich ist, stellt es in der Alltagseinstellung eine ziemlich banale Bewußtseinsleistung dar: Es ist für Menschen „normalerweise“ so normal, daß es gar nicht Gegenstand ihres alltäglichen Interesses wird.

Menschliches Verhalten, als beobachtbare Form menschlichen Handelns - sei es nichtsprachlicher oder sprachlicher Art - ist von und für Menschen interpretierbar, weil es neben vielen anderen Eigenschaften immer die der (Proto-) Zeichenhaftigkeit aufweist. Von der Geste bis zum ‚signifikanten‘ Symbol, vom Anzeichen und Symptom bis zum konstruierten und eindeutig definierten mathematischen Zeichen, vom Körper- und Gesichtsausdruck bis zur Kleidung, vom Natureindruck bis zum menschlichen Produkt ordnen wir uns und unserer Umwelt Zeichenqualitäten zu und konstituieren damit den menschlichen Interpretationshorizont (Wundt, 1921; Mead, 1934; Bühler, 1934). Dabei korrespondieren den verschiedenen Zeichenarten und ihrer jeweils unterschiedlichen Semantik und Verknüpfungsform auch unterschiedliche Deutungsprozeduren (dazu auch Schütz & Luckmann 1984, S. 178-200).

Verstehen ist also durchaus keine Erfindung der Sozialwissenschaften. Es geschieht zunächst auch nicht in einer besonderen theoretischen Einstellung, sondern es ist für Menschen einfach Alltagsroutine. Das permanente Problem verstehender Wissenschaftler besteht deshalb darin, zu plausibilisieren, was ihr Tun eigentlich zu einem wissenschaftlichen Unternehmen mache, obwohl es doch explizit auf einem ganz alltäglichen, allgemeinmenschlichen Vermögen aufruht. D.h., sie haben in aller Regel ganz praktische Schwierigkeiten, nicht nur Nicht-Wissenschaftlern, sondern auch ihren Fachkollegen zu erklären, wozu das, was sie tun, als **wissenschaftliches** Tun nützlich sei.

## 1.2 Geschichte des hermeneutischen Verstehens

Die Tradition der **Hermeneutik**, der Lehre vom interpretativen Vorgehen, hat sich ursprünglich als Hilfswissenschaft zur Auslegung von Texten, vor allem in der Theologie und in der Jurisprudenz, entwickelt.<sup>1</sup> Besonders der Typus theologischer Auslegungsarbeit vieler Deuter an einer „Heiligen Schrift“

<sup>1</sup> Neuere Überblicke zur Entwicklung der Hermeneutik z.B. bei Gadamer & Boehm (1976, 1978), Grondin (1991), Nassen (1979) und Riedel (1978).

drückte lange Zeit den verschiedenen Denk- und Interpretationsrichtungen unübersehbar ihren Stempel auf. Es war die Suche nach dem im Text repräsentierten **einen** (einzigem) Subjekt und nach dessen Willen. Verwiesen sei hierzu lediglich exemplarisch auf die mittelalterliche Bibelhermeneutik, die sich insbesondere darauf konzentriert hat, den sogenannten vierfachen Sinn der Heiligen Schrift auszulegen: 1. den literarischen Sinn (durch Explikation des historischen Berichts), 2. den allegorischen Sinn (durch Explikation der Glaubensgehalte), 3. den tropologischen Sinn (durch Explikation der moralischen Implikationen für das Verhalten der Menschen untereinander), und 4. den anagogischen Sinn (durch Explikation der Offenbarung des Zukünftigen, durch Hinführung zum ‚höheren‘ Sinn).

Gleichzeitig allerdings pflegten schon früh Philosophie und Juristerei auch mit ganz anderen als dem **einen** Subjekt und dem einen Sinn umzugehen. Durch diesen ganz anderen Umgang mit ihren Subjekten und Objekten stellten sie dem Singular der einen Wahrheit den Plural der vielen (möglichen) Wahrheiten und Wahrscheinlichkeiten an die Seite: Der Erschließung der **einen** Bedeutung, des **einen** Sinns die Entfaltung von Bedeutungs- und Sinn-Potentialen. Die Suche nach Deutungsalternativen, der methodische Zweifel und damit die hermeneutische Skepsis waren aber auch hier keineswegs von Anfang an zugleich mit den bewußten Praktiken und Methoden der Auslegung da. Sie entwickelten sich historisch ‚hinter dem Rücken‘ der Interpreten, bis sie schließlich deren Bewußtsein einholten.

Hermeneutische Praxis im Rahmen einer Kunstlehre der Auslegung explizierte und präziserte ihre Regeln und ihre Zielsetzungen zunächst, indem sie sich an ‚heiligen Schriften‘, ehrwürdigen Codices und ‚Quellen‘ abarbeitete. Das wirkte sich aus bis auf die Verfahren der Rechtshermeneutik des frühen 19. Jahrhunderts, wie sie sich insbesondere um Friedrich Karl von Savigny entwickelt hat. Hier ging es noch insbesondere um eine vierdimensionale Gesetzesauslegung: 1. im Hinblick auf die Grammatik (also auf den Wortsinn des Textes), 2. im Hinblick auf die Systematik (also auf die Verortung des einzelnen Gesetzes im Rahmen des Gesetzeszusammenhangs), 3. im Hinblick auf die Historik (also auf die geschichtliche Entstehung und die geschichtliche Situation des Textes), und 4. im Hinblick auf die Teleologie (also auf den Zweck des Gesetzes hin).

Aus der philologischen Detailarbeit, aus der Fortschreibung theologischer Predigt- und Deutungspraxis, aus der Filigranarbeit der Gesetzeskommentierung und aus der tradierten Rechtsprechungspraxis entwickelte sich gleichwohl allmählich ein historisch-kritisches, rekonstruktives Verstehen als Methodologie **und** als wissenschaftliche Haltung ‚wertfreier‘ Distanz. An dieser Zwischenstation der Entwicklung einer wissenschaftlichen Methodologie des Verstehens und des bewußten Versuches, die wissenschaftliche Auslegung um

eine Phänomenologie der Deutungsakte und Deutungsleistungen zu ergänzen, wurde die Hermeneutik als verstehende Wissenschaft dann zwangsläufig umgeformt und erweitert zu einer Wissenschaft vom Verstehen.

Auslegung wurde nun nicht länger als bloße Methode, Hermeneutik nicht mehr ausschließlich als Methodologie oder Kunstlehre des Verstehens begriffen: das Verstehen, seine Konstitution, Regeln und Motive wurden vielmehr selbst Gegenstand der Aufmerksamkeit. Die Hermeneutik wurde **selbstreflexiv** insbesondere durch die Pionier-Arbeiten von Friedrich Schleiermacher (1777), der Hermeneutik als Technik der divinatorisch-komparativen Auslegung von Texten durch ein Sich-Hineinversetzen des Interpreten in die ‚Welt‘ des Autors propagiert, und die von Wilhelm von Humboldt, welcher die berühmte ‚Zirkelstruktur des Verstehens‘ (d.h. des Auslegens auf der Grundlage von Vorausgelegtem) formuliert hat (z.B.: Ich erkenne aktuell ein Geräusch als ein „Lied“. Ich weiß also bereits, was ein Lied ist. Zudem ‚erkenne‘ ich die ‚**Melodie**‘ von ‚Hänschen klein‘ und assoziiere den Text über die Melodie. Diese bestimmte Melodie erkenne ich auf der Grundlage meines Vorverständnisses einer bestimmten, rhythmischen Aufeinanderfolge von Tönen als „die Melodie ‚Hänschen klein‘“. Jedes Wiederhören und Wiedererkennen von ‚Hänschen klein‘ vertieft wiederum meine Kenntnis über die ‚richtige‘ Tonfolge dieser Melodie, usw.).

Über das lebensphilosophisch-dialektische Konzept des Ausdrucksverstehens von Wilhelm Dilthey (1970), die phänomenologisch-hermeneutische Existenzialontologie von Martin Heidegger (1972) und die historisch-philosophische Hermeneutik von Hans-Georg Gadamer (1960) ist dann schließlich aus der philosophischen Hermeneutik eine hermeneutische Philosophie entstanden, innerhalb derer die methodologische Sonderstellung der Geisteswissenschaften ihre Begründung findet vor der Frage, wie Verstehen überhaupt möglich sein kann: Der Betrachtung und Erklärung natürlicher Ereignisse ‚von außen‘ wird die Teilhabe an und das Verstehen von kulturellen Phänomenen ‚von innen‘ gegenübergestellt.

Diese im hermeneutischen Erkenntnisstil angelegte Selbstreflexion führte auch zur Entdeckung und Problematisierung des Lesers und des Interpreten als des ‚Koautors von Texten‘ (dazu Jauß, 1982; Iser, 1972). Dieser implizite Mitgestalter füllt durch seine Interpretation den interpretierten Text (im weitesten Sinne) mit seinen eigenen Bedeutungen und Bedeutungshorizonten auf. Er ‚öffnet‘ - im Rekurs auf seine Biographie und seine geschichtliche Situation - systematisch und historisch das Bedeutungspotential des Textes. Dabei vollzieht sich sowohl auf der Seite des Interpreten als auch auf der des jeweiligen Dokumentes der Wandel von einer absoluten zu einer historischen Semantik (dazu bereits Simmel, 1918; und Bultmann, 1954). Und gleichzeitig wächst die Einsicht in unsere unaufhebbare Involviertheit in Sinn und Bedeutung:

die Einsicht in die Tatsache, daß wir immer schon deuten, deuten müssen, uns in einer immer schon gedeuteten Welt und in scheinbar kaum kontrollierbaren Bedeutungsüberschüssen bewegen. Wissenschaftliche Hermeneutik hat dementsprechend die Maxime, **alles** sehen, denken und auslegen zu wollen.

Die **sozialwissenschaftliche** Hermeneutik steht nun in dieser Tradition der (klassischen) Hermeneutik der Geisteswissenschaften einerseits und in der der Verstehenden Soziologie im Sinne Max Webers andererseits. Weber hat ja, und das war und ist für die weitere Entwicklung essentiell, jegliche Form **intuitiven** Verstehens für die empirischen Wissenschaften zurückgewiesen und statt dessen typische Rekonstruktionen vermittelt rationaler Urteilsvollziehung eingefordert (dazu auch Schütz, 1932). ‚Verstehen‘ heißt, so Weber (1972), „deutende Erfassung: a) des im Einzelfall real gemeinten . . . oder b) des durchschnittlich und annäherungsweise gemeinten . . . oder c) des für den reinen Typus (Idealtypus) einer häufigen Erscheinung wissenschaftlich zu konstruierenden (idealtypischen) Sinnes oder Sinnzusammenhangs“ (S. 4).

Darüber hinaus bezieht sich die sozialwissenschaftliche Hermeneutik in ihren verschiedenen Strömungen zum Teil auf die idealistisch-strukturalistische Theorietradition, auf die Theorie des Symbolischen Interaktionismus, auf die phänomenologische Protosoziologie und auf die anthropologisch und historisch informierte Wissenssoziologie. Unter diesem ‚Dach‘ lassen sich derzeit wohl vor allem die folgenden Ansätze versammeln: **Objektive Hermeneutik** (z. B. Oevermann, 1991; Garz/Kraimer, 1992), **Rekonstruktive Hermeneutik** (z.B. Soeffner, 1989, 1992; Reichertz, 1991; Lau 1992; Voß 1992), **Deutungsmusteranalyse** (z. B. Matthiesen, 1992; Lüders, 1991), **Dokumentarische Methode** (z.B. Bohnsack, 1991), **Milieuanalyse** (z.B. Grathoff, 1989; Hildenbrand, 1991), **Lebensweltanalyse** (z.B. Honer, 1993; Knoblauch, 1992), **Empirischer Konstruktivismus** (z. B. Knorr Cetina, 1984, 1989), **Typologische Analyse** (z. B. Gerhardt, 1986, 1991), **Geschichtenanalyse** (z. B. Vonderach, 1986), **Narrationsanalyse** (z.B. Riemann, 1987; Schütze, 1989; Hauptert, 1991), **Gattungsanalyse** (z. B. Luckmann, 1988; Bergmann, 1987), **Konversationsanalyse** (z.B. Bergmann, 1991; Knauth/Wolff, 1991).<sup>2</sup>

So heterogen diese Ansätze sind, sie haben ein gemeinsames Postulat, das des **Zweifels**: 1. des Zweifels als Grundeinstellung des Interpretieren, 2. des Zweifels auch (und insbesondere) an den Vor-Urteilen des Interpretieren, und 3. des Zweifels an (monistischen) theoretischen Erklärungen. Als geistes- und sozialwissenschaftliche Variante einer allgemeineren wissenschaftlichen Einstellung repräsentiert die sozialwissenschaftliche Hermeneutik somit insgesamt eine distanzierte, rekonstruktiv-konstruktive Haltung gegenüber sogenannten Fak-

2 Zu weiteren, gegenwärtig in den Sozialwissenschaften diskutierten Hermeneutik-Konzepten siehe Müller-Doohm (1990), vgl. aber - sozusagen zum ‚Paradigma‘ - auch Giddens (1984, S. 27-84) und Habermas (1981, S. 152-203, 1982, S. 89-366).



ten und eine dauerhafte, methodisch eingesetzte Skepsis gegenüber ‚positivem Wissen‘. Denn „durch die zunehmende Beschleunigung (...) sozialen Wandels wird immer schneller Vertrautes zu Fremdem, zum Unverständlichen: nicht nur das, was neu ist in der Welt, sondern auch und gerade das in ihr Alte“ (Marquard, 1981, S. 584).

### 1.3 Besonderheiten sozialwissenschaftlichen Verstehens

Jeder Sozialwissenschaftler hat es, bevor er sich an Prognosen wagt, zunächst einmal mit der Beschreibung und Analyse jener Konstruktionen zu tun, auf die sich das Handeln und Planen von Gesellschaftsmitgliedern in alltäglicher, pragmatischer Perspektive beziehen: der Konstruktionen ‚erster Ordnung‘ (Schütz, 1971, S. 3-54) - der alltäglichen, soziohistorisch verankerten Typen, Modelle, Routinen, Plausibilitäten, Wissensformen, Wissensbestände und (oft impliziten) Schlußverfahren. Indem er sich damit beschäftigt, verdoppelt er in seinen Rekonstruktionen nicht einfach die jeweiligen Konstrukte alltäglichen Handelns. Vielmehr überzieht er in den Prozessen des Beschreibens, Verstehens und Erklärens die alltäglichen Konstruktionen mit einem Netz von Kategorisierungen, idealtypischen Annahmen, Modellen, ex-post-Schlüssen und Kausalisierungen oder Finalisierungen.

Das bedeutet vor allem, daß die „Daten“ des Sozialwissenschaftlers, anders als die Daten des Naturwissenschaftlers, vorinterpretiert sind, daß seine Konstruktionen eben Konstruktionen von Konstruktionen sind, die so beschaffen sein müssen, „daß ein Handelnder in der Lebenswelt dieses typisierte Handeln ausführen würde, falls er völlig klares und bestimmtes Wissen von allen Elementen, und nur von diesen Elementen hätte, die der Sozialwissenschaftler als für sein Handeln relevant voraussetzt, und falls er die konstante Neigung hätte, die angemessensten zur Verfügung stehenden Mittel zur Erreichung seiner vermittels der Konstruktion definierten Zwecke einzusetzen“ (Schütz, 1971, S.51). Kurz: Der Sozialwissenschaftler entwirft Konstruktionen ‚zweiter Ordnung‘. Diese sind (wissenschaftstheoretisch auch formal modellhaft darstellbare) kontrollierte, methodisch überprüfte und überprüfbare, verstehende Rekonstruktionen der Konstruktionen ‚erster Ordnung‘ (dazu auch Helling, 1979).

Der wissenschaftliche Interpret macht also nichts prinzipiell anderes, als das, was Menschen im Alltag auch tun: Er deutet Wahrnehmungen als Verweise auf einen ihnen zugrundeliegenden Sinn hin. Aber anders als der Alltagsmensch versucht der wissenschaftliche Interpret, sofern er hermeneutisch reflektiert arbeitet, sich über die Voraussetzungen und die Methoden seines Verstehens Klarheit zu verschaffen. Denn dadurch, und nur dadurch, wird Verstehen zu einer wissenschaftlichen Methode. Dadurch auch erst wird Verstehen systematisch lehr- und lernbar.

Zwischen den Konstruktionen ‚erster Ordnung‘ und ‚zweiter Ordnung‘ besteht somit eine logische Differenz (dazu auch Carnap, 1928) und mehr als nur diese. Das Handeln, auf das sich die Rekonstruktionen beziehen, ist, wenn jene beginnen, längst vorüber, ein für allemal vergangen und nicht wiederholbar. Es muß - sofern es überhaupt der Interpretation zugänglich sein kann - in bestimmten ‚Daten‘ (Spuren) repräsentiert sein, und es ‚präsentiert‘ sich in den Daten als abgeschlossene Handlung. Da es ihm um überprüfbare, d.h. intersubjektiv verstandesmäßig nachvollziehbare Rekonstruktionen geht, kann der Sozialwissenschaftler diese Handlungen letztlich weder kongenial nachvollziehen, noch empathisch in die Seelen und Gemüter, Gedanken und Empfindungen der (damals) Handelnden einziehen wollen: er wird ‚rekonstruktiv-hermeneutisch‘ Möglichkeitsmodelle der Handlungsabläufe **und** der Handelnden entwerfen.

Die überlieferten Daten allerdings sind nicht die ‚ursprünglichen‘ Handlungssituationen, sondern deren Protokolle. Ebenso wenig sind die Interpretationen die in der Rekonstruktion wiederholte und ‚rational explizierte‘ Ursprungshandlung, sondern Modelle objektiv möglicher Sinnfiguren, die aus Handlungsprotokollen (letztlich haben für das Interpretieren alle menschlichen Produkte den Status von Handlungsprotokollen) gewonnen werden und sich nur auf diese beziehen. Dementsprechend enthalten die Interpretationen auch nicht mehr die real existierenden Handelnden (bzw. die Konstruktionen ‚erster Ordnung‘, die Handlungspartner in konkreten Situationen von einander entwerfen), sondern **Modelle** von Handelnden, Homunculi. Diese wiederum bekommen - nicht zuletzt durch die Fragestellung des Sozialwissenschaftlers, der sich mit ihnen beschäftigt - eine Situation zugeschrieben, die nicht von ihnen selbst, sondern eben vom Wissenschaftler definiert wird.

D.h.: Verstehende Sozialwissenschaftler beschäftigen sich zwar im wesentlichen mit Dingen, mit denen die Menschen sich ohnehin alltäglich beschäftigen. Aber sie beschäftigen sich mit diesen Dingen systematisch anders, als man dies im Alltag normalerweise tut. Verstehende Sozialwissenschaftler reklamieren für sich, daß ihre Sichtweise eine zwar auf der alltäglichen Erfahrung aufbauende, von dieser aber sich unterscheidende Perspektive der Weltwahrnehmung darstellt: Sozialwissenschaftliches Verstehen unterscheidet sich vom alltäglichen Verstehen dadurch, daß die Interpretationsleistungen hier nicht unter Rückgriff auf den Alltagsverstand, sondern unter Rückgriff auf extensiv aktiviertes Wissen und auch auf einen Vorrat an professionellem Sonderwissen beruhen. Dieses Verstehen ist, anders als das alltägliche, nicht bezogen auf pragmatische Bedürfnisse des Lebensvollzugs, sondern auf das Relevanzsystem eines pragmatisch desinteressierten Beobachters (Schütz, 1932, S.313-340).

Das Verstehen des Sozialwissenschaftlers geschieht also in einer besonderen, eben nicht alltäglichen Einstellung, die Schütz als theoretische Subsinnwelt

bezeichnet. Dies ist eine Einstellung des prinzipiellen Zweifels an sozialen Selbstverständlichkeiten. Diese Einstellung ist dadurch geprägt, daß sie die Sorge um die eigene Existenz ausklammert und nur daran interessiert ist, die Wirklichkeit zu durchschauen, die ‚Wahrheit‘ (d.h. das Zustandekommen) der Wirklichkeit zu erkennen. In dieser Einstellung gibt es keine sozialweltliche Präsenz, kein In-Situation-sein, keine lebendigen Mitmenschen, sondern nur idealisierte Modelle sozialer Erscheinungen und vom Sozialwissenschaftler konstruierte künstliche Geschöpfe.

Sozialwissenschaftliches Verstehen zielt ab auf die Erkenntnis der Konstitutionsbedingungen für ‚Wirklichkeit‘, auf die Entzauberung gesellschaftlicher Konstruktionen. Sozialwissenschaftliches Verstehen soll Phänomene, die der Wissenschaftler in den Blick genommen hat, sinnentsprechend, problemadäquat und logisch konsistent rekonstruieren und es dadurch ermöglichen, sie - im Sinne Webers (1917) - ‚kausal‘ zu erklären. Der praktische **gesellschaftliche** Nutzen dieses Unternehmens liegt darin, die Menschen auf die vom Alltagsverstand gemeinhin nicht thematisierten Umstände, Zusammenhänge und Regeln aufmerksam zu machen, in deren Rahmen sie ihr Leben vollziehen (dazu auch Luckmann, 1980).

Sozialwissenschaftliches Verstehen ist mithin notwendigerweise immer auch ein Verstehen des Verstehens, ein Verstehen ‚zweiter Ordnung‘. Das heißt nun nicht, daß der praktische Vollzug des Wissenschaftsbetriebs nicht auch zum Alltag gehören würde. Aber der institutionelle wissenschaftliche Alltagsbetrieb dient dazu - oder sollte wenigstens dazu dienen - dem Sozialwissenschaftler den ‚Rückzug‘ in jene besondere Einstellung zu ermöglichen, welche zeitweise die pragmatischen Interessen des Alltagsverstandes ausklammert und durch ein rein kognitives Interesse ersetzt, durch das Interesse nämlich, einen Sachverhalt nicht praktisch zu bewältigen, sondern ihn ‚sine ira et studio‘ rational zu analysieren und z.B. aus den - wie auch immer entstandenen - Handlungsprotokollen des alltäglichen Wissens dieses Wissen und darüber hinaus die Bedingungen und Möglichkeiten dieses Wissens zu rekonstruieren. In der vollendeten Auslegung **bat** der Sozialwissenschaftler dann dieses Wissen, aber er **lebt es** nicht.

## 2. *Hermeneutik als selbstreflexives Unternehmen:*

### *Die Relativität der Deutung und der Verlust des Einmaligen*

Die Mehrzahl menschlicher Deutungsleistungen vollzieht sich keineswegs methodisch kontrolliert, sondern fraglos - und sozusagen beiläufig - vor dem Hintergrund eines unproblematischen Vorwissens, besser: eines impliziten Wissens um das, was ‚hier und jetzt‘ ist und getan werden muß. Zu diesem

implizit gewußten, immer schon gedeuteten und in die Deutung von Handlungen einbezogenen Bereichen gehört das, was in der phänomenologisch orientierten Sozialphilosophie bzw. ‚Protozoologie‘ als „alltägliche Lebenswelt“ (Luckmann, 1980, 1990) und in der Sozialforschung als „Milieu“ (Gratthoff, 1989) oder „kleine soziale Lebens-Welt“ (B. Luckmann, 1970; Hitzler & Honer 1984, 1988) bezeichnet wird: Die konkrete Umgebung eines Menschen, die Gesamtheit dessen, was von ihm als auf ihn wirksam erlebt wird, ungeachtet der Frage nach dem, was ‚objektiv‘ auf ihn einwirkt (Gurwitsch, 1977, S. 86).

Wenn hier von ‚Milieu‘ gesprochen wird, so kommt es zunächst darauf an, wenigstens zwei der größten Mißverständnisse, die häufig mit diesem Ausdruck verknüpft werden, abzubauen. Es geht dabei weder um naiv in die Sozialwissenschaften verschleppte Biologismen und sich daran anknüpfende kybernetische Wunschvorstellungen von elegant abbildbaren Mensch-Umwelt-Relationen noch um die ebenso naive Determinationshypothese, wonach der Mensch zu dem wird, ‚was das Milieu aus ihm macht‘, was immer hier mit dem Ausdruck Milieu gemeint sei. Für eine Sozialwissenschaft der gesellschaftlichen Orientierungs-, Handlungs-, Produktions- und Wissensformen gilt vielmehr ebenso wie für die phänomenologisch orientierte Philosophie, daß ‚Umwelt‘ ein Begriff ist, „der ausschließlich in der geistigen Sphäre seine Stelle hat“ (Husserl, 1936, S. 317), d.h. ein Begriff, der die spezifisch menschlichen zeichen- und symbolhaft organisierten Wahrnehmungs-, Deutungs- und Handlungsformen repräsentiert, ‚hinter‘ die - zumindest - Menschen nicht zurückgreifen können.<sup>3</sup>

Menschliche Umwelt, Lebenswelt, läßt sich dementsprechend weder nach einem Modell von ‚außen/innen‘ oder ‚Subjekt/Objekt‘ noch mit Hilfe räumlicher Abmessungen und territorialer Verteilungen beschreiben. Sie ist für uns kein Gegenüber, weder Käfig noch unbegrenzter Raum, sondern eher Wahrnehmungs-, Orientierungs- und Handlungshorizont. Sie bewegt sich mit uns, wenn wir uns bewegen, sie verändert uns - unser Handeln -, wenn wir sie verändern. Sie existiert nicht ohne uns, und wir existieren nicht ohne sie. Aber wir **sind** nicht unsere Umwelt, wir **haben** sie. Unser Verhältnis zu ihr und zu uns ist - um mit Plessner (1970) zu sprechen - bestimmt durch unsere „exzentrische Positionalität“, durch den „doppeldeutigen Charakter“ unserer Existenz (S.41f.), die zwei unterschiedliche und dennoch ineinandergreifende Ordnungen repräsentiert. Empirische Milieuanalyse bzw. Lebensweltanalyse stellt somit den Versuch dar, die konkreten Orientierungs-, Handlungs- und

<sup>3</sup> Dementsprechend ist es ein Unding, umweltliche Natur oder ‚unbeseelte‘ Umwelt als in sich Geistesfremdes zu begreifen und so die Geistes- und Gesellschaftswissenschaften „durch Naturwissenschaft (...) unterbauen und so vermeintlich exakt“ (Husserl, 1936, S.317) machen zu wollen.

Organisationsformen von Individuen in und mit ihrer Umwelt zu beschreiben und konkretes Handeln vor diesem Hintergrund zu interpretieren:

**Deskriptionen** von Milieus bzw. ‚kleinen Lebens-Welten‘ dokumentieren also - neben gesprochenen und transkribierten Texten - die Modi der Orientierung eines Menschen im Raum, in der konkreten Umgebung, in der gelebten Zeit, gegenüber der eigenen Leiblichkeit und gegenüber anderen Personen: Sie dokumentieren damit die weitgehend nichtsprachliche Produktion und Reproduktion eines sozialen Interaktionsgefüges, dessen Singularität in die kollektiven semantischen Typen der Sprache übersetzt und dabei immer schon auch gedeutet werden muß.<sup>4</sup> Dabei verweist die Mühsal der kontrollierten, sprachlichen Deskription auf den sprachlich nicht zu bearbeitenden Handlungsrest. Bei der interpretierenden Verknüpfung a) der aus der sprachlichen Deskription nichtsprachlicher Handlungs- und Milieunetze gewonnenen Texte und b) der transkribierten sprachlichen Texte verweisen a) und b) wechselseitig so aufeinander, daß der Fall in seiner Konkretion, d.h. in potentieller **Unabhängigkeit von seiner Vertextung** erkennbar bleibt: Die Unterschiedlichkeit in der Erarbeitung der beiden Textebenen **appräsentiert** das Nichtvertextete, die Ebene der Konkretion von Handlung und Milieu des Falles außerhalb der Texte.

Die Funktion und die Problematik all dessen, was mit dem ebenso umfassenden wie dunklen Ausdruck ‚außersprachlicher Kontext‘ umschrieben und in der Regel in Nebenbemerkungen abgelegt wird, wird - zumindest in einem Teilbereich - durch die Milieu- und Lebensweltanalyse konkretisiert und als unverzichtbare Interpretationsaufgabe in das Bewußtsein der Interpreten gehoben. Zugleich damit entsteht aber auch das Grundproblem der Protokollierung und Deskription von Milieus und/oder Situationen: der Versprachlichung nicht-sprachlicher Zusammenhänge.

Über der sinnlich wahrnehmbaren Welt baut sich darüber hinaus - diese ordnend, klassifizierend und deutend - eine eigene Welt kollektiver Zeichen und Symbole auf (Cassirer, 1953; Langer, 1965), sowohl in der Sprache als auch in Handlung und Orientierung (dazu Berger & Luckmann 1969, S.89-138). Kleine Alltagswelten, Milieus und die in ihnen stattfindenden und sie gestaltenden Handlungen sind symbolhaft konstituiert: Leben in sozialer Ordnung und in Milieus als Bestandteilen dieser Ordnung bedeutet Leben in Symbolen (Hitzler, 1988). Insofern sind Milieu- und Sprachanalyse gleichermaßen

4 Mit deutlichem Zeitabstand gegenüber den einschlägigen Forschungsaktivitäten in den USA entwickelt sich seit den achtziger Jahren auch im deutschsprachigen Raum eine Alltagsethnographie im hier gemeinten Verstande. Beispielhaft sei hier verwiesen auf Studien zum bäuerlichen Milieu (Hildenbrand, 1983), zur kleinen Lebens-Welt des Bodybuilders (Honer, 1985) und des Wüschelrutengängers (Knoblauch, 1992), zur Laborsituation (Knorr Cetina, 1984, 1988), zur Konstruktion von Transsexualität (Hitschauer, 1993), zur Kultur des Punk (Lau 1992), zur Polizeiarbeit (Reichertz, 1991) und zur Spendenpraxis (Voß 1992). - Siehe auch die materialen Beiträge in Soeffner (1988).

Symbolanalyse, und insofern bestimmt auch erst der Symbolzusammenhang als ganze Formen und Typik menschlichen Handelns. Wissenschaft als Symbolanalyse (Soeffner, 1991) besteht demnach in dem Versuch der Rekonstruktion des symbolischen Gesamtzusammenhanges menschlicher Handlungs-, Orientierungs- und Wissensformen (materiale Beispiele dazu in Soeffner, 1992).

Jenseits der (Sprach-)Texte und in der wissenschaftlichen Hermeneutik oft von ihnen überlagert, wird mit der Blickwendung auf Milieu bzw. alltägliche Lebenswelt zugleich der universale Auslegungsanspruch der Hermeneutik erkennbar. Für sie gibt es keine ‚materia nuda‘, keine ‚brute facts‘, keine ‚nackten Tatsachen‘. Statt dessen wird jetzt das Problem der Abgrenzbarkeit von Texten und/oder Deutungsgegenständen sichtbar, mit anderen Worten: das Problem des Kontextes, der Einbettung des Sinnhorizonts von **Deutendem, Deutung und Deutungsgegenständen** (dazu auch Bühler, 1934). Daraus folgt: 1. Als immer schon sich alltäglich vollziehender (und daher wissenschaftlich auch rekonstruierbarer und methodisierbarer), deutender menschlicher Zugriff auf Welt und die menschliche Existenz in ihr ist die Hermeneutik ihrem Anspruch nach **universal**; 2. wegen der Abhängigkeit des Deutenden, der Deutung und der Deutungsobjekte von ihrer jeweiligen Einbettung in Milieus, Geschichte, Geschichten und Deutungsgemeinschaften sind die jeweiligen ‚Resultate‘ hermeneutischer Auslegung jedoch **relativ**. D.h., sie stehen in Relation zu einem je gegebenen soziohistorischen Sinnzusammenhang und erlangen in bezug auf diesen ihre Geltung.

Diese Relativität hat also nichts zu tun mit Beliebigkeit. Sie schließt - in der wissenschaftlichen Hermeneutik - kontrollierte Überprüfungsprozesse gerade **nicht** aus: Die Überprüfungen richten sich auf den - intersubjektiv plausibilisierbaren - Zusammenhang zwischen der Deutung und ihren spezifischen Randbedingungen. Dadurch, daß hier die prinzipielle Relativität und das konkret Relative bewußt in Rechnung gestellt werden, wird der Anspruch auf Intersubjektivität des Verfahrens und der Ergebnisse aufrechterhalten und durchgesetzt: Beliebigkeit wird ausgeschlossen, indem Relativität und Intersubjektivität aufeinander bezogen werden (zur Methodik siehe Teil 3. dieses Textes).

Aus milieutheoretisch-lebensweltanalytischer Perspektive, die eine der historischen Perspektive vorgelagerte objektiv wirksame Sinnschicht subjektiver Orientierung sichtbar werden läßt, wird der gesamte Umkreis des Wahrnehmbaren durch die kulturellen Relevanzen eines Wahrnehmenden bestimmt. Die Annäherung an intersubjektiv nachvollziehbares Verstehen müßte dementsprechend folgendes leisten (Scheler, 1923; auch Srubar, 1981): 1. das bewußte und kontrollierte Abstrahieren des Interpreten von den eigenen kulturellen Fraglosigkeiten und der eigenen historischen Perspektive (Reflexion der eige-

nen Vor-Urteile); 2. die Rekonstruktion (so weit wie möglich) der Struktur des ‚fremden‘ Milieus und der historischen Bindung eines überlieferten Dokumentes oder ‚records‘ und der ‚anderen‘ Lebenswelt seines Produzenten (das Fremde zum Sprechen bringen); 3. die Zuordnung der eigenen und der fremden Erfahrungsstruktur sowie der eigenen Deutung und des Deutungsgegenstandes zu einem wissenschaftlichen ‚Universe of discourse‘ objektiv möglicher, d.h. intersubjektiv nachvollziehbarer Milieus, Kontexte und Bedeutungen (Verortung im Bedeutungsraum).

Diese wissenschaftliche Konstruktion der Intersubjektivität und Perspektivenneutralität weist hin auf die unüberbrückbare Differenz zwischen wissenschaftlicher Ex-post-Interpretation eines Sinnes und der ursprünglichen Sinn-schicht - einer alltäglichen Interaktionssituation, in der die ‚records‘ aufgenommen wurden, die der Interpretation zugrundeliegen. Der Interpret kann weder diese Ursprungsszene wieder erstehen lassen noch sich in das konkrete Milieu und die Perspektiven der darin agierenden Personen versetzen. Seine Interpretation der ‚records‘ und Dokumente besteht letztlich darin, das Unverstehbare (Singuläre) des Einzelfalles in das Verstehbare (Allgemeine) einer intersubjektiven Perspektive zu übersetzen; d.h. die wissenschaftliche Interpretation verzichtet auf die Deutung der Konkretion des Einzelfalles, weil sie darauf verzichten muß. Was nicht übersetzbar, verallgemeinert und damit sozial verstehbar gemacht werden kann, ist für die Interpretation verloren - was aber nicht heißt, daß es nicht vorhanden und wirksam gewesen wäre und auch noch wirksam sein kann.

Alle dokumentierten bzw. dokumentierbaren vergangenen, ‚gegenwärtigen‘ und zukünftigen Erscheinungen der sozialen Welt hingegen sind **potenziell** sozialwissenschaftliche Daten. Das Dokumentieren bleibt jedoch prinzipiell ebenso hinter der Vielzahl der Erscheinungen zurück wie das Interpretieren hinter der Vielzahl der Dokumente und der möglichen Interpretationsgegenstände. Schon daraus folgt, daß sozialwissenschaftliche Auslegung notwendig exemplarisch arbeitet. Sie ist per se Fallanalyse und zielt auf das Typische, Verallgemeinerungsfähige von historischen ‚Einzel‘-Erscheinungen (wobei der oder die Auslegenden entscheiden, wo die ‚Grenze‘ der Einzelercheinung liegt, d.h. wo sie in eine andere Erscheinung übergeht), d.h. sie kann Intersubjektivität und Verallgemeinerbarkeit ihrer Ergebnisse niemals dadurch erreichen, daß sie alles in ‚Daten‘ umwandelt und bearbeitet. Daraus wiederum folgt, daß die Qualität ihrer Aussagen und Interpretationen prinzipiell nicht von der Quantität ihrer Daten, wohl aber von der Intention, der Fragerichtung und den Prinzipien und Verfahren der Sinnzumessung durch den Wissenschaftler abhängt. Denn diese wiederum präformieren (mit **prinzipiell offenem Fragehorizont**), was und ‚wieviel‘ an Daten für die Interpretation einer ‚Einzelercheinung‘ für erforderlich gehalten wird.

### 3. Probleme methodisch kontrollierten Verstehens

#### 3.1 Der konkrete Fall und der ‚ideale‘ Typus

Sozialwissenschaftliche Auslegung ist notwendigerweise exemplarische Arbeit am Fall. Sie vollzieht sich auf zwei Ebenen: 1. im Aufsuchen, Erproben und Absichern ihrer Interpretationsregeln und ihrer Verfahren; 2. in der Rekonstruktion einer Fallstruktur, in der sie Bedingungen und Konstitutionsregeln sozialer Erscheinungen und Gebilde in ihrer Konkretion, ihrer konkreten Wirksamkeit und Veränderbarkeit sichtbar macht. Dabei sollen einerseits der Fall in seiner Besonderheit und die Bedingungen seiner Individuierung sichtbar werden (exemplarisch Soeffner, 1992, Kap. 1). Andererseits sollen diese Typik und Vergleichbarkeit aus der Analyse der Formen und Strukturen der Typenbildung und -Veränderung entwickelt und ‚erklärt‘ werden.

Die Interpretation des Falles erhebt Anspruch auf Objektivität in zwei Richtungen: 1. im Hinblick auf die Überprüfbarkeit, d.h. Offenlegung der Auslegungsverfahren und des in sie eingehenden Vorwissens, sowie - damit verbunden - auf die Überprüfungspflicht, die der Interpret sich und anderen wissenschaftlichen Interpreten auferlegt; 2. im Hinblick auf Richtung und Ziel des Verfahrens: auf die Analyse des sozial ‚objektiv‘ Wirksamen - auf die gesellschaftlichen Institutionen sowie deren historisch gültigen Sinn als Handlungsdeterminanten und auf für den Handelnden möglicherweise verborgene, ‚latente‘ Sinnstruktur des Handelns (Oevermann u.a., 1979).

Ziel der Analyse ist die Rekonstruktion eines objektivierten Typus sozialen Handelns in seinen konkreten, fallspezifischen Ausprägungen. Dieser objektivierte Typus ist insofern ‚Idealtypus‘, als er mit dem Zweck konstruiert wird, einerseits gegenüber der Empirie insofern systematisch unrecht zu haben, als er das Besondere im Einzelfall nur unzulänglich wiedergibt, andererseits aber gerade dadurch dem Einzelfall zu seinem Recht zu verhelfen, daß er das historisch Besondere vor dem Hintergrund struktureller Allgemeinheit sichtbar abhebt (Weber, 1917a, S.190-202; dazu auch Schütz & Luckmann 1979, S.277-290).

Die Rekonstruktion eines objektivierten Typus gesellschaftlichen Handelns baut sich auf von - jeweils extensiven - Einzelfallanalysen über Fallvergleich, Deskription und Rekonstruktion fallübergreifender Muster bis hin zur Deskription und Rekonstruktion fallübergreifender und zugleich fallgenerierender Strukturen. Der so rekonstruierte Typus enthält und veranschaulicht die strukturelle Differenz von evolutionär und historisch sich verändernden Strukturinformationen einerseits und ihren konkret historisch-kulturspezifischen Ausdifferenzierungen andererseits.<sup>5</sup>

<sup>5</sup> Als ‚berühmtes‘ Beispiel seien hier die Krankenhaus-Studien von Barney Glaser und Anselm Strauss genannt, in deren Kontext die sogenannte ‚grounded theory‘, das Prinzip der schritt-



Die Einzelfallanalysen dienen so der schrittweisen Entdeckung allgemeiner Strukturen sozialen Handelns, während der Einzelfall selbst als historisch-konkrete Antwort auf eine konkret-historische Situation und Strukturformation interpretiert wird: Mit den Einzelercheinungen wird die Strukturentwicklung, mit den Einzelfallanalysen die Theorieentwicklung historisch fortgeschrieben (dazu auch Bude, 1985).

Der Weg vom deutenden Verstehen zum ‚ursächlichen‘ **Erklären** (zum Problem unterschiedlicher Typen des Erklärens vgl. Kap. 9: Theoriebewertung) des Ablaufs und der Wirkungen sozialen Handelns führt mithin über die Konstruktion eines **begrifflich reinen Typus** von dem oder den als Typus gedachten Handelnden und dem von ihnen subjektiv gemeinten Sinn (Weber 1972, S. 1-4), d.h. über eine Konstruktion ‚zweiter Ordnung‘. Erst und nur im Reiche der idealtypischen zweckrationalen Konstruktionen läßt sich entscheiden, wie ein Akteur im Falle ‚idealer Zweckrationalität‘ disponiert und gehandelt haben wurde. Erst mit Hilfe dieser idealtypischen Konstruktionen, die terminologisch, klassifikatorisch und heuristisch um so besser ihren Dienst leisten, je ‚weltfremder‘ sie sind, lassen sich Vergleiche mit dem dokumentierten Handeln anstellen. Erst dann ist es auch möglich, den ‚Abstand‘ zwischen dem Handeln in idealtypischer Zweckrationalität einerseits und dokumentiertem Handeln andererseits dadurch ‚kausal‘ zu erklären, daß die Elemente benannt werden können, die sich im untersuchten Fall in die ‚reine Zweckrationalität‘ eingemischt haben.

Der konkrete Einzelfall wird also ausschließlich im Hinblick auf seinen Abstand vom und seine Differenz zum begrifflich ‚reinen‘ zweckrationalen Idealtypus kausal erklärt. Nicht durch diese kausale Erklärung der Differenz läßt sich der Einzelfall deutend verstehen, sondern umgekehrt: durch deutendes Verstehen sozialen Handelns gelangt man zur Konstruktion von Idealtypen, die ihrerseits den Einzelfall als solchen sichtbar machen und ihm zu seinem Recht verhelfen. Indem sie **seine Differenz** zum Idealtypus **erklären**, tragen sie dazu bei, ihn in seiner Singularität und Konkretion zu **verstehen**.

Verstehende Sozialwissenschaft in diesem Sinne ist die fortschreitende Rekonstruktion, das fortschreitende, den Einzelfall und damit die Menschen, ihre Ordnungen und ihre Geschichte ernst nehmende, deutende Verstehen sozialen Handelns. Die wissenschaftlichen Konstruktionen ‚zweiter Ordnung‘, die historisch-genetischen Idealtypen, zielen auf eben dieses historische Verstehen des Einzelfalls und auf das Verstehen der Historie gleichermaßen.

---

weise abstrahierenden Theoriebildung auf der Basis von Einzelfallanalysen, entwickelt worden ist (Glaser & Strauss, 1965, 1968; dazu auch Glaser & Strauss, 1967; Strauss, 1991).

## 3.2 Die Datenkonstitution (vgl. auch Kap.12: Kontrolltechniken)

**Voraussetzung** dafür, daß etwas zum ‚Datum‘ sozialwissenschaftlicher Analyse werden kann, ist, daß es als ‚Dokument‘ einer Handlung oder Lebensäußerung ‚diskursiv‘ vorliegt, d.h. daß es fixiert ist, immer wieder und in gleicher ‚Gestalt‘ von jedem beliebigen Interpreten angesehen, hin- und hergewendet und damit kontrolliert werden kann, und daß schließlich aufgrund dieser dokumentarischen Diskursivität des Interpretationsgegenstandes die Interpretation sowie deren Methoden ihrerseits kontrolliert, verifiziert oder falsifiziert werden können.

„Natürlich“ vorliegende oder aufgefundene Daten und Dokumente (Briefe, Urkunden, Gemälde, Gebäude, Musikstücke etc.) sind von solchen zu unterscheiden, die von Sozialwissenschaftlern methodisch mit Hilfe spezifischer Verfahren erhoben worden sind; im zweiten Fall sind die Verfahren **Teil** der Daten und ihrer Aussagekraft. Explorative Ethnographen, sozusagen die ‚Reporter‘ unter den interpretativ orientierten Sozialwissenschaftlern, sammeln deshalb vorzugsweise das, was man ‚natürliche Daten‘ nennen könnte, Daten, die in situ von den Handelnden erzeugt werden. Selbstverständlich bleiben auch diese Daten nicht ‚natürlich‘, da sie verschriftet, also künstlich hergestellt werden wie andere Daten auch. Aber immerhin geht es um den Versuch, die Struktur des beobachteten Handelns in der Aufbereitung der Daten so wenig wie möglich zu verändern, sie so weit wie möglich ‚selbst sprechen‘ zu lassen. Genauer: die Struktur der Transkription und der Datenaufbereitung hat sich nach der Struktur des beobachteten Handelns zu richten - und nicht umgekehrt.

Gleichwohl: Auch die sogenannten qualitativen Verfahren der Datenerhebung formen - graduell sicher unterschiedlich stark - die durch sie produzierten Daten. Die Chancen der Kontrolle dieser Überformung schwinden dabei im gleichen Maße, in dem der Sozialwissenschaftler - sei es als teilnehmender Beobachter oder als ‚offener‘ Interviewer eines ‚offenen Interviews‘ - der Illusion verfällt, daß ‚Nähe zum Feld‘ oder abnehmende Standardisierung des Erhebungsverfahrens von sich aus bereits ‚natürliche‘ Daten produziert: Auch das ‚offene‘ Interview bleibt ein Interview, d.h. eine spezielle Technik, die eine besondere Interaktionssituation schafft. Und: Der teilnehmende Beobachter konzentriert sich auf die Beobachtung, nicht auf die eigene Teilnahme am Interaktionsgeschehen; er handelt nicht in demselben Sinne, wie die von ihm Beobachteten dies tun.<sup>6</sup>

---

<sup>6</sup> Auch der sogenannte Aktionsforscher bildet dabei durchaus keine Ausnahme. Bei seinem Bemühen, sich ‚quasi natürlich‘ im ‚Feld‘ zu bewegen und ‚mitzuhandeln‘, geht ihm das, was er an Handlungskapazität aufbringt, an Beobachtungs-, Kontroll- und Interpretationskapazität verloren, sofern er sich nicht ex post als ‚ein anderer‘ mit den von und mit ihm produzierten Daten auseinandersetzt.

### 3.2.1 Divergenzen und Konvergenzen standardisierter und nicht-standardisierter Verfahren

Wo also liegen die tatsächlichen Unterschiede zwischen den standardisierten und den nicht-standardisierten Verfahren? - Zunächst und wesentlich darin, daß bei ersteren die Standards ‚künstlich‘ erarbeitet und die auf ihnen basierenden Daten sowie deren Auswertung in überprüfbare Relationen zu diesen Standards gebracht werden. (Das gilt etwa für die in den anderen Kapiteln dieses Bandes genannten Methoden). Die nicht-standardisierten Verfahren dagegen beziehen sich auf **natürliche Standards und Routinen der Kommunikation**, die zunächst einmal gewußt und in ihrer Funktionsweise bekannt sein müssen, bevor die auf ihnen basierenden Daten kontrolliert interpretiert werden können.’ Und schon auf dieser ersten Ebene scheitert eine ganze Anzahl der qualitativen Untersuchungen, deren zweifelhafte Qualitäten entweder in naivem Intuitionismus und Empathie sowie in der unkontrollierten Anwendung alltäglicher Deutungsroutinen und Plausibilitätsmaximen bestehen, oder aber auf der mehr als problematischen Übertragung geborgter und nicht fall-spezifisch angewandter Interpretationsmuster und Vorwegdeutungen (z. B. psychoanalytischer Provenienz) beruhen. Beide Verfahrensweisen sind Versuche, sich an den Problemen der impliziten ‚Standards‘ alltäglicher Interaktion vorbeizumogeln.

Auf der Grundlage dieser ersten Unterscheidung lassen sich nun einige weitere aufführen: Der Befragung und gesteuerten Darstellung im standardisierten Verfahren steht die Selbstdarstellung der Informanten im nicht-standardisierten Verfahren gegenüber; der Strukturierung der Befragungs- und Beobachtungsanlage die Interaktions-, Darstellungs- und Redestrukturierung durch ‚alltägliche Routinen‘.\* Bei standardisierten Verfahren sind Themenfestlegung und -eingrenzung durch das Untersuchungsdesign, bei nicht-standardisierten durch situative, interaktionsstrukturelle und biographische Faktoren gegeben - und zu berücksichtigen. Andererseits erzielen standardisierte Verfahren eine Themenerweiterung durch Fragevielfalt und Kontrollfragen, während die nicht-standardisierten dieses Ziel durch das Animieren von Erzählungen und die Eigendynamik der kommunikativen Gattungen erreichen.

7 Beispiele hierfür wären etwa beobachtende Teilnahme, offene bzw. narrative Gesprächsführung beim Interviewen sowie hermeneutische Varianten der Textauswertung (dazu auch verschiedene Beiträge im 6. Kapitel von Flick, Kardoff, Keupp, Rosenstiel & Wolff, **1991**, S. **177-281**).

8 Fragt man z.B. mit Hilfe von Fragebögen oder Leitfadeninterviews Menschen im Hinblick auf ihre Alltagsverrichtungen danach, warum sie tun, was sie tun, so bekommt man mit Sicherheit Antworten, die genau das nicht betreffen, nämlich das implizite Routinewissen, auf das man zielt. Der Grund dafür besteht darin, daß, wie wir alle wissen, zwischen Einstellung und Defactohandeln, zwischen gelerntem, explizit darstellbarem Wissen und habituellem Handeln ein tiefgehender Unterschied besteht (vgl. dazu auch Knorr Cetina, 1984; Honer, 1993).

Beide Verfahrenstypen basieren also prinzipiell auf der kontrollierten Erhebung und Interpretation von Daten. Unterschiedlich ist jedoch ihre Kontrollbasis. Der erste beruht auf der kontrollierten Erarbeitung und Relationierung ‚künstlicher‘ Erhebungs- und Auswertungsverfahren. Der zweite basiert auf dem zuvor herzustellenden Wissen über die Verfahren alltäglicher, ‚relativ natürlicher‘ Interaktions-, Darstellungs-, Rede- und Deutungsstrukturierung und auf deren kontrolliertem Einsatz in der Interpretation. Die Schwäche des zweiten besteht darin, daß der Wissenschaftler sich - was die Bekanntheit jener vorgängigen Alltagsroutinen angeht - auf Neuland bewegt und sich in noch unbekanntes Gebiet vorarbeiten muß, während er interpretiert; die des ersten, daß sich die ihn verwendenden Wissenschaftler in ihrer überwiegenden Mehrzahl keine Rechenschaft darüber ablegen, wie sehr ihre Erhebungs- und Meßinstrumente auf Alltagswissen und -routinen beruhen, die sie zwar formalisiert, aber als ‚ursprünglich‘ alltägliche nicht erkannt haben und daher auch nicht kontrollieren können.

### 3.3 Der Text und das Vertextete

Längst nicht alle, nicht einmal die Mehrzahl der Erscheinungen der sozialen Welt sind sprachlicher Natur oder sprachlich gefaßt. Und wir können - bezogen auf unser alltägliches Leben - von Glück sagen, daß dies so ist. Sozialwissenschaftliche Analysen - wissenschaftliche Analysen ganz allgemein - haben jedoch in der Regel ein sprachlich gefaßtes Endprodukt: einen Text. Im allgemeinen gehen diesem Endtext nicht nur andere Texte voraus: Auch die - zunächst nichtsprachlichen - Beobachtungen und ihre Elemente werden schon auf einer sehr frühen Stufe der Analyse in Sprache überführt, d.h. von einem nichtsprachlichen in ein sprachliches Zeichensystem übersetzt. Mehr noch: Nichtsprachliche Ordnungsprinzipien und Abläufe werden in ein sprachliches und begriffliches Ordnungssystem überführt. Ganzheitlich-gleichzeitige Wahrnehmungen, Empfindungen und Erfahrungen verwandeln sich in sprachlich gefaßte Erinnerungschiffren, die individuellen Erfahrungseinheiten in kollektive semantische Typen.

Sozialwissenschaftliche Analyse - und nicht nur sie - ist das Produzieren von Texten und das Produzieren von Texten über Texte. Sie geht oft genug von der naiven Prämisse aus, daß alles Relevante - nicht nur das wissenschaftlich Relevante - sprachlich ausdrückbar ist. Sprache und Erfahrung wachsen so zusammen und werden naiv in eins genommen: Die Tradierung und Speicherung von Erfahrung wird für die Erfahrung selbst gehalten. Für die wissenschaftliche Analyse selbst wird - noch jenseits des Mangels an primären Erfahrungen - eine solche Naivität dann verhängnisvoll, wenn das ‚künstlich‘ hergestellte Datum mit der Lebens-Realität und das sprachliche Ordnungs-

system mit der ‚Ordnung der Dinge‘ verwechselt wird. Sozialwissenschaftliche Analysen gehen jedoch nur dann kontrolliert gegenüber der Differenz von beobachteter Erscheinung und sprachlichem Ausdruck für die Erscheinung vor, wenn sie den Verlust und die Veränderung artikulieren können, die sich aus der sprachlichen und/oder mathematischen Übersetzung des Phänomens ergeben.

Texte sind Interaktionsprodukte, nicht Produkte eines isolierten Bewußtseins. Sie sind dies selbst im Grenzfall des Monologs, des Gesprächs einer Person mit sich selbst. Sie werden in Interaktionen produziert, sind Bestandteil der Interaktion, haben dementsprechend Handlungscharakter und bilden Handlungsreihen ab, die als Aktions- und Reaktionsgefüge irreversibel sind (Austin, 1962; Searle, 1969). Aufgezeichnete Texte sind zu verstehen als Handlungsprotokolle, genauer: als Protokolle von Handlungen, die unwiderruflich vorbei sind, die sich aus der verschrifteten Textform als Real- oder Ursprungshandlungen nie wieder hervorzaubern lassen, sondern nur noch durch Protokolle repräsentiert sind.

Der Protokolltext stellt eine feststehende, unumkehrbare Sequenz von Aktion und Reaktion dar. Dabei können Aktion und Reaktion nicht so bewertet werden, daß das eine nur dies und das andere nur jenes ist, vielmehr ist jede Äußerung innerhalb eines Gesprächs immer gleichzeitig Aktion und Reaktion. Darüber hinaus enthält jede Äußerung entsprechend ihrer doppelten Funktion als Aktion und Reaktion Interpretationen des Sprechers gegenüber eigenen Handlungszügen und -planungen sowie Deutungen der Äußerungen von Interaktionspartnern. Da somit die ersten Interpreten dieses Textes die Gesprächs- und Interaktionspartner selbst sind, besteht der erste Schritt der Analyse im wesentlichen darin, diese Interpretationsleistungen der Handlungs- und/oder Sprechpartner zu rekonstruieren, sie gleichzeitig als Handlungszüge und Interpretationsleistungen hervorzuheben und zu beschreiben, von welchen Darstellungsregeln sie geleitet sind.

Nochmals: Voraussetzung jeder wissenschaftlichen Interpretation ist, daß der Text diskursiv aufgezeichnet ist. Die Betonung der Notwendigkeit der Diskursivität von Texten für die wissenschaftliche Analyse hat schon Tradition. Gemeint ist schlicht die Tatsache, daß bei einer wissenschaftlichen Textinterpretation die Texte entweder schriftlich oder neuerdings auf Ton- oder Videobänder aufgezeichnet vorliegen müssen. Denn jede absichtsvolle Interpretation verlangt große Aufmerksamkeit des Interpreten. „Aber auch angestrengteste Aufmerksamkeit“, so stellt schon Dilthey (1900) fest, „kann nur dann zu einem kunstmäßigen Vorgang werden, in welchem ein kontrollierbarer Grad von Objektivität erreicht wird, wenn die Lebensäußerung fixiert ist und wir so immer wieder zu ihr zurückkehren können. Solches kunstmäßiges Verstehen von dauernd fixierten Lebensäußerungen nennen wir Auslegung oder In-

terpretation“ (S. 318f). Die Fixierung der Lebensäußerungen ist deswegen also Voraussetzung für jede Art von Interpretation, weil nur durch diese Diskursivität gewährleistet ist, daß der Interpret einen Text hin- und herwenden, mehrfach abrufen und frei vom Handlungsdruck konkreter Interaktion interpretieren kann.

### 3.4 Die Sequenzanalyse

Wenn wir Texte als Protokolle vergangener, irreversibler Interaktions- und Interpretationssequenzen ansehen, so ist in dieser Annahme die Prämisse enthalten, daß diese Sequenzen einen Handlungszusammenhang repräsentieren, innerhalb dessen die Einzeläußerungen grundsätzlich über sich selbst hinausweisen und immer den Handlungsrahmen als Ganzen, als Horizont mit in Rechnung stellen. Für die Interpretation folgt daraus, daß es - gleich in welchem Sinne - keine Einzeläußerungen innerhalb eines Textes gibt, die ‚für sich‘, als einzelne interpretierbar wären. Jede der Äußerungen ist interaktionstheoretisch in folgende Bezüge eingebettet: sie bezieht sich 1. auf die ihr vorausgehenden Äußerungen und den Handlungskontext insgesamt, 2. auf die unmittelbar vorangehende Äußerung, sei es des Gegenübers oder des Sprechers selbst, 3. auf die erwarteten oder erwartbaren Nachfolgeäußerungen und 4. auf den Handlungs- und Sinnhorizont des Interaktionszusammenhangs als Ganzem. (Diese Feststellung muß unabhängig von der oben genannten Abgeschlossenheit des ‚ursprünglichen‘ Handlungssinnes getroffen werden.) Gleichzeitig repräsentiert und reproduziert 5. jeder Interaktionsprozeß eine ihm zugrundeliegende Interaktionsstruktur in einer historisch konkreten, die historischen Rahmenbedingungen mitbeinhaltenen Textform.

Interpretieren ist somit die Rekonstruktion der Textbedeutung „in der Linie des Geschehens“ (Dilthey, 1976, S.214). Oder zeitgenössisch-technisch ausgedrückt: Interpretieren ist Sequenzanalyse. Die Interpretation von aus dem Kontext gerissenen Einzeläußerungen - ob sie nun sprechakttheoretisch, inhaltsanalytisch oder sonstwie durchgeführt wird - mag über viele Dinge etwas aussagen, kaum aber etwas **interaktionstheoretisch** Brauchbares über den Herkunftstext und über den spezifischen Sinn dieser Äußerung innerhalb ihres spezifischen Kontextes.

Das ursprüngliche, konkrete Ereignis, innerhalb dessen der aufgezeichnete Text entstand, ist, wie gesagt, unwiderruflich vorüber - und mit ihm neben den nicht- sprachlichen Handlungselementen die konkret geäußerten, von den Interpretationspartnern im Ereignisaugenblick wechselseitig wahrgenommenen affektiven ‚Beimischungen‘, das Gewebe wechselseitig unterstellter Subjektivität und Intentionalität. Was bleibt, ist eben das Handlungsprotokoll in seiner sprachlich objektivierten und nun Objekthaft diskutierten Form. **Dieser**

Text in seiner objektiven Struktur ist Ausgangspunkt und Verifikationsmaßstab der wissenschaftlichen Interpretation, die folglich also **nicht** auf künstlich im nachhinein erzeugter Empathie gegenüber den als ‚ursprünglich‘ nur noch imaginierbaren Interaktionspartnern und ihren Intentionen beruht.

An die Stelle des mit Empathie besetzten ‚intentionalistischen Vorurteils der (Umgangs-)Sprache‘ (Oevermann, Allert, Konau & Krambeck, 1979, S. 359) stellt die sozialwissenschaftliche Interpretation die für jeden kompetenten Sprecher (Chomsky, 1965; Katz & Fodor, 1963) einer Sprache objektiv möglichen und realisierbaren Textbedeutungen einer Textabfolge; an die Stelle empathischer Rekonstruktion singulärer, subjektiver Intentionalität stellt sie die Perspektivenneutralität; an die Stelle des unmittelbaren Zugriffs auf eine Textbedeutung die Explikation der Kriterien für die Wahl eines sprachlichen Ausdrucks und die Ausschlußkriterien gegenüber anderen sprachlichen Möglichkeiten. Durch diese notwendige und methodisch eingesetzte Perspektivenneutralität ergibt sich zwangsläufig eine Differenz zwischen der objektiven Sinnstruktur eines Textes und der in diesem Text aufscheinenden und sich dem Interpretieren aufdrängenden subjektiven Intentionalität, die die Sprecher für sich in Anspruch nehmen und der sie eine subjektive sprachliche Ausdrucksqualität verleihen.

### 3.4.1 *Die Methodik der sequentiellen Interpretation*

Somit ergibt sich das methodische Vorgehen bei der Sequenzanalyse: Es wird in der Abfolge- und Reaktionsstruktur des Textes, des Interaktionsfalles interpretiert. Dies bedeutet:

1. Die Interpretation des ersten Interaktes (z. B. einer Begrüßungsformel wie: „Seien Sie mir aufs herzlichste willkommen!“) - und die der nach ihm folgenden Interakte - darf bei der Rekonstruktion objektiv möglicher Bedeutungen, d.h. möglichst sinnvoller Kontexte für diese Äußerungen (z.B. Begrüßung zweier Liebender, eines alten Ehepaares, zwischen Geschäftspartnern, auf einer Party, während eines Festaktes, usw.), nicht auf Informationen aus den nachfolgenden Interakten vorgreifen (die z.B. die Bedeutung der Begrüßungsformel ruckblickend **scheinbar** vereindeutigen), wenn sie der Gefahr entgegen will, das objektiv mögliche Bedeutungspotential des Interaktes unzulässig einzuschränken (im Hinblick z.B. auf Doppeldeutigkeiten). Denn **nur** vor dem Hintergrund dieses so ausführlich wie möglich zu entfaltenden Bedeutungspotentials werden sich in der Text- und Interpretationsabfolge die textspezifische Bedeutungsselektion und damit die Fallspezifik abheben.

2. Das Bedeutungspotential des ersten Interaktes, d.h., seine möglichen sinnvollen Kontexte werden mit dem Interaktionsrahmen, dem faktischen Kontext

- sofern dieser bekannt ist -, verglichen (z.B. zwei miteinander befreundete Studenten begegnen sich nach den Semesterferien in der ersten Sitzung eines Seminars wieder). Ist der faktische Kontext in der Menge der interpretatorisch konstruierten enthalten, so sind fallspezifische Ausschlußkriterien für die restlichen zu analysieren. Ist der faktische Kontext nicht in dieser Menge enthalten, so kann bereits beim ersten Interakt auf eine fallspezifische Abweichung von common-sense-Normen handlungs- und sprachkompetenter Gesellschafter der everyday- und der scientific community geschlossen werden (z. B. hinsichtlich unserer kulturspezifischen Vorstellungen darüber, wie sich Studenten heutzutage typischerweise begrüßen - dazu auch Kalimeyer, 1977, 1980). Das Aufsuchen der Gründe für diese Abweichung wird zur Interpretationsaufgabe. Dabei ist ein Argument von vornherein ausgeschlossen: der Hinweis oder die Spekulation auf eine spezifische innere Verfaßtheit des Sprechers. Ein ‚Inneres‘, das sich nicht im ‚Außen‘, in der Äußerung ausdrückt und repräsentiert, ist für die Interpretation verloren. Es ist ein ‚schweigsames Inneres‘; und was nicht ausdrückbar ist, worüber man also nicht reden kann, darüber soll man bekanntlich schweigen - insbesondere bei der wissenschaftlichen Interpretation. Ist schließlich der faktische Kontext des ‚records‘ nicht bekannt und daher nicht mit dem Interpretationsertrag aus 1. korrelierbar, so besteht die Aufgabe der Interpreten in der Konstruktion möglicher alternativer Kontexte, multipler Welten als möglicher Sinnhorizonte der Äußerung, die in der Textabfolge präzisiert oder ausgeschlossen werden (z.B. die Begrüßungsformel könnte als Zitat gedeutet werden).

3. Das Interpretationsergebnis von 1. und 2. wird zum inneren Kontext für den nächsten Interakt (z.B.: „Salü! Hast Du ihn denn schon gesehen?“), der als Reaktion hierauf angesehen und begründet wird. Bisher herausgearbeitete Grundannahmen (z.B.: Jemand mit ‚Heimrecht‘ begrüßt relativ ‚formell‘ einen oder mehrere Neuankömmlinge) gelten als Handlungs- und Sinnhorizonte des Folgeinteraktes so lange weiter, bis sie durch den Text selbst ‚expressis verbis‘ aufgehoben oder widerlegt werden. D.h., sie fungieren bis zu ihrer Widerlegung als Handlungsrahmen für Folgeinterakte.

Der von den Interaktionspartnern in den ersten Äußerungen eingesetzte Handlungsrahmen und der damit unterstellte Sinnhorizont für die Folgehandlungen enthalten bereits die Handlungsperspektive(n) des nachfolgenden Interaktionsprozesses. Dies bedeutet: die Eröffnungssequenzen einer Interaktion sind auch zu verstehen als Reaktion der Interaktionspartner auf ein im Vorhinein angenommenes Handlungsziel oder Handlungsergebnis. Das in der Zukunft erwartete Ergebnis steuert die Aktionen der Gegenwart (Mead, 1934). In ihnen wird der Sinnhorizont als Erwartungshorizont mitartikuliert (z. B. dahingehend, daß auf eine Begrüßungsformel eine kulturell als ‚entsprechend‘ geltende Reaktion erfolgt).



Als methodische Konsequenz aus diesen Annahmen folgt, daß das Aufsuchen von Deutungsmöglichkeiten und hypothetischen Kontexten für die ersten Interakte so ausführlich wie nur eben möglich sein muß. Und es folgt weiterhin daraus - unsere konkreten Interpretationserfahrungen bestätigen dies -, daß mit der Erschließung des Erwartungshorizontes und des Handlungsrahmens sowie der diesen zugrundeliegenden Interaktionsstruktur die Erzeugungsmechanismen für die folgenden Handlungszüge und die sie repräsentierenden ‚Texte‘ weitgehend ermittelt sind (dazu z.B. Bergmann, 1980; Oevermann, 1983; Schröer, 1992). Vor diesem Hintergrund kann der Interpret seine Deutungen zugleich als Prognosen für den weiteren Textverlauf und dessen Sinnstruktur verstehen: als Prognose über die ‚Linie des Geschehens‘, auf der sich die Interaktionspartner vorwärtsbewegen (z. B.: Der zweite Student scheint durch die Begrüßungsformel des ersten nicht irritiert zu sein. Er ‚antwortet‘ darauf mit einer scheinbar ‚unsinnigen‘ Rückfrage. Mögliche ‚Prognose‘: Die beiden Studenten kennen sich gut genug, um indexikal, d.h. auf in den Interakten selber nicht explizierte Kontexte rekurrierend, miteinander interagieren zu können).

Entsprechend dem Verifikationsgebot wissenschaftlicher Interpretation besteht nun die Hauptaufgabe des Interpretieren darin, im Text Widersprüche zum bisherigen Interpretationsertrag aufzufinden, gemäß der Maxime, daß der Text selbst als Korrekturinstanz für die Interpretation anzusetzen ist, und daß diese nicht dadurch gesichert wird, daß man durch die Wiederholung und Auflistung von Bestätigungsargumenten die eigene Deutung ‚wasserdicht‘ macht, sondern dadurch, daß man sie den im Text enthaltenen Inkonsistenzen und Widersprüchen aussetzt und sie an ihnen testet (z.B. die mögliche Deutung des „Sie“ im 1. Interakt als Hinweis auf eine formelle Begrüßung wird durch das reaktive „Du“ im zweiten Interakt zumindest problematisiert).

4. Die Abfolge des Textes, seine Sequenzierung - soviel wird aus dem bisher Gesagten deutlich -, wird nun in einer ganz bestimmten Richtung zum Interpretationsgegenstand. Die Abfolge von Aktion und Reaktion verweist auf die fallspezifischen **Selektionsmechanismen** und die fallspezifische Bedeutungskonstitution des Interaktionsprozesses, der in seinem Fortschreiten deutlich macht, welche der im ersten Interpretationsschritt genannten, für ihn objektiv möglichen Welten tatsächlich seine ist. Im Fortschreiten des Textes konkretisiert sich der Fall (z.B. im 3. Interakt: ‚Ja, ich war gestern in seiner Vorlesung. Und er wie immer: ‚Seien Sie mir herzlich willkommen, meine Damen und Herren!‘ Du hältst das im Kopf nicht aus!‘). Die Textabfolge repräsentiert die fallspezifische Bedeutungs- und Handlungsselektion. In ihr vollzieht sich die Individuierung des Falles in der Produktion des konkret singulären Textes.

Mit der Aufdeckung der Abfolge- und Selektionsmechanismen des Textes ist die Interpretation an ihrem ersten Ziel. In der hypothetischen Rekonstruktion einer Handlungs- und Problemsituation und in der Konfrontation der in dieser Situation objektiv möglichen Handlungs- und Bedeutungsalternativen mit den fallspezifisch gewählten und konkretisierten Bedeutungen zeigt sich die objektive Struktur und Bedeutung des Falles (z.B.: Student begrüßt anderen Studenten unter Zitation einer als in ihrer Stereotypie bekannt vorausgesetzten Begrüßungsformel eines Dozenten. Der andere Student zeigt durch eine indexikal formulierte Nachfrage an, daß er den ‚Sinn‘ der Zitation verstanden hat, usw.). Die Spezifik des Falles, seine ‚Subjektivität‘, besteht in der selektiven Konkretisierung einer der Welten aus dem gesellschaftlichen Kosmos der objektiv möglichen. Die Interpretation rekonstruiert diese Welt, ihre Aufbauprinzipien und die interaktionsstrukturellen und historischen Gründe ihrer Wahl.

### 3.5 Prinzipien der sinnschließenden Rekonstruktion

Die hier skizzierte **sequenzanalytische Interpretationslehre** im Rahmen sozialwissenschaftlicher Hermeneutik zielt also ab auf die Erschließung sozial objektivierbaren Sinnes von Interaktionen, auf die Ausgrenzung konkreter, typisierbarer - von konkreten anderen anders typisierbaren - Interaktionskonfigurationen. Interpretation besteht somit im Ausformulieren der umgangssprachlichen, d.h. alltagssprachlichen Kompetenz und des Regelwissens am konkret analysierten Interaktionsfall. Es ist das Ausformulieren von Kompetenzen, die von den Handelnden in ihren Interaktionen immer schon als ‚tacit knowledge‘ eingesetzt werden.

Dies setzt voraus: 1. die Entlastung der Interpreten und der Interpretation vom aktuellen Handlungsdruck, wie er in konkreten Interaktionsprozessen vorliegt. Und dies erfordert 2. die systematische Aufsuche jeder denkbaren möglichen ‚Lesart‘ (Oevermann et al. 1979) gegenüber einem Interaktionsprodukt, sowie 3. die komparative Überprüfung und Aussonderung von Lesarten (hierzu allgemeiner Kap. 10: Induktion). Diese Überprüfung vollzieht sich durch das Fällen von Plausibilitätsentscheidungen, die ihrerseits immer wieder in der Rückkopplung an den Analysetext als Urtext begründet werden und dem **Konsistenzgebot** (Repräsentation einer Sinn-Einheit) gerecht werden müssen. Die immer schon unterstellte Sinnkonsistenz fungiert dabei als Generierungsinstrument von Textkonsistenz (dazu und zum Folgenden auch Soeffner, 1980).

Methodisch arbeitet die sinnschließende Interpretation zunächst als Rekonstruktion der von den Interaktionspartnern während des Interaktionsprozesses vorgenommenen Konstruktion von Sinnkonsistenz. Für die Analysepraxis

bedeutet das, daß der Interpret die für ihn auffälligen oder von den Interagierenden selbst akzentuierten Inkonsistenzen als Verdeckung von Sinnkonsistenz behandelt, als Interaktionsstörung, deren latenter Sinn auf einer noch verdeckten, zu erschließenden Konsistenzebene liegt, von der aus die Unbestimmtheitsphänomene Sinnhaft produziert werden.

Zunächst werden die als konsistent unterstellten Intentionen eines Sprechers (dessen egologisch-monothetische Perspektive) rekonstruiert, um sie dann in einem zweiten Schritt mit den objektiv möglichen Textbedeutungen zu konfrontieren. In diesem zweiten Schritt, der Aufdeckung der **Inkonsistenzen** zwischen den Intentionen eines Sprechers und dem Spektrum möglicher Bedeutungen (also einer polythetisch-interaktionsbezogenen Perspektive), wird das generelle Ziel wissenschaftlichen Textverstehens deutlich: es besteht in der Rekonstruktion einer interaktiv entstandenen Problemsituation. Objektivierendes wissenschaftliches Verstehen ist also das Verstehen einer Interaktion als Problemsituation sowie der in ihr und mit ihr verbundenen, objektiv denkbaren Handlungs- und Lösungsmöglichkeiten. Daran schließt sich dann die Deskription der im Text vollzogenen Handlungswahl und die hermeneutische Rekonstruktion der Gründe für diese Wahl an (wodurch sich der einheitliche Interaktionssinn konstituiert).

Die Interpretation gliedert sich also auf in drei Analyseschritte, von denen jeder im Kontext der konkreten Interpretation der vorangegangenen theoretischen Konzeption zugeordnet und methodisch erläutert wird.

### **3.5.1 Rekonstruktion der egologisch-monothetischen Perspektive eines Sprechers**

Der erste Analyseschritt wird verstanden als methodisch eingesetzte, interpretierende Übernahme der ‚idealisierten Perspektive‘ des Sprechers und der Rekonstruktion des egologisch-monothetischen Sinnes der Äußerung aus dieser Perspektive. Diese Ebene der Sinnschließung ist im Anschluß an die von Schütz (1932) erarbeitete Phänomenologie der ‚Schemata der Erfahrung als Deutungsschemata‘ methodisch entwickelt worden. Ansatzpunkt der Interpretation ist die aus der ‚Alltagshermeneutik‘ phänomenologisch erschlossene idealtypische Konstruktion der egologisch-monothetischen Perspektive der Selbstdeutung von Individuen und der mit dieser verbundenen einheitlichen ‚subjektiven‘ Sinnggebung von Interaktionsprodukten. Diese in der Selbstdeutung des Individuums fundierte Sinneinheit ergibt sich nicht zufällig. Es existiert vielmehr ein Zwang zur Gestalt- und Sinnschließung von Handlungen, der dem unüberschreitbaren Identitätspostulat von Individuen entspringt: eins zu sein mit sich selbst, seinen Erfahrungen, Handlungen und seiner Hand-

lungsperspektive. Dieses Deutungsregulativ fungiert vorweg immer schon sinnstiftend.

Die Analyse rekonstruiert nun aber, wie immer wieder zu betonen ist, nicht das Individuelle, ‚das eigentlich und in Wahrheit subjektiv Intendierte‘ einer Handlung, das einer interaktionstheoretischen Analyse *per se* gar nicht zugänglich wäre. Sie rekonstruiert interpretierend vielmehr anhand der methodischen Übernahme der ‚idealisierten Perspektive‘ des Sprechers eine bereits in einem Interaktionsbeitrag typisierte, egologisch-monothetisch konstituierte Sinnfigur einer ebenfalls typisierten Individualität. Durch die Übernahme der ‚idealisierten‘ egologischen Perspektive des Sprechers schließen sich die Äußerungen als subjektive monothetisch zu einer Sinneinheit zusammen, die objektiv zugänglich wird, insofern bei der Rekonstruktion die strukturell vorgegebenen, objektiven Regeln der Sinnkonstitution eingesetzt werden. Solche objektiven Regeln der Konsistenzbildung, der Konstruktion und Rekonstruktion von egologisch-monothetischen Sinneinheiten sind - so Schütz (z.B. 1932, 1971, S. 3-54) - repräsentiert in der objektiven Wirkungsweise von „Um-zu-“ und „Weil-Motiven“. In ihnen werden konsistente Deutungs- und Selbstdeutungsmuster generiert.

Bei der Formulierung der Um-zu- und Weil-Motive kommt es notwendig zu einer Verkürzung des Sinnhorizonts der einzelnen Äußerungen. Diese Verkürzung baut jedoch auf der interpretativen Erschließung der ‚idealisierten Perspektive‘ des Sprechers auf und muß sich an ihr bewähren. D.h. jedes Detail der vorausgegangenen Interpretation muß sich mit der Motivangabe widerspruchsfrei verrechnen lassen. Die interpretierende Rekonstruktion des Motivzusammenhangs hat demnach vor allem folgende Aufgaben: a) Angabe der manifesten Weil-Motive, b) Angabe der mit ihnen verbundenen und damit noch weiterwirkenden Um-zu-Motive, und c) Analyse des Verknüpfungs- und Sinnzusammenhangs von Um-zu- und Weil-Motiven.

### ***3.5.2 Übernahme der polythetisch-interaktionsbezogenen Perspektive der Alltagshermeneutik***

Das aus monothetischer Perspektive sinnkonsistente Selbstverstehen wird in polythetischer Sicht gebrochen und Interpretationsaufgabe für das Fremdverstehen: Aus einer Äußerung als Repräsentation monothetisch geschlossenen Sinnes wird im Interaktionszusammenhang ein von den anderen Partnern zu interpretierender Text. Damit wird die Aufgabe der Konsistenzherstellung auf den bzw. die Interaktionspartner verlagert. Diese interpretative Konsistenzherstellung ist ein alltäglicher Vorgang innerhalb jeder Interaktion. Sie beruht auf der hermeneutischen Kompetenz alltäglich Handelnder und ist funktional bestimmt durch das in der Sozialisation eingeübte Regelwissen der Interak-

tionspartner. Treten innerhalb alltäglicher Interaktion (z. B. im Gespräch) Inkonsistenzen auf, so existiert ein Interaktionsrepertoire zur Korrektur, zur Reparatur oder zum Auffüllen der Lücken in den jeweiligen Interaktionsbeiträgen: Nachfragen, Erstellung und Äußerung eines Interpretationsangebots, Aufrechterhaltung des Gesprächs bis zur praktikablen Klärung von Unbestimmtheiten, usw.

Jede interpretierende Interaktionsaktivität (und jede Interaktionsaktivität ist immer schon - ob latent oder manifest - interpretierend) besteht aus zwei Basisaktivitäten: 1. der Sinnzumessung und 2. der Sinnüberprüfung (d.h. der Korrektur usw.). Sinnüberprüfung tritt im Fall spürbarer Inkonsistenzen oder Sinndefizite sowie zur Absicherung von Sinnzumessung auf. Sie wird in alltäglicher Interaktion ‚automatisch‘ eingesetzt, fungiert jedoch im Rahmen des vorgestellten Interpretationsverfahrens als **methodisches Prinzip** der Aufdeckung von Inkonsistenzen, auf das der ex-post-interpretierende Sozialwissenschaftler, der als solcher ja, wie wir gesehen haben, auf das Alltagsrepertoire der aktuellen, interaktiv korrigierenden Sinnerschließung verzichten muß, angewiesen ist.

Die Übernahme der polythetisch-interaktionsbezogenen Perspektive der Alltagshermeneutik fungiert somit generell als methodisches Prinzip sozialwissenschaftlicher Interpretation von Interaktion. Der reflektierte Einsatz der Alltagskompetenz des Sozialwissenschaftlers legt dabei die systematische Trennung der Sinn- **Überprüfung**, d. h. der Aufdeckung von Inkonsistenzen, und der Sinn-Schließung, d.h. der Konsistenzherstellung, nahe.

### **3.5.3 Konstitution des einheitlichen Interaktionssinnes**

Methodisch kann die Interpretation nun zum einen zurückgreifen auf die Rekonstruktion der idealisierten egologischen Perspektive des Sprechers (bei dieser Rekonstruktion wird die Fähigkeit und Alltagskompetenz des Interpreten eingesetzt, die er **selbst** in alltäglicher Interaktion zur Selbstdarstellung und Selbsterklärung anwendet), zum anderen hat sie die monothetische Perspektive der rekonstruktiv idealisierten Subjektivität polythetisch gebrochen, d.h. sie hat die Fähigkeit alltäglich Interagierender zur wechselseitigen multiplen Perspektivenübernahme methodisch eingesetzt und damit auch die monothetische Sinneinheit polythetisch gebrochen. Sie hat darüber hinaus das alltäglich regelhaft verwendete Interaktionsrepertoire zur Aufdeckung von Lücken und Inkonsistenzen im Text genutzt (dazu bereits Mead, 1934; aber auch Schütz & Luckmann, 1979). Dabei werden ‚Bedeutungslücken‘, ‚Inkonsistenzen‘ und Sprecherstrategien sichtbar, die eine monothetisch-egologische Interpretation des Textes überschreiten und zugleich deren Grenzen zeigen: Der interaktionsbezogene Sinn eines Textes ist auf Polythetik angelegt und auch in seinem

Interaktionssinn, d.h. in seiner intersubjektiven Bedeutung nur polythetisch zu erschließen. Polythetische Bedeutungszumessung und Sinnkonstruktion ist demnach immer perspektivisch gebrochen, also ein Angebot von Lesarten und Deutungen eines Textes, das allein aus sich heraus nicht zu einer Sinneinheit zusammengesgeschlossen werden kann: Polythetik stiftet keine Sinneinheit, sondern verhindert sie eher.

Einen einheitlichen Interaktionssinn erhält ein Text **nicht** durch die jeweilige konkrete, subjektiv-perspektivische Sinnzumessung der einzelnen Interaktanten. Die Sinneinheit wird vielmehr gestiftet durch die alle Interaktanten in einem Interaktionszusammenhang zusammenschließende, von allen Beteiligten gemeinsam ausgehandelte Interaktionskonfiguration innerhalb eines für alle gemeinsamen Interaktionsprozesses. Die **interaktionstheoretische** Interpretation eines Textes hat dementsprechend nicht lediglich die jeweiligen perspektivisch variierenden, polythetisch strukturierten, möglichen Lesarten der Interaktionsbeiträge zu rekonstruieren, sondern die spezifische Bedeutung der sinnstiftenden Einheit. D.h., sie rekonstruiert das spezifische Handlungs- und Sinnsystem der jeweiligen Interaktionskonfiguration, so wie es sich innerhalb eines konkreten Interaktionsprozesses, einer Interaktionssequenz, äußert.

Die Mitglieder dieses spezifischen Handlungs- und Sinnsystems unterstellen - und konstituieren durch diese Unterstellung - eine gemeinsame Wahrnehmung- und Handlungssituation und ein gemeinsames Relevanz- und Bedeutungssystem. Diese bilden zusammen das Sinngefüge, den Interaktionsrahmen, innerhalb dessen die Interaktionsbeiträge in einer bestimmten Abfolge produziert werden. Diese Abfolge strukturiert den Interaktionsprozeß, wobei die Abfolgestructur ein allgemeines Prinzip interaktiver Bedeutungszumessung widerspiegelt: die sequenzielle Bedeutungskonstitution und die Selektion von Bedeutungsalternativen durch den Filter der Textabfolge.

Aus dem Interaktionsrahmen und der Abfolgestructur des Interaktionsprozesses wird die Bedeutung der Interaktionsprodukte abgeleitet. Dadurch erhalten sie ihren ‚objektiven‘, d.h. intersubjektiv gültigen und rekonstruierbaren Sinn - sowohl für die in einem konkreten Interaktionsrahmen Handelnden als auch für die späteren Interpreten der in diesem Rahmen produzierten Handlungstexte. Herauszuarbeiten ist also die konkrete, von allen Beteiligten latent als gemeinsam unterstellte Situations- und Interaktionstypisierung und der ‚objektive‘, d.h. intersubjektive (implizite, stillschweigend vorausgesetzte) Sinn dieser Typisierung. Erst durch die Aufdeckung dieses Sinnes können Verhaltensstrategien, Sprechertaktiken, bewußte und ‚unbewußte‘ Interaktionsangebote und Reaktionsmuster wie Zustimmung oder Abwehr als Netz Sinnhaft aufeinander bezogener Bedeutungseinheiten dargestellt werden.

Der interpretative Sinnschließungsprozeß hat demnach die Aufgabe, über die Darstellung möglicher Lesarten des Textes den einheitlichen Sinn der Inter-

aktionskonfiguration zu erschliessen, und zwar so, daß die polythetisch möglichen Lesarten eines Textes in einen durch die Interaktionskonfiguration gestifteten einheitlichen Interaktionsinn des Textes überführt werden können.

#### 4. *Hermeneutik als selbstreflexives Unternehmen: Die Unumgänglichkeit der Interpretation in den Sozialwissenschaften*

Die Interessen oder auch die Interessengegensätze nicht nur der Wissenschaftler, sondern auch der Alltagsmenschen orientieren sich vornehmlich an der **Deutung der wahrgenommenen Akte** - am Deutungsergebnis. Sie übergehen dabei in aller Regel jedoch die - oft nicht bewußt wahrgenommenen und registrierten - **Akte der Deutung**, mit deren Hilfe das Deutungsergebnis erarbeitet wird. Anders ausgedrückt: Menschen sind üblicherweise so intensiv damit beschäftigt, unentwegt zu verstehen, daß sie sich mit dem Problem des Verstehens selber gar nicht beschäftigen (können).

Gerade für die auslegenden, interpretierenden Wissenschaften kommt es aber - und je empirischer sie vorgehen, um so mehr - darauf an, die Differenz zwischen der Deutung der Akte und anderer ‚Daten‘ einerseits und den Akten der Deutung und ihren spezifischen Ausformungen und Arbeitsweisen andererseits herauszuarbeiten und - vor allem - aus dieser Unterscheidung bei der Interpretationsarbeit Konsequenzen zu ziehen: Wer über die Akte der Deutung nichts weiß und sich über ihre Prämissen und Ablaufstrukturen keine Rechenschaftspflicht auferlegt, interpretiert auf der Grundlage impliziter alltäglicher Deutungsroutinen und Plausibilitätskriterien, d.h. - aus der Sicht wissenschaftlicher Überprüfungspflicht - einfältig. Anders ausgedrückt: Wer Strukturen und Arbeitsweisen alltäglicher Deutung nicht kennt, ist weder imstande, alltägliche - ‚naive‘ - Deutungen zu kontrollieren, noch sie zu widerlegen.

Es ist also nicht der „Universalitätsanspruch der Hermeneutik“ (Habermas, 1982, S. 331-366) und der damit schon nahezu rituell verbundene Idealismusverdacht ihr gegenüber, sondern vielmehr die Unausweichlichkeit, mit der wir im Alltag zur Deutung und in der Wissenschaft zur theoretischen Erfassung der Deutungsarbeit gezwungen sind (u.a. Popper, 1972), die zu der Einsicht geführt hat, daß zum wissenschaftlichen ‚Verstehen von etwas‘ die **Beschreibung und das Verstehen des Verstehens** selbst gehören. Festzuhalten ist demnach, daß jede Form von Forschung - und damit neben bzw. mit den Geisteswissenschaften auch und gerade die Sozialforschung - auf Akten der Deutung basiert (König, 1962, S. 109). Insofern ist jede Form von Sozialforschung in einem sehr allgemeinen Sinn ‚interpretativ‘.

Ebenso deutlich ist, daß die unterschiedlichen Produktionen und Qualitäten sozialwissenschaftlicher Daten nicht nur von den Wissenschaftlern selbst auf Verstehen hin organisiert, sondern schon vorweg allgemein ‚verstehensmäßig‘ konstituiert sind. Denn während natürliche Ereignisse keinen Sinn ‚in sich‘ tragen und deshalb ihre Bedeutung vom Beobachter definiert wird, sind kulturelle Phänomene eben ‚immer schon‘ mit Sinn besetzt. Und diesen Eigen-Sinn gilt es zu rekonstruieren. Kurz: Sozialforschung basiert auf Akten der Deutung, erarbeitet und bezieht sich auf verstehensmäßig konstituierte Daten und gewinnt ihre Erklärungen durch Dateninterpretation.

Aus **dieser** Perspektive ergibt sich also kein Fundament für eine Konfrontation zwischen ‚quantitativer‘ und ‚qualitativer‘ Sozialforschung. Setzt man diese allgemeine Reflexionsebene voraus, so unterscheiden sie sich zwar in ihren Methoden, nicht aber in ihren Prämissen und Zielen: Beide beziehen sich auf historisch-soziales menschliches Handeln, auf seine Organisation und Orientierung, auf Dokumente und Produkte des Handelns sowie auf Deutungen von Handlungen in ‚Texten‘ etc. Die kontrollierte, d.h. die eigenen Prämissen, Verfahren und Variationskriterien überprüfende Auslegung von Daten, die sich - seien sie zeitlich auch noch so nah an der Gegenwart -**prinzipiell** auf vergangene Planungen, Ereignisse und Handlungen beziehen, deren Ergebnisse und Dokumente sie sind und die sie repräsentieren - dieser Typus der Auslegung ist keine Spezialität der Sozialwissenschaften. Es ist vielmehr die allgemeine Form wissenschaftlichen Verstehens.

Dabei sind die wissenschaftlichen Verstehensleistungen weitgehend ähnlich strukturiert wie die alltäglichen, aus denen sie herrühren und deren Verfahren und Kriterien sie mehr unreflektiert und implizit als bedacht und kontrolliert ausleihen. Auch die Resultate des inexpliziten wissenschaftlichen und des vorwissenschaftlichen Verstehens weisen eine Reihe von Ähnlichkeiten auf: Beide münden in der Artikulation von Erklärungen für etwas, wobei diese Erklärungen zumeist auf Plausibilitätsstandards beruhen. Diese leiten sich ihrerseits aus den nicht oder nicht mehr gewußten Routinen des Typisierens, Verknüpfens von Standarderfahrungen und Prozessen des ‚Deutens-wie-immer‘ ab.

Vor diesem theoretischen Hintergrund wird deutlich, daß die Diskussionen um eine **sozialwissenschaftliche** Hermeneutik - allgemeiner: um die hermeneutischen Grundlagen der Sozialwissenschaften - weder geeignet sind noch eine Grundlage bieten für wissenschaftstheoretische und rhetorische Feldschlachten zwischen den ‚Quantitativen‘ und den ‚Qualitativen‘. Im Gegenteil: Bei diesen Diskussionen geht es um das gemeinsame Fundament sozialwissenschaftlicher Auslegungs- und Analyseverfahren.

Allerdings ist, wie wir im Vorhergehenden zu zeigen versucht haben, die hermeneutische Argumentation grundsätzlich anticartesianisch. Sie akzeptiert weder den **formalmethodischen** Subjektivismus (Descartes, 1637) noch die hieraus



folgende Aufteilung der Welt und des ihr gegenüberstehenden Erkenntnissubjekts in Außendinge (ausgedehnte Materie, „res extensa“) einerseits und deutende Erkenntnis (denkende Substanz, „res cogitans“) andererseits (Descartes, 1641 und 1644) - sowie die daraus abgeleitete Hypothese von der ‚Mathematisierbarkeit‘ der Welt und dem daraus wiederum folgenden ‚Objekt‘- und ‚Objektivitäts‘-Begriff. Sie sieht vielmehr die miteinander agierenden Individuen und deren a prioriische Intersubjektivität **in** der sozial gedeuteten Welt (Strauss, 1991a; Goffman, 1974) und nicht dieser **gegenüber**. Sie zielt nicht nur ab auf das Beobachten, Beschreiben, Verstehen und Erklären des Sozialen, sondern in eins damit auf das Soziale der artspezifischen, historisch sich verändernden Wahrnehmungs- und Artikulationsmuster und Zwecke des Beobachtens, Beschreibens, Verstehens und Erklärens (Durkheim, 1912, S. 557-571).

Das Beschreiben und auslegende Verstehen sozialer Orientierung, sozialen Handelns, sozialer Handlungsprodukte und des jeweiligen historischen ‚subjektiven‘ oder ‚kollektiven‘ Selbstverständnisses menschlicher Individuen, Gruppen oder Gesellschaften ist demnach grundsätzlich verbunden mit der Auffindung, Beschreibung und Auslegung der ‚Praktiken‘, ‚Regeln‘, ‚Muster‘, derer wir uns bedienen, wenn wir uns orientieren, vergewissern, artikulieren, verständigen - wenn wir handeln, wenn wir produzieren und interpretieren. Neben dem, worüber wir uns verständigen, woraufhin wir handeln, was wir erklären und verstehen, werden so die Verständigungsprozesse, die impliziten Regeln der Konsenskonstitution und -herstellung, die gestischen, bildhaften und sprachlichen Artikulationsinstrumente, ihr historisch sich wandelnder Verwendungs- und Bedeutungszusammenhang und das historisch jeweils als gesichert geltende Regel- und Wissenssystem notwendig in die Untersuchung des Sozialen mit einbezogen.

Die erkenntnistheoretische Bedeutung verstehender Ansätze liegt also sozusagen in einer ‚Reform‘ sozialwissenschaftlichen Denkens im Allgemeinen, die mit der Aufklärung ihrer eigenen Praxis beginnt - denn diese Praxis ist vor allem anderen (und jenseits aller methodischen Raffinessen quantifizierender Forschung) ein unreflektiertes, fragloses, quasi-selbstverständliches Fremdverstehen.

## 5. Resümee

Sozialwissenschaftliches Verstehen soll dazu dienen, gesellschaftliche Wirklichkeit(en) angemessen und stimmig, zuverlässig, gültig und überprüfbar zu rekonstruieren. Sozialwissenschaftliches Verstehen zielt, anders als andere artifizielle Verstehensformen (wie z. B. intuitionistische, mystische, existentialistische) auf die Erkenntnis des Typischen, und zwar sowohl des typischen

Handelns als auch des mit diesem zusammenhängenden typischen Wissens, wie schließlich auch des typischen alltäglichen Verstehens. Erst von einem ‚Verstehen des Verstehens‘ aus lassen sich dann auch systematisch Ähnlichkeiten und Unterschiede zwischen alltäglichem und wissenschaftlichem Verstehen aufweisen. Tatsächlich liegen diese nicht im Verfahren des Verstehens selbst, sondern betreffen eher 1. Organisationsform, 2. Reflexionsgrad und 3. Zielsetzung der Deutung.

Sozialwissenschaftliche Hermeneutik beruht auf der Prämisse, daß Menschen versuchen, ihrem Handeln einen einheitlichen Sinn zu geben, weil sie grundsätzlich bestrebt sind, mit sich selber eins zu sein, weil sie **ihre** Sichtweisen als Teil ihrer selbst betrachten. Diese Sinn-‚Stiftung‘ ist zu rekonstruieren. Das methodische und methodologische Problem des Interpretieren besteht also wesentlich in der Frage, wie er sicherstellen kann, daß er die andere Perspektive auch tatsächlich einholt, denn der vom anderen subjektiv gemeinte Sinn liegt eben nicht offen zutage, bzw. genau genommen versteht ego den subjektiv gemeinten Sinn von alter ego überhaupt nicht ‚genau‘. Zugänglich ist grundsätzlich **nicht** dessen Bewußtsein; erfassbar und damit interpretierbar sind lediglich seine intersubjektiv wahrnehmbaren - intendierten wie unbeabsichtigten - (Ent-)Äußerungen.

Wichtige ‚Hilfsmittel‘ der wissenschaftlichen Annäherung an den Handlungsentswurf eines anderen Menschen ist mithin die **Fixierung** von dessen (Ent-) Äußerungen, denn das interpretative Vorgehen bezieht sich grundsätzlich auf fixierte Handlungssedimente, also auf Texte im weitesten Sinne. Damit beansprucht die sozialwissenschaftliche Hermeneutik einerseits, für alle Arten von ‚Texten‘ zuständig zu sein und andererseits, menschliche Wirklichkeiten generell erfassen zu können.

Um diese ihrem typischen Sinngehalt nach verstehen zu können, müssen sie dem Interpretieren zunächst einmal relevant genug erscheinen, daß er sich ihnen überhaupt - mit welchen pragmatischen Absichten auch immer - zuwendet. D.h., der Interpret muß seine zwischen lebenspraktischen Vordringlichkeiten und rein kognitiven Erkenntnisinteressen angesiedelten situativen Relevanzen klären. Strebt er nun systematisch Wissen über das sich (ent-)äußernde andere Individuum an, dann muß er seine Vor-Urteile über den anderen Menschen reflektierend in die Deutung miteinbeziehen. Und schließlich muß er eruieren, was die infrage stehenden Objektivationen im Hinblick auf kulturell bereitstehende ‚Optionen‘ bedeuten bzw. bedeuten können. D.h., er muß den sozio-kulturellen Zusammenhang, auf den der sich (Ent-)Äußernde bezieht, erkennen können und hinreichend - was immer das heißt - kennen, und er muß die Differenzen zwischen diesem und seinem eigenen kulturellen Kontext reflektieren und bei seinen Deutungen mitberücksichtigen.

Das hier vorgestellte interpretative Vorgehen ist mithin ein Interaktionsver-  
**stehen** in grundsätzlicher (zeitlicher) Distanz zum Interaktionsgeschehen selber. D.h., daß das Verstehen des Sozialwissenschaftlers sich nicht auf eine aktuelle, lebendige Umwelt, sondern auf das erfahrungsgesättigte und kompetente Imaginieren einer Mit- oder Vorwelt bezieht. Sein Verstehen geschieht in einer besonderen Einstellung, die sich einerseits auf die allgemeinen Bedingungen und Verfahren sowohl alltäglicher als auch wissenschaftlicher Auslegungen und andererseits auf „alle Probleme der Vernunft - der Vernunft in allen ihren Sondergestalten richtet“ (Husserl, 1936, S. 7). Das aber impliziert eine Einstellung des prinzipiellen Zweifels an sozialen Selbstverständlichkeiten, eine Einstellung, die man als ‚methodischen Skeptizismus‘ oder, etwas pointiert, auch als ‚künstliche Dummheit‘ bezeichnen könnte (vgl. Hitzler 1986).

Die herkömmliche sozialwissenschaftliche Forschungspraxis klärt hingegen **nicht**, zumindest nicht genügend, wie ihr Wissen und Erkenntnis über die Lebenswelt des Anderen überhaupt möglich ist. Die herkömmliche Forschungspraxis setzt vielmehr ihre Gegenstände, ohne Rechenschaft darüber zu geben, wie sie sich konstituieren. Sie perpetuiert somit unreflektiert den ‚gesunden Menschenverstand‘ des Alltags im sogenannten ‚Fachverstand‘ der Sozialwissenschaften. Der Unterschied zwischen ‚verstehenden‘ und ‚nicht-verstehenden‘ Verfahren besteht demnach also keineswegs darin, daß es die letzteren mit ‚nackten Tatsachen‘ zu tun hätten, sondern eher darin, wie **reflektiert** bzw. unreflektiert von Beginn an die einzelnen Deutungsschritte ab-  
 solviert werden.

Zur wissenschaftlichen, rational verfahrenen Hermeneutik wird Auslegungsarbeit also erst, wenn genügend Distanz zum dokumentierten Deutungsgegenstand besteht und wenn darüber hinaus die Verfahren der Sinnkonstitution, der Bildung von Alternativen, der Sinnselektion, der Verständigung und Konsensbildung beschrieben und aus ihrem Strukturzusammenhang heraus erklärt werden können. Das aber bedeutet programmatisch für jedes sich als ‚interpretativ‘ verstehende Vorgehen, daß das Verhältnis von alltäglichem und sozialwissenschaftlichem Verstehen beständig neu zu klären ist - und zwar im hier propagierten Verstande einer ‚Soziologie des Verstehens‘ (vgl. bereits Soeffner 1981).

## Literatur

- Austin, J.L. (1962). **How to do Things with Words**. Cambridge (Mass.)
- Berger, P.L. & Luckmann, Th. (1969). **Die gesellschaftliche Konstruktion der Wirklichkeit**. Frankfurt a.M.: Fischer.
- Bergmann, J. (1980). **Interaktion und Exploration. Eine konversationsanalytische Studie zur sozialen Organisation der Eröffnungsphase von psychiatrischen Aufnahmegesprächen**. Unveröff. Diss., Universität Konstanz.
- Bergmann, J. R. (1987). **Klatsch. Zur Sozialform der diskreten Indiskretion**. Berlin: de Gruyter.
- Bergmann, J. R. (1991). Konversationsanalyse. In U. Flick, E. v. Kardorff, H. Keupp, L. v. Rosenstiel & S. Wolff (Hrsg.), **Handbuch Qualitative Sozialforschung** (S.213-219). München: Psychologie Verlags Union.
- Bohnsack, R. (1991). **Rekonstruktive Sozialforschung**. Opladen: Leske + Budrich.
- Bude, H. (1985). Die individuelle Allgemeinheit des Falls. In H.-W. Franz (Hrsg.), **22. Deutscher Soziologentag 1984. Beiträge der Sektions- und Ad-hoc-Gruppen** (S. 84-86). Opladen: Westdeutscher.
- Bühler, K. (1934/1965). **Sprachtheorie. Die Darstellungsfunktion der Sprache** (2. Aufl. mit einem Vorwort von Friedrich Kainz). Stuttgart.
- Bultmann, R. (1954). Zum Problem der Entmythologisierung. In R. Bultmann: **Glauen und Verstehen**, Gesammelte Aufsätze, Viertes Band (S. 128-137). Tübingen: Mohr.
- Carnap, R. (1928/1961). **Der logische Aufbau der Welt**. Hamburg.
- Cassirer, E. (1953). **Philosophie der symbolischen Formen**. 3 Bände. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Chomsky, N. (1965). **Aspects of the Theory of Syntax**. Cambridge (Mass.)
- Cooley, C. H. (1902). **Human Nature and the Social Order**. New York.
- Descartes, R. (1637). **Discours de la Méthode** (Zweisprachige Ausgabe). Hamburg 1960.
- Descartes, R. (1641). **Meditationes de prima philosophia** (Zum akademischen Gebrauche neu hrsg. und eingeleitet von C. S. Barach). Wien 1866.
- Descartes, R. (1644). **Principia philosophiae** (=Oeuvres, publ. par C. Adam. Band 8,I.). Paris 1905.
- Dilthey, W. (1900). Die Entstehung der Hermeneutik. In W. Dilthey, **Gesammelte Schriften (Band 5, S.317-331)**. Leipzig, Berlin 1914-1936, fortgeführt Stuttgart, Göttingen 1962ff.
- Dilthey, W. (1970). **Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften**. Frankfurt a. M.: Suhrkamp.
- Dilthey, W. (1976). Entwürfe zur Kritik der historischen Vernunft. In H.-G. Gadamer & G. Boehm (Hrsg.), **Seminar: Philosophische Hermeneutik** (S. 189-220). Frankfurt: Suhrkamp.
- Durkheim, E. (1912). **Die elementaren Formen des religiösen Lebens**. Frankfurt 1981.
- Eberle, T.S. (1984). **Sinnkonstitution in Alltag und Wissenschaft**. Bern: Haupt.

- Flick, U., Kardoff, E.v., Keupp, H., Rosenstiel, L. v. & Wolff, S. (Hrsg.) (1991). **Handbuch Qualitative Sozialforschung**. München: Psychologie Verlags Union.
- Gadamer, H.-G. (1960). **Wahrheit und Methode**. Tübingen: Mohr/Siebeck.
- Gadamer, H.-G. & Boehm, G. (Hrsg.) (1976). **Seminar: Philosophische Hermeneutik**. Frankfurt a.M.: Suhrkamp.
- Gadamer, H.-G. & Boehm, G. (Hrsg.) (1978). **Seminar: Die Hermeneutik und die Wissenschaften**. Frankfurt a. M.: Suhrkamp.
- Garz, D. & Kraimer, K. (Hrsg.) (1992). Diskussion: **Objektive Hermeneutik**. Frankfurt a.M.: Suhrkamp.
- Gerhardt, U. (1986). **Patientenkarrieren**. Frankfurt a. M.: Suhrkamp.
- Gerhardt, U. (1991). Typenbildung. In U. Flick, E.v. Kardorff, H. Keupp, L. v. Rosenstiel & S. Wolff (Hrsg.), **Handbuch Qualitative Sozialforschung (S.435-440)**. München: Psychologie Verlags Union.
- Giddens, A. (1984). **Interpretative Soziologie. Eine kritische Einführung**. Frankfurt a.M.: Campus.
- Glaser, B. G. & Strauss, A. L. (1965). **Awareness of Dying**. Chicago: Aldine.
- Glaser, B. G. & Strauss, A. L. (1967). **The Discovery of Grounded Theory**. Chicago: Aldine.
- Glaser, B. G. & Strauss, A. L. (1968). **Time for Dying**. Chicago: Aldine.
- Goffman, E. (1974). **Frame Analysis. An Essay on the Organization of Experience**. New York: Harper & Row.
- Grathoff, R. (1989). **Milieu und Lebenswelt**. Frankfurt a.M.: Suhrkamp.
- Grondin, J. (1991). **Einführung in die philosophische Hermeneutik**. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Gurwitsch, A. (1977). **Die mitmenschlichen Begegnungen in der Milieuwelt** (Hrsg. u. eingeleitet v. A. Metraux). Berlin.
- Habermas, J. (1981). **Theorie des kommunikativen Handelns** (Bde. 1 u. 2). Frankfurt a.M.: Suhrkamp.
- Habermas, J. (1982). zur **Logik der Sozialwissenschaften** (5. erweiterte Aufl.). Frankfurt a. M.: Suhrkamp.
- Hauptert, B. (1991). Vom narrativen Interview zur biographischen Typenbildung. In D. Garz & K. Kraimer (Hrsg.), **Qualitativ-empirische Sozialforschung (S.213-254)**. Opladen: Westdeutscher.
- Heidegger, M. (1972). **Sein und Zeit**. Tübingen: Niemeyer.
- Helling, I. (1979). zur **Theorie der Konstrukte erster und zweiter Ordnung bei Alfred Schütz: Einige Probleme der Explikation und Anwendung**. Unveröffentlichte Dissertation, Konstanz.
- Hildenbrand, B. (1983). **Alltag und Krankheit**. Stuttgart: Klett-Cotta.
- Hildenbrand, B. (1991). **Alltag als Therapie**. Bern: Huber.
- Hirsehauer, S. (1993). **Die soziale Konstruktion von Transsexualität**. Frankfurt a.M.: Suhrkamp.

- Hitzler, R. (1986). Die Attitüde der künstlichen Dummheit. **Sozialwissenschaftliche Informationen (SOWI)**, 33, 53-59.
- Hitzler, R. (1988). **Sinnwelten**. Opladen: Westdeutscher.
- Hitzler, R. & Honer, A. (1984). Lebenswelt, Milieu, Situation. **Kölner Zeitschrift für Soziologie und Sozialpsychologie**, 36(1), 56-74.
- Hitzler, R. & Honer, A. (1988). Der lebensweltliche Forschungsansatz. Neue **Praxis**, 18(6), 496-501.
- Honer, A. (1985). Beschreibung einer Lebens-Welt. Zur Empirie des Bodybuilding. **Zeitschrift für Soziologie**, 14(2), 131-139.
- Honer, A. (1993). **Lebensweltliche Ethnographie**. Wiesbaden: DUV.
- Husserl, E. (1936). **Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie** (Husserliana, Bd. IV). Den Haag: Nijhoff 1969.
- Iser, W. (1972). **Der implizite Leser**. München: Fink.
- Jauß, H. R. (1982). **Ästhetische Erfahrung und literarische Hermeneutik**. Frankfurt a. M.: Suhrkamp.
- Kallmeyer, W. (1977). Verständigungsprobleme in Alltagsgesprächen. **Der Deutschunterricht**, 29(6), 52-69.
- Kallmeyer, W. (1980). Aushandlung und Bedeutungskonstitution. In P. Schröder & H. Steger (Hrsg.), **Dialogforschung** (S. 89-127). Düsseldorf: Schwarm.
- Knauth, B. & Wolff, S. (1991). Zur Fruchtbarkeit der Konversationsanalyse für die Untersuchung von Texten. **Zeitschrift für Soziologie**, 20(1), 36-49.
- Katz, J.J. & Fodor, J. A. (1963). **The Structure of Semantic Theory**. Language 39, 170-210.
- Knoblauch, H. (1992). **Die Welt der Wüschelrutengänger und Pendler**. Frankfurt a.M.: Campus.
- Knorr-Cetina, K. (1984). **Die Fabrikation von Erkenntnis**. Frankfurt a. M.: Suhrkamp.
- Knorr-Cetina, K. (1988). Das naturwissenschaftliche Labor als Ort der ‚Verdichtung‘ von Gesellschaft. **Zeitschrift für Soziologie**, 17(2), 85-101.
- Knorr-Cetina, K. (1989). Spielarten des Konstruktivismus. Soziale **Welt**, (1-2), 86-96.
- König, R. (1962). Die Beobachtung. In R. König (Hrsg.), **Handbuch der empirischen Sozialforschung** (Band 2, S. 1-65). Stuttgart: Enke.
- Langer, S.K. (1965). **Philosophie auf neuem Wege. Das Symbol im Denken, im Ritus und in der Kunst**. Frankfurt: Suhrkamp.
- Lau, T. (1992). **Die heiligen Narren. Punk 1976 - 1986**. Berlin, New York: de Gruyter.
- Luckmann, B. (1970). The Small Life-Worlds of Modern Man. **Social Research**, 4, 580-596.
- Luckmann, T. (1980). **Lebenswelt und Gesellschaft**. Paderborn: Schöningh.
- Luckmann, T. (1988). Kommunikative Gattungen im kommunikativen ‚Haushalt‘ einer Gesellschaft. In G. Smolka-Koerdt, P.M. Spangenberg & D. Tillmann-Bartylla (Hrsg.), **Der Ursprung von Literatur** (S. 279-288). München: Fink.
- Luckmann, T. (1990). Lebenswelt: Modebegriff oder Forschungsprogramm. **Grundlagen der Weiterbildung (GdW)**, 1, 9-13

- Lüders, C. (1991). Deutungsmusteranalyse. Annäherungen an ein risikoreiches Konzept. In D. Garz & K. Kraimer (Hrsg.), **Qualitativ-empirische Sozialforschung** (S. 377-408). Opladen: Westdeutscher.
- Matthiesen, U. (1992). Standbein - Spielbein: Deutungsmusteranalysen im Spannungsfeld von Objektiver Hermeneutik und Sozialphänomenologie. In D. Garz & K. Kraimer (Hrsg.), **Diskussion: Objektive Hermeneutik**. Frankfurt a.M.: Suhrkamp.
- Mead, G. H. (1934). **Geist, Identität und Gesellschaft aus der Sicht des Sozialbehaviorismus** (Mit einer Einleitung von C. W. Morris). Frankfurt 1969.
- Müller-Doohm, S. (1990). Vom Positivismusstreit zur Hermeneutikdebatte - Die Aktualität des interpretativen Paradigmas. **KulturAnalysen**, 2(3), 292-307.
- Nassen, U. (Hrsg.) (1979). **Texthermeneutik - Aktualität, Geschichte, Kritik**. Paderborn u. a.: Schöningh.
- Oevermann, U. (1983). Zur Sache. In L. v. Friedeburg & J. Habermas (Hrsg.), **Adorno-Konferenz 1983**. Frankfurt a.M.: Suhrkamp.
- Oevermann, U. (1991). Genetischer Strukturalismus und das sozialwissenschaftliche Problem der Erklärung der Entstehung des Neuen. In S. Müller-Doohm (Hrsg.), **Jenseits der Utopie** (S.267-336). Frankfurt a.M.: Suhrkamp.
- Oevermann, U., Allert, T., Konau, E. & Krambeck, J. (1979). Die Methodologie einer ‚objektiven Hermeneutik‘ und ihre allgemeine forschungslogische Bedeutung in den Sozialwissenschaften. In H.-G. Soeffner (Hrsg.), **Interpretative Verfahren in den Sozial- und Textwissenschaften** (S. 352-434). Stuttgart: Metzler.
- Plessner, H. (1928). **Die Stufen des Organischen und der Mensch** (Um Vorwort, Nachwort und Register erweiterte 2.Aufl.). Berlin 1965.
- Plessner, H. (1970). **Philosophische Anthropologie** (Hrsg. mit einem Nachwort von G. Dux). Frankfurt: Suhrkamp.
- Popper, K. (1963). Conjectures and **Refutations. The Growth of Scientific Knowledge**. London.
- Popper, K. (1972). **Objektive Erkenntnis. Ein evolutionärer Entwurf** (2. Aufl.). Hamburg 1974.
- Reichertz, J. (1986). **Probleme qualitativer Sozialforschung. zur Entstehungsgeschichte der objektiven Hermeneutik**. Frankfurt: Campus.
- Reichertz, J. (1991). **Aufklärungsarbeit**. Stuttgart: Enke.
- Riedel, M. (1978). **Verstehen oder Erklären? Zur Theorie und Geschichte der hermeneutischen Wissenschaften**. Stuttgart: Klett-Cotta.
- Riemann, G. (1987). **Das Fremdwerden der eigenen Biographie. Narrative Interviews mit psychiatrischen Patienten**. München: Fink.
- Scheler, M. (1923). **Wesen und Formen der Sympathie. Die deutsche Philosophie der Gegenwart** (Hrsg. mit einem Anhang von M. S. Frings, 6. durchgesehene Aufl. v. ‚Phänomenologie und Theorie des Sympathiegefühls‘, Deutsche Philosophie der Gegenwart, 2. durchgesehene Aufl.). Bern: Haupt.
- Schleiermacher, F. D. E. (1977). **Hermeneutik und Kritik**. Frankfurt a. M.: Suhrkamp.
- Schröer, N. (1992). **Der Kampf um Dominanz. Hermeneutische Fallanalyse einer polizeilichen Beschuldigtenvernehmung** Berlin, New York: De Gruyter.

- Schütz, A. (1932). **Der sinnhafte Aufbau der sozialen Welt**. Frankfurt a.M.: Suhrkamp 1974.
- Schütz, A. (1971). **Gesammelte Aufsätze** (Bd. 1). Den Haag: Nijhoff.
- Schütz, A. & Luckmann, Th. (1979/1984). **Strukturen der Lebenswelt** (Bde. 1 u. 2). Frankfurt a.M.: Suhrkamp.
- Schütze, F. (1989). Kollektive Verlaufskurve oder kollektiver Wandlungsprozeß - Dimensionen eines Vergleichs von Kriegserfahrungen amerikanischer und deutscher Soldaten im zweiten Weltkrieg. **BIOS**, 1,
- Searle, J.R. (1969). **Speech Acts**. Cambridge (Mass.)
- Simmel, G. (1918). **Vom Wesen des historischen Verstehens**. Berlin: Mittler & Sohn.
- Soeffner, H.-G. (1980). Überlegungen zur sozialwissenschaftlichen Hermeneutik am Beispiel der Interpretation eines Textausschnitts aus einem ‚freien‘ Interview. In T. Heinze, H.W. Klusemann & H.-G. Soeffner (Hrsg.), **Interpretationen einer Bildungsgeschichte** (S. 70-96). Bensheim: päd-extra.
- Soeffner, H.-G. (1981). Verstehende Soziologie - Soziologie des Verstehens. In J. Mattes (Hrsg.), **Lebenswelt und soziale Probleme. Verhandlungen des 20. Deutschen Soziologentags zu Bremen** (S. 329-331). Frankfurt: Campus.
- Soeffner, H.-G. (Hrsg.) (1988). **Kultur und Alltag** (Sonderband 6 von ‚Soziale Weh‘). Göttingen: Schwartz.
- Soeffner, H.-G. (1989). Auslegung **des Alltags - Der Alltag der Auslegung**. Frankfurt a.M.: Suhrkamp.
- Soeffner, H.-G. (1991). Zur Soziologie des Symbols und des Rituals. In J. Oelkers & K. Wegenast (Hrsg.), **Das Symbol - Brücke des Verstehens**. Stuttgart: Kohlhammer.
- Soeffner, H.-G. (1992). **Die Ordnung der Rituale - Auslegung des Alltags 2**. Frankfurt a.M.: Suhrkamp.
- Srubar, I. (1981). Max Scheler: Eine wissenssoziologische Alternative. In N. Stehr & V. Meja (Hrsg.), **Wissenssoziologie** (Sonderheft der Kölner Zeitschrift für Soziologie und Sozialpsychologie) (S. 343-359). Opladen: Westdeutscher.
- Strauss, A. L. (1991). **Grundlagen qualitativer Sozialforschung**. München: Fink.
- Strauss, A. L. (1991a). **Creating Sociological Awareness**. New Brunswick: Transaction.
- Ungeheuer, G. (1987). **Kommunikationstheoretische Schriften Z: Sprechen, Mitteilen, Verstehen**. Aachen: Rader.
- Vonderach, G (1986). **Verstricktsein in Sprachspiele, Metaphern und Geschichten**. Leer.
- Voß, A. (1992). **Betteln und Spenden - Eine soziologische Studie über Rituale freiwilliger Armenunterstützung, ihre historischen und aktuellen Formen sowie ihre sozialen Leistungen**. Berlin, New York: De Gruyter.
- Weber, M. (1917). Der Sinn der ‚Wertfreiheit‘ der soziologischen und ökonomischen Wissenschaften. In M. Weber, **Gesammelte Aufsätze zur Wissenschaftslehre** (Hrsg. v. J. Winckelmann, 4. erneut durchgesehene Aufl.) (S.489-540). Tübingen: Mohr/Siebeck 1973.
- Weber, M. (1917a). Die ‚Objektivität‘ sozialwissenschaftlicher und sozialpolitischer Erkenntnis. In M. Weber, **Gesammelte Aufsätze zur Wissenschaftslehre** (Hrsg. v. J.



Winckelmann, 4. erneut durchgesehene Aufl.) (S. 146-214). Tübingen: Mohr/Siebeck 1973.

Weber, M. (1972). ***Wirtschaft und Gesellschaft. Grundriß der verstehenden Soziologie*** (5. revidierte Aufl., hrsg. v. J. Winckelmann). Tübingen: Mohr.

Wundt, W. (1921). ***Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte.*** Stuttgart.

#### 4. Kapitel

## Methodologische Grundlagen und Probleme der Psychophysik

***Rainer Mausfeld***

Dieses Kapitel beschäftigt sich mit einigen methodologischen und metatheoretischen Fragen der Psychophysik; dabei ergänzen psychologiehistorische und ideengeschichtliche Aspekte gelegentlich die Darstellung. Es hat weitgehend einen einführenden Charakter; doch wird sich die Berechtigung der behandelten metatheoretischen Probleme eher demjenigen erschließen, der mit Grundkonzepten der Psychophysik, wie sie beispielsweise bei Tack (1984) dargestellt sind, und mit wahrnehmungspsychologischen Fragestellungen vertraut ist.

In der Psychophysik, mit der die experimentelle Psychologie ihren historischen Anfang nahm, erweisen sich metatheoretische Betrachtungen psychologischer Fragen als besonders lohnend und ertragreich. In ihr verbinden sich nämlich eine Vielzahl methodologischer und philosophischer Grundfragen der Psychologie in einer Weise miteinander, die es gestattet, sich ihnen in einer präziseren Weise zu nähern, als es in anderen Bereichen der Psychologie möglich ist.

Für eine Behandlung derartiger Fragen wird eine Differenzierung von zwei Auffassungen der Psychophysik nützlich sein. Es lassen sich bei einer geschichtlichen Betrachtung der Theorieentwicklung in der Psychophysik zwei unterschiedliche Positionen abstrahieren, deren Spannungsverhältnis die gesamte Geschichte der Psychophysik von Fechner bis heute durchzieht. Richtet die eine Perspektive das Augenmerk auf die Frage der Meßbarkeit des Psychischen, so sucht die andere Perspektive Mechanismen der Wahrnehmung unter dem funktionalen Aspekt der Kopplung des Organismus an seine Umwelt zu verstehen. Als Vertreter der erstgenannten Tradition der Psychophysik, die stark durch das Vorbild der Physik geprägt ist, seien Fechner, Stevens, Luce und Falmagne genannt; herausragende Vertreter der zweiten Tradition, in der die Psychophysik, sei es aus eher neurophysiologischer oder aus funktionalistischer Perspektive, der Biologie näher steht als der Physik, sind Hering, Bühler, Brunswik, Gibson und Marr. Die Behandlung allgemeiner methodologischer Fragen, wie beispielsweise derjenigen, was in der Psychophy-

sik unter einer ‚Theorie‘ oder einem ‚Modell‘ zu verstehen sei und was als ‚Daten‘ anzusehen sei, wie auch die Diskussion grundlegender spezifischer Fragen, etwa der Definition des ‚Reizes‘, hängt davon ab, von welcher der beiden Positionen aus man sie betrachtet; einige Probleme gar werden nur innerhalb einer einzelnen Perspektive sichtbar, nicht hingegen in der jeweils anderen. Der Versuch einer Gegenüberstellung dieser theoretischen Perspektiven und Traditionen, der zwar Idealisierungen und den Verzicht auf einige wichtige Nuancierungen notwendig macht, läßt hoffen, daß die jeweiligen methodologische Probleme besonders deutlich hervortreten.

Welcher Art nun sind die methodologischen Fragen, die sich mit diesen beiden Traditionen verbinden? Einige Beispiele vorweg mögen verdeutlichen, daß hinter substanzwissenschaftlichen Problemen der Psychophysik tatsächlich weitreichende metatheoretische Probleme verborgen sind.

Grundprobleme der erstgenannten Tradition sind etwa: Kann es im Bereich der Psychophysik eine Messung subjektiver Größen in gleich strengem Sinne geben, wie es in der Physik eine Messung physikalischer Größen gibt? Was ist das ‚Meßobjekt‘ einer psychophysikalischen Skala: der physikalische Reiz, die neurale Erregung oder die Empfindung? Vermögen derartige Skalen die Kluft zwischen ‚Außenwelt und ‚Innenwelt‘ zu überbrücken und somit zu einer Klärung des Leib-Seele-Problems beizutragen? Zu welchen Aspekten der physikalischen Welt soll man eine solche Skala in Beziehung setzen: Zu den eindimensionalen physikalischen Größen, die sich beim Aufbau der Physik als fruchtbar erwiesen haben, oder zu komplexeren physikalischen Eigenschaften, die grundlegend für die funktionalen Leistungen des Wahrnehmungssystems sind? Was also ist ein Reiz? Führen psychophysikalische Skalen zu einem theoretischen Gefüge ähnlich der dimensional Struktur der Physik?

In der funktionalistischen und ‚mechanistischen‘ Tradition der Psychophysik werden methodologische Fragen zunächst durch das gegenwärtig vorherrschende Paradigma perzeptuell-kognitiver Informationsverarbeitung angeregt: In welchem Sinne ist es sinnvoll, das Wahrnehmungssystem als informationsverarbeitendes System anzusehen? Wie läßt sich dabei der Begriff ‚Information‘ bestimmen? Lassen sich die mit einer funktionalistischen Perspektive einhergehenden teleologischen Betrachtungsweisen epistemologisch rechtfertigen? Kann man von Hypothesen oder Interpretationen des Wahrnehmungssystems reden? (Was wurde es dann heißen, ein Wahrnehmungsurteil anzuzweifeln?) Können die Empfindungsqualitäten, die sog. Qualia, theoretische Bestandteile psychophysikalischer Modelle sein? Worauf sonst zielt hier der Erklärungs-begriff? Die spezifischen konzeptuellen Probleme des informationstheoretischen Paradigmas sind damit freilich verlassen, und man stößt wieder auf Varianten des Leib-Seele-Problems: Läßt sich ein neuroreduktionistisches Programm rechtfertigen? Lassen sich die Funktionsweisen des

Wahrnehmungssystems unabhängig von seiner biologischen Realisierung verstehen?

Bereits diese Aufzählung macht die Grenzen des Unterfangens, methodologische Probleme der Psychophysik zu behandeln, deutlich; denn eine Abgrenzung zu weiterreichenden Fragen, wie sie in der Philosophie der (Wahrnehmungs)Psychologie, in der Wissenschaftstheorie und Epistemologie ebenso intensiv wie tiefgehend diskutiert werden, muß stets zu einer Verkürzung der Perspektive führen (stellvertretend für die umfangreiche philosophische Literatur hierzu seien Rorty, 1979, Putnam, 1988, Carrier & Mittelstraß, 1989, und Dennett, 1991, genannt). Einige Einschränkungen, denen die hier gegebene Darstellung unterliegt, sollen daher vorweg genannt werden: Ihr Ziel liegt darin, die Aufmerksamkeit auf konzeptuelle und methodologische Fragen der Psychophysik zu lenken, ohne daß sie dabei freilich beanspruchen könnte, diese umfassend zu erörtern oder gar zu beantworten. Zudem sollen methodologische Probleme der Psychophysik nur insoweit behandelt werden, als sie unmittelbar mit dem Forschungsparadigma der Psychophysik verknüpft sind. Unter diesem Aspekt erweist sich die erstgenannte Tradition als vergleichsweise einheitlich. Anders die zweite Tradition: Sie ist in ihren theoretischen Perspektiven und Methoden hoch ausdifferenziert und befindet sich in einer dynamischen Entwicklung. Ihre metatheoretischen Probleme reichen weit über die Psychophysik hinaus und sind Gegenstand multidisziplinärer Bemühungen im Kontext der sog. cognitive sciences; diese Tradition kann folglich nur in einigen Grundzügen behandelt werden.

## **1. Der Beginn der Psychophysik**

### 1.1 Die Entstehung der psychophysikalischen Zugangsweise

Die erste psychophysikalische Gesetzmäßigkeit wurde von den Pythagoräern formuliert (vgl. v. Fritz, 1971, S. 47ff.). Sie entdeckten, daß die Länge einer Saite in einem bestimmten Verhältnis zur Tonhöhe steht und eine Verkürzung der Saite auf die Hälfte die Tonhöhe um eine Oktave hebt, unabhängig von der Länge der Saite und ihrer Spannung. Warum es nach dieser Entdeckung mehr als zweitausend Jahre brauchte, bis man sich in systematischer Weise der Untersuchung von Gesetzmäßigkeiten der Wahrnehmung zuwandte, ist angesichts der Entwicklung, welche die Physik im selben Zeitraum genommen hat, eine interessante, doch bislang unbeantwortete Frage der Wissenschaftsgeschichte.

Wissenschaftshistorisch betrachtet nehmen Psychophysik und Physik bei denselben Phänomenen ihren Anfang. Die **Phänomene** und die **perzeptuellen Qualitäten** bildeten bei der Entstehung der Wissenschaften den Ausgangs-

punkt und das Rohmaterial des Theoretisierens. Die Mechanik entstand aus Beobachtungen im Zusammenhang mit der Muskelkraft, die Optik aus Phänomenen des Gesichtsinns, die Akustik aus solchen des Gehörsinns und die Wärmelehre aus Phänomenen der Temperaturempfindung. Ein Verständnis der Welt wurde seit den Vorsokratikern indessen nicht auf der Ebene der Phänomene erwartet, sondern nur auf jener Ebene der dem unmittelbaren Schauen verborgenen einheitlichen Kräfte und Entitäten, welche die tiefere Ordnung hinter den Erscheinungen bestimmen. In der Wahrnehmungswelt sollte sich eine Ordnung im Bereich der Außenwelt zuerst widerspiegeln. Ziel physikalischer Theoriebildung war aber der Aufbau eines Bildes der hinter den Erscheinungen stehenden ‚wahren‘ Welt. Für diesen Aufbau mußte der das ‚Chaos der Erscheinungen‘ ordnende Geist seinen Ausgangspunkt zwar bei den Sinnen nehmen, doch suchte er deren ‚Unzuverlässigkeiten‘ und Eigenheiten zu eliminieren und Regelmäßigkeiten einem physikalischen, d.h. unpersönlichen Objekt zu attribuieren. Die - in den Worten von Max Planck - fortschreitende Ausmerzung des Anthropomorphen aus dem entstehenden Bild einer ‚hinter den Sinnesqualitäten‘ liegenden ‚wahren‘ Welt, die Entstehung eines physikalischen Weltbildes also, verstellte den Blick für die Regelmäßigkeiten, die der menschlichen Wahrnehmung selbst zugrunde liegen. Erst sehr viel später, mit Beginn des neunzehnten Jahrhunderts, entdeckte man, daß dem, was sich aus physikalischer Perspektive als Wahrnehmungsfehler erwies, selbst Gesetzmäßigkeiten unterliegen. Die Korrespondenz zwischen physikalischen Reizen und den durch sie ausgelösten psychischen Reaktionen wurde zu einem eigenständigen, zu einem psychologischen Untersuchungsgegenstand.

Phänomenologisch haben Physik und Psychophysik also denselben wissenschaftshistorischen Ausgangspunkt, epistemologisch unterliegen ihnen - pointiert ausgedrückt - antithetische Betrachtungsweisen, aus denen heraus sie sich als eigentlich theoretische Disziplinen konstituierten. Beginnt die Physik mit der Untersuchung von Invarianzen unter Abstraktion der Wahrnehmung, so beginnt - ausgehend von der Physik - die Psychophysik mit der Untersuchung von Invarianzen der Wahrnehmung unter Abstraktion von den jeweils spezifischen Objekten. Aufgabe der Psychophysik wird es, strukturelle Invarianzen der Wahrnehmung über Variationen physikalischer Größen hinweg aufzufinden und so die Struktur der Wahrnehmung selbst zu erhellen.

## 1.2 Fechners Psychophysik

G.Th. Fechner, auf den zugleich der Forschungsgegenstand dieser Disziplin wie auch ihre Bezeichnung zurückgehen, verstand die Psychophysik als die „exakte Lehre von den functionellen oder Abhängigkeitsbeziehungen zwi-

schen Körper und Seele“ und suchte durch sie den Nachweis zu führen, daß „das Geistige überhaupt quantitativen Verhältnissen unterliegt“ (Fechner, 1860, I, S. 8, S. 55). Zwischen den Erscheinungen unseres Bewußtseins und den ihnen entsprechenden physischen Vorgängen müsse sich eine mathematische Gesetzmäßigkeit feststellen lassen, konstatierte Fechner, durch Spinozas monistische Weltansicht geprägt, 1851 in seinem philosophischen Hauptwerk **Zend-Avesta**. Dort skizzierte er unter der Überschrift **Kurze Darlegung eines neuen Prinzips mathematischer Psychologie** eine solche Gesetzmäßigkeit: Die „im Bewußtsein zu schätzende Intensität geistiger Tätigkeit“ sei nicht von der **absoluten** Änderung des äußeren Reizes abhängig, sondern ihre Änderung sei der **verhältnismäßigen** Änderung des physikalischen Reizes proportional. Mit der Unterschiedsschwelle als „Einheit“ der Empfindung erhält Fechner das (formale) Resultat, daß sich die Empfindung als Logarithmus des physikalischen Reizes darstellen läßt (oder vorsichtiger: daß sie sich durch eine solche Definition bestimmen läßt). Die **Elemente der Psychophysik** von 1860 dienten der eigenständigen Behandlung und Begründung dieses „Prinzips“. Fechner hoffte, dadurch seiner im **Zend-Avesta** entworfenen Weltanschauung eine wissenschaftliche Grundlage geben zu können. Zugleich suchte er aber alle seine Untersuchungen von jeglichen metaphysischen Gesichtspunkten frei zu halten und sie „nur auf den erfahrungsmäßigen Beziehungen zwischen Leib und Seele“ zu gründen (Fechner, 1860, I, S.6). Ein neues Forschungsprogramm, die Psychophysik, war geschaffen, und es wurde zur Keimzelle der experimentellen Psychologie.

„Der Psychologie und Physik schon durch den Namen verbunden, hat die Psychophysik einerseits auf der Psychologie zu fassen und verspricht andererseits, derselben mathematische Unterlagen zu gewähren“ (a.a.O., S. 11). Mit der Psychophysik sollte die Psychologie aus dem wissenschaftlich unfruchtbaren Dickicht philosophischer Fragestellungen, die sich vornehmlich um das Leib-Seele-Problem ranken, herausgeführt und einer rein erfahrungsmäßigen Behandlung zugänglich gemacht werden. Die Psychophysik teilt sich nach Fechner in zwei Bereiche: in das Gebiet der „äußeren Psychophysik“, welche die Beziehungen zwischen den psychischen Phänomenen und den äußeren, physikalischen Reizen behandelt, und das Gebiet der „inneren Psychophysik“, welche die Beziehungen zwischen den psychischen Phänomenen und den inneren physikalischen Vorgängen (d.h. der durch die Reize ausgelösten „psychophysischen Tätigkeit“) untersucht. Die psychophysische Tätigkeit, in modernen Worten: die neurale Codierung, ist die Verbindungskette zwischen Reiz und Empfindung. Die Untersuchung der Abhängigkeit der Empfindung von der psychophysischen Tätigkeit sieht Fechner als die eigentliche Aufgabe der Psychophysik. Doch nur die äußere Psychophysik sei der Erfahrung zugänglich, die Beziehungen der inneren Psychophysik seien aus ersterer zu folgern (Fechner, 1860, II, S. 377ff.). In Fechners Unterscheidung

von innerer und äußerer Psychophysik wird bereits eines der zentralen metatheoretischen Probleme der Psychophysik deutlich: die logische Kluft zwischen neuraler Erregung und Empfindung. Die Ideengeschichte der Psychophysik führt mit mannigfaltigen Beispielen vor Augen, daß scheinbar einfache und in der Alltagssprache vertraute Begriffe, hier der Begriff ‚Empfindung‘, als wissenschaftliche Begriffe ontologische Implikationen haben, durch die man immer wieder in die Aporien des Leib-Seele-Problems, einem Erbe der abendländischen Philosophie, geführt wird. Fechner freilich sah die Psychophysik als einen Weg zur Klärung gerade dieser philosophischen Frage an.

Im Zentrum der Fechnerschen Psychophysik stand der Versuch, „ein Mass der Empfindung“ zu entwickeln. Dazu bedarf es einer Maßeinheit, durch die sich - da wir „Gleichheit im Empfindungsgebiete“ zu beurteilen in der Lage sind - die Stärke einer Empfindung als eine „Summierung eines Soundsovielmal des Gleichen“ (Fechner, 1860, I, S. 56) messen läßt. Dieses Maß braucht nicht selbst im Gebiet des Psychischen zu liegen, denn „es wird niemals möglich sein, eine Empfindung unmittelbar so über eine andere zu legen, dass ein Mass der einen durch die andere erwüchse; aber es kann durch Zuziehung von etwas anderem, woran die Empfindungen so gut geknüpft sind, als die Ausdehnung der Elle an die Materie der Elle, möglich sein, ein Mass der Empfindung zu gewinnen“ (a.a.O., S. 57). Dieses Andere, auf dem sich ein Maß der Empfindung gründen ließe, fand Fechner in der von ihm **Webersches Gesetz** genannten Beziehung. Damit stelle die äußere Psychophysik die Grundlage einer Empfindungsmessung bereit. Es war diese Idee einer Empfindungsmessung, die die weitere Entwicklung der Psychophysik stark beeinflusste und die sich in gewisser Weise verselbständigte. Ein weiteres entscheidendes Konzept war das der Schwelle, das sich schon in Leibniz' und Herbarts Behandlung von Bewußtseinsvorgängen findet. Ihm kam in Fechners metaphysischen Betrachtungen eine besondere Rolle zu, und er wurde grundlegender Baustein seiner Psychophysik. Nach dem Schwellengesetz (s. Fechner, 1860, I, S. 238ff.) schwindet eine Empfindung oder ein Unterschied zwischen Empfindungen nicht erst bei einem Reiz der physikalischen ‚Intensität‘ Null, sondern bereits bei einer bestimmten Reizgröße, der Schwelle. Aus dem Schwellengesetz und aus dem Weberschen Gesetz leitet Fechner sein „Massgesetz der Empfindung“ und seine „Unterschiedsmassformel“ ab, denen zufolge die Größe der Empfindung zu dem Logarithmus der Größe des auf seinen Schwellenwert bezogenen Reizes in Beziehung steht (Fechner, 1860, II, S. 13). Diese Ableitung vollzog sich unter der Voraussetzung der Gültigkeit des Weberschen Gesetz sowie unter der Annahme, daß den ebenmerklichen Reizunterschieden gleiche Empfindungsunterschiede entsprechen, wodurch gleichsam eine Einheit zur Messung der Empfindung ausgezeichnet wird.

Keine dieser Voraussetzungen blieb unwidersprochen. Die grundsätzliche Möglichkeit einer Empfindungsmessung wurde ebenso in Abrede gestellt

(z. B. durch von Kries, 1882) wie die Berechtigung einzelner Teile der Fechnerschen Argumentation. Helmholtz, Mach, Brentano und Hering, um nur die bekanntesten Namen zu nennen, brachten zahlreiche Einwände gegen Fechners Auffassungen vor, denen Fechner in einer Verteidigungsschrift (Fechner, 1877; s. a. 1888) zu begegnen suchte. Die zeitgenössischen Einwände betrafen oftmals konzeptuelle Fragen, die damals wie heute auf die grundsätzlichen Probleme hindeuten, die mit dem Versuch einer ‚Empfindungsmessung‘ verbunden sind. Wie berechtigt mancher Einwand auch gewesen sein mag, die Voraussetzungen einer experimentellen Psychologie, wie sie schließlich durch Wundt institutionalisiert wurde, waren geschaffen. Was die Psychophysik zu dieser beitrug, waren nicht so sehr ihre Fragestellungen, sondern vor allem ihre Methodik.

Der Fechnersche Entwurf einer Psychophysik betraf nicht nur die Frage nach dem quantitativen Verhältnis von physikalischem Reiz und Empfindung, sondern war seiner Intention nach sehr viel umfassender. Die Psychophysik sollte Phänomene der Aufmerksamkeit, Gefühle und Triebe ebenso behandeln wie Träume, Erinnerungs- und Vorstellungsbilder, mit denen sich Fechner (1860, II, S.445ff.) beschäftigte. Fechner sah in der Psychophysik mehr als lediglich eine quantitativ fundierte oder erweiterte Wahrnehmungspsychologie. Ihm ging es um die Möglichkeit einer wissenschaftlichen Psychologie insgesamt. Als experimenteller Wahrnehmungspsychologe hatte er schon, seine methodischen Erfahrungen aus der Experimentalphysik nutzend, im Bereich der Untersuchung von Nachbildern und farbigen Schatten gearbeitet (Fechner, 1838, 1840). In der Psychophysik nun ging es aber um mehr: Ein für allemal sollte die prinzipielle Möglichkeit einer quantitativen Behandlung des Psychischen nachgewiesen und die Psychologie in gleicher Weise einem experimentellen Vorgehen und einer mathematischen Begriffsbildung zugänglich gemacht werden wie zuvor die Physik. Doch Fechner selbst trug dazu bei, daß dieser Entwurf sich zunehmend auf Fragen der Methodik reduzierte und das eigentlich Psychologische in den Hintergrund trat. Der größte Teil seiner Arbeiten zur Psychophysik befaßte sich mit methodischen Fragen; in ihrer akribischen, detaillierten und subtilen Behandlung lag Fechners Stärke, so daß die Psychophysik bald mit ihren Methoden identifiziert wurde (s. Boring, 1942, S. 36) und diese zum eigentlichen Gegenstand der Psychophysik wurden (etwa Müller, 1903). Wie man auch den psychologischen Ertrag der Fechnerschen Psychophysik bewerten mag: Fechners Beschäftigung mit Fragen der Meßbarkeit des Psychischen und mit damit verbundenen Problemen der Behandlung von ‚Meßfehlern‘ initiierte ein Forschungsprogramm, dessen methodische Ergebnisse weit über die Psychophysik hinaus von großem Einfluß waren. Aus so verschiedenen Bereichen wie der Abstrakten Meßtheorie, der Latent-Trait-Testtheorie, der Entscheidungstheorie und der Lerntheorie sind die Konzepte, die auf formale Impulse dieser Tradition zurückgehen, kaum mehr weg-



zudenken. Dieses Forschungsprogramm präzisierte die Konzepte der absoluten und der Unterschiedsschwelle und machte sie einer statistischen Behandlung zugänglich; es erfaßte allgemeine Regularitäten des Diskriminationsverhaltens und stellte zahlreiche gut studierte probabilistische Modelle für den Umgang mit Urteilsfluktuationen bereit.

Zu weiteren wissenschaftshistorischen Aspekten des Fechnerschen Forschungsprogrammes sei auf Boring (1950) sowie auf den Sammelband von Brozek & Gundlach (1988) verwiesen. Scheerer (e. g. 1992) behandelt in einer Reihe von Arbeiten die vernachlässigte Ideengeschichte der inneren Psychophysik und macht deutlich, wie sehr deren Voraussetzungen denen einer biophysikalisch orientierten Psychologie ähneln.

## ***2. Vom Status psychophysikalischer Skalen***

Für Fechner bildete noch das Leib-Seele-Problem, das Verhältnis von Innenwelt und Außenwelt, das Ausgangsproblem. Seine Psychophysik jedoch hatte er bewußt hiervon frei zu halten gesucht und die (für ihn als empirisch entscheidbar angesehene) Frage nach der Meßbarkeit des Psychischen gestellt. Mit der Präzisierung sowohl des Meß- wie auch des Größenbegriffs - eine Präzisierung, die immer zugleich auch eine Neubestimmung des jeweiligen Begriffes bedeutet - ist die Kluft zwischen dem so gefaßten Problem und dem philosophischen Kernproblem Fechners zunehmend größer geworden. In der Tat haben die Untersuchungen zur Meßbarkeit die ursprüngliche philosophische Frage Fechners nach der Leib-Seele-Beziehung nicht beantwortet, ja nicht einmal tangiert, und sie können dies auch nicht. Aus einer philosophischen Perspektive mag die Wendung, welche die Fechnersche Frage nach den quantitativen Abhängigkeitsbeziehungen zwischen physischer und psychischer Welt genommen hat, einem Verlust gleichkommen, für die Theorieentwicklung in der Psychologie bedeutet sie jedoch eine wichtige Klärung.

In einem ursprünglichen Sinne suchte Fechner gleichsam introspektiv Empfindungen zu messen und sah die äußere Psychophysik dabei lediglich als einen notwendigen Umweg an. Das Subjekt selbst war der Beobachter, das Objekt der Beobachtung waren die Empfindungen. Die extreme Gegenposition zu dieser Auffassung psychophysikalischer Skalenkonstruktion markierte Stevens: Der Versuchsleiter war der Beobachter, das Objekt der Beobachtung waren Reiz-Response-Beziehungen. Aus dieser Sicht spiegeln psychophysikalische Skalen nicht eine interne, subjektive Entität der untersuchten Person wider, sondern lediglich Regularitäten in der durch bestimmte Instruktionen vermittelten Reaktionen auf physikalische Reize. Modernen Auffassungen zur Skalenkonstruktion hingegen scheint am ehesten die Vorstellung zugrunde zu liegen, daß physikalische Reize Objekte sind, die von einer Person gleichsam

durch die Brille der neuralen Transduktion und anderer sensorischer Mechanismen beurteilt werden. Was bei der psychophysikalischen Skalenkonstruktion das Objekt und was der Beobachter ist, hängt also von der theoretischen und metatheoretischen Perspektive ab, so daß auch bei gleichen formalen Techniken der Skalenkonstruktion die erstellte Skala eine jeweils andere Interpretation erfährt.

So unterschiedlich die genannten Perspektiven auch sind, so ist ihnen doch durch die Bedeutung, die sie der Skalenkonstruktion beimessen, eine Auffassung gemeinsam: nämlich die Psychophysik gleichsam als perzeptuelle Physik anzusehen.

## 2.1 Die Psychophysik als perzeptuelle Physik

Fechners Psychophysik war zwangsläufig durch die physikalische Wissenschaft seiner Zeit geprägt; hinzu kam der Elementarismus, der die Philosophie des Geistigen seit Locke, Berkeley und Hume beherrschte und es naheliegend erscheinen ließ, als Fundament der Psychophysik isolierte eindimensionale Größen der Physik heranzuziehen. So bildeten in der Geschichte der Psychophysik bis in jüngere Zeit die eindimensionalen Größen, die sich zur Beschreibung der Welt im Rahmen einer physikalischen Theoriebildung als fruchtbar erwiesen haben, den Leitfaden zur Konstruktion psychophysikalischer Skalen. Im Mittelpunkt der Psychophysik standen folglich jene perzeptuellen Attribute, die sich als unmittelbares Korrelat physikalischer Größen ansehen lassen; zumeist waren dies intensitäts/energie-korrelierte Attribute wie Lautstärke, Helligkeit, Wärme oder Druckempfindung. Für Fechner bedeutete dies keine wesentliche Beschränkung, denn ihm ging es darum, auf der Grundlage einer einzigen empirischen Gesetzmäßigkeit und gewisser, für ihn plausibler Annahmen den theoretischen Nachweis einer **prinzipiellen Meßbarkeit** des Psychischen zu führen.

Dem Forschungsprogramm der Fechnerschen Psychophysik, wie es später im behavioristischen Gewande von Stevens fortgeführt wurde, liegt vereinfacht folgende Vorstellung zugrunde: Zu den physikalischen Reizen gehören Empfindungsattribute, die in struktureller Hinsicht den physikalischen Dimensionen vergleichbar sind. Diese Empfindungsattribute gilt es unter geeigneten Bedingungen in numerischen Skalen zu erfassen, die sich dann wieder untereinander zu einer Struktur ähnlich der dimensional Struktur der Physik verbinden (vgl. Luce & Krumbhansl, 1988, S.4f.). Daher ist es Ziel dieses Forschungsprogrammes, a) Techniken der Skalenkonstruktion bereitzustellen, b) die Beziehungen dieser Skalen untereinander zu untersuchen und c) ihre Beziehungen zu den entsprechenden physikalischen Skalen zu erfassen.

Die Auffassung, die Psychophysik gleichsam als perzeptuelle Physik zu denken und sich bei der Skalenkonstruktion durch die Größen der Physik leiten zu lassen, ist eng damit verbunden, das Wahrnehmungssystem selbst als eine Art Meßinstrument für physikalische Größen anzusehen und funktionale Aspekte einer adaptiven Kopplung des Organismus an seine Umwelt zu vernachlässigen. Hierauf werden wir in den Abschnitten 3 und 5 zu sprechen kommen. Wenden wir uns zunächst kurz der Frage nach dem Status psychophysikalischer Skalen zu.

## 2.2 Was ‚messen‘ eindimensionale psychophysikalische Skalen?

Das geschilderte Vorhaben bringt eine Vielzahl grundsätzlicher begrifflicher und methodologischer Probleme mit sich: War das Fechnersche Gesetz in ähnlicher Weise von Nutzen wie beispielsweise das Gesetz, das Druck und Volumen miteinander verbindet? Hat man etwa ‚Empfindung‘ gemessen, wenn man physikalische Reize mißt und sie logarithmiert oder potenziert? Oder schaut man bei diesem Versuch stets durch die Brille zumeist verborgener (und oft schwer zu rechtfertigender) theoretischer Annahmen - vergleichbar jenen, die man beispielsweise auch bei der ‚Messung‘ der Konzentrationsleistung durch die physikalische Reaktionszeit machen muß?

Betrachten wir den Status, den eine Aussage der Art „Die Empfindung  $E$  ergibt sich als Logarithmus/Potenzfunktion des physikalischen Reizes  $R$ “ hat. Wird damit ein Gesetz formuliert oder nur eine empirische Regularität, d.h. eine empirische Gesetzmäßigkeit beschrieben? Von einem Gesetz wurden wir nur dann sprechen, wenn die beiden beteiligten Größen  $E$  und  $R$  unabhängig meßbar wären und wenn zudem die durch solche unabhängigen Messungen bestimmte Beziehung gewisse Invarianzeigenschaften hätte. Lassen sich die beiden Größen zwar unabhängig bestimmen, ist jedoch ihre postulierte Beziehung an die Wahl einer bestimmten Skala gebunden, so wurden wir von einer empirischen Regularität sprechen. Beide Fälle kommen indes nicht in Betracht, da wir  $E$  nicht unabhängig messen können (zumindestens gibt es keinen Konsens über ein geeignetes Kriterium hierfür). Eine genauere Betrachtung zeigt, daß wir es bei der obigen Aussage mit einer verkappten **Definition** von  $E$  zu tun haben; dieser Definitionscharakter verbirgt sich hinter der empirischen Relation, der zufolge beispielsweise gleiche Dominanzwahrscheinlichkeiten gleiche Empfindungsdifferenzen bzw. gleiche Zahlenverhältnisse bei der Größenschätzung gleichen Empfindungsverhältnissen entsprechen sollen. Dieses sind willkürliche Setzungen, durch die das Konzept der Empfindung überhaupt erst eingeführt wird. Auf diesem Argument gründete bereits v. Kries (1882) seinen Versuch aufzuzeigen, daß eine eigenständige, fundamentale Messung von Empfindungen prinzipiell nicht möglich sei, sondern

daß hier lediglich eine physikalische Messung zusammen mit einer willkürlichen Konvention vorliege. „Wir können **festsetzen**, dass die eben merklichen Empfindungszuwüchse zu einer ganzen Intensitätsreihe als gleich betrachtet werden sollen. Thun wir dies, so können wir nun eine Anzahl beobachtbarer Thatsachen so ausdrücken, dass wir den Empfindungen ein Wachstum mit dem Logarithmus des Reizes zuschreiben. Aber dieses Gesetz bedeutet, wie man nun sieht, **ohne** jene Festsetzung gar nichts, **mit ihr zusammen** aber nichts weiter als die beobachteten Thatsachen. Wir können auch ebensogut festsetzen, dass Empfindungszuwüchse als gleich gelten sollen, welche gleichen Reizzuwüchsen entsprechen. Dann können wir dieselben Thatsachen so ausdrücken, dass, je stärker die Empfindung wird, um so grösser die eben merklichen Zuwüchse werden. Eines ist so richtig wie das Andere. Es kann sich nur um die Frage handeln, was zweckmäßiger ist“ (a.a.O., S. 276). Seine ausführlichen methodologischen Analysen zum Status psychophysikalischer Skalen schließt von Kries so: „Von einer Messung der Empfindungen kann ebenfalls nur nach einer willkürlichen Festsetzung dessen, was als gleich betrachtet werden soll, die Rede sein. Der Streit, ob die Empfindung mit dem Logarithmus des Reizes wachsen oder dem Reize proportional ist, ist daher keine sachliche Differenz, sondern ein auf Missverständniss beruhender Streit um Worte.“ (a.a.O., S. 294) Zu ähnlichen Schlußfolgerungen kamen später u.a. Falmagne (1985, S.321f.) und Laming (1991).

Der behavioristische Zugang zur Skalenkonstruktion in der Psychophysik scheint auf den ersten Blick von diesem Einwand nicht berührt zu sein, da er gar nicht erst vorgibt, ‚Empfindungen‘ messen zu wollen. Versteht man hier konsequent den Versuchsleiter (statt die Versuchsperson selbst) als Beobachter und die Reiz-Response-Beziehung (statt der Empfindung) als den Beobachtungsgegenstand, so verschwindet in der Tat jede Frage nach dem Status der so erhaltenen Skalen. Mit ihr verschwindet jedoch auch die theoretische Motivation, sich der Untersuchung derartiger Skalen zuzuwenden: Erst die Intuition, daß ‚interne Zustände‘, zu denen die Person selbst Zugang hat und die mit dem physikalischen Reiz zusammenhängen, für die entdeckten Regularitäten verantwortlich sind, könnte aus theoretischer Perspektive das Studium psychophysikalischer Skalen als lohnend erscheinen lassen.

Der Einwand, daß eine eigenständige, fundamentale Messung der Empfindung nicht möglich sei, trifft also gleichermaßen Fechners ‚subjektivistische‘ wie in eher indirekter Weise die behavioristische Interpretation psychophysikalischer Skalen. Daß sich über eine ‚Empfindungsmessung‘ erst auf der Grundlage einer Konvention sprechen läßt, muß jedoch weder notwendig eine eventuelle theoretische Bedeutung noch einen möglichen praktischen Nutzen solcher Skalen ausschließen.

### **3. Die Meßinstrumentkonzeption der Wahrnehmung**

Die zuvor beschriebene Tradition der Psychophysik beschäftigt sich typischerweise mit solchen Reizen, die sich durch eine eindimensionale Größe der Physik beschreiben lassen. Dieser Sachverhalt kann geradezu als ein Unterscheidungskriterium dieser Tradition der Psychophysik und der Wahrnehmungspsychologie angesehen werden: Je einfacher und präziser man den der Wahrnehmung zugrunde liegenden Reiz beschreiben kann, um so eher wird man geneigt sein, diesen Untersuchungsbereich der (traditionellen) Psychophysik zuzurechnen; je komplexer eine physikalische Beschreibung des Reizes wird und je schwerer die Wahrnehmungsphänomene mit der physikalischen Reizbeschreibung in Beziehung gesetzt werden können, um so eher wird man von ‚Wahrnehmungspsychologie‘ sprechen. Die (zumeist implizite) Auffassung von der Rolle und Funktionsweise des Wahrnehmungssystems, die jener Tradition der Psychophysik unterliegt, kann als **Meßinstrumentkonzeption der Wahrnehmung** bezeichnet werden. Pointiert läßt sie sich so charakterisieren: Das von der ‚Außenwelt‘ auf die Sinnesrezeptoren treffende Energiemuster wird vom Wahrnehmungssystem Punkt für Punkt ‚ausgewertet‘ und führt punktweise zu ‚Empfindungen‘, aus denen sich dann auf ‚höherer Ebene‘ ‚Wahrnehmungen‘ konstituieren. In der Betonung der Punkt-für-Punkt-Auswertung ist diese Konzeption **lokal** orientiert (**global** bedingte Phänomene werden den sog. Kontexteffekten zugerechnet); durch ihre Annahme, daß gleichsam Empfindungsatome die Grundlage der Wahrnehmung bilden, ist sie **atomistisch**. Wir wollen sie daher auch als **lokal-atomistische Konzeption der Wahrnehmung** bezeichnen. Ideengeschichtlich geht diese Konzeption zurück auf die empiristische Tradition von Locke, Berkeley und Hume, in der 1765 Thomas Reid in seinen **Essays on the Intellectual Powers of Man** die einflußreiche Unterscheidung zwischen „sensations“, als dem Rohmaterial der Erfahrung, und „perceptions“, als etwas auf externe Objekte Bezogenes, explizit eingeführt hatte. Diese Unterscheidung, die Bezüge zu der später im logischen Empirismus diskutierten Sinnesdaten-Theorie aufweist, prägte nachhaltig das wahrnehmungspsychologische Denken. Wenn dieser Auffassung zufolge die Empfindung das Primäre ist, das dann durch Aufmerksamkeit, Gedächtnis und andere kognitive Prozesse zur Wahrnehmung umgebildet werde, so könne auch die Psychophysik nur durch ein Verständnis des Einfachen, d.h. der Empfindungen, zu einem Verständnis des darauf aufbauenden Komplexen, d.h. der Wahrnehmungen, gelangen. Diese elementaristische Auffassung von Empfindungen als Grundbausteinen der Wahrnehmung führt in natürlicher Weise zu einer Fokussierung auf den proximalen Reiz, zu seiner Charakterisierung in Termini elementarer physikalischer Größen und zur Konzentration auf entsprechende Primärrezeptoren. Der **lokal-atomistischen Konzeption der Wahrnehmung** zufolge stehen an der Basis des Wahrnehmungssystems gleich-

sam sensorische Meßinstrumente, die das Wahrnehmungssystem über den physikalischen Input ‚unterrichten‘.

Es scheint nur natürlich zu sein, daß aus einer solchen Perspektive die klassische Psychophysik vorrangig von bekannten Größen der Physik ausging, wie Schallenergie, Lichtintensität und Wellenlänge des Lichtes, und deren Beziehung zu den korrespondierenden Empfindungen, wie Lautstärke, Helligkeit und Farbe, untersuchte. Untersuchungsbereiche wie die Bewegungswahrnehmung waren hingegen weniger durch elementaristische Haltungen dieser Art gefährdet, da hier eine lokal-atomistische Konzeption des Reizes auf der Grundlage einer einfachen physikalischen Größe nicht möglich war. Die noch zu behandelnde funktionalistische Psychophysik indessen ist gerade dadurch gekennzeichnet, daß sie den komplexen physikalischen Input erst zu identifizieren sucht, der für eine bestimmte Wahrnehmungsleistung den ‚Reiz‘ konstituiert. (Über den Begriff des **perzeptuellen Codes** lassen sich beide Perspektiven in Beziehung setzen: Der klassische Zugang der skalenorientierten Tradition ist durchaus auch als ein Bemühen zu verstehen, die Form elementarer interner Codes zu bestimmen. Faßt man das wichtige Konzept perzeptueller Codes hinreichend abstrakt, so werden hinter der Verschiedenheit beider Traditionen wiederum aufschlußreiche Gemeinsamkeiten sichtbar.)

Für die Untersuchung der primären Codierung und für die sog. subjektive Sinnesphysiologie, die mit oftmals gleichen Methoden wie die Psychophysik auf eine neurophysiologische Theoriebildung zielt, erwies sich der metaphorische Vergleich der Wahrnehmung mit einem physikalischen Meßinstrument als überaus fruchtbar und brachte u. a. die Young-Helmholtz-Theorie des Farbsehens und die Theorie auditiver Primärcodierung hervor. Sucht man aber Einsichten in die psychologische Struktur des Wahrnehmungssystems zu gewinnen, so erweist sich die Meßinstrumentkonzeption als irreführend; zahlreiche klassische Phänomene legen hiervon Zeugnis ab, wie beispielsweise die virtuelle Polyphonie in der Musikwahrnehmung oder das Phänomen der Farbkonstanz: Würde der mit einer bestimmten Stelle einer visuellen Szene verbundene Farbeindruck nur von der Wellenlängenzusammensetzung des Lichtes abhängen, das von dieser Stelle zum Auge kommt, so müßte sich der Farbeindruck mit der Farbigkeit der Beleuchtung ändern. Tatsächlich bleibt er aber weitgehend konstant (ein Farbdia, mit einem Tageslichtfilm unter Kunstlicht aufgenommen, führt das Ausmaß der physikalischen Änderung vor Augen). Hier zeigt sich, daß die Meßinstrumentkonzeption geradezu den Weg zu einem Verständnis komplexerer psychologischer Wahrnehmungsfunktionen verstellt.

So bildete für eine Hauptströmung der Psychophysik die Meßinstrumentkonzeption der Wahrnehmung den Hintergrund, vor dem erst die Konstruktion von Skalen zur zentralen Aufgabe werden konnte. Zugleich jedoch ist die

gesamte Geschichte der Wahrnehmungspsychologie auch die Geschichte einer fortwährenden Auseinandersetzung mit dieser Meßinstrumentkonzeption der Wahrnehmung (deren moderne komplexere Variante die algorithmischen Bemühungen im Rahmen einer sog. „inversen Optik“ sind). Fechner selbst (in seinen frühen Arbeiten zu den farbigen Schatten), Mach, Hering, Katz und Gelb haben ebenso auf ihre Unangemessenheit hingewiesen wie die Gestaltpsychologen. Ihre mangelnde theoretische Fruchtbarkeit wurde besonders deutlich mit dem Aufkommen einer funktionalistischen und evolutionstheoretischen Betrachtungsweise in der Psychophysik, welche die Wahrnehmung im Dienste einer komplexen Anbindung des Organismus an seine Umwelt sieht.

#### **4. Die Entwicklung von Theorien und Modellen der perzeptuell-kognitiven Informationsverarbeitung**

Die skalenorientierte Tradition der Psychophysik ging mit ihrer Fechnerschen Grundlegung aus dem philosophischem Unterfangen hervor, mit der Psychophysik zur Lösung des Leib-Seele-Problems beizutragen. Jedoch fand Fechners monistische Philosophie in der sich rein empirisch und antimetaphysisch verstehenden Psychologie keinen Platz - was Fechner selbst am Ende seines Schaffens als tragisch erlebte. Die Psychophysik ignorierte die Intentionen ihres Begründers und verselbständigte sich als Disziplin der experimentellen Psychologie. In den Worten von Stevens (1951, S. 30): „Fechner, like his fellow mystic Pythagoras, advanced a science by trying to prove a theology.“

Doch das Leib-Seele-Problem zog, pointiert formuliert, hundert Jahre nach dem Erscheinen von Fechners **Elementen der Psychophysik** und nach Dekaden einer behavioristischen Abstinenz wieder die Aufmerksamkeit der Psychologie auf sich. Intuitionen ganz unterschiedlicher Forschungsbereiche wie der Informationstheorie, der Kybernetik, der Computerwissenschaften, der Linguistik und der beginnenden kognitiven Psychologie bündelten sich in einer Haltung zum Leib-Seele-Problem, die unter dem Stichwort **Funktionalismus** in unterschiedlichsten Varianten die **sog. cognitive sciences** beherrscht. Putnams (1960/1975) Arbeit **Minds and Machines** markiert in der expliziten Ausgestaltung dieses Standpunktes den Beginn einer mehr als zwei Jahrzehnte vorherrschenden Denkweise, deren Prämissen innerhalb psychologischer Theoriebildung erst durch den jüngeren Konnektionismus eine grundlegende Herausforderung erfuhren.

In einer gängigen Spielart des Funktionalismus werden mentale Zustände als funktionale Zustände aufgefaßt, die wiederum als logische Zustände einer Turing-Maschine definiert werden. (Die Turing-Maschine ist eine idealisierte ab-

strakte Rechenmaschine, die den Begriff der Berechenbarkeit für Funktionen auf Mengen von diskreten Zeichenreihen präzisiert, d.h. Input wie Output sind jeweils als diskret vorausgesetzt. (Siehe Penrose, 1989, für eine nicht-technische Einführung.) Psychologische Eigenschaften, wie Motive, Bedürfnisse, Vorstellungen oder Erinnerungen, können demnach als funktionale Zustände niemals mit physikalischen oder physiologischen Zuständen identifiziert werden. Die These des Turing-Maschinen-Funktionalismus, daß mentale Zustände den logischen Zuständen einer Turing-Maschine entsprechen oder gar mit ihnen zu identifizieren seien, legt die Vorstellung nahe, daß sich Psychisches und Physisches zueinander verhalten wie Software und Hardware. Mechanismen (im Sinne einer Turing-Maschine) perzeptuell-kognitiver Informationsverarbeitung lassen sich folglich als eine - zumeist komplexe - Verarbeitung einer spezifischen Input-Information zu einer Output-Information auffassen. Die durch ein solches Input-Operationen-Output-System bestimmten Symbolfolgen oder Codes werden oft Repräsentationen genannt. Zudem wird angenommen, daß sich jeder komplexe Verarbeitungsprozeß in eine zeitliche Abfolge einfacherer Verarbeitungsprozesse zerlegen läßt. **Representation** und **Computation** bilden damit zentrale Begriffe dieses Paradigmas perzeptuell-kognitiver Informationsverarbeitung. So klar und fest die formalen Grundlagen, auf denen dieses Paradigma gründet, auch zu sein scheinen: Die Versuche einer Übertragung entsprechender Konzepte in die Psychologie sind begleitet von zahlreichen begrifflichen Kontroversen. Zum einen erweist sich die Unterscheidung dessen, was als Software und was als Hardware anzusehen ist, so unmißverständlich sie auf den ersten Blick scheinen mag, als fließend und nur relativ zu den jeweiligen Forschungsintentionen formulierbar (s. Kalke, 1969). Auch widersetzt sich bislang der häufig als semantisch charakterisierte Informationsbegriff, der sich auf Information **über etwas** bezieht und nicht lediglich ein quantitatives Maß der Unsicherheitsreduktion ist, einer präzisen Fassung. (Als Beispiele für die umfangreiche Literatur hierzu seien genannt Putnam, 1988, Penrose, 1989, sowie die Sammelbände Graubard, 1988, und Mohyeldin Said et al., 1990.)

Das Paradigma interner Informationsverarbeitung entfaltete sich innerhalb der Psychologie zunächst im Bereich jener Phänomene, die man als prototypisch für intelligentes Verhalten ansah, wie Denken, Problemlösen, Planen, Entscheiden, Konzeptbildung, Sprache und Gedächtnis. Hier legen unsere Intuitionen über Denkprozesse nahe, daß die beteiligten mentalen Prozesse sich in der Art einer kognitiven Informationsverarbeitung auffassen oder zumindest modellieren lassen. Die Leistungen hingegen, die unser Wahrnehmungssystem (ebenso wie das sensumotorische System) erbringt, sind einer solchen Introspektion nicht zugänglich und erscheinen uns einfacher und müheloser vorstatten zu gehen als kognitive Leistungen. Daß jedoch diese Intuition für die Komplexität der resultierenden Theoriebildung ein irreführender Hinweis



ist, zeigt sich beispielsweise darin, daß sich die subjektiv als schwierig empfundene Leistung intelligenten Schachspielens so viel leichter modellieren läßt als die als mühelos erlebte Leistung, Schatteninformation von Informationen über die Objektremission zu trennen. (Allgemein wird man natürlich die Schwierigkeit der Ausübung einer perzeptuell-kognitiven Tätigkeit nicht als Maßstab der Schwierigkeit ihrer theoretischen Modellierung verwenden wollen.) Diese Einschätzung könnte dazu beigetragen haben, daß in der kognitiven Psychologie eine Überbetonung des Kognitiven zu Lasten des Perzeptuellen in einem perzeptuell-kognitiven Gesamtzusammenhang zu erkennen ist. Indem die kognitive Psychologie die klassische Trennung von Wahrnehmungs- und Denkprozessen nicht nur beibehielt, sondern noch vertiefte, erschwerte sie lange Zeit die Einsicht, in wie komplexer Weise Wahrnehmungs- und Denkprozesse miteinander verwoben sein müssen (vgl. Shepard, 1984; Pylyshyn, 1984; Smolensky, 1988). Besonders dadurch, daß im Bereich der **Künstlichen Intelligenz** (KI) mit den Forschungen zur Erstellung künstlicher Wahrnehmungssysteme (**artificial perception**) deutlich wurde, daß den Leistungen des Wahrnehmungssystems hochgradig komplexe ‚intelligente‘ Prozesse zugrunde liegen müssen, wurde die Psychophysik, wenn auch in neuer Gestalt, zu einem zentralen Bereich der Kognitionswissenschaft.

Im Paradigma der perzeptuell-kognitiven Informationsverarbeitung wird die Wahrnehmung als ein Prozeß gesehen, der durch eine Abfolge von Verarbeitungsstufen gekennzeichnet ist, durch welche der proximale Reiz sukzessive in eine abstrakte und mit dem distalen Reiz verbundene interne Repräsentation transformiert wird. Dabei ist vor der Frage, wie etwas erreicht wird, die Frage zu stellen, was eigentlich erreicht werden soll. Die Aufgaben und Ziele, die mit einer bestimmten Wahrnehmungsleistung verbunden sind, müssen also zunächst bestimmt werden, bevor untersucht werden kann, auf welche Weise das Wahrnehmungssystem aus dem verfügbaren Input die jeweilige Leistung erbringen kann. Globale und funktionale Aspekte der Wahrnehmung rücken folglich in den Vordergrund. Am einflußreichsten wurde diese Auffassung in der Marrschen Variante (Marr, 1982): Aus der Sicht des Funktionalismus durchaus naheliegend, trennte Marr in idealisierter Weise drei Analyseebenen: **computational theory**, **algorithmic representation** und **hardware implementation**. Die Aufgaben der Psychophysik beziehen sich Marr zufolge vorrangig auf die Ebene der **computational theory**. In dieser sucht man für eine spezifische Leistung des Wahrnehmungssystems in präziser Weise Ziele und Strategien des Systems zu formulieren. In dem oben genannten Beispiel der Farbkonstanz besteht die Leistung des Wahrnehmungssystems darin, aus dem auf das Auge treffenden Licht den Effekt der Beleuchtung gleichsam wieder herauszurechnen und so den Farbeindruck stärker an die Remissionscharakteristik eines Objektes, d.h. gleichsam an seine ‚wirkliche‘ Farbe, zu binden. Aufgabe der Psychophysik ist neben einer Beschreibung des sensorischen Input

und einer Charakterisierung der erbrachten Leistung, relevante Beobachtungen bereitzustellen, aus denen sich interessante Randbedingungen für die Erstellung einer formalen **computational theory** ergeben. Die internen Größen derartiger Theorien sind zumeist, anders als in der skalenorientierten Tradition der Psychophysik, nicht direkt perzeptuell oder kognitiv faßbar; sie beziehen sich etwa auf Gradienten, Ortsfrequenzfilter, Autokorrelationsmuster oder Zerlegungen von Farb-Codes in solche für Lichteigenschaften und solche für Objektremissionen.

Auch wenn sich die Grenzziehung zwischen den drei Ebenen bereits konzeptuell als sehr viel weniger scharf erweist, als Marr dies postulierte (vgl. Kalke, 1969), und sie bei der Untersuchung des Wahrnehmungssystems umso mehr verschwimmen müssen, je komplexer und umfassender die betrachteten Wahrnehmungsleistungen werden, so hat diese Unterscheidung doch die konzeptuelle Klärung damit verbundener Probleme angeregt. Wichtiger noch ist jedoch aus Sicht der Psychophysik, daß die Erstellung von **computational theories** zu isolierten Wahrnehmungsleistungen etwa im Bereich der Bewegungswahrnehmung, der Farbkonstanz, der Texturwahrnehmung und des Stereosehens zu wichtigen neuen Einsichten in Wahrnehmungsprozesse und zugleich zur Entwicklung neuer Arten von Experimenten führte.

Mit der ‚computistischen‘ Wende erreicht die Psychophysik ein anderes Stadium der Theorieentwicklung: Sie differenziert sich entsprechend den verschiedenen Bereichen aus, und die scheinbare Einheitlichkeit der frühen psychophysikalischen Theoriebildung weicht einer Vielzahl spezieller, hoch ausdifferenzierter Theorien in eng umschriebenen Phänomenbereichen. Anders als die skalenorientierte Tradition der Psychophysik, deren Bedeutung für die komplexere Wahrnehmungspsychologie als eher marginal erachtet wurde, wird die neue, ‚computistische‘ Psychophysik zu einem wesentlichen Bestandteil der Kognitionswissenschaft. Zugleich behauptet sie nicht nur ihre Unabhängigkeit gegenüber der Neurophysiologie, sondern erhält in der Marrschen Unterscheidung dieser gegenüber eine epistemologische Priorität bei der Untersuchung der menschlichen Wahrnehmung. Daß der Psychophysik aus der Perspektive des Funktionalismus eine prinzipielle Bedeutung innerhalb der Kognitionswissenschaft zukommt, erklärt jedoch noch nicht, warum sie sich im Paradigma perzeptuell-kognitiver Informationsverarbeitung als so erfolgreich erweist. Dieses Paradigma erst bot einen theoretischen und methodologischen Rahmen, in dem sich alte Intuitionen einer komplexen und funktionalistischen Psychophysik, wie sie sich neben der skalenorientierten Tradition entwickelt hatten, fruchtbar machen ließen. (Der Begriff **Funktionalismus** ist mit vielfach verschiedenen und sich gelegentlich kreuzenden Bedeutungen belegt. So ist auch die funktionalistische Perspektive in der Psychophysik zunächst unabhängig von dem oben genannten Funktionalismus und bezieht sich darauf, das Wahrnehmungssystem von seinen adaptiven Aufgaben und

Funktionen im Wechselspiel Organismus-Umwelt her zu verstehen; nur in diesem Sinne sei das Attribut ‚funktionalistisch‘ im folgenden verwendet.) In der Sprache der Informationsverarbeitung zeigten sich die in der funktionalistischen Tradition bereitgestellten und vorher schwer zu präzisierenden Intuitionen plötzlich als machtvolle Heuristiken bei der Untersuchung der menschlichen Wahrnehmung.

Es bleiben dennoch einige Vorbehalte zu nennen: Im neuen Paradigma der Psychophysik wurde es beinahe selbstverständlich, sich das perzeptuell-kognitive System als ein informationsverarbeitendes zu denken und es in einer entsprechenden theoretischen Sprache zu behandeln. Doch anders als bei der Theoriebildung über Denkprozesse scheint sich in der Psychophysik der Schritt von einer Beschreibung und Simulation dieser Prozesse als Prozesse der Informationsverarbeitung zu der Auffassung, daß diese Prozesse tatsächlich Prozesse der Informationsverarbeitung **sind**, nur schwer rechtfertigen zu lassen. Das Wahrnehmungssystem findet nämlich keineswegs ‚Informationen‘ vor, sondern macht gleichsam aus einem Input erst Information. Dabei setzt der inhaltliche bzw. semantische Informationsbegriff für ‚Sender‘ und ‚Empfänger‘ eine Art gemeinsamen Gegenstandsbereich, d.h. etwas Vorstrukturiertes, voraus, über das informiert wird; das Wahrnehmungssystem konstruiert aber diesen Gegenstandsbereich (etwa Kanten, Flächen, Objekte) erst aus dem Input (und muß dazu gleichsam bestimmte ‚Vorannahmen‘ über diesen Gegenstandsbereich machen). Die Übertragung theoretischer Vorstellungen aus dem Denkbereich in den Wahrnehmungsbereich wird noch problematischer, wenn man Denkprozesse mit dem Begriff der Turing-Maschine in Zusammenhang bringt: Anders als im Bereich des sprachgebundenen Denkens, wo der Idee einer derartigen diskreten Symbolverarbeitung eine gewisse Plausibilität zuzukommen scheint, ist die Angemessenheit eines solchen Bildes in der Wahrnehmungspsychologie fragwürdig. Denn man wird kaum den Input in das Wahrnehmungssystem als diskrete Zeichenreihen betrachten wollen; auch spricht auf verschiedenen Ebenen des Wahrnehmungssystems einiges für die Annahme analoger Prozesse. Schließlich steht die Annahme stetiger Codes in der Psychophysik in einem gewissen, wenn auch nicht notwendigerweise unüberbrückbaren Spannungsverhältnis zu einer Turing-Maschinen-Konzeption der Wahrnehmung. Jedoch ist eine solche Konzeption nicht notwendiger Teil von **computational approaches** in der Psychophysik: Verfolgt man einen stetigen Ansatz, so ist für die Berechenbarkeit nur wichtig, daß es einen physikalischen Prozeß gibt, der etwas in gleichsam analogischer Weise realisiert.

Grundsätzlicher ist jedoch der Einwand, daß sich biologische Systeme, sei es die Leber, das Immunsystem oder das Wahrnehmungssystem, die sich im Verlaufe der Evolution physico-chemisch gerade so herausgebildet haben, daß sie spezifische Aufgaben erfüllen, nicht in einem tieferen Sinne als symbolische Repräsentationen in einem informationsverarbeitenden System charakterisie-

ren lassen, auch wenn sie auf theoretischer Ebene als solches Input-Output-System rekonstruierbar sind (vgl. Searle, 1984; Ramachandran, 1990, S. 23f.)

## **5. Von der Meßinstrumentkonzeption zu einer funktionalistischen Perspektive**

### 5.1 Das Problem der Reizbestimmung

In der Psychophysik wie auch in anderen Wissenschaften erweisen sich die am einfachsten erscheinenden Fragen oftmals als die hintergründigsten und schwierigsten. So ist es auch mit der Frage nach dem Reiz. Hängt die Definition des Reizes von der Bestimmung dessen ab, was als ‚Reizempfänger‘ angesehen wird? Verhält es sich nicht so, daß wir ohne Kenntnis der Funktionsweisen und Aufgaben des Organismus nicht wissen, was ein Reiz ist, und ohne Charakterisierung dessen, was einen Reiz konstituiert, Funktionen oder Aufgaben des Organismus nicht bestimmen können? Ist die Definition des Reizes somit verschieden, je nachdem, ob man am Gesamtorganismus, am Rezeptor, an einem bestimmten ‚Mechanismus‘ des Wahrnehmungssystems etc. interessiert ist? Ist ein Reiz in Termini der physikalischen Objektwelt zu charakterisieren oder in Termini der auf den Rezeptor treffenden Energie? Hängt die Definition des Reizes mit Aufgaben und Zielen des Wahrnehmungssystems zusammen oder mit Transduktionseigenschaften der Sinnesorgane? Ist das zweidimensionale retinale Bild der Reiz oder das dreidimensionale Objekt? Sind in der Tonpsychologie die Einzeltöne, die Melodie, der Akkord oder die zeitlichen Intervalle zwischen den Tönen der Reiz? Sind physikalische Eigenschaften, die vom Organismus nicht unterschieden werden können, verschiedene Reize oder nur ein Reiz, sind also metamere Lichter verschiedene oder identische Reize? Fragen dieser Art lassen sich ohne Mühe in verschiedenen Bereichen der Psychologie stellen (und Gibson, 1960, zeigt an einer Auswahl von Definitionen auf, wie groß die begriffliche Unklarheit hierüber tatsächlich ist). So erstaunt es nicht, wenn Stevens (1951, S. 31) konstatiert: „In a sense there is only one problem of psychophysics, namely, the definition of the stimulus.“

Die skalenorientierte Tradition der Psychophysik trivialisierte gewissermaßen durch die Meßinstrumentkonzeption der Wahrnehmung das Problem der Bestimmung des Reizes, indem sie einfache Größen der physikalischen Theorie, wie Schallenergie und Lichtintensität zugleich auch als Reize betrachtete, die einer psychophysikalischen Theoriebildung zugrunde zu legen seien. Warum aber sollten sich die Größen, die bei der Entwicklung einer **physikalischen** Theoriebildung als nützlich erwiesen, auch bei der Konstruktion von Theorien über das Wahrnehmungssystem als entscheidend erweisen? Sicherlich verspra-

chen sie, in der Sinnesphysiologie eine lohnende Ausgangsbasis für Transduktionsprozesse, d.h. Prozesse primärrezeptoraler Codierung zu sein. Die Psychophysik jedoch muß erst die physikalische Beschreibung der Situation, die dem Wahrnehmungssystem bestimmte Leistungen ermöglicht, schrittweise mit dem Verständnis dieser Leistung gewinnen. (Erst als mit der formalen Analyse des Problems der Farbkonstanz deutlich war, daß nicht einfach Punkt-für-Punkt die Wellenlängenzusammensetzung des zum Auge kommenden Lichtes den Reiz bilden konnte, bezog man komplexere physikalische Eigenschaften der Szene in die Definition des Reizes mit ein.) Nicht elementare Größen der Physik intern abzubilden wird hier als Aufgabe der Sinne gesehen, sondern eine angemessene Orientierung des Gesamtorganismus in seiner physikalischen Umgebung zu gewährleisten. Hier wird das elementaristische Mißverständnis der klassischen Psychophysik mit der Einsicht überwunden, daß physikalisch beschreibbar nicht heißt, elementaristisch, d.h. im Sinne der Meßinstrumentkonzeption beschreibbar zu sein.

## 5.2 Psychophysik aus physikalistischer vs. funktionalistischer Perspektive

Neben dem skalenorientierten Forschungsprogramm hatte sich in der Psychophysik ein weiteres entwickelt, in dem man **spezifische** Funktionsweisen und Leistungen des Wahrnehmungssystems zu verstehen und in geeigneten Theorien und Modellen zu erfassen suchte. Auch diese Tradition nahm ihren Ausgangspunkt bei Fechner, nämlich in seinen Arbeiten zur binokularen Integration von Helligkeiten, zu farbigen Schatten und Nachbildern sowie zu Gemeinsamkeiten des bildhaften Vorstellens mit dem Wahrnehmen. Doch band sich diese in gewisser Weise **mechanistische** Tradition, die sehr eng an den spezifischen Besonderheiten des jeweiligen Wahrnehmungsbereiches orientiert blieb, zunehmend an die Sinnesphysiologie an; zugleich grenzte die Psychologie aus einem neuroreduktionistischen Mißverständnis heraus große Teile der damit verbundenen Forschungsfragen aus der Psychologie aus. So entstand bereits in der Anfangszeit der Psychophysik zwischen einer physikalistisch und einer psychologisch orientierten Perspektive eine sich stetig vergrößernde Kluft. Durch ihre enge Anbindung an Physik und Physiologie konnte die physikalistische Tradition ihre Befunde in das naturwissenschaftliche Weltbild einbinden, der psychologischen Tradition einer komplexen Psychophysik fehlte indes eine geeignete ‚Sprache‘, die sie für eine Theoriebildung hätte heranziehen können und durch die sie ihre Einsichten hätte fruchtbar machen können. Erst das Paradigma perzeptuell-kognitiver Informationsverarbeitung bot eine solche ‚Sprache‘ an. In der nun möglichen Neuformulierung klassischer Fragen läßt sich das, was gegenüber den scheinbar härteren

Daten der Neurophysiologie zumeist als bloße qualitative Phänomenologie geringgeschätzt wurde, als oftmals entscheidender Baustein der jeweils spezifischen Theorie perzeptuell-kognitiver Informationsverarbeitung betrachten.

Mit dem zunehmenden Einfluß evolutionstheoretischer Betrachtungsweisen im vergangenen Jahrhundert gewann auch die funktionalistische Perspektive in der Psychophysik an Bedeutung. Deutlich trat sie bei Ernst Mach und Ewald Hering hervor und fand durch Karl Bühler, Fritz Heider und Egon Brunswik Eingang in die Wahrnehmungspsychologie. In der Gibbonschen Variante erfuhr sie schließlich ihre umfassendste Ausgestaltung, bevor sie in das Paradigma perzeptuell-kognitiver Informationsverarbeitung integriert wurde.

### 5.3 Internalisierte Regularitäten in der Wahrnehmung

Beim Versuch der Bestimmung des Reizes stand ein Problem im Vordergrund, das bereits in der Gestaltpsychologie im Zentrum stand und zudem in analoger Weise in der Erkenntnistheorie (vgl. Quine, 1969) eine große Rolle spielt: das Mißverhältnis zwischen dem mageren Input und der erbrachten Leistung. Betrachten wir als Beispiel folgende experimentelle Situation: Versieht man in einem vollständig abgedunkelten Raum ein Rad (dessen Konturen nicht sichtbar sind) in seinem Zentrum mit einem kleinen Lichtpunkt und läßt es an der Stirnseite des Raumes entlang rollen, so sieht der sich auf der gegenüberliegenden Raumseite befindende Beobachter einen sich auf einer Geraden bewegenden Lichtpunkt. Bringt man nun an der Peripherie des Rades einen weiteren Lichtpunkt an, so sieht der Beobachter nicht etwa neben dem eine Gerade beschreibenden Lichtpunkt einen weiteren, sich auf einer Epizykloiden bewegenden Lichtpunkt, sondern er sieht ein sich drehendes Rad. Der Organismus sieht hier also ‚mehr‘, als ihm physikalisch an visueller Stimulierung angeboten wird, nämlich statt der Überlagerung zweier sich bewegender Lichtpunkte ein sich drehendes Rad. Er sieht gleichsam etwas in die physikalische Situation hinein.

Wenn wir von ‚mehr sehen‘, ‚etwas hineinsehen‘ sprechen, so sind dies freilich nur aus der Sicht des Experimentators bequeme Sprechweise, da ihnen ein naives Bild unterliegt, das wir explizit machen müssen. Was bedeutet es denn, **nicht** ‚mehr‘ zu sehen? Der Vergleichsmaßstab dieses ‚Mehr-Sehens‘ ist ein fiktives veridikales und unmittelbares Sehen der physikalischen Situation. Nur wenn man sich dieses naiven Bildes bedient, ist es verwunderlich, daß der Organismus in einer bestimmten Situation ‚mehr‘ sieht. Die Verwunderung stellt sich erst ein, wenn das Sehen anders als erwartet ist. So ist das naive Bild hilfreich, weil es in dieser Situation das Selbstverständliche des Sehens in Frage stellt; es zeigt - und ohne die Physik vorauszusetzen, gäbe es keine Möglichkeit, sich überhaupt Gedanken hierüber zu machen -, daß das ver-

meintlich Einfache gar nicht einfach ist und daß wir auch im ‚Normalfall‘ nicht verstehen, wie wir wahrnehmen. Bleibt man indes auf die Täuschung fixiert, und somit dem naiven Bild verhaftet, so verstellt man sich den Blick für die Frage, was eigentlich Wahrnehmen im ‚Nicht-Täuschungsfall‘ bedeutet. Die Verwunderung über die ‚Täuschungen‘ der Wahrnehmung verbirgt, daß die Wahrnehmung im ‚Normalfall‘ in keiner Weise weniger verwunderlich und erklärungsbedürftig ist als im ‚Täuschungsfall‘. Man erinnere sich nur an die Tatsache, daß wir feinere räumliche Unterschiede aufzulösen vermögen, als es dem Durchmesser eines einzelnen Rezeptors entspricht (*sog. hyperacuity*): Würde man hier von einer Täuschung sprechen wollen?

Von einem funktionalistischen Standpunkt könnten wir formal von einer Underdeterminiertheit der Wahrnehmung durch die physikalische Reizsituation sprechen. Zu diesem theoretischen Aspekt lassen sich auf empirischer Seite Phänomene aus den verschiedenen Bereichen der Wahrnehmungspsychologie in Beziehung setzen, die auf oftmals drastische Weise eine solche Unterbestimmtheit vor Augen führen. Korrespondierend dazu zeigen die formalen Analysen im Bereich *artificial perception*, daß ein künstliches Sehsystem nicht allein durch einen aprioristisch-universellen *bottom up approach* erstellt werden kann, auch wenn dieser in überraschender Weise weiter führt als man erwartet hatte (Marr, 1982). Die mit der physikalischen Reizsituation konsistente Klasse möglicher Interpretationen muß durch implementierte Restriktionen reduziert werden, damit überhaupt eine angemessene ‚Interpretation‘ des Input erfolgen kann. Ein künstliches Sehsystem, das die Welt ‚sehen‘ soll, muß bereits - dies belegen die formalen Analysen sehr deutlich - von spezifischen physikalischen Regularitäten, die im Zusammenhang mit den jeweiligen Wahrnehmungsleistungen auftreten, Gebrauch machen.

Doch in welcher Weise kann der Organismus über eine solche verfügen? Was ermöglicht dem Organismus, sich auf der Basis eines scheinbar so mageren sensorischen Inputs so schnell, so sicher und angemessen in seiner Umwelt zurechtzufinden? Jeder Versuch, das Wesen der Wahrnehmung zu erfassen, somit Wahrnehmungspsychologie und Psychophysik zu betreiben, führt zurück auf diese grundlegende Frage. Die Intuition, daß da mehr sein muß, als sich ‚dem Auge enthüllt‘, finden wir folglich allenthalben in der Wahrnehmungspsychologie; sie ist, mal mehr, mal weniger explizit, in unterschiedlichster Weise ausgestaltet. Hierzu gehören Helmholtz’ „unbewußte Schlüsse“, die so „blitzschnell“, „genau“ und „sicher“ sind, daß sie mit den „schwerfälligen“ gedanklichen Schlußfolgerungen „nicht im Entferntesten verglichen werden können“; ihnen unterliegen, so Helmholtz, „angeborene, und durch die Organisation des Menschen selbst wesentlich bedingte Verknüpfungen von Vorstellungen“ (Helmholtz, 1855, 1894). In besonderer Weise hat Mach sich mit diesem Problem befaßt; es durchzieht viele seiner psychologischen Arbeiten. Mach hat sehr klar die entwicklungsgeschichtlichen Beziehungen zwi-

schen der Struktur der ‚Außenwelt‘ und der Struktur des Wahrnehmungssystems erfaßt: „Man könnte sagen die Netzhaut **schematisirt** und **karriert**. Die teleologische Bedeutung dieses Processes ist für sich klar. Er ist ein Analogon der Abstraction und der Begriffsbildung.“ (Mach, 1868, S. 19; siehe auch Mach, 1911/1985, S. 178). Auf das gleiche Problem bezogen finden wir bei Hering die Annahme von durch die Evolution in zweckmäßiger Weise angelegten „Wechselwirkungen im somatischen Sehfeld“. Hering verglich die Reaktionen des Organismus mit der Resonanz eines Klaviers, das „nicht bloß abhängt von den Schwingungen der Saiten, welche der Schlag direkt trifft, sondern auch von der Resonanz des ganzen Instrumentes“ (Hering, 1920, S.210). Entsprechend führe auch ein äußerer Reiz zu einer „Resonanz unseres ganzen Sensoriums“. Der Organismus hat sich im Kontakt mit seiner physikalischen Umwelt so entwickelt, daß sich ihm deren Gesetzmäßigkeiten eingeschrieben haben; daher reichen vergleichsweise minimale ‚Informationen‘, die ein ‚Mitschwingen‘ seiner ‚Kenntnisse‘ über die physikalische Welt hervorbringen, für eine angemessene Wahrnehmung.

Die Frage, was der Organismus der sensorischen Reizung bei der Wahrnehmung hinzufügt, ruckte bei den Gestaltpsychologen in den Mittelpunkt. Sie nahmen an, daß der Organismus durch angeborene neurale Strukturen global und in gestalthafter Weise auf einen Reiz reagiert. Der Aspekt der Koppelung des Organismus an seine Umwelt trat dabei jedoch zugunsten einer internen relationalen Analyse des Reizes in den Hintergrund. Die Gestaltpsychologie gleicht der klassischen Psychophysik darin, daß sie sich nur mit dem Organismus beschäftigte, nicht aber mit seiner Beziehung zur natürlichen Umwelt; so wurden zwar interne Organisationsweisen entdeckt, diese jedoch überwiegend auf der - wenn auch gegenüber der klassischen Psychophysik komplexer gewordenen - Reizebene beschrieben. (Diese Beschränkung hat sicherlich dazu beigetragen, daß die Gestaltpsychologie später zunehmend in Verruf geriet, doch trug hierzu auch eine oftmals nur noch deskriptive, atheoretische Orientierung und eine bisweilen dunkle Gestaltmetaphorik bei.)

Gegen die gestaltpsychologische Position wurde der ‚funktionalistische‘ Aspekt der Reize, ihre Bedeutung für den Organismus, insbesondere durch Bühler, Heider, Brunswik und Gibson hervorgehoben. Heider (1926) suchte in seiner wichtigen Arbeit **Ding und Medium** aufzuzeigen, in welcher Weise Wahrnehmungs- und Denkprozesse durch die Struktur der physikalischen Welt bestimmt sind. Er spricht von der „Umweltgemäßheit des Wahrnehmungssystems“, die „nur durch eine Berücksichtigung der weiteren Umwelt verständlich wird“ (Heider, 1930, S. 371ff.). „Dingkonstanz“ und die Beziehung von „Wahrnehmung und Gegenstandswelt“ bilden dann die Kernkonzepte für Brunswiks „probabilistischen Funktionalismus“. Jeder proximale Reiz - ein retinales Bild beispielsweise - ist demnach nur mit einer gewissen Wahrscheinlichkeit ein korrekter Indikator für einen bestimmten Aspekt der



„Außenwelt“. Der Organismus nutze daher die in einem komplexen Reiz vorhandenen Beziehungen und gewichte diese nach ihrer „ökologischen Validität“, d.h. nach ihrer Wahrscheinlichkeit, mit der sie für einen Aspekt der Außenwelt stehen. (Dies veranschaulicht Brunswik, 1934, S.97, durch sein „Linsenmodell“.) Das Wahrscheinlichkeitslernen, als dessen Resultat erst der Organismus von uneindeutigen Reizkonstellationen zu „ökologisch validen“ Wahrnehmungen gelangen kann, sei zu einem großen Teil im Verlaufe der Evolution erfolgt und dem Wahrnehmungssystem fest eingeschrieben. Dennoch wiesen Wahrnehmung und Denken eine große Ähnlichkeit auf. Brunswik nennt die Wahrnehmung „ratiomorph“, er sieht Wahrnehmung und Denken als „analoge Funktionen“, als „mit grundsätzlich gleichen Mitteln“ arbeitend an. „Wahrnehmung und Denken dienen beide der gleichen Aufgabe des Organismus: der Erkenntnis der Umwelt.“ (Brunswik, 1934, S.224; vgl. Shepard, 1982) Wo das Denken auf die individuelle oder sprachlich vermittelte Erfahrung angewiesen ist, hat die Wahrnehmung durch die Evolution des Organismus in seiner Umwelt „gelernt“; wo das Denken flexibel, doch langsam ist, ist die Wahrnehmung stereotyp, aber schnell. Hier erinnern wir uns wieder an Helmholtz' Beschreibungen, und tatsächlich könnte man Brunswiks Theorie als eine Ausformulierung der Helmholtzschen Theorie „unbewußter Schlüsse“ lesen.

Eine weitere Antwort auf die eingangs gestellte Frage gibt Gibson. Ihm zufolge liegen alle zur Orientierung des Organismus in seinem Umfeld benötigten Invarianzen im proximalen Reiz; sie sind keine Zutat des Organismus. Setzt man die beiden vagen, intuitiven Konzepte der ‚Invarianzen‘ und der ‚Gestalten‘ miteinander in Beziehung, so läßt sich Gibsons Auffassung als komplementär zur gestaltpsychologischen ansehen. Diese lokalisierte die Invarianzen vollständig im Organismus. Untersucht man, worin eigentlich der Unterschied zwischen den Sprechweisen „Die Invarianzen liegen im proximalen Reiz“ und „Der Organismus analysiert sie aus diesem heraus“ besteht, so wird deutlich, daß hier zwei Bilder aufeinanderstoßen, die in der Sache auf möglicherweise nicht so Verschiedenes hindeuten, wie sie suggerieren mögen. Im übrigen zeigen sich die Unterschiede in der theoretischen Perspektive auch hier wieder deutlich in der Auswahl des Reizmaterials. Die Gestaltpsychologen haben, das metaphorische Bild eines ‚Hinzufügens‘ betonend, das Reizmaterial so ausgewählt, daß der Effekt des ‚Hinzufügens‘ maximal wird, während er bei Gibsons Reizmaterial kaum auffällt. Auch bei Gibson bestimmt nicht das retinale Bild, sondern die „optische Situation“ (**Optical array**) die Wahrnehmung. In ihr seien Invarianzen enthalten, die nahezu perfekt mit dem distalen Objekt verbunden sind. Diese Invarianzen wurden durch einen aktiven Organismus erkundet und zur Orientierung benutzt; der Organismus zeige eine Resonanz auf die für ihn wichtigen Invarianzen. Wie der Organismus diese Invarianzen bestimmt und auswählt, bleibt in dieser Theorie eigen-

artig vage (Gibson betrachte, so Denett, 1984, das visuelle System als „wonder tissue“), ebenso die Konzepte des „information pick up“ und der „affordances“ (vgl. Ullman, 1980; Koenderink, 1980; Fodor & Pylyshyn 1981). In den letzten Jahren hat sich jedoch eine Reformulierung Gibsonscher Positionen im Paradigma perzeptueller Informationsverarbeitung als fruchtbar erwiesen (s. Marr, 1982; Shepard, 1984).

Die Intuition, daß die Wahrnehmung durch die physikalische Reizsituation gewissermaßen unterdeterminiert ist, durchzieht also die Geschichte der Psychophysik, und alle Versuche einer Antwort beruhen in der einen oder anderen Weise auf der Vorstellung, daß die fehlenden Bindeglieder durch interne Restriktionen des Organismus bereitgestellt werden (vgl. Sloman, 1983; Shepard, 1987). Aus dieser Perspektive läßt sich pointiert die Wahrnehmung als ein inputkontrolliertes Halluzinieren auffassen: Der sensorische Input gleicht einem „Stichwortgeber“, der für ein wesentlich durch interne Strukturen bestimmtes komplexes Geschehen eine stabile Anbindung an die biologisch relevante physikalische Umwelt garantiert. In diesem Zusammenspiel von Inputstruktur und interner Struktur für spezifische Wahrnehmungsleistungen die jeweiligen Anteile auszumachen ist Aufgabe der Psychophysik. Ihr Ziel, interne Restriktionen experimentell zu entdecken und ihre Struktur zu erhehlen, läßt sich am Beispiel der Arbeiten von Reuman & Hoffman (1986), Ramachandran (1990) und Anstis (1991) verdeutlichen.

In der KI-orientierten Wahrnehmungsforschung finden sich in diesem Zusammenhang entsprechende Intuitionen: Versucht man nämlich, die Leistungen der visuellen Wahrnehmung im Rahmen künstlicher Sehsysteme zu simulieren, so lassen sich der jeweilige physikalische Input und die spezifische erbrachte Leistung zumeist nur dann durch eine mathematische Theorie, eine **sog. computational theory**, in Beziehung setzen, wenn die Kluft zwischen Input und Output, die sich als Uneindeutigkeit der Lösung zeigt, durch empirisch angemessene Zusatzannahmen hinreichend reduziert werden kann. Sieht man in dieser Weise die Wahrnehmung metaphorisch als eine Art idealisiertes Rechenproblem, so stellt sich die Frage nach den Zusatzannahmen, durch die sich erst aus dem sensorischen Input gleichsam rückwärts (**inverse optics**) die einer Szenenrepräsentation zugrunde liegenden Codes bestimmen lassen (vgl. Poggio, 1990). Eine solche Auffassung der visuellen Wahrnehmung als „inverse Optik“ stellt eine komplexe, funktionalistisch gefärbte Variante der Meßinstrumentkonzeption der Wahrnehmung dar und wird kaum eine fruchtbare Heuristik für die wahrnehmungspsychologische Theorienbildung darstellen können. Beispielsweise bietet sie, wie jede Meßinstrumentkonzeption der Wahrnehmung, keinen Raum für eine natürliche Behandlung einer zentralen theoretischen (nicht jedoch phänomenalen) Eigenschaft der Wahrnehmung, die sich mit ‚Vagheit‘ oder ‚Unbestimmtheit‘ nur unzureichend umschreiben läßt. Es ist wenig plausibel, als Basis des Wahrnehmungseindrucks ein kom-

plexes Gewebe von **präzisen** Codes und **eindeutigen** Szenendesignatoren anzunehmen, auf denen dann ‚auf höherer Ebene‘ kognitive Aufmerksamkeits- und Entscheidungsprozesse operieren. Schon für frühe Stufen der internen Codierung finden sich Hinweise, daß sich, metaphorisch gesprochen, das Wahrnehmungssystem ‚fast überall‘, d.h. mit Ausnahme eines kleinen Szenenausschnitts, in seinen Interpretationen soweit eben möglich nicht festlegt (selbst im Fokus der Aufmerksamkeit kann sich eine solche Vagheit erhalten, wie etwa eine Betrachtung von Objektfarben unter chromatischer Beleuchtung zeigt). Dadurch kann es seine Interpretation schnell, flexibel und angemessen an neue Evidenzen anpassen, wie sie durch Szenenänderungen bereitgestellt oder durch aktives Suchen gewonnen werden. Ein solcher Schwebeszustand muß jedoch in dem Sinne optimiert werden, daß eine Art Voreinstellung in Richtung der momentan jeweils ‚wahrscheinlichsten‘ Interpretation erfolgt; die Bandbreite möglicher Interpretationen wird dabei, so ist anzunehmen, durch weitere Restriktionen eingeschränkt, wie sie sich etwa aus der Forderung einer globalen Konsistenz der Szeneninterpretation sowie aus verschiedenen Arten von Stetigkeitsannahmen ergeben. Daß sich Konzepte dieser Art der theoretischen Aufmerksamkeit so lange entziehen konnten, scheint auf phänomenaler Ebene mit der Geschlossenheit und Stabilität des Wahrnehmungseindrucks, auf theoretischer Ebene mit der Schwierigkeit einer angemessenen Präzisierung derartiger Intuitionen zusammenzuhängen (zur Vagheit siehe jedoch James, 1890, S. 254).

Ungeachtet solcher Vorbehalte ergibt sich zwischen formalen Analysen KI-orientierter Wahrnehmungspsychologie und der Psychophysik ein interessantes Wechselspiel: Die Psychophysik kann zum einen die psychologische Angemessenheit einer solchermaßen erstellten Theorie untersuchen und wird oftmals durch diese Analysen zu neuen Arten von Experimenten angeregt. Auf der anderen Seite bilden die experimentellen Befunde der Psychophysik die Grundlage, auf der sich erst Ideen über die Art jener spezifischen Bedingungen gewinnen lassen, die dem Wahrnehmungssystem seine besonderen Leistungen ermöglichen.

## 5.4 Was ist ein Reiz?

In der genannten Beschreibung stellen sich also die Ziele einer funktionalistisch orientierten Psychophysik als ganz verschieden von denen einer skalenorientierten Psychophysik dar: Anders als diese gründet jene ihre Theoriebildung über eine Wahrnehmungsleistung zunächst auf einer Aufgabenanalyse der betrachteten Wahrnehmungsfunktion. Wie der physikalische Reiz zu charakterisieren ist, auf Grund dessen eine bestimmte Wahrnehmungsleistung erfolgt, ist nicht von vornherein vollständig festgelegt, sondern wird erst durch

experimentelle und theoretische Untersuchung des Zusammenwirkens von verfügbarem Input und internalisierten Restriktionen sukzessive bestimmt. Steht die Reizdefinition in der skalenorientierten Psychophysik am Anfang der Untersuchung, so bildet sie in der funktionalistischen Psychophysik geradezu den Abschluß der Theoriebildung.

Die Definition des Reizes ist somit theorieabhängig. Auch bei gleichem Reizmaterial kann sie je nach theoretischer Perspektive ganz verschieden ausfallen; so lassen sich etwa die bekannten Balkenmuster (spatial frequency gratings) einerseits im Rahmen eines merkmalsorientierten Modelles durch eine Kombination von Ecken, Linien und Kanten beschreiben oder aber im Rahmen eines Filter-Modelles durch die Ortsfrequenzkomponenten ihrer Fourier-Zerlegung. Im allgemeinen können wir sagen, daß die Neurophysiologie primär an einer Beschreibung des Reizes in den üblichen physikalischen Dimensionen interessiert ist - worin ihr die klassische Psychophysik gleicht -, während die funktionalistische Psychophysik eine solche Beschreibung von den ‚intentionalen‘, aufgabenbezogenen Aspekten her versucht. (Der Begriff des Reizes, der in seiner Verwandtschaft mit dem der Reizung noch seine sinnesphysiologische Herkunft erkennen läßt, wandelt sich in der funktionalistischen Perspektive zu dem des Input.) Eine aufgabenbezogene Reizbeschreibung kann natürlich wieder in eine in weiterem Sinne physikalische, d.h. nicht mit symbolischen, rein kognitiven Prozessen zusammenhängende Charakterisierung münden, sofern die mit bestimmten Zielen und Aufgaben des Wahrnehmungssystems verbundenen Objekte und Prozesse der Umwelt entsprechend der Abstraktionsleistung des Wahrnehmungssystems durch ihre Invarianten beschreibbar sind.

Aus dieser Perspektive erfahren auch die Kontroversen um ökologische vs. reduzierte Reize eine neue Bewertung. Das Attribut ‚ökologisch‘ läßt sich nämlich nicht nach der Plausibilität des Augenscheines den Reizen zuschreiben, sondern allein relativ zu einer Theorie über die betrachtete Leistung. Reize, die jene physikalischen Invarianten enthalten, die für die spezifische Leistung eines Wahrnehmungsmechanismus vorhanden sein müssen, können dem vor-theoretischen Augenschein nach überaus reduziert erscheinen, bezogen auf die Anforderungen des Mechanismus jedoch gleichsam ökologisch sein. Dies wird besonders deutlich, wenn man ethologische und neuroethologische Aspekte mit einbezieht, da sich aus dieser Perspektive ein dem Konzept des Schlüsselreizes analoges Konzept formulieren läßt (Mausfeld, 1993). Dieses bezieht sich aber nicht mehr, wie das des Schlüsselreizes, auf ein vom Gesamtorganismus erbrachtes Verhalten, sondern auf das Verhalten eines (idealisierten) Mechanismus. Nicht, wie bei Gibson, möglichst reichhaltige ökologische Reize konstituieren dann das Reizmaterial, sondern reduzierte Reize, die jedoch gerade so reichhaltig sind, daß sie die physikalischen Invarianten enthalten, die den betrachteten Mechanismus, etwa der Farbkonstanz (vgl. Mausfeld & Niederée, 1993), zu der ihm eigenen Leistung anregen.

Die Extraktion der für die Anpassung an die Umwelt geeigneten physikalischen Invarianten kann man aus evolutionstheoretischer Perspektive als das eigentliche Ziel des Wahrnehmungssystems sehen. So natürlich eine solche Auffassung auch erscheinen mag, so sind doch die hinter dem Begriff der Invarianten verborgenen Intuitionen nur schwer präzisierbar. Zwar läßt sich dieser Begriff rein mathematisch bestimmen als das, was unter einer bestimmten Gruppe von Transformationen invariant bleibt, womit die Invariante gleichsam ein abstraktes Objekt konstituiert; doch verlagert dies in der Psychophysik das Problem lediglich auf die Frage, was geeignete Transformationen sind. Wir wollen dennoch versuchen, den Invariantenbegriff zu erhellen, indem wir die Äquivalenzklassenbildung perzeptueller Codes betrachten. Haben wir nämlich geleitet durch eine Aufgabenanalyse eines perzeptuellen Mechanismus eine Idee über seine mögliche Funktionsweise und dieser postulierten Funktionsweise entsprechend eine vorläufige Präzisierung des Reizes gewonnen, so haben wir zugleich festgelegt, was äquivalente Reize sind. Die formale Bestimmung des Reizes führt uns zu dem wichtigen Konzept der Äquivalenzklassenbildung durch das Wahrnehmungssystem: Unterschiedliche physikalische Reize können nämlich bezogen auf den erbrachten Output eines Mechanismus denselben perzeptuellen Effekt haben. Bekanntestes Beispiel hierfür ist die Metamerie in der Farbwahrnehmung, wo physikalisch unterschiedliche Energieverteilungen über das Spektrum zu perzeptuell ununterscheidbaren, d.h. metameren Farben führen können. In anderen Fällen, etwa der Bewegungswahrnehmung, kann dies nicht mehr in so elementarer Weise explizit gemacht werden, da hier die Beschreibung des physikalischen Reizes ungleich schwieriger ist. Doch bezieht sich auch in diesen wie in anderen Bereichen eine wichtige Klasse von Experimenten auf die Frage, über welche interessanten physikalischen Transformationen des Inputs der Output konstant bleibt (als Beispiel siehe Ratliff & Sivovich, 1978). Dieser Gedanke läßt sich in natürlicher Weise durch den von Krantz (Suppes, Luce, Krantz, Tversky, 1989, S.257) entwickelten Begriff des perzeptuellen Codes präzisieren. Vereinfacht gesagt ist ein solcher Code eine reellwertige Funktion auf den physikalischen Reizen mit der Eigenschaft, daß zwei Funktionswerte dann gleich sind, wenn die untersuchten perzeptuellen Attribute gleich sind. Da sich eine Äquivalenzklassenbildung auf jeder Ebene des Wahrnehmungssystem findet, ist es insbesondere auf höherer Ebene zweckmäßig, die Äquivalenzklassenbildung statt auf physikalischer Ebene auf der Ebene der internen Codes zu beschreiben. Beispielsweise läßt sich das sog. Urgelb, d.h. ein Gelb, das weder rötlich noch grünlich erscheint, physikalisch als Äquivalenzklasse von Energieverteilungen beschreiben oder auf der Ebene von Codes durch den Nullpunkt des sich durch eine Linearkombination von Grassmann-Codes ergebenden Rot-Grün-Codes.

Setzt man auf diese Weise das aus einer funktionalistischen Betrachtungsweise wichtige Konzept der Invariantenbildung mit dem Konzept eines abstrakten

perzeptuellen Codes in Verbindung, so läßt sich das Ziel psychophysikalischer Theoriebildung als die Bestimmung und Untersuchung der für bestimmte Wahrnehmungsleistungen verantwortlichen perzeptuellen Codes und ihrer Beziehungen charakterisieren. Eine solche Betrachtungsweise fügt sich in die neurophysiologische Auffassung (e.g. Barlow, 1985, 1992), der zufolge der Cortex gleichsam durch ein allgemeines ‚Datenformat‘ charakterisiert ist, durch das erst eine intermodale sensumotorische Integration, Aufmerksamkeitssteuerung, Wissensrepräsentation, etc. gewährleistet wird. Statt der zumeist eindimensionalen Codes der klassischen Psychophysik werden bei der Analyse derartiger Leistungen des perzeptuell-kognitiven Systems ‚höherstufige‘ Codes mit einer hinreichend reichhaltigen internen Struktur im Vordergrund der Theoriebildung stehen (vgl. v.d. Malsburg, 1986).

## 5.5 Die Psychophysik in ihrem Verhältnis zur Neurophysiologie und subjektiven Sinnesphysiologie

Psychophysik und Sinnesphysiologie haben historisch, mit Fechner, Helmholtz, Mach, Hering u. a., einen gemeinsamen Ausgangspunkt. Während sich jedoch die Sinnesphysiologie als biologische Disziplin mit den sich ausdifferenzierenden Methoden der Anatomie und Neurophysiologie weitgehend kumulativ entwickelte - freilich unter zunehmender Ausgrenzung von Problemen komplexerer Wahrnehmung, wie sie noch von Helmholtz und Hering untersucht worden waren -, unterlagen Psychophysik und Wahrnehmungspsychologie vielfältigen Paradigmenwechseln mit sich jeweils ändernden Auffassungen darüber, was eigentlich den Gegenstandsbereich und die Grundfragen einer Psychologie der Wahrnehmung ausmache. Die eigenständige Berechtigung einer Wahrnehmungspsychologie gegenüber einer sich rasch entwickelnden Sinnesphysiologie wurde vielfach auf der Grundlage eines kranken Neuroreduktionismus bezweifelt; in der Folge zog sich die Psychologie lange Zeit weitgehend aus der Wahrnehmungspsychologie zurück.

Die grundsätzliche logische Lücke, die zwischen dem Verhalten eines Neurons und einer - dazu in Beziehung gesetzten - Leistung des Organismus klafft, gestaltet die Beziehung von Neurophysiologie und Psychophysik sowie allgemeiner das Problem des Neuroreduktionismus äußerst subtil. Was nämlich lehren Eigenschaften des neuralen Substrates über die Funktion und in welcher Weise läßt sich umgekehrt aus Eigenschaften der Funktion auf solche des neuralen Substrates schließen? Welche perzeptuelle Bedeutung soll man einzelnen Zellen beimessen, die unter bestimmten Bedingungen eine Spezifität für Helligkeit, Farbe, Disparität, räumliche oder zeitliche Frequenzen etc. zeigen? Der Wunsch, neurophysiologischen Befunden auf jeder Ebene des Wahrnehmungssystems eine perzeptuelle Bedeutung zu geben, scheint auch damit

zusammenzuhängen, daß man hofft, dadurch umgekehrt für entsprechende psychophysikalische Befunde eine ‚Erklärung‘ verfügbar zu haben. Darin zeigt sich eine Variante des in Abschnitt 6.2 behandelten mechanistischen Erklärungsbegriffes: Ein Wahrnehmungsphänomen gilt erst dann als verstanden, wenn man glaubt, die für dieses Phänomen verantwortlichen neuronalen Mechanismen gefunden zu haben. Diese oder ähnliche Haltungen finden sich in Varianten des Neuroreduktionismus; als Beispiel sei Churchland (1986) angeführt. Fragen eines Neuroreduktionismus bilden ein wichtiges metatheoretisches Problemfeld der Psychophysik, das sie mit der Psychologie teilt, in der diese Fragen oftmals versteckter auftreten. Vor ihrer Erörterung stellen sich Fragen wie: Ist ein metaphysischer Neuroreduktionismus gemeint, im Sinne einer letztendlichen Reduktion der Psychophysik auf die Neurophysiologie? Oder die These der Reduzierbarkeit auf eine Neurophysiologie, wie sie sich beim gegenwärtigen Stand ihrer Theoriebildung zeigt? (Die Geschichte der Gestaltpsychologie zeigt die Gefahren, die eine zu frühzeitige Orientierung psychophysikalischer Forschung an der zeitgenössischen Neurophysiologie birgt.) Wird für die Neurophysiologie wiederum eine Reduzierbarkeit auf die Physik behauptet, und wurde sich daraus auch ein physikalistischer Reduktionismus für die Psychophysik ergeben? In welchem Sinne will man jeweils Reduzierbarkeit verstehen?

Neuroreduktionistischen Perspektiven steht entgegen, daß aus prinzipiellen Gründen wenig Hoffnung besteht, die durch die Neurophysiologie bereitgestellten Kenntnisse über die lokale Codierung könnten zu einem tiefergehenden Verständnis der globalen Leistungen und der funktionalen Aspekte des Wahrnehmungssystems führen (e.g. Putnam, 1973). Anti-reduktionistischen Perspektiven zufolge können neurophysiologische Daten keine ‚Fundierung‘ psychophysikalischer Theorien leisten, sondern sie können diese ergänzen und somit zu umfassenderen Modellen biologischer Informationsverarbeitung führen, wobei aber die psychophysikalische Theorie als solche unabhängig von einer spezifischen hardware-Realisierung ist. Als Beispiel für eine solche Haltung sei Fodor (1974) angeführt: „There are no firm data for any but the grossest correspondence between types of psychological states and types of neurophysiological states, and it is entirely possible that the nervous system of higher organisms characteristically achieves a given psychological end by a variety of neurophysiological means. If so then the attempt to pair neurophysiological structures with psychological functions is foredoomed. Physiological psychologist of the stature of Karl Lashley have held precisely this view“ (Fodor, 1974, S.105). Nach einer solchen Auffassung können die entscheidenden ‚Strategien‘ und die ‚Logik‘ des Wahrnehmungssystems nur auf physico-phänomenologischer Ebene, d.h. auf der Ebene der Psychophysik studiert werden.

Unabhängig von den mit neuroreduktionistischen Programmen verbundenen metatheoretischen Kontroversen, die wesentlich mit unterschiedlichen Auf-

fassungen über den Erklärungs begriff verbunden sind, geht wissenschaftsgeschichtlich betrachtet die Psychophysik der Neurophysiologie voraus: Die Neurophysiologie der Wahrnehmung bedarf der Psychophysik als eines Leitfadens ihrer Forschungsorientierung, worauf gerade Neurophysiologen vielfach hingewiesen haben (e. g. Barlow, 1983, S. 11). Psychophysikalische Einsichten und Befunde zu globalen Wahrnehmungsleistungen haben der Neurophysiologie seit jeher in entscheidender Weise als Forschungsheuristiken gedient.

## **6. Theorien in der Psychophysik**

### 6.1 Metaphern, Heuristiken und Metaprinzipien der Theoriebildung

Glauht man, beim Prozeß wissenschaftlicher Theoriebildung streng zwischen ideengeschichtlichen Ursprüngen und psychologischer Entstehung einerseits und logischer Rekonstruktion andererseits eine klare Trennung ziehen zu können, so wird man Metaphern, ebenso wie Heuristiken, in den exakten Wissenschaften lediglich dem Entstehungszusammenhang wissenschaftlicher Theorien zuordnen, keineswegs jedoch ihrem Begründungszusammenhang. In einer reifen Wissenschaft dienen Metaphern dann lediglich dem Zwecke didaktischer Veranschaulichung. Daß jedoch eine solche Trennung sehr viel weniger klar ist, als man sie sich in dem Bemühen um eine Rechtfertigung und Fundierung wissenschaftlicher Tätigkeit erhoffte, hat die jüngere Wissenschaftstheorie vielfach deutlich gemacht. Interessiert man sich nämlich weniger für die Entwicklung von aprioristischen Vorschriften und Normen wissenschaftlicher Tätigkeit, sondern lenkt den Blick auf die komplexen dynamischen Prozesse, aus denen heraus erfolgreiche wissenschaftliche Theorien entstanden sind, so erscheint die Untersuchung der Rolle von Metaphern in diesem Prozeß mehr zu sein als ein psychologischer Neben aspekt.

Ohne Metaphern scheint Sprache geradezu undenkbar zu sein (nur innerhalb eines metaphorisch als formale ‚Sprache‘ charakterisierten Kalküls werden Metaphern zwangsläufig eliminiert). Das vorwissenschaftliche, mythische Denken wird wesentlich durch Gleichnisse und Bilder bestimmt, und in dem Maße, wie mit dem wissenschaftlichen Denken Sprache als Mittel der Erkenntnis betrachtet wird, bildet sie das Vehikel, durch das Unterschiedliches verglichen und nach Ähnlichkeit gebündelt wird. Die sprachliche Form des Vergleichens erlaubt erst die wissenschaftliche Begriffsbildung. Sie führt beispielsweise zu Gattungs- und Artbegriffen und läßt so das Spezifische unter dem Schema des Typischen sehen. Im Fortgang der Wissenschaften mußte sich die Sprache auf Bereiche ausdehnen, die sich dem Denken neu erschlossen und für die ein



angemessenen Vokabular nicht bereitstand. (In der Physik ist in Begriffen wie Kraft, Widerstand, Welle oder Naturgesetz die metaphorische Ausdehnung der Sprache deutlich zu erkennen.) In der Übertragung der Sprache auf neue Bereiche stiftet sie dort neuartige Beziehungen und vermag so, den Bereich dessen, worüber sie redet, wiederum zu erweitern. Dieser Prozeß läßt sie selbst jedoch nicht unverändert: Sie paßt sich gewissermaßen den erschlossenen Bereichen an. Die Übertragung eines als klar empfundenen Begriffs oder Konzeptes auf etwas Unklares, Unfertiges ist die Quelle wissenschaftlicher Theorieentwicklung. In den Möglichkeiten einer solchen Übertragung, d.h. in ihren Metaphern liegt somit die Reichhaltigkeit und das schöpferische Potential einer wissenschaftlichen Sprache. Gleichzeitig stellt die Existenz ebendieser Metaphern aber auch eine Schwäche und Gefährdung der theoretischen Sprechweise dar. Metaphern vermögen Leitbild und Irbild zugleich zu sein, worüber im vorhinein zumeist schwerlich zu urteilen ist.

Metaphern treten also in den Wissenschaften in ganz unterschiedlichem Gewande auf (s. Mac Cormac, 1976, sowie den Sammelband von Ortony, 1979), und ihre Bedeutung für eine Theorie kann von einer psychologisch-heuristischen Beiläufigkeit bis zu einer theoriekonstitutiven Kernidee reichen. Der letzte Fall ist von besonderem Interesse, da er einer Auffassung, die bis auf Aristoteles zurückgeht, widerspricht, daß nämlich eine Metapher eine Benennung ist, die eigentlich etwas anderes bedeutet und durch wörtlichen Gebrauch des eigentlich Gemeinten auflösbar ist. Wie aber steht es mit metaphorischen Sprechweisen wie der Überzeugung Platons, das Buch der Natur sei in mathematischen Lettern geschrieben, oder Vorstellungen, wie sie einer Geometrisierung und Mechanisierung des Weltbildes unterliegen? Wie steht es mit der Vorstellung, der menschliche Geist sei eine Turing-Maschine, ein informationsverarbeitendes System oder dergleichen? Was wurde es in solchen Fällen bedeuten, metaphorische Sprechweisen in wörtlich Gemeintes zu übersetzen? Offensichtlich stoßen wir hier auf Metaphern, die so fest mit einer Theorieperspektive verwoben sind, daß sie diese geradezu konstituieren und nicht durch eine nicht-metaphorische Sprechweise substituiert werden können. Auf solche Metaphern werden wir im folgenden das Augenmerk richten, ohne dabei zwischen der metaphorischen und der analogen Art des Vergleichens eine scharfe Grenze zu ziehen.

Die älteste Metapher der Psychophysik entstammt der Alltagssprache: das Konzept der Schwelle. Durch die Leibnizsche Trennung von *perceptio* und *apperceptio* vorbereitet wurde der Schwellenbegriff von Herbart in die Psychologie eingeführt: verstanden als Schwelle, die Bewußtes und Unterbewußtes trennt. Bei Fechner wandelte sich dieser Begriff; die Schwelle, bestimmt als der „Punct, wo die Mercklichkeit eines Reizes oder eines Reizunterschiedes beginnt und schwindet“ (Fechner, 1860, I, S.238), wurde einer empirischen Bestimmung zugänglich gemacht. Über diese Metapher hinaus ist die Psycho-

physik selbst historisch geradezu aus einer Metapher hervorgegangen, steht doch Fechners Übertragung des Bernoullischen Nutzenbegriffs auf Empfindungsskalen, die zudem in Analogie zu physikalischen Skalen gedacht sind, an ihrem Anfang.

Ebenfalls sehr alt ist eine Metapher, die ihrerseits ihren Ursprung in der Psychologie hat. Sie drückt sich in dem Versuch aus, die Wahrnehmung in Konzepten und Attributen des Denkens zu beschreiben. Mit seinem Konzept „unbewußter Schlüsse“ hat Helmholtz als erster in expliziter Weise denkpsychologische Begriffsweisen auf die Wahrnehmung übertragen. Von ihr als einem ratiomorphen Vorgang, einem Hypothesentesten oder einem Inferenzprozeß zu reden, ist seitdem innerhalb mancher Theorievorstellungen selbstverständlich geworden und wird durch das Paradigma der Informationsverarbeitung in besonderer Weise nahegelegt. Wir sind auf diese Metapher in Abschnitt 5.3 ausführlich zu sprechen gekommen.

Eine weitere sehr alte Metapher hängt mit Vorstellungen einer Mechanisierung des Geistes zusammen. Eine ideengeschichtliche Entwicklung, die mit Leibniz, La Mettrie und Hobbes ihren Ausgang nahm und im gegenwärtigen Paradigma der Denk- und Wahrnehmungspsychologie, sei es kognitivistisch oder konnektionistisch, kulminiert, bildet den Hintergrund für eine der einflußreichsten Metaphern in der Geschichte der Psychologie: die Metapher des Geistes als eines Computers oder eines informationsverarbeitenden Systems. Mit dem Begriff des Mechanismus ist vor allem die Vorstellung von etwas Regelmäßigem verbunden, das konkrete Mechanismen ebenso einschließt wie einen Kalkül. Für Leibniz bedurfte jedes Denken notwendigerweise der Zeichen und war somit als symbolischer Prozeß in einem normativen Sinne durch einen Kalkül charakterisierbar (vgl. Webb, 1980; Krämer, 1991). Die materialistische Lesart eines gesetzmäßigen Wirkens des Geistes verkörperte die Auffassung von La Mettrie, der im Verlaufe seiner Untersuchungen zur Frage, ob der Mensch eine Maschine sei, über die Natur der Denkvorgänge bemerkte: „Ich halte das Denken für so wenig unvereinbar mit der organisierten Materie, daß es mir geradezu eine ihrer Eigenschaften zu sein scheint“ (La Mettrie, 1748/1988, S.87). So liegt der Gedanke nicht fern, den abstrakten Kalkül des Denkens material zu realisieren zu suchen: Seit dem 17. Jahrhundert wird das menschliche Denken von dem Wunsch beherrscht, sich selbst in Form von Denkmaschinen technisch zu reproduzieren (zur Ideengeschichte siehe Vartanian, 1953, insbes. S. 203ff.).

Sieht man in dieser Weise den Geist als eine Funktion der organisierten Materie an, so liegt es nahe, sich seine Funktionsweise in Analogie zu bereits Bekanntem vorzustellen. Wie das Herz einer Pumpe gleicht und die Niere einem Filter, so gleiche der menschliche Geist: einem Uhrwerk, einem durch Triebe bestimmten hydrodynamischen System, einem Relaisystem, einem

Nachrichtenübermittlungssystem, einem kybernetischen Regelsystem, je nach Verfügbarkeit technischer Analogien. Das in der Natur Vorgefundene erscheint uns verständlicher, wenn wir es mit etwas von uns Erstelltem vergleichen. Doch während in der Tat der Vergleich des Herzens mit einer Pumpe den wesentlichen Mechanismus beschreibt, scheint das perzeptuell-kognitive System so einzigartig zu sein, daß uns passendere technische Bilder fehlen.

Heute wird vielfach angenommen, daß mit der Verfügbarkeit des Konzepts der Turing-Maschine und dem Aufkommen des Funktionalismus ein neues Bild für die Funktionsweise des Geistes bereit steht. In einem folgenschweren Abstraktionsschritt wurde nun die ‚Mechanik des Geistes‘ nicht mehr material, sondern funktional bestimmt: Das perzeptuell-kognitive System wurde als informationsverarbeitendes System aufgefaßt, und man spricht von Inputs, die sich in einer zeitlichen Abfolge von Verarbeitungsschritten durch Operationen der Symbolmanipulation zu jeweiligen Outputs transformieren. Der Entwurf dieses neuen Bildes über die Arbeitsweise des menschlichen Geistes geht zwar der eigentlichen Konstruktion von digitalen Computern voraus (Craik, 1943), Doch mußte allein die Existenz des Computers eine vollständig andersartige Theorie geistiger Prozesse nahelegen. Man begann nun, in mentalistischer Weise über interne Verarbeitungsprozesse von Computern zu reden und ihnen interne Zustände zuzuschreiben: die Alltagssprache wirkt hier metaphorisch auf einen technischen Bereich zurück. Dies könnte zu der Annahme verführen, daß sich durch diese Verknüpfung auch den entsprechenden Sprechweisen in Theorien über Wahrnehmungs- und Denkprozesse ein gesichertes Fundament geben ließe. Begriffe wie ‚Denken‘ und ‚Sehen‘ erhalten aber ihre Bedeutung erst durch Bezüge zu anderen psychologischen Ausdrücken und Prädikaten, welche sich nicht zuletzt auch auf Denk- und Wahrnehmungserfahrungen beziehen. Keinesfalls wurde man bei einem Fotoapparat von ‚Sehen‘ sprechen wollen, und auch bei einem Roboterauge, das den Sensorinput in ‚intelligenter‘ Weise weiterverarbeitet, kann man nicht anders als in metaphorischen Sinne von ‚Sehen‘ reden.

Die einer solchen Perspektive der Informationsverarbeitung zugrunde liegende Metapher ist als theoriekonstitutive Metapher (Boyd, 1979) so fest mit dem theoretischen Gegenstand, auf den sie zielt, verwoben, daß sie nicht durch eine nicht-metaphorische Sprechweise ersetzt werden kann. Um ihre Bedeutung und ihren Status ranken sich einige der bedeutendsten metatheoretischen Kontroversen der Kognitionswissenschaft (siehe etwa Lucas, 1961; Searle, 1984; Putnam, 1988; Penrose, 1989). Während einige den metaphorischen Charakter der genannten Sprechweisen gänzlich bestreiten und sie als eine testbare empirische Hypothese zur Natur geistiger Prozesse (etwa Pylyshyn, 1984, S. 55, 1989) verstehen, sehen andere in ihnen lediglich eine Art Beschreibungsinstrument, das neue Heuristiken bereitstellt und so die Forschung sti-

muliert (vgl. Daugman, 1990); auch können bestimmte Metaphern eine Metamorphose von Instrumenten zu Theorien erfahren (vgl. Gigerenzer, 1988).

Es ist kaum verwunderlich, daß die Metapher perzeptueller Informationsverarbeitung, in all ihren unterschiedlichen Varianten, die gegenwärtig bedeutendste und umfassendste Metapher der Theoriebildung in der Psychophysik ist, denn: „In der Wahrnehmungsforschung spiegelt sich die zeitgenössische Technik“ (Kohler, 1957). Wie kein Leitbild zuvor hat sie neue Forschungsstrategien eröffnet, sich ausdifferenziert und expliziert und selbst auf den Bereich, dem sie entstammt, zurückgewirkt. Sie hat, der Aufgabe einer Metapher entsprechend, unsere theoretische Sprache zur Erfassung neuer Phänomenbereiche erweitert und vielleicht, aus der Sicht eines wissenschaftlichen Realismus, zu einer „accomodation of language to the causal structure of the world“ (Boyd, 1979, S. 358) beigetragen.

Einen ebenso mächtigen Einfluß entfaltet eine der Evolutionsbiologie entlehnte Perspektive: die adaptationistische bzw. evolutionistische Metapher. Dieser Metapher zufolge lassen sich bestimmte Eigenschaften von Wahrnehmungsmechanismen dadurch erklären, daß im Verlaufe der Evolution eine optimale Anpassung an spezifische Anforderungen der Umwelt erfolgt ist (vgl. Gould & Lewontin, 1984, S.256; Dennett, 1983). Daß hier die theoretischen Einheiten weder Gene noch Gesamtorganismen sind, sondern Teilfunktionen des Wahrnehmungssystems, deutet darauf hin, daß es sich hier tatsächlich um eine eher metaphorische Anwendung biologischer Sprechweisen handelt - die zudem auch in der Evolutionsbiologie explikationsbedürftig sind. Als zentrale Metaphern der Psychophysik haben sich die Metapher perzeptueller Informationsverarbeitung und die evolutionistische Metapher überlagert und vermischt. In ihrem Gefolge kehrten teleologische Sprechweisen in die wissenschaftliche Sprache zurück: Von Zielen und Funktionen des Wahrnehmungssystem zu reden läßt sich oftmals rechtfertigen als elliptische Redewendungen über nicht-teleologische Erklärungen oder in bestimmten Fällen auch als ein Sprechen über Eigenschaften sog. selbstorganisierender Systeme. Teleologische Sprechweisen sind oftmals eine bequeme Façon de parler einer **externen** Zuschreibung einer Als-ob-Teleologie: Sie sind ein theoretisches Mittel, gleichsam ‚von außen‘ das Verhalten eines als Submechanismus isolierten Teiles des Wahrnehmungssystems zu beschreiben, ohne daß durch diese Sprechweisen impliziert wird, daß ein Wahrnehmungsmechanismus ein eigenes Ziel im Sinne eines gleichsam ‚von innen‘ gesetzten Zieles hat. Siehe hierzu die Sammelbände Sober (1984) und Rescher (1986).

Neben diesen beiden Metaphern, die nicht allein als theoriekonstitutiv, sondern mehr noch als paradigmakonstitutiv ansehen kann und die eher allgemeine Metaphern darstellen, bedient sich die Psychophysik einer Fülle spezifischer Metaphern bei der Theoriebildung. Computerwissenschaft und

Nachrichtentechnik werden dabei ebenso als Metaphernquellen verwendet wie Physik, Neurophysiologie und Mathematik. Betrachten wir im folgenden einige Beispiele aus den Bereichen, aus denen die Psychophysik Metaphern ihrer Theoriebildung vorrangig gewonnen hat.

Waren es zunächst experimentelle Methoden und metatheoretische Prinzipien, welche die klassische Psychophysik aus der Physik gewann, so traten später theoretische Vorstellungen und Perspektiven der Physik hinzu und dienten der Psychophysik auch in vielfacher Hinsicht als Leitfaden ihrer Theoriebildung. Eine aus der Physik stammende Metapher haben wir bereits in Abschnitt 3 behandelt: die Meßinstrumentkonzeption der Wahrnehmung. In den Teilen der Psychophysik, die weniger an der Erstellung von Skalen als an der Theoriebildung über einzelne Wahrnehmungsfunktionen interessiert sind, prägt jedoch eine andere Metapher das Denken: das Konzept des perzeptuellen Mechanismus. Dies ist sicherlich die einfachste und undifferenzierteste Metapher aus der Physik; sie ist kaum noch als solche erkennbar und gleicht eher schon einer idiomatischen Redewendung. Dieses Konzept, das zu den am häufigsten verwendeten in der Psychophysik gehört, beinhaltet die Vorstellung, daß sich aus dem perzeptuell-kognitiven System ein Teil, der für eine klar umschriebene Funktion verantwortlich ist, experimentell und theoretisch isolieren läßt.

Eine weitere allgemeine Metapher aus der Physik, die bislang keine spezifische Ausarbeitung erfahren hat und doch seit langem als anregendes Leitbild die Forschungsintuition prägt, ist die Resonanz-Metapher, die in Abschnitt 5.3 dargestellt wurde. Sie taucht m.W. zum ersten Mal bei Hering auf, danach wieder unabhängig bei Heider (1926) und in vielen Arbeiten von Gibson, in dessen Theorievorstellungen sie eine wichtige Rolle spielt. In jüngerer Zeit findet sie sich bei Shepard (1984) sowie im Kontext der Modellierung nicht-linearer neuraler Netze bei Grossberg (e.g. Cohen & Grossberg, 1984). Wie sich Resonanz im Sinne einer direkten Erfassung komplexer Variablen durch Analogie zu einem speziellen Meßinstrument veranschaulichen läßt, zeigt Runeson (1977; vgl. Pomerantz & Kubovy, 1981, S.451ff.). Neurophysiologisch lassen sich die mit dieser Metapher verbundenen Phänomene mit Beobachtungen einer Synchronisation von Oszillationen cortikaler Neuronenverbände in Beziehung setzen (Singer, 1990), die insbesondere Grundlage einer präattentiven Segmentierung bzw. Figur-Grund-Gliederung visueller Szenen sein könnten.

Zu den Metaphern aus der Physik ist die neuroelektrische Feldtheorie zu zählen, wie sie insbesondere von Wolfgang Köhler entwickelt worden war, und als moderne Variante die Hologramm-Theorie von Pribram, Nuwer & Baron (1974). Diese Theorie nimmt an, daß bestimmte Eigenschaften perzeptuell-kognitiver Leistungen auf neuroelektrischen Interferenzmustern beruhen, die

sich formal in gleicher Weise behandeln lassen wie die optischen Interferenzmuster des Hologramms. Weitere Beispiele spezifischer Theorien, die in substantieller Weise physikalische Vorstellungen in metaphorisch-analogischer Weise verwenden, sind die Theorien zum Stereosehen von Julesz & Chang (1976) und Sperling (1970, 1981). Die erstgenannte Theorie nimmt an, daß beim stereoskopischen Sehen die Erfassung von Disparitäten in einer Weise erfolgt, die sich durch ein System von magnetischen Dipolen modellieren läßt. Sperling zieht Konzepte aus der Potentialtheorie heran, um beispielsweise das binokulare Vergenzverhalten durch Gleichgewichtsdynamiken in Kraftfeldern zu modellieren.

Derartige Übertragungen physikalischer Konzepte in die Psychophysik dienen nicht vorrangig einer reinen Beschreibung oder Simulation der jeweils untersuchten Phänomene, sondern sollen vor allem neue Einsichten in den betrachteten Bereich eröffnen und zur Erklärung der Phänomene beitragen. Indem man die Tragweite des physikalischen Bildes in der Psychophysik erkundet, hofft man auf neue Fragestellungen und Heuristiken zu stoßen. Bereits an diesen Beispielen wird deutlich, daß es ein ganzes Spektrum unterschiedlicher Verwendungsweisen von Metaphern gibt, in Abhängigkeit davon, wie weit man die substantielle Interpretation einer Metapher treiben will. Metapher können mehr oder weniger substantiell sein, je nachdem bis zu welcher Detailauflösung die Bestandteile der Metapher einer substantiellen Interpretation unterzogen werden. Bleibt die Interpretation einer Metapher auf einer Art Oberflächenebene, ohne daß ihre Komponenten substantiell interpretiert werden, so haben solche Metapher einen eher instrumentellen Charakter: Man verwendet sie, um Vorhersagen zu machen oder etwas zu simulieren; ein substantieller Gebrauch von Metaphern ist hingegen stärker mit dem Konzept der Erklärung verbunden. Oft wird aus einem konkreten physikalischen Modell eine mathematische Struktur abstrahiert und in einem neuen Bereich interpretiert. Mit einer solchen Loslösung von der Herkunftsstruktur verliert die jeweilige Struktur den Charakter einer Metapher: Man hat es dann mit einer Mathematisierung zu tun, die lediglich heuristisch noch an eine metaphorische Herkunft geknüpft ist. Schließlich lassen sich auch substantiell leere Metaphern finden: Sind nämlich die aus Physik oder Mathematik übernommenen Konzepte in ihrer Anwendung so wenig restringiert, daß sie sich als reine Sprechweisen auf eine Vielzahl disparater Phänomenbereiche anwenden lassen (wie etwa eine vordergründige Anwendungen der Katastrophentheorie), so besteht die Gefahr, daß die Metapher theori degenerierend wirkt.

Dem Spektrum unterschiedlicher Verwendungsweisen von Metaphern entsprechen auch die unterschiedlichen Grade, in denen sich die empirische Angemessenheit von Metaphern experimentell prüfen läßt. Hiermit sind komplexe Probleme verbunden, doch kann man stärker substantiell interpretierte Metaphern als in Grenzen testbar ansehen, während eher instrumentell ver-

wendete Metaphern in der Regel dem Bereich, in den sie übertragen werden, nur schwache Restriktionen auferlegen und daher nicht eigentlich testbar sind; gleichwohl können sie sich natürlich als mehr oder weniger fruchtbar erweisen.

Zu den Metaphern aus der Physik, die eher zu allgemeinen mathematischen Konzepten statt zu spezifischen substantiellen Theorien der Psychophysik Anlaß gaben, läßt sich die Verwendung von Konzepten aus statistischer Mechanik und Festkörperphysik zählen, beispielsweise im Fall der sog. Spinglas-Systeme. Das energetische Verhalten solcher wechselwirkender Vielteilchensysteme wird durch mathematische Modelle beschrieben, die sich auch für die unter spezifischen Fehler- und Wechselwirkungsbedingungen stattfindenden Minimierungs- und Gleichgewichtsprobleme neuraler, perzeptueller oder kognitiver ‚Vielteilchensysteme‘ nutzbar machen lassen. Daher findet sich die Übertragung derartiger Konzepte zumeist in den mit einer konnektionistischen Modellbildung befaßten Bereichen der Psychophysik (s. Kemke, 1988; Ritter, Martinez & Schulten, 1990).

Spricht man in der Psychologie von metaphorischer Anwendung der Mathematik, so ist damit eine eher negative Bewertung einer mathematischen Konzeptbildung verbunden, Metaphorische Verwendungen mathematischer Konzepte haben weniger den Charakter spezifischer und grundsätzlich empirisch testbarer Modelle, sondern gleichen eher der Einführung einer allgemeinen mathematischen Sprechweise. A priori ist dabei zumeist schwer zu beurteilen, inwieweit sich die jeweilige Sprechweise als fruchtbar erweisen könnte. Zu den metaphorischen Anwendungen der Mathematik in der Psychologie könnte man Herbarts mathematische Psychologie ebenso rechnen wie Lewins topologische Feldtheorie oder Cassirers (1944) Behandlung perzeptueller Invarianz- und Konstanzphänomene aus einer durch das Kleinsche Programm inspirierten gruppentheoretischen Betrachtungsweise.

Als Übertragungen mathematischer Konzepte, die eher der Bereitstellung einer Sprache psychophysikalischer Theoriebildung dienen, kann man die folgenden Beispiele ansehen: Garners (1962) Anwendung der Informationstheorie als ein mathematisches Rüstzeug zur Organisation und Diskussion experimenteller Befunde und Probleme der Wahrnehmungspsychologie; die Verwendung differentialgeometrischer Konzepte in der Bewegungswahrnehmung; Hoffmans (1978) Beschreibung von Invarianzphänomene der Wahrnehmung durch bestimmte topologische Transformationsgruppen, die in der Mathematik als Lie-Gruppen bekannt sind; die Verwendung von stochastischen Differentialgleichungen zur Modellierung dynamischer Systeme; das zur Formulierung und Klärung konzeptueller Probleme der Wahrnehmungspsychologie entwickelte formale System von Bennett, Hoffman & Prakash (1989). Instrumentellen Charakter haben ferner mathematische Konzepte, die

vorrangig auf eine perzeptuell angemessene Analyse des physikalischen Reizes zielen, wie beispielsweise gewisse algebraisch-geometrische Konzepte für die Untersuchung der Formzerlegung bei Hoffman und Richards (1984) oder die Differentialgeometrie in Koenderinks (1990) Analyse geometrischer Körper (auch wenn man hofft, auf diese Weise auf gleichsam passende perzeptuelle Mechanismen zu stoßen, die sich in der gleichen formalen Sprache behandeln lassen).

Oftmals ist bei Anwendungen mathematischer Konzepte schwer zu entscheiden, ob es sich um eine dem Gegenstand angemessene mathematische Modellbildung oder lediglich um allgemeine, nicht an spezifische Restriktionen des Gegenstandsbereichs gebundene Techniken handelt; dies ist selbst dann der Fall, wenn sich diese Konzepte auf einen eng umgrenzten Gegenstandsbereich beziehen. Beispielsweise ist die Rolle der Fourieranalyse in der Untersuchung der visuellen Wahrnehmung durchaus schillernd: Sie kann einerseits eine reine (und mathematisch sehr allgemeine) Analysetechnik zur Reizbeschreibung sein, andererseits aber auch eine mechanistische Filter-Metapher, wenn sie nämlich mit der Vorstellung spezifischer neuraler Kanäle verbunden ist, die wie Ortsfrequenzfilter wirken (DeValois & DeValois, 1988; Graham, 1989). Entsprechendes gilt für mathematische Konzepte der Signalverarbeitung aus der Nachrichtentechnik und Bildverarbeitung, Bereiche, die für die Psychophysik eine reichhaltige Quelle von Konzepten bereitstellen.

Beispiele, in denen die Verwendung mathematischer Konzepte eher als eine Modellbildung über spezifische, durch den Phänomenbereich bestimmte Eigenschaften und Prozesse gedacht ist, sind die Anwendung des Vektorraumkonzeptes auf die Farbwahrnehmung in der Grassmann-Struktur (Krantz, 1975a), die Modellierung bestimmter perzeptueller Prozesse durch Autokorrelationsfunktionen, wie sie von Reichardt (1957) eingeführt wurde und beispielsweise durch van Santen und Sperling (1984) zur Modellierung der menschlichen Bewegungswahrnehmung weiterentwickelt wurde, oder die neural inspirierten Übertragungen statistischer Konzepte von Barlow (1985, 1992).

Die obigen Beispiele machen deutlich, daß eine Behandlung mathematischer Konzeptbildungen in der Psychophysik unter dem Aspekt der Metaphernbildung wenig aufschlußreich ist; sie wird vielmehr unter allgemeinen Gesichtspunkten einer adäquaten Mathematisierung zu erfolgen haben,

Noch schwieriger als im Falle mathematischer Konzepte gestaltet sich die Bewertung der Rolle neurophysiologischer Konzepte in der Psychophysik. Psychophysik und Neurophysiologie stellen zunächst zwei logisch getrennte Analyseebenen dar, deren Beziehungen überaus subtil sind. Werden also ohne explizite und detaillierte Rechtfertigung neurophysiologische Konzepte, mögen sie in der Neurophysiologie eine präzise Bedeutung haben oder auch dort



wiederum eine Metapher sein, in die Psychophysik übertragen, so läßt sich auch eine solche Übertragung als eine metaphorische bezeichnen. Nun liegen aber vielen Modellen in der Psychophysik neurophysiologische Vorstellungen zumeist unausgesprochen zugrunde. Zum Teil hängt dies damit zusammen, daß Psychophysik auch von ausschließlich an neurophysiologischer Modellbildung Interessierten als Hilfswissenschaft betrieben wird. Hauptsächlich ist es jedoch durch eine natürliche Tendenz bedingt, die psychophysikalische Modellbildung, die ihrem Status nach hinsichtlich interner Mechanismen weitgehend neutral ist, durch Vorstellungen über solche Mechanismen zu ergänzen, eine Tendenz, die in einer neuroreduktionistischen Perspektive ihren deutlichsten Ausdruck findet. Daher vermischen sich oft in Darstellungen psychophysikalischer Modelle psychophysikalische und neurophysiologische Sprechweisen. Spricht man beispielsweise von der Modularität des Wahrnehmungssystems oder unabhängigen Pfaden für Bewegungs- und Farbinformation, weil sich neben neurophysiologischen Hinweisen auch in psychophysikalischen Experimenten Bedingungen finden lassen, durch die beide Aspekte gleichsam separiert werden können (s. Hubel & Livingstone, 1987), so wird man solchen Sprechweisen innerhalb psychophysikalischer Theoriebildung metaphorischen Charakter zuschreiben können; sie haben eine heuristische Funktion für das experimentellen Auffinden entsprechender Befunde. Gleiches gilt für Sprechweisen, die auf Detektoren, Kanäle oder Mechanismen Bezug nehmen. Doch können diese Konzepte mit zunehmender Einbindung eine neue Präzisierung in der Psychophysik erfahren und so, wie dies auch allgemein für Metaphern in der Sprache gilt, zu idiomatischen Redeweisen eines Wissenschaftsbereiches werden.

Damit sind wichtige Bildfelder umrissen, aus denen die Psychophysik Anregungen für die Theoriebildung gewinnt. Ob eine Metapher sich in ihrem strukturellen Kern als empirisch adäquat, als fruchtbare Heuristik oder als theoriodegenerierend erweist, kann oft erst im Fortgang der Theoriebildung entschieden werden. Von einer präzisen Metapher zu einer dunklen Metaphorik ist es oft nur ein kleiner Schritt: Beispielsweise kann die Untersuchung von Wechselwirkungen in nicht-linearen dynamischen Systemen sowohl zu spezifischen und substantiell interpretierten Modellen führen (s. Grossberg, 1988) wie auch zu einer leeren Chaos-Metaphorik. Auch die Rolle von Metaphern wird in den Beispielen erkennbar: In erster Linie haben sie eine katchrestische Funktion, d.h. sie weisen auf eine Mangelercheinung des Sprachsystems hin, die sie durch eine Übertragung aus einem anderen Bereich zu beheben suchen. Wegen der Einzigartigkeit des perzeptuell-kognitiven Systems mangelt es der Sprache, in der man sich über dieses System ein theoretisches Bild zu machen sucht, an geeigneten Begriffen zu seiner Beschreibung. Durch Rückgriff auf vertraute Begriffe aus Bereichen, zu denen man Ähnlichkeitsbezüge vermutet, sucht man den Mangel des theoretischen Vo-

kabulars zu mildern. Doch überbrücken Metaphern nicht nur eine Notlage des theoretischen Vokabulars, sondern regen zugleich an, die Tragweite des übernommenen Bildes auszuloten. Zudem werden durch den Gebrauch einer Metapher bestimmte Aspekte des Bereiches, auf den sie zielt, hervorgehoben und so die Aufmerksamkeit auf diese gelenkt. Metaphern sind damit eine wesentliche Quelle von Heuristiken und steuern den Forschungsprozeß.

Die für die Theoriebildung interessantesten Metaphern sind jene, die in ihrer Entwicklung offen sind und gerade durch ihre Ausgestaltung den Forschungsprozeß anregen. Dies kann freilich auch für konkrete Modelle gelten - womit hier nicht als Theorieelemente verwendete mathematische Modelle oder Modelle im Sinne der Modelltheorie gemeint sind -, und in der Tat haben beide Konzepte vielerlei Berührungspunkte. Die theoriekonstitutiven Metaphern unterscheiden sich jedoch von konkreten Modellen darin, daß letztere etwas theoretisch Komplexes anschaulich begreifbar machen sollen. Sie dienen somit zunächst didaktischen Zwecken; auch läßt sich ihre Tragweite zumeist präzise angeben. Daher finden sich, etwa in der Physik, ohne weiteres zum gleichen theoretischen Konzept miteinander logisch unverträgliche Modelle. Je nach Problemzusammenhang läßt sich dadurch eine komplexe theoretische Struktur nicht nur vorstellungsmäßig, sondern auch hinsichtlich ihrer mathematischen Behandlung handhabbarer machen (vgl. Readhead, 1980).

Andere Heuristiken und Regulativa des Forschungsprozesses werden aus Metaprinzipien (vgl. Tack, in diesem Band, Kap. 1, 2.2.2; Gadenne, in diesem Band, Kap. 9) gewonnen, die als metatheoretische und somit nicht empirisch prüfbare Richtlinien zumeist implizit die Theorieentwicklung bestimmen. Das pythagoräische Erkenntnisideal, die Einheit der verschiedenen Wissenschaften in der Einheit des Mathematischen zu suchen, gehört ebenso hierzu wie Aspekte der Theorieästhetik: Einfachheit, Ökonomie, Optimalität, Harmonie (am bekanntesten ist deren Rolle in Keplers astronomischen Untersuchungen), Symmetrie und Invarianz sind bekannte Beispiele. Solche Metaprinzipien, etwa der Einfachheit, sind selbst hochgradig theoretische Konzepte, für die es keine einfachen theorieunabhängigen Kriterien gibt; so hängt insbesondere das Prinzip der Einfachheit stark mit Vorstellungen über theoretische Plausibilität zusammen.

Die psychophysikalische Theoriebildung unterscheidet sich aber hinsichtlich der Rolle dieser Metaprinzipien in keiner Weise von der Theoriebildung anderer naturwissenschaftlicher Disziplinen, so daß diese Aspekte einer allgemeinen Wissenschaftslehre hier nicht weiter verfolgt werden. Als Beispiel expliziter Symmetrie- und Invarianzbetrachtungen in der Psychophysik sei auf die in Mausfeld (in diesem Band, Kap. 13.5.) behandelten Untersuchungen zur **sog. *Meaningfulness*** verwiesen.

## 6.2 Theorietypen der Psychophysik

Metaphern können, wie die angeführten Beispiele deutlich machen sollten, in verschiedenem Grade die Theoriebildung prägen: Sie können theoriekonstitutiv sein, Theoriebausteine bereitstellen oder lediglich einen mathematischen Formalismus nahelegen. Über Metaphern bei der Theoriebildung zu reden beinhaltet daher auch, über Theorietypen zu reden. Unterschiedliche Theorietypen der Psychophysik lassen sich jedoch auch aus einer anderen Perspektive ausmachen; einige von ihnen wollen wir im folgenden betrachten. Zudem werden wir kurz darauf zu sprechen kommen, welche Ziele sich mit einer psychophysikalischen Theoriebildung verbinden lassen.

Der Theoriebegriff kann sehr unterschiedlich gefaßt sein (s. Gadenne, in diesem Band, Kap. 7); wir wollen ihn hier im Sinne der naturwissenschaftlichen Umgangssprache verwenden und eine inflationäre Verwendung im Sinne isolierter ad-hoc-Ideen vermeiden. Eine Theorie in diesem Sinne ist beispielsweise die Young-Helmholtz-Theorie der Farbwahrnehmung in der Grassmannsehen Formulierung; die Informationstheorie hingegen ist keine Theorie in diesem Sinne, da sie einen allgemeinen mathematischen Formalismus darstellt, der nicht an eine spezifische substantielle Interpretation gebunden ist. Als Theorien bezeichnet man häufig aber auch jene formalen Theorietypen und Theorieschemata, deren substantielle Interpretation von Fall zu Fall variieren kann und die eher einen universell-methodologischen Charakter haben. Auf dieses wollen wir zunächst einen Blick werfen, auch wenn die Unterscheidung von spezifischen substantiellen und methodologischen Theorietypen nur eine sehr grobe Orientierung erlaubt, da die Grenze zwischen beiden unscharf ist und von Intentionen des Forschers abhängt.

Vorrangig methodologischen Charakter haben die meisten Theoriebildungen der klassischen Psychophysik. Diese wird nicht zu Unrecht schon früh in ihrer Entwicklungsgeschichte mit ihren Methoden identifiziert. Ihre Theoriebildung hatte weniger den Charakter spezifischer substantieller Theorien über einzelne Wahrnehmungsleistungen, sondern suchte eher allgemeine Aspekte, wie beispielsweise das Diskriminations-, Signalentdeckungs- und Entscheidungsverhalten, zu modellieren.

In ähnlicher Weise hat die Theoriebildung in der mit der Messung von Reaktionszeiten befaßten Tradition einen wesentlich methodologischen Charakter (s. Pachella, 1974, S. 42). Reaktionszeitindizes lassen sich ebenso wie die bei der Skalenkonstruktion erhaltenen Indizes nur unter weitreichenden und oftmals schwer zu rechtfertigenden Annahmen zu wesentlichen Konzepten einer substantiellen psychologischen Theoriebildung in Beziehung setzen. Insbesondere tritt bei der Interpretation von globalen Indizes wie Reaktionszeiten und evozierten Potentialen in verschärfter Weise das genannte Invertie-

rungsproblem auf, da mit einer empirisch erhaltenen Struktur dieser Indizes unendlich viele Strukturmuster interner Komponenten verträglich sind (s. Utal, 1967, 1990).

Anders als die universellen Theorieschemata der klassischen Psychophysik sind die Theorien der mit spezifischen Leistungen des Wahrnehmungssystems befaßten Psychophysik zugeschnitten auf die jeweiligen Charakteristiken eines Phänomenbereichs. Bestimmte Theorietypen, die sich freilich wiederum in verschiedenem Grade durchdringen können, lassen sich hinter der Unterschiedlichkeit solcher Theorien schematisch ausmachen. Diese Stile der Theoriebildung in der Psychophysik sind keineswegs psychophysikspezifisch, sondern charakteristisch für unterschiedliche Zugangsweisen des naturwissenschaftlichen Denkens allgemein. Auch in der Psychophysik fallen nämlich zunächst jene Theoriestile ins Auge, die als **esprit géométrique** und **esprit mécanique** durch Duhem (1908/1978, S. 67ff.) bekannt wurden.

Prominentes Beispiel für den auf geometrischen Intuitionen beruhenden Theorietyp ist Newtons Theorie der Metamerie von Farben, wie sie sich in seinem geometrischen Modell des Farbkreises niederschlägt. Diese Theorie - die historisch erste quantitative psychophysikalische Theorie - wurde von Newton, der an psychophysikalischen Fragen nur am Rande interessiert war, noch weitgehend spekulativ entworfen; bereits Grassmann (1853) jedoch hat ihren psychophysikalischen Kern formal klar herausgearbeitet und gezielten empirischen Überprüfungen zugänglich gemacht. Zuvor hatte man sich schon in der Philosophie Sinnesqualitäten und Qualia abstrakt-geometrisch in Räumen lokalisiert vorgestellt. In Grassmanns Ausgestaltung findet dies erstmals eine formale Präzisierung, und zwar in einem Bereich, dem der Farben, der nicht ohne weiteres eine geometrische Betrachtungsweise nahelegt. Das sich an eine solche Annahme anschließende Konzept eines mit einem perzeptuellen Code assoziierten Farbraumes suchte man in der Folge durch Distanzkonzepte geometrisch anzureichern. Beruhen diese Konzepte auf theoretischen Intuitionen über lokale Schwellenrelationen, so erhält man beispielsweise die sog. Linielement-Theorien, wie sie durch Helmholtz, Schrödinger und Stiles entwickelt worden sind (s. Wyszecki & Stiles, 1982). Beruhen sie auf Intuitionen über globale Ähnlichkeitsrelationen ergeben sich Modelle wie die durch Konzepte der nichtmetrischen multidimensionalen Skalierung inspirierten (Shepard, 1978).

In anderen Bereichen der Psychophysik, wie der Bewegungs-, Form- oder Raumwahrnehmung, ist eine geometrische Modellbildung weniger erstaunlich, da hier bereits eine perzeptuell angemessene physikalische Reizbeschreibung in natürlicher Weise geometrische Konzepte beinhaltet.

Algebraisch orientierte Theorietypen in der Psychophysik wollen wir nicht als eigenen Typ den geometrisch orientierten gegenüberstellen, da beide Zu-

gangswesen, die in der Mathematik zwei Spannungspole des Denkens kennzeichnen, sich je nach Vorlieben des Forschers vermischen und kaum zu trennen sind.

Abstrakt-geometrisch orientierte Modelle der Wahrnehmung zeichnen zunächst keinen spezifischen Mechanismus aus und sind oft nur schwer mit Vorstellungen von konkreten Mechanismen in Beziehung zu setzen. So tritt in ihnen die Zeitvariable, die den eigentliche Prozeßcharakter von Wahrnehmungsprozessen ausmacht, in der Regel nicht explizit auf. In der Grassmann-Struktur beispielsweise gibt es die Zeitvariable nicht, obwohl der zugrunde liegende Mechanismus als Prozeß gedacht ist. Einer rein psychophysikalischen Betrachtungsweise ist es zunächst fremd, in expliziten Prozeßstufen zu denken; erst wenn man zusätzlich neurophysiologische Intuitionen heranzieht oder wenn andere Restriktionen eine Abfolge von Verarbeitungsschritten festlegen, kann man zu der Vorstellung eines stufenweise zergliederten Prozesses kommen. Anders in mechanistisch orientierten Theorien: Bei diesem Theorietyp, der neben konkret-mechanistischen auch abstrakt-mechanistische Theorien umfaßt, ergibt sich der zeitliche Prozeßcharakter in natürlicher Weise. Die Attraktivität mechanistisch orientierter Theorietypen ist aber vor allem darin begründet, daß einem mechanistischen Bild ein sehr viel höherer ‚Erklärungswert‘ zugeschrieben wird als einem geometrischen Bild. Berühmt ist Kelvins Ausspruch, daß wir ein physikalisches Phänomen erst verstehen, wenn wir ein entsprechendes mechanisches Modell konstruieren können (vgl. Duhem, 1908/1978, S. 89). Genau dieser Erklärungsbegriff beherrscht - erweitert auf abstrakt-mechanistische Vorstellungsweisen - die Psychophysik. Daher gehören zu diesem Theorietyp die meisten theoretischen Ansätze der gegenwärtigen Psychophysik perzeptuell-kognitiver Informationsverarbeitung; dies wird bereits in Sprechweisen ‚modulare Organisation des visuellen Systems‘, ‚Mechanismen der Bewegungswahrnehmung‘, ‚Ortsfrequenzkanal‘ u.ä. deutlich. Der mechanistische Theorietyp wird durch das Paradigma interner Informationsverarbeitung in besonderer Weise nahegelegt. Dabei hat das Attribut ‚mechanistisch‘ eine Bedeutungsänderung erfahren und wird nicht mehr im Sinne der Mechanik verstanden, sondern in dem Sinne, daß sich die Theorie im Prinzip durch einen entsprechenden Apparat (etwa mechanisch, elektronisch oder als Programm) realisieren läßt.

Werden **dynamische** Aspekte und temporale Charakteristika, wie Augenbewegungen, zeitliche Adaptationsprozesse u.ä., in expliziter Weise zu zentralen Bestandteilen der Theorie, so ergeben sich mechanistisch-dynamische Theorien als spezieller Theorietyp. Als Beispiel seien Cohen & Grossberg (1984) genannt, die eine Vielzahl von Phänomenen der monokularen und binokularen Helligkeitswahrnehmung durch zwei dynamisch interagierende Konturauswertungsprozesse zu simulieren suchen.

Ob man eher einem geometrischen oder eher einem mechanistischen Theorietyp zuneigt, kann zudem durch metatheoretische Vorstellungen über den Gegenstandsbereich der Psychophysik und die Natur ihrer Gesetze bestimmt sein. Betrachtet man die Psychophysik als eine Art Erweiterung der Physik durch psychologische Relationen, so wird man auch ihre Gesetze in ähnlicher Weise ansehen wie physikalische Gesetze. In der Physik treten Gesetze erst durch geeignete Idealisierungen, etwa die des idealen Gases oder die der gleichförmig beschleunigten Bewegung, deutlich hervor; die Theorie bezieht sich nur auf diese Idealisierungen und hängt entscheidend von der Formulierung von Randbedingungen ab, unter denen sie mit Phänomenen in Beziehung gesetzt wird. Nimmt man eine solche Haltung gegenüber der Psychophysik ein, wird man ihre Gesetze auf der Ebene entsprechender Idealisierungen suchen. Zwar wird man nicht die universelle Gültigkeit psychophysikalischer Gesetze behaupten, doch beispielsweise bereit sein, in zur Physik analoger Weise Invarianzbetrachtungen anzustellen, die wiederum mit Invarianzbetrachtungen der physikalischen Struktur gekoppelt sein können (vgl. Mausfeld, in diesem Band, Kap. 13.5.). Derartige Invarianzbetrachtungen lassen sich zu abstrakt-geometrischen Konzepten in Beziehung setzen. Sieht man den Gegenstandsbereich der Psychophysik als in wesentlichen Teilen dem der Biologie ähnlich - womit psychophysikalische Gesetzmäßigkeiten biologischen gleichen - wird man solchen Invarianzbetrachtungen nur einen sehr begrenzten Nutzen zusprechen. Denn in diesem Fall werden die Gesetzmäßigkeiten der Psychophysik durch spezifische Eigenschaften der zufälligen Realisation biologischer Mechanismen bestimmt sein (vgl. Marrs Typ II-Theorien, Marr, 1977, S. 38). Eine solche Auffassung liegt etwa der „utilitaristischen Theorie“ der Wahrnehmung von Ramachandran (1990) zugrunde, und sie legt einen mechanistischen Theorietyp nahe. Sie wird gestützt durch evolutionstheoretische Auffassungen, wie sie sich in der Metapher ausdrücken, daß die Evolution einem Kesselflicker gleich zu Werke gehe und für die jeweiligen Zwecke mit den jeweils verfügbaren Mittel ad hoc passende Lösungen, Strategien und Tricks entwerfe (Jacob, 1977). Von einer solchen Flickwerk-Strategie wird man kaum erhoffen können, daß sie zu einer theoretisch geschlossenen und mathematisch ästhetischen computational **theory** der Wahrnehmung Anlaß gibt, wenn sie sich auch in gewissen Segmenten durch eine solche beschreiben läßt. Aus dieser Perspektive kann man in der Psychophysik keine Gesetze in einem der Physik vergleichbaren Sinne erwarten, sondern lediglich bestimmte lokale Gesetzmäßigkeiten. Anders als physikalische Gesetze stellen psychophysikalische Gesetze dieser Auffassung zufolge eher generalisierte bzw. idealisierte Fakten dar (vgl. Smart, 1964, S. 52). Auch wenn man, wie in der KI, auf ideale Modelle statt auf die Modellierung realer Mechanismen zielt, stellen diese Modelle spezifische Ingenieurslösungen für einen aus der Psychophysik abstrahierten Aspekt dar und rechtfertigen nicht die Annahme universeller Wahrnehmungsgesetze.

Werfen wir abschließend - freilich ohne eine Antwort geben zu können - einen kurzen Blick auf Fragen, die sich mit der Bewertung und den möglichen Zielen einer psychophysikalischen Theorie verbinden. Hierzu gehören Fragen wie: Was unterscheidet eine gute von einer schlechten psychophysikalischen Theorie? Etwa der Grad ihrer Anbindung an die Neurophysiologie? Oder der Umfang ihres (deskriptiven) Geltungsbereichs (z. B. eine quantitative Beschreibung des Zusammenspiels von Form-, Farb- und Bewegungsinformationen)? Ihre Vorhersagegenauigkeit im Einzelfall? Oder die Möglichkeit, sie als KI-System zu realisieren? Muß eine gute psychophysikalische Theorie die tatsächliche Detailfunktion des Systems erkennen lassen, oder soll sie idealisiert die für eine globale Leistung wesentlichen Charakteristika der Wahrnehmung beschreiben? Gibt es nicht-reduktionistische Erklärungen in der Psychophysik? Können neurophysiologische Beobachtungen psychophysikalische Phänomene erklären?

Die Antwort auf derartige Fragen wird durch die Vorstellungen bestimmt sein, die man sich vom Gegenstandsbereich der Psychophysik und von den Zielen ihrer Theoriebildung macht. Trägt man der logischen Kluft zwischen Psychophysik und Neurophysiologie Rechnung und sucht die Theoriebildung zunächst frei von Spekulationen über neurophysiologische ‚Korrespondenzen‘ zu halten, so wird man als vorrangiges Ziel psychophysikalischer Theoriebildung ansehen, ein kohärentes theoretisches System zu schaffen, in dem möglichst viel vom ‚tatsächlichen‘ Sinneseindruck erfaßt wird. Dabei ist die Theoriebildung keineswegs auf Konstrukte über interne Verarbeitungsschritte beschränkt, sondern es ist durchaus möglich, auch Aspekte von Empfindungseindrücken, d.h. von Qualia, mit einzubeziehen, indem diese als ‚Konstrukt von außen‘ in die Theorie eingeführt werden. Ein solchermaßen theoretisches Konzept von Qualia wird zwar vielfältige Plausibilitätsbezüge zum alltags-sprachlichen haben, doch logisch ist es davon unabhängig, wie Qualia gleichsam von innen betrachtet erscheinen mögen. Eine ähnliche Auffassung fand ihren frühen Ausdruck in Machs (1863, S.202) Dictum: „Was die Seele sei, das kann für eine mathematische Psychologie so gleichgültig sein, als für die Physik das Wesen der Materie.“

Was auch immer man als fiktives Endziel psychophysikalischer Theoriebildung ansehen mag, das Ziel gegenwärtiger Theoriekonstruktion ist die phänomenadäquate Erfassung bestimmter Teilfunktionen des Wahrnehmungssystems. Dabei stehen in der gegenwärtigen Diskussion fast ausschließlich funktionale Leistungsaspekte im Vordergrund der Theoriebildung, kaum jedoch Erlebnisaspekte.

## **7. Experiment, Beobachtung und Daten in der Psychophysik**

In wissenschaftstheoretischen Lehrmeinungen wird oftmals das Experiment als eine ‚Magd‘ der Theorie angesehen, wodurch es eine unverdiente Geringschätzung erfährt. Während man sich in der allgemeinen Wissenschaftslehre der Komplexität der Theoriedynamik zunehmend bewußt geworden ist, scheint die Komplexität der Rolle von Experimenten, von Beobachtungen, Daten und Phänomenen immer noch unterschätzt zu werden (Franklin, 1986). Dies und die häufig bemerkbare Tendenz, Wissenschaftsgeschichte im nachhinein zugunsten eines theoriegeleiteten Experimentierens umzuschreiben, müssen natürlich ein verzerrtes Bild vom Verhältnis von Experiment und Theorie mit sich bringen. Verkürzt formuliert ist der entscheidende Punkt dies: Das Experiment führt im wissenschaftlichen Erkenntnisprozeß gegenüber der Theorie durchaus ein Eigenleben und kann auf vielerlei Weise zur Theoriebildung beitragen. Wie in anderen Disziplinen ist auch in der Psychophysik das Experiment zumeist nicht theoriegeleitet, sondern - häufig im Sinne von Theoriefragmenten - theoriebezogen; in der Regel geht ihm keineswegs eine eigentliche Theorie wegweisend voraus (vgl. Herrmann, 1990, S. 10). Vielmehr zeigt sich auch hier die ganze Vielfalt möglicher Funktionen des Experimentes, wie sie für die Physik durch Hacking (1983) dargelegt wurde. Dieser Vielfalt Rechnung zu tragen bedeutet nicht, einem naiven Induktivismus oder blinden Empirismus zu erliegen.

Betrachten wir einige Beispiele psychophysikalischer Experimententypen, die von aufschlußreichen Zufallsbeobachtungen, über qualitative Experimente zur Herstellung und Untersuchung neuer Phänomene bis hin zu eigentlich theoriegeleiteten Experimenten - sei es zur qualitativen oder quantitativen Widerlegung bestimmter Vorhersagen, sei es zum Schätzen numerischer Parameter oder zum Testen qualitativ-algebraischer Bedingungen - reichen.

Experimente von wesentlich qualitativer Natur spielen in der Psychophysik eine besondere Rolle. Sie reichen von einfachen, aber folgenreichen Demonstrationen (wie den Landschen Mondriaan-Experimenten oder den Juleszschen **random** dot-Stereogrammen) bis hin zu apparativ äußerst aufwendigen Versuchen. Hier sind etwa die auf dem Auge zu befestigenden Miniaturspiegelsysteme erwähnenswert, durch die sich ein Bild netzhautstabilisieren läßt. Ganze Theorieklassen können durch derartige Experimente verworfen werden, wohingegen keine eigentliche Theorie diese Experimente zu leiten braucht. Eine Vielfalt solcher primär qualitativ orientierter Experimente wurde von Livingstone & Hubel (1987) durchgeführt, um zu erkunden, inwieweit sich getrennte interne Kanäle für verschiedene Wahrnehmungsdimensionen annehmen lassen. Eine wichtige apparative Technik der Reizerzeugung stellt dabei die nach der individuellen Luminanzfunktion durch den Computer er-



stellte Äquilibriumsebene dar. In dieser kann eine Szenensegmentierung zwar durch Tiefen-, Farb- oder Bewegungsinformation erfolgen, jedoch gibt es keine Information über relative Luminanzen, da alle Teile die gleiche Helligkeit (i. S. der Luminanz) aufweisen. Geleitet werden derartige Experimente durch die Idee, daß das visuelle System modular in ‚Kanälen‘ aufgebaut ist und sich einer dieser Kanäle, der Luminanzkanal, auf diese Weise sozusagen ausschalten läßt; dadurch läßt sich feststellen, welche perzeptuelle Information über diesen Kanal geleitet wird.

Dieses Beispiel zeigt, daß durch Experimente erst gezielt neue Phänomene erzeugt werden. Die in Äquilibriumsebenen auftretenden Phänomene existierten nämlich solange nicht, bis mit computergesteuerten Monitoren der Apparat zu ihrer Erzeugung bereitstand (s. jedoch Liebmann, 1927). Gleiches gilt für die Phänomene, die bei netzhautstabilisierten Bildern auftreten. In der Physik heißen besonders lehrreiche und bemerkenswerte Phänomene Effekte, und es ist dort ein wissenschaftlicher Gemeinplatz, daß eine der Hauptaufgaben des Experimentes das Auffinden und Erzeugen von Effekten ist (s. Hacking, 1983; vgl. Herrmann, 1990). Auch in der Psychophysik gehören viele der bedeutendsten Experimente zu dieser Klasse phänomenerzeugender Experimente.

Ist ein neues Phänomen gefunden oder erzeugt, so ist es Funktion des Experimentes, durch die Untersuchung von Randbedingungen eine **stabile** Erzeugung der erhaltenen Phänomene zu garantieren. Hierzu gehört auch, die in der Regel komplexe experimentelle Apparatur zu erkunden; denn viele der sich aus Experimenten ergebenden Beobachtungen sind nicht ‚direkt‘, sondern zumeist über komplexe Instrumente, Apparaturen und Tabellenwerke gewonnen; in diese selbst ist wiederum in vielfältiger Weise Theorie verkörpert (s. Kuhn, 1961). In manchen Fällen braucht es sehr lange, bis eine geeignete experimentelle Apparatur entwickelt und getestet ist, während für die eigentlichen Beobachtungen eine vergleichsweise nur kurze Zeitspanne genügt. Die Wissenschaftsgeschichte ist reich an Beispielen, die zeigen, daß häufig erst die Bereitstellung eines geeigneten Instrumentariums einen experimentellen Weg eröffnete, der dann seinerseits die Theoriebildung nach sich zog. Auch der Beginn der experimentellen Psychologie war wesentlich durch die Erfindung und Konstruktion neuer Apparate bestimmt; in Wundts **Physiologischer Psychologie** finden sich viele dieser Apparate in Radierungen dargestellt und sorgfältig beschrieben. Die Entwicklung geeigneter apparativer Instrumente des Experimentierens ist eine überaus bedeutende und eigenständige Kunstfertigkeit. Die - außerhalb der Physik zumeist gravierend unterschätzte - Bedeutung dieser Kunstfertigkeit läßt sich auch in der Psychophysik in vielfacher Weise belegen, von Wundts brillantem Werkzeugmacher E. Zimmermann bis hin zu jüngeren Entwicklungen, wie den genannten Apparaten zur Erzeugung netzhautstabilisierter Bilder.

Auch die Fähigkeit zur Beobachtung ist eine Kunstfertigkeit, der innerhalb des experimentellen Zugangs eine eigenständige Rolle zukommt. Diese Rolle ist aber nur schwer einer theoretischen Analyse zugänglich; einige Aspekte hängen mit dem vielfach diskutierten Problem des Unterschiedes von Sehen **Daß** und **Sehen Als** (s. Hanson, 1969, S. 111ff.) zusammen. Nicht nur die theoretische Brille, durch die man wahrnimmt, formt die Beobachtung, sondern auch Erfahrung und Übung. Jedem Weinkenner ist der erstaunliche Grad an Expertentum, den man in der Wahrnehmung erreichen kann, bekannt. Für das Auffinden bemerkenswerter Phänomene in der Psychophysik ist ein solches Expertentum in der ‚Kunst‘ der Wahrnehmung, d.h. ein hoher Differenzierungsgrad verbunden mit der Fähigkeit, ‚Wesentliches‘ aus einem Wahrnehmungseindruck zu abstrahieren, von entscheidender Bedeutung. Ein einschlägiges Beispiel sind die experimentellen Studien von Katz (1911) zu Erscheinungsweisen der Farben.

Mit dem Konzept der Beobachtung verbindet sich ein tief liegendes methodologisches Problem, das für die Untersuchung der menschlichen Wahrnehmung als geradezu charakteristisch angesehen werden muß: die Unterscheidung einer internen und einer externen Beobachtungskategorie. Indem man nicht nur Reaktionsweisen der Versuchsperson protokolliert, sondern ihre Berichte - implizit oder durch eigene Beobachtungen - mit der eigenen Wahrnehmung vergleicht und beides wiederum mit den Größen der Theorie verbindet, werden unter gewissen stillschweigenden Annahmen die von ‚innen‘ und die von ‚außen‘ gemachten Beobachtungen innerhalb einer psychophysikalischen Theorie in Beziehung gesetzt. Auf diese ‚internen Beobachtungen‘ - man bezeichnet die Versuchsperson häufig auch als Beobachter - kann man dann nicht verzichten, wenn man die theoretischen Größen an eigene Wahrnehmungserfahrungen anbinden möchte. Zwar hat der Experimentator zu diesen Empfindungen oder Qualia der Versuchsperson keinen Zugang, aber er kann sie durch eigene ‚interne Beobachtungen‘ zu dieser Wahrnehmungssituation gleichsam auf Plausibilität prüfen. Diese Andeutungen mögen hier genügen, da das Problem, wie sich innerhalb einer Theorie ein Sprechen über Qualia - und ohne ein solches kommt eine Theorieentwicklung in der Psychophysik menschlicher Wahrnehmung nicht aus - rechtfertigen läßt, weit über die Psychophysik hinausführt (s. etwa Rorty, 1979).

Durch die sich im Paradigma perzeptuell-kognitiver Informationsverarbeitung eröffnenden Möglichkeiten einer Theoriebildung in der Psychophysik steht eine zunehmende Anzahl von Experimenten in einem dialogischen oder, strenger, in einem theoriekritischen Verhältnis zu einzelnen Theorien oder Theoriebausteinen. Mit dem **computational approach** entsteht ein neues Gerüst zur Formulierung von Theorien, das zudem eine theoriegeleitete Konstruktion kritischer und oft neuartiger Experimente ermöglicht. Die formale Ausarbeitung eines Theorieansatzes erlaubt nämlich, interessante Vorhersagen über

perzeptuelle Phänomene abzuleiten, auf die man ohne eine solche formale Ausarbeitung nicht hätte stoßen können (bei Ramachandran, 1990, finden sich entsprechende Beispiele). Dieser Klasse von Experimenten, die gezielt diejenigen Eigenschaften des Wahrnehmungssystems untersuchen, die sich beim Aufbau einer Theorie als besonders wichtig erweisen, sind auch Experimente zuzurechnen, die nicht über eine **computational theory**, sondern durch mathematische Analysen anderer Art testbare kritische Bedingungen isolieren. Ein lehrreiches Beispiel ist die meßtheoretische Analyse einer Version der opponierten Farbtheorie durch Krantz (1975b), die zu kritischen Bedingungen geführt hat, deren empirische Gültigkeit Voraussetzung für eine lineare Beschreibbarkeit der Opponenten Codierung ist.

Quantitative Experimente, die mit bestimmten mehr oder weniger explizierten Theoriebausteinen zusammenhängen, bilden die weitaus größte Klasse von Experimenten der Psychophysik. Das Quantitative in derartigen Experimenten hängt freilich zumeist nur lose mit der zugrunde liegenden Theorie zusammen. In der Regel ist es nur der notwendige Weg, etwas Qualitatives herauszufinden: nämlich den Gültigkeitsbereich und die Stabilität theoretisch interessanter Phänomene. Das Quantitative drückt sich in den im Experiment erhaltenen Daten aus, das Qualitative in den durch eine Theorie zu erklärenden Phänomenen. Die Aufgabe der Theorie wird man nicht darin sehen wollen, die jeweils spezifischen Daten zu erklären, denn diese werden durch die jeweilige apparative Konstellation und ihre Eigenschaften geprägt; sie sind an den experimentellen Kontext gebunden, außerhalb dessen sie keine Bedeutung haben. Vielmehr zielt die Theorie auf eine Erklärung der erhaltenen Phänomene, die in den Daten ihren Ausdruck finden. Theoretisch interessante Phänomene sind über verschiedene experimentelle Untersuchungsweisen stabil, während sich die Daten jeweils ändern. Nicht die Replikation von Daten ist Ziel des Experimentes, sondern die Replikation von Phänomenen (s. hierzu Woodward, 1989). Da das Experiment dem experimentellen Einkreisen aufschlußreicher Phänomene dient, ist seine bloße Wiederholung kaum von Interesse, sondern man sucht mit seiner Hilfe Randbedingungen für das Auftreten von Phänomenen zu studieren und einzuzugrenzen, um so stabilere Bedingungen für die Erzeugung der Phänomene zu finden.

Phänomene bringen nun nicht einfach Daten hervor. Vielmehr ist die Erzeugung und Behandlung brauchbarer Daten wiederum eine eigenständige Fertigkeit des Experimentators; diese bezieht sich ebenso auf Aspekte apparativer Fehlerkontrolle und Kalibrierung wie auf die Aufbereitung der Daten für mathematische Analyseverfahren. Die Verwendung überaus komplexer Apparate zur Reizerzeugung und Reizdarbietung unterscheidet die Psychophysik von den meisten anderen Gebieten der Psychologie: Die Apparate treten zwischen Theorie und Daten und vergrößern die Distanz zwischen beiden. Angesichts der Komplexität von kaum explizit zu machenden Randbedingungen und

Fehlerquellen in einem technisch aufwendigen Experiment kann der Wert von Daten aufgrund der publizierten Beschreibungen allein nicht beurteilt werden. (In diesem Zusammenhang sei auch an die gängige Praxis der Elimination ‚unbrauchbarer‘ Daten erinnert.) Eine solche Bewertung wurde nicht zuletzt eigene Erfahrung im jeweiligen Bereich voraussetzen. Bereits hieran wird erkennbar, daß Daten Teil eines komplexen wissenschaftlichen Argumentationsmusters sind und im wissenschaftlichen Diskurs durchaus auch Teil einer Persuasions-Rhetorik sein können: Da sich bei technisch aufwendigen Experimenten die Güte der Daten nicht ohne weiteres bewerten läßt, ist man auf die Bewertung der wissenschaftlichen Reputation des Untersuchenden angewiesen. Der wissenschaftssoziologische Erhebungskontext (vgl. Herrmann, in diesem Band, Kapitel 6) ist für die Bewertung von Daten ebenso wichtig wie der theoretische Erhebungskontext.

So dient die Mehrzahl quantitativer Experimente nicht dem Selbstzweck einer quantitativen Bestimmung irgendwelcher Theoriegrößen oder Modellparameter, auch wenn solche Bestimmungen bei Entscheidungen zwischen konkurrierenden Theoriebausteinen in der Forschungspraxis eine wichtige Rolle spielen können. Quantitative Analysen stellen hier vielmehr ein Mittel dar, durch das man hofft, die jeweils im Forschungsinteresse stehenden Eigenschaften des Wahrnehmungssystems in noch feinerer Auflösung zu erfassen und somit entscheidende neue **qualitative** Einsichten zu gewinnen; sie sind daher ein oft notwendiger Weg zu einem tieferen Verständnis der globalen funktionalen Struktur des Wahrnehmungssystems.

Die numerische Bestimmung bestimmter Größen ist eher selten das eigentliche Ziel quantitativer Experimente. Derartige genuin quantitative Experimente können beispielsweise der Bestimmung zeitlicher Konstanten für verschiedene Adaptationsprozesse dienen, der Schätzung spektraler Sensitivitätsfunktionen für das Farbsehen, der Bestimmung des feinsten räumlichen Auflösungsvermögens oder der beim skotopischen Sehen für einen Wahrnehmungseindruck minimalen Anzahl absorbierter Lichtquanten.

Abschließend bleibt ein Forschungsmittel zu nennen, das kaum je explizit angeführt wird (s. jedoch Kubovy, 1981, S. 78ff.), dem aber für die Theorieentwicklung auch der Psychophysik entscheidende Bedeutung zukommen kann: das Gedankenexperiment. Als rein vorstellungsmäßige Durchführung eines als interessant erachteten qualitativen Experimentes kann das Gedankenexperiment natürlich weder neue Daten noch gänzlich neue Phänomene produzieren. Doch kann es Anomalien, Widersprüche zwischen theoretischen Vorstellungen und Phänomenen oder auch Inkonsistenzen der theoretischen Konzepte selbst aufzeigen (vgl. Kuhn, 1964/1981). Als Gedankenexperimente zum menschlichen visuellen System lassen sich viele der innerhalb des Forschungsbereiches der **Künstlichen Intelligenz** durchgeführten Untersuchungen

zum **computational vision** auffassen, da derartige formale Untersuchungen erkunden, auf welchen Wegen bestimmte Wahrnehmungsziele erreicht werden könnten.

## **8. Implizite Annahmen der experimentellen Psychophysik**

Der Prozeß der Theoriebildung ist in einem steten Fluß begriffen, wenn auch die nachträgliche Rekonstruktion diese Dynamik der Theorieentwicklung einmal eher stockend, einmal eher stürmisch oder gar diskontinuierlich erscheinen läßt. Konfrontiert ein Wissenschaftler an einem bestimmten Punkte dieser Entwicklung Theorieelemente mit Ausschnitten aus einer - wiederum theoretisch gedachten - Empirie, so wird er es kaum als wünschenswert ansehen, alle dabei eingehenden Annahmen, die ihn bei der Konstruktion, Durchführung und Bewertung von Experimenten leiten, explizit zu machen. Ein solches Vorgehen, das schon aus logischen Gründen nicht möglich ist, würde die Theorieevolution nahezu zum Stillstand bringen. Liegen schon die theoretischen Vorstellungen nicht in vollständig explizierter Form vor, so gilt dies mehr noch für die vielfältigen Bindeglieder, durch welche allein Theorieelemente mit spezifischen Experimenten in Beziehung gesetzt werden können. Sie werden erst in dem Maße zum Gegenstand eigenständiger Behandlung, in dem Zweifel an ihrer Brauchbarkeit aufkommen. Theoretische Konzepte, wie sie etwa einem Verweis auf ‚Normalbedingungen‘ zugrunde liegen, appellieren an gewisse Intuitionen und stillschweigende Konventionen, die Teil eines Forschungsprogrammes sind. Auf allen Ebenen des wissenschaftlichen Erkenntnisprozesses bilden stillschweigende Annahmen ganz unterschiedlicher Art ein notwendiges und wesentliches Element dieses Prozesses. Sie betreffen die oben genannten metatheoretischen Aspekte ebenso wie das theoretische Paradigma selbst (z. B. Wahrnehmung als Informationsverarbeitung anzusehen), die Anbindung von Phänomenen an die Theorie, das Experiment und die experimentelle Technik, die verwendeten Apparate, die Daten und ihre Analyse oder die Frage, was als zulässige Abweichung der Daten von den theoretischen Vorhersagen noch zu akzeptieren ist.

Betrachten wir einige Beispiele aus der Psychophysik. Kennzeichnend für diese sind neben bestimmten Arten von Idealisierungen insbesondere die Brückenannahmen, durch die psychophysikalische Befunde mit neurophysiologischen in Beziehung gesetzt werden sollen. Auf die Fülle der auf mehr theoriespezifischer Ebene benötigten stillschweigenden Annahmen können wir ohne Darstellung des jeweiligen theoretischen Kontextes nicht eingehen; sie sind zumeist in komplexer Weise mit einem Forschungsparadigma verwoben. Als Beispiel möge folgendes genügen: Graham (1989) identifiziert in einem Bereich der visuellen Psychophysik 92 solcher Annahmen, die man im Verlauf

der theoretischen Modellierung machen muß, will man etwa Teilmechanismen durch bestimmte Arten von Schwellenexperimenten isolieren.

Zunächst zu Idealisierungen: Als eine nomothetische Disziplin der Psychologie ist die Psychophysik an allgemeinen Gesetzmäßigkeiten und Prinzipien der Wahrnehmung interessiert. Diese lassen sich nur dann aus experimentellen Befunden abstrahieren, wenn bestimmte Idealisierungen zugrunde gelegt werden. Die wichtigste dieser Idealisierungen hängt mit der Fiktion einer ‚Normversuchsperson‘ zusammen, einer Versuchsperson also, für die es auf der Grundlage des verfügbaren Wissens über den untersuchten Gegenstandsbe- reich keine Hinweise dafür gibt, daß sie in ‚entscheidender Weise‘ im unter- suchten Bereich Abweichungen von der überwiegenden Mehrzahl aller an- deren Personen zeigt. Annahmen einer prinzipiellen Strukturgleichheit der Wahrnehmungsprozesse bei ‚normalen‘ Personen erlauben, Experimente mit nur sehr wenigen Versuchspersonen (oftmals nur dem Autor einer Studie) durchzuführen. Zugleich liefern sie eine Rechtfertigung dafür, daß sich die Theoriebildung auf idealisierte **qualitative** Regularitäten stützt, die möglicher- weise für keine einzige Person eine **quantitative** Gültigkeit haben.

Andere Idealisierungen in der Psychophysik hängen mit dem Problem zu- sammen, ‚Anfangsbedingungen‘ präzise festzulegen - in der Physik die wich- tigste Idealisierung zur Formulierung von Modellen - bzw. allgemeiner mit der Schwierigkeit, ceteris paribus-Bedingungen zu formulieren; stets sind nämlich Gedächtnis, Aufmerksamkeitsprozesse, motorische Prozesse, sprach- liche Kategorisierungsprozesse etc. mitbeteiligt (dies sind selbst wiederum theoretische Idealisierungen, denn bei jedem Versuch ist der Mensch als Gan- zes tätig). Die Vorstellung, Wahrnehmungsprozesse ließen sich aus den Tätig- keiten des menschlichen Geistes isolieren und experimentell wie theoretisch getrennt untersuchen, muß selbst dann, wenn sie rein forschungsmethodischen Charakter hat, als eine sehr grobe Idealisierung angesehen werden. Zwar mag sie in dem Maße, wie sich als ähnlich gedachte Prozesse in künstlichen Wahr- nehmungssystemen realisieren lassen, an Plausibilität gewinnen. Je komplexer aber der betrachtete Wahrnehmungszusammenhang ist, umso mehr muß in der Psychophysik die Idee des Isolierens, die so typisch für die Physik ist, als unangemessene Idealisierung erscheinen. Solchermaßen idealisierte Vorstel- lungen einer Isolierbarkeit, in denen noch einmal das oben behandelte Prob- lem einer angemessenen Bestimmung des Reizes anklingt, bringen jedoch für die Theoriebildung eine Vereinfachung mit sich, ohne die die Psychophysik nicht auskommt.

Brückenannahmen stellen eine für die Psychophysik charakteristische Klasse impliziter Annahmen dar. Durch sie wird versucht, psychophysikalische Re- gularitäten und Gesetzmäßigkeiten zu internen Prozessen in Beziehung set- zen. Das Problem, wie sich aus Funktionseigenschaften etwas über Struktur-

eigenschaften erschließen läßt, gleicht abstrakt dem zuvor genannten Invertierungsproblem. Dies betrifft besonders die Interpretation globaler Indizes, wie Reaktionszeiten (s. etwa Pylyshyn, 1984, S. 120ff., Uttal, 1990). Durch (zumeist implizite) Brückenannahmen sollen nun rein psychophysikalische Gesetzmäßigkeiten mit neurophysiologischen Beobachtungen (die selbst wiederum durch Idealisierungen und implizite Annahmen zu gewinnen sind) in Beziehung gesetzt werden (s. Abschnitt 5.5), um so die Klasse der mit bestimmten beobachteten Funktionseigenschaften verträglichen Strukturen einzuschränken. Die Reizspezifität einzelner Neurone, die Organisation rezeptiver Felder, postulierte Mechanismen lateraler Inhibition, Funktionsausfälle bei bestimmten Läsionen u. a. werden zu solchen perzeptuellen Phänomenen in Beziehung gesetzt, von denen man annimmt, daß sie ‚wesentlich‘ durch die jeweiligen neurophysiologischen Eigenschaften ‚bedingt‘ seien. Grobe Brückenannahmen können durchaus eine gewisse Plausibilität haben; auch gibt es einzelne, die gegenwärtig als gut gestützt anzusehen sind. Ein Beispiel hierfür ist die Annahme, daß zwei foveal dargebotene farbige Lichtreize genau dann als gleich erscheinen, wenn die drei Erregungswerte der Farbrezeptoren in beiden Fällen gleich sind. In der Mehrzahl sind solche Brückenannahmen indes mehr oder weniger spekulativ und drücken die Haltung aus, daß Wahrnehmungsphänomene erst mit der Bestimmung eines neuralen Korrelates eine ‚Erklärung‘ finden könnten.

Andere implizite Annahmen beziehen sich auf die Bewertung der Daten: Die Psychophysik unterscheidet sich von anderen Disziplinen der Psychologie nicht zuletzt darin, daß die Kontrolle von Zufallsfehlern größtenteils durch die für die Reizerzeugung und Reizdarbietung verwendete Apparatur erfolgt. Zeigen sich in einem Experiment die erwarteten Effekte, so werden sie in den Daten den Einfluß von Zufallsfehlern in der Regel um Größenordnungen übertreffen. Die Verlagerung der Fehlerkontrolle in die Apparatur bringt mit sich, daß in vielen psychophysikalischen Experimenten eine statistische Fehlerkontrolle unnötig ist. Bei geeigneter graphischer Darstellung der Daten kann man oft mit dem Auge abschätzen, ob sie den theoretischen Vorhersagen entsprechen. Dabei gehen stillschweigende Konventionen darüber ein, welche Abweichungen man - in einem bestimmten Forschungsbereich und nach den gegenwärtigen Kenntnissen über die untersuchten Prozesse - zu tolerieren bereit ist (s. hierzu Kuhn, 1961).

Eine weitere Klasse impliziter Annahmen bezieht sich darauf, wie sich Berichte und Protokolle der Versuchsperson zu den interessierenden Größen der Theorie in Beziehung setzen lassen, ein Problem, das wir bereits im Zusammenhang mit dem Konzept der Beobachtung angesprochen hatten. In psychophysikalischen Experimenten betrachtet man die Urteile einer Versuchsperson - unter stillschweigenden Annahmen über ihre Zuverlässigkeit - so, als könnten es eigene Wahrnehmungsurteile sein. Man setzt also die Reaktio-

nen und Berichte der Versuchsperson zu eigenen Wahrnehmungserfahrungen in Beziehung und sieht die theoretischen Größen als Quasi-Observablen an. Dabei wird man von der Idee geleitet, daß die theoretischen Größen etwas mit eigenen Beobachtungen und eigener Wahrnehmung zu tun haben und nicht nur an externe Beobachtungen des Responseverhaltens von Versuchspersonen geknüpft sind. Erst die Anbindung des ‚internen Beobachters‘ an die Größen der Theorie stiftet die Verbindung der Theorie mit den eigenen Erfahrungen und den Intuitionen über Wahrnehmungsprozesse. Würde man darauf verzichten, verlöre die Theorie eine solche Verankerung und es bliebe nur ein ‚Konstrukt von außen‘ (ein in mancher Hinsicht analoges Problem kognitivistischer Modellbildung diskutiert Herrmann, 1982).

Selbst dort, wo man lediglich einen äußeren Beobachter hat, wie bei Wahrnehmungsuntersuchungen mit Tieren und Säuglingen, bindet man diese äußeren Beobachtungen implizit an Intuitionen an, die erst die Bezeichnung ‚Wahrnehmung‘ rechtfertigen. Wenn man etwa über ‚Wahrnehmung‘ bei Tieren spricht und beispielsweise darüber spekuliert, wie ein Tier, das vier Rezeptortypen für die Farbwahrnehmung hat, wohl Farben wahrnimmt (s. Thompson, Palacios & Varela, 1992; vgl. auch Griffin, 1991), unterstellt man stillschweigend ein gewisse Analogie zur menschlichen Wahrnehmung.

Berichte der Versuchsperson im Sinne ‚interner Beobachtungen‘ gehen also - zumeist indirekt auf der Basis stillschweigender Annahmen - in die Theorieentwicklung mit ein und werden nicht als theoretisch uninteressantes Epiphänomen ihrer Reaktionsweisen angesehen. (Dies bedeutet indes nicht, die introspektiven Beobachtungen als Berichte über die theoretisch zu erfassenden internen Zustände und Wahrnehmungsmechanismen anzusehen.) Die Verständigung von Experimentator und Versuchsperson über reizkorrelierte ‚interne Zustände‘ und die Verknüpfung dieser ‚internen Beobachtungen‘ mit den betrachteten theoretischen Größen setzt implizit weitreichende Annahmen über die Natur mentaler Prozesse voraus. Abermals werden hier die Konturen des Leib-Seele-Problems sichtbar, das jenseits methodologischer Reflektionen zur Psychophysik eigenständiger Gegenstand philosophischer Bemühungen ist. Mögen sich auch wesentliche seiner Aspekte bereits durch logisch-philosophische Analysen als Scheinprobleme erweisen, muß dennoch jeder Versuch, dieses Rätsel ohne engen Kontakt zu den Kognitionswissenschaften, ihren empirischen Befunden und ihrer Theoriebildung zu lösen, steril bleiben.

## **Literatur**

Anstis, S. (1991). Hidden assumptions in seeing shape from shading and apparent motion. In A. Gorea (Ed.), **Representations of vision. Trends and tacit assumptions in vision research.** (S. 279-293). Cambridge: Cambridge University Press.



- Barlow, H. B. (1983). Understanding natural Vision. In O. J. Braddick & A. C. Sleight (Eds.), **Physical and biological information processing of images** (S. 2-14). Berlin: Springer.
- Barlow, H.B. (1985). Perception: What quantitative laws govern the acquisition of knowledge from the senses? In C. W. Coen (Ed.), **Functions of the brain** (S. 11-43). Oxford: Clarendon.
- Barlow, H. B. (1992). The biological role of neocortex. In A. Aertsen & V. Braitenberg (Eds.), **Information processing in the cortex** (S. 53-80). Berlin: Springer.
- Bennett, B.M., Hoffman, D.D. & Prakash, C. (1989). **Observer mechanics. A formal theory of perception**. San Diego: Academic Press.
- Boring, E. G. (1942). **Sensation and perception in the history of experimental psychology**. New York: Appleton-Century.
- Boring, E.G. (1950). **A history of experimental psychology**. (2nd. Ed.). Englewood Cliffs: Prentice Hall.
- Boyd, R. (1979). Metaphor and theory change: What is „metaphor“ a metaphor for? In A. Ortony, (Ed.), **Metaphor and thought** (S. 356-408). Cambridge: Cambridge University Press.
- Brozek, J. & Gundlach, H. (Eds.) (1988). G. I: **Fechner and psychology**. Passau: Passavia Universitätsverlag.
- Brunswik, E. (1934). **Wahrnehmung und Gegenstandswelt. Grundlegung einer Psychologie vom Gegenstand** her. Leipzig: Deuticke.
- Carrier, M. & Mittelstraß, J. (1989). Geist, Gehirn, Verhalten. Das Leib-Seele-Problem und die Philosophie der Psychologie. Berlin: de Gruyter.
- Cassirer, E. (1944). The concept of group and the theory of perception. **Philosophy and Phenomenological Research**, 5, 1-35.
- Churchland, P. S. (1986). **Neurophilosophy**. Cambridge, Mass.: MIT Press.
- Cohen, M.A. & Grossberg, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. **Perception and Psychophysics**, 36, 428-456.
- Craik, K. J. W. (1943). **The nature of explanation**. Cambridge: Cambridge University Press.
- Daugman, J. G. (1990). Brain metaphor and brain theory. In E. L. Schwartz (Ed.), **Computational neuroscience** (S. 9-18). Cambridge, Mass.: MIT Press.
- Dennett, D. C. (1978). **Brainstorms**. Cambridge, Mass.: MIT Press.
- Dennett, D.C. (1983). Intentional Systems in cognitive ethology: The „Panglossian paradigm“ defended. **Behavioral and Brain Sciences**, 6, 343-390.
- Dennett, D.C. (1984). Cognitive wheels: The frame problem of AI. In C. Hookway (Ed.), **Minds, Maschine, and Evolution: Philosophical Studies** (S. 129-151). Cambridge: Cambridge University Press.
- Dennett, D. C. (1991). **Consciousness explained**. London: Penguin.
- DeValois, R. L. & DeValois, K.K. (1988). **Spatial Vision**. Oxford: Oxford University Press.

- Duhem, **P. (1908/1978). Ziel und Struktur der physikalischen Theorien.** Hamburg: Meiner.
- Falmagne, J.C. (1985). **Elements of psychophysical theory.** Oxford: Clarendon.
- Fechner, G. T. (1838). Ueber die subjektiven Complementarfarben. **Poggendorffs Annalen der Physik und Chemie, 44, 221-245.**
- Fechner, G. T. (1840). über die subjective Nachbilder und Nebenbilder. **Poggendorffs Annalen der Physik und Chemie, 50, 193-221; 427-470.**
- Fechner, G. Th. (1851). **ZendAvesta, oder über die Dinge des Himmels und des Jenseits.** Leipzig.
- Fechner, G.Th. (1860). **Elemente der Psychophysik. Bd. I. & II.** Leipzig: Breitkopf & Härtel.
- Fechner, G.Th. (1877). **Zn Sachen der Psychophysik.** Leipzig.
- Fechner, G.Th. (1888). Über die psychischen Maßprinzipien und das Webersche Gesetz. **Philosophische Studien, 4, 161-230.**
- Fodor, J. A. (1974). Special sciences. **Synthese, 28, 75-115.**
- Fodor, J. A., & Pylyshyn, Z. W. (1981). How direct is visual perception? Some reflections on Gibson's 'ecological approach'. **Cognition, 9, 139-196.**
- Franklin (1986). **The neglect of experiment.** Cambridge: Cambridge University Press.
- v. Fritz, K. (1971). **Grundprobleme der Geschichte der antiken Wissenschaft.** Berlin: de Gruyter.
- Garner, W. R. (1962). **Uncertainty and structure as psychological concepts.** New York: Wiley.
- Gigerenzer, G. (1988). Woher kommen Theorien über kognitive Prozesse? **Psychologische Rundschau, 39, 91-100.**
- Gibson, J. J. (1960). The concept of the Stimulus in psychology. **American Psychologist, 15, 694-703.**
- Gould, S. J. & Lewontin, R. C. (1984). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist Programme. In E. Sober (Ed.), **Conceptual issues in evolutionary biology** (S.252-270). Cambridge, Mass.: MIT Press.
- Graham, N. (1989). **Visual pattern analysers.** New York: Oxford University Press.
- Grassmann, H. (1853). Zur Theorie der Farbmischung. **Poggendorffs Annalen der Physik und Chemie, 89, 69-84.**
- Graubard, S. R. (Ed.) (1988). **The artificial intelligence debate. False starts, real foundations.** Cambridge, Mass.: MIT Press
- Griffin, D.R. (1991). Progress toward a cognitive ethology. In C. A. Risteau (Ed.), **Cognitive ethology. The mind of other animals** (S. 3-17). Hillsdale: Erlbaum.
- Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. **Neural Networks, 1, 17-61.**
- Hacking, I. (1983). **Representing and intervening. Introductory topics in the philosophy of the natural sciences.** Cambridge: Cambridge University Press.
- Hanson, N.R. (1969). **Perception and discovery.** San Francisco: Freeman.

- Heider, F. (1926). Ding und Medium. **Symposium**, 1, 109-157.
- Heider, F. (1930). Die Leistung des Wahrnehmungssystems. **Zeitschrift für Psychologie**, **114**, 371-394.
- Helmholtz, H. v. (1855). Ueber das Sehen des Menschen. In H. v. Helmholtz, **Vorträge und Reden**. Bd. 1. (S. 85-117). Braunschweig: Vieweg.
- Helmholtz, H. v. (1894). Über den Ursprung der richtigen Deutung unserer Sinnes-eindrücke. **Zeitschrift für Psychologie und Physiologie der Sinnesorgane**, 7, 81-96.
- Hering, E. (1920). Grundzüge **der Lehre vom Lichtsinn**. Berlin: Springer.
- Herrmann, Th. (1982). Über begriffliche Schwächen kognitivistischer Kognitionstheorien: Begriffsinflation und Akteur-System-Kontamination. **Sprache & Kognition**, **1**, 3-14.
- Herrmann, Th. (1990). Die Experimentiermethodik in der Defensive? **Sprache & Kognition**, **9**, 1-11.
- Hoffman, D.D. & Richards, W. A. (1984). Parts of recognition. **Cognition**, **18**, 65-96.
- Hoffman, W.C. (1978). The Lie transformation group approach to visual neurophysiology. In E. L.J. Leeuwenberg & H. F. Buffart (Eds.), **Formal theories of visual perception** (S. 27-66). Chichester: Wiley.
- Jacob, F. (1977). Evolution as tinkering. **Science**, 196, 1161-1166.
- James, W. (1890). **Principles of psychology**. Vol. 1. New York: Holt.
- Julesz, B. & Chang, J. J. (1976). Interaction between pools of binocular disparity detectors tuned to different disparities. **Biological Cybernetics**, **22**, 107-119.
- Kalke, W. (1969). What is wrong with Fodor and Putnam's functionalism? **Nûs**, 3, 83-93.
- Katz, D. (1911). **Die Erscheinungsweisen der Farben und ihre Beeinflussung durch die individuelle Erfahrung**. Leipzig: Barth.
- Kemke, C. (1988). Der Neuere Konnektionismus. Ein Überblick. **Informatik-Spektrum**, **11**, 143-162.
- Koenderink, J. J. (1980). Why argue about direct perception. **Behavioral and Brain Sciences**, 3, 390-391.
- Koenderink, J. J. (1990). **Solid shape**. Cambridge, Mass.: MIT Press.
- Kohler, I. (1957). Psychophysik heute? **Studium Generale**, **10**, 340-347.
- Krämer, S. (1991). Denken als Rechenprozedur: Zur Genese eines kognitionswissenschaftlichen Paradigmas. **Kognitionswissenschaft**, **2**, 1-10.
- Krantz, D.H. (1972). Measurement structures and psychological laws. **Science**, 175, 1427-1435.
- Krantz, D.H. (1975a). Color measurement and **color** theory: I. Representation theorem for Grassmann structures. **Journal of Mathematical Psychology**, **12**, 283-303.
- Krantz, D. H. (1975b). Color measurement and color theory. II: Opponent-colors theory. **Journal of Mathematical Psychology**, 12, 283-303.
- von Kries, J. (1882). über die Messung intensiver Größen und über das sogenannte psychophysische Gesetz. **Vierteljahresschrift für wissenschaftliche Philosophie**, 6, 257-294.

- Kubovy, M. (1981). Concurrent pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. R. Pomerantz (Eds.), **Perceptual organization** (S. 55-98). Hillsdale, N.J.: Lawrence Erlbaum.
- Kuhn, T. S. (1961). The function of measurement in modern science. In H. Wolf (Ed.), **Quantification. A history of the meaning of measurement in the natural and social sciences** (S. 31-63). Indianapolis: Bobbs-Merrill.
- Kuhn, T.S. (1964/1981). A function for thought experiments. In I. Hacking (Ed.), **Scientific revolutions** (S. 6-27). Oxford: Oxford University Press.
- de La Mettrie, J. O. (1748/1988). **Der Mensch als Maschine**. Nürnberg: LSR-Verlag.
- Laming, D. R.J. (1991). Reconciling Fechner and Stevens? **Behavioral and Brain Sciences**, 14, 188-191.
- Liebmann, S. (1927). über das Verhalten farbiger Formen bei Helligkeitsgleichheit von Figur und Grund. **Psychologische Forschung**, 9, 300-353.
- Livingstone, M. S. & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement and depth. **The Journal of Neuroscience**, 7, 3416-3468.
- Lucas, J.R. (1961). Minds, machines, and Gödel. **Philosophy**, 36, 120-124.
- Luce, R.D. & Krumhansl, C.L. Measurement, scaling, and psychophysics. In R.C. Atkinson, R. J. Herrnstein, G. Lindzey & R. D. Luce (Eds.), **Stevens' handbook of experimental psychology** (S. 1-74). New York: Wiley.
- Mach, E. (1863). Vorträge über Psychophysik. **Oesterreichische Zeitschrift für praktische Heilkunde**, IX, 146-148, 167-170, 202-204, 225-228, 242-245, 260-261, 277-280, 294-298, 316-318, 335-338, 352-354, 362-366.
- Mach, E. (1868). Über die physiologische Wirkung räumlich vertheilter Lichtreize. **Sitzungsberichte der Wiener Akademie der Wissenschaft. Mathematisch-naturwissenschaftliche Classe**, Bd. 57, II. Abt., 11-19.
- Mach, E. (1911/1985). **Die Analyse der Empfindungen**. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Mac Cormack, E. R. (1976). **Metaphor and myth in science and religion**. Durham: Duke University Press.
- Malsburg, C. v. d. (1986). Am I thinking assemblies? In G. Palm & A. Aertsen (Eds.) **Brain theory** (S. 161-176). Berlin: Springer.
- Marr, D. (1977). Artificial intelligence: A personal view. **Artificial Intelligence**, 9, 37-48.
- Marr, D. (1982). **Vision. A Computational investigation into the human representation and processing of visual information**. San Francisco: Freeman.
- Mausfeld, R. (1993). Die Untersuchung von Segmentierungsleistungen in der Farbwahrnehmung durch minimale Reize. Unveröffentlichtes Manuskript.
- Mausfeld, R., & Niederée, R. (1993). An inquiry into relational concepts of colour based on incremental principles of colour coding for minimal relational stimuli. **Perception**.
- Mohyeldin Said, K. A., Newton-Smith, W. H., Viale, R. & Wilkes, K.V. (Eds.) (1990). **Modelling the mind**. Oxford: Clarendon Press.

- Müller, G. E. (1903). Die Gesichtspunkte und die Tatsachen der psychophysischen Methodik. In L. Asher & K. Spiro (Hrsg.), **Ergebnisse der Physiologie, Biophysik und Psychophysik** (S. 267-516). Wiesbaden: Bergmann.
- Ortony, A. (Ed.) (1979). **Metaphor and thought**. Cambridge: Cambridge University Press.
- Pachella, R. G. (1974). The Interpretation of reaction time in information-processing research. In B.H. Kantowitz (Ed.), **Human information processing: Tutorials in performance and cognition** (S.41-82). New York: Wiley.
- Penrose, R. (1989). **The emperor's new mind**. Oxford University Press: New York.
- Poggio, T. (1990). Vision: The 'other' face of AI. In K.A. Mohyeldin Said, W. H. Newton-Smith, R. Viale & K. V. Wilkes, (Eds.) **Modelling the mind** (S. 139-154). Oxford: Clarendon Press.
- Pomerantz, J.R. & Kubovy, M. (1981). Perceptual organization: An overview. In M. Kubovy & J.R. Pomerantz (Eds.), **Perceptual organization** (S. 423-456). Hillsdale, N.J.: Lawrence Erlbaum.
- Pribram, K., Nuwer, M. & Baron, R. (1974). The holographic hypothesis of memory structure in brain function and perception. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), **Contemporary developments in mathematical psychology. Vol. II. Measurement, psychophysics, and neural information processing** (S. 416-457). San Francisco: Freeman.
- Putnam, H. (1960/1975). Minds and machines. In H. Putnam (1975), **Mind, language and reality. Philosophical papers. Vol. 2.** (S. 362-385). Cambridge: Cambridge University Press.
- Putnam, H. (1973). Reductionism and the nature of psychology. **Cognition**, 2, 131-146.
- Putnam, H. (1988). **Representation and reality**. Cambridge, Mass.: MIT Press.
- Pylyshyn, Z. W. (1984). **Computation and cognition. Towards a foundation for cognitive science**. University Press.
- Quine, W.V. (1969). **Ontological relativity and other essays**. New York: Columbia University Press.
- Ramachandran, V.S. (1990). Visual perception in people and machines. In A. Blake & T. Troscianko (Eds.), **AI and the eye** (S.21-77). Chichester: Wiley.
- Ratliff, F., & Sirovich, L. (1978). Equivalence classes of visual stimuli. **Vision Research**, 18, 845-851.
- Readhead, M. (1980). Models in physics. **British Journal of Philosophy of Science**, 31, 145-163.
- Reichardt, W. (1957). Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems. **Zeitschrift für Naturforschung**, 12b, 448-457.
- Reuman, S. R. & Hoffman, D.D. (1986). Regularities of nature: The interpretation of visual motion. In A. P. Pentland (Ed.), **From pixels to predicates: Recent advances in computational and robotic vision** (S. 201-226). Norwood: Ablex.
- Rescher, N. (Ed.) (1986). **Current issues in teleology**. Lanham: University Presses of America.
- Ritter, H., Martinez, T. & Schulten, K. (1990). **Neurale Netze. Eine Einführung in die Neuroinformatik selbstorganisierender Netzwerke**. Bonn: Addison-Wesley.

- Rorty, R. (1979). ***Philosophy and the mirror of nature***. Princeton: Princeton University Press.
- Runeson, S. (1977). On the possibility of „smart“ perceptual mechanisms. ***Scandinavian Journal of Psychology***, *18*, 172-179.
- van Santen, J. P. H. & Sperling, G. (1984). A temporal covariance model of motion perception. ***Journal of the Optical Society of America***, *A1*, 451-473.
- Searle, J. (1984). ***Minds, brains and science***. Cambridge, Mass.: Harvard University Press.
- Scheerer, E. (1992). Fechner's inner psychophysics: Its historical fate and present status. In H. G. Geissler, S. W. Link & J. T. Townsend (Eds.), ***Cognition, information processing, and psychophysics: Basic issues*** (S. 3-21). Hillsdale: Erlbaum.
- Shepard, R.N. (1978). The circumplex and related topological manifolds in the study of perception. In S. Shye (Ed.), ***Theory construction and data analysis in the behavioral sciences*** (S.29-80). San Francisco: Jossey-Bass.
- Shepard, R.N. (1982). Perceptual and analogical bases of cognition. In J. Mehler, E. C.T. Walker & M. Garrett (Eds.), ***Perspectives on mental representations*** (S. 49-67). Hillsdale: Lawrence Erlbaum.
- Shepard, R. N. (1984). Ecological constraints in internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. ***Psychological Review***, *91*, 417-447.
- Shepard, R.N. (1987). Evolution of a mesh between principles of the mind and regularities of the world. In J. Dupré (Ed.), ***The latest on the best. Essays on evolution and optimality*** (S.251-275). Cambridge, Mass.: Bradford.
- Singer, W. (1990). Search for coherence: A basic principle of cortical self-organization. ***Concepts of Neuroscience***, *1*, 1-26.
- Sloman, A. (1983). Image interpretation: The way ahead? In O. J. Braddick & A. C. Sleigh (Eds.), ***Physical and biological processing of images*** (S.380-401). Berlin: Springer.
- Smart, J. J. C. (1964). ***Philosophy and scientific realism***. London: Routledge.
- Smolensky, P. (1988). On the proper treatment of connectionism. ***Behavioral and Brain Sciences***, *11*, 1-23.
- Sober, E. (Ed.) (1984). ***Conceptual issues in evolutionary biology***. Cambridge, Mass.: MIT Press.
- Sperling, G. (1970). Binocular Vision: A physical and neural theory. ***American Journal of Psychology***, *83*, 461-534.
- Sperling, G. (1981). Mathematical models of binocular Vision. In S. Grossberg (Ed.), ***Mathematical psychology and psychophysiology*** (S. 281-300). Providence: American Mathematical Society.
- Stevens, S.S. (1951). Mathematics, measurement, and psychophysics. In S.S. Stevens (Ed.), ***Handbook of experimental psychology*** (S. 1-49). New York: Wiley.
- Suppes, P., Krantz, D.H., Luce, R. D. & Tversky, A. (1989). ***Foundations of measurement. Vol. II. Geometrical, threshold, and probabilistic representations***. New York: Academic Press.

- Tack, W.H. (1983). Psychophysische Methoden. In H. Feger & J. Bredenkamp (Hrsg.), **Enzyklopädie der Psychologie. Messen und Testen** (S. 346-426). Göttingen: Hogrefe.
- Thompson, E., Palacios, A. & Varela, F. J. (1992). Ways of coloring. **Behavioral and Brain Sciences**, *15*, 1-74.
- Ullman, S. (1980). Against direct perception. **Behavioral and Brain Sciences**, *3*, 373-415.
- Uttal, W.R. (1967). Evoked brain potentials: Signs or codes? **Perspectives in Biology and Medicine**, *10*, 627-639.
- Uttal, W. R. (1990). On some two-way barriers between models and mechanisms. **Perception & Psychophysics**, *48*, 188-203.
- van Fraassen, B. C. (1989). **Laws and symmetry**. Oxford: Clarendon.
- Vartanian, A. (1953). **Diderot and Descartes. A study of scientific naturalism in the enlightenment**. Princeton: Princeton University Press.
- Webb, J. C. (1980). **Mechanism, mentalism, and metamathematics**. Dordrecht: Reidel.
- Woodward, J. (1989). Data and phenomena. **Synthese**, *79*, 393-472.
- Wyszecki, G. & Stiles, W. S. (1982). **Color science. Concepts and methods, quantitative data and formulae**. (2nd. Ed.). New York: Wiley.

## Modellierung mit Hilfe wissensbasierter Systeme

***Klaus Opwis und Hans Spada***

Darum ist ja die künstliche Intelligenzforschung für die kognitive Psychologie so interessant: Ihre Programme bilden Handlungssysteme . . . so exakt und in so durchsichtiger Weise ab, daß sie Theorien gleichen, die man laufen lassen kann.

(Aebli, 1980)

### ***1. Prolog: Von Brücken aus Bausteinen, Ökosystemen in Form von Differentialgleichungen und Problemlöseprozessen auf Rechnern***

Ein Kind erzählt von einer Brücke. Sie sei groß und verbinde die Ufer eines Flusses; sie glänze in der Sonne und wenn schwere Lastwagen über sie führen, dann bebe sie. Daheim, in seinem Spielzimmer, beginnt das Kind, eine Brücke zu bauen. Es hat einen Baukasten mit vielen verschiedenen Elementen, so daß es in seinem kleinen Modell Pfeiler nachbauen kann, tragende Balken verwenden, Verzierungen anbringen usw. Es ist ganz in sein Tun vertieft, denn es macht Spaß, Brücken zu konstruieren und sei es auch nur aus Bausteinen. Ohne Zweifel hat das Kind sein Wissen über Brücken und über diese eine, besonders faszinierende, sowohl in der verbalen Kommunikation wie auch in der Rekonstruktion mit Elementen des Baukastens zum Ausdruck gebracht. Im letzteren Fall schafft es sich eine Welt, mit der es experimentieren kann; so prüft es beispielsweise, unter welchen Bedingungen sein Bauwerk zusammenbricht. Falls das Kind tatsächlich versucht, die gesehene Brücke nachzubauen, wird der Erfolg umso größer sein, je besser es sich die Brücke eingepägt hat, je reichhaltiger der Baukasten mit verschiedenen Bestandteilen ausgestattet ist und je kompetenter das Kind damit umgehen kann. Natürlich bleibt die Rekonstruktion ein Modell, zugleich ist es aber ein Beispiel für eine Brücke. Das Modell ist eine Brücke.

Ein Biologe untersucht die Belastung eines Binnensees durch Abwässer. Er ist insbesondere an Fragen der Eutrophierung interessiert. Umfassende Da-



tenreihen zu wichtigen biotischen und abiotischen Variablen werden erhoben. Der Biologe spezifiziert ein System von Differentialgleichungen, das die Abhängigkeit der verschiedenen Komponenten abbildet und damit Aspekte der Veränderung des Ökosystems See über die Zeit zu simulieren gestattet. Prognosen über den zukünftigen Entwicklungsverlauf werden möglich. Modell-daten können auch in dem Sinne experimentell generiert werden, als unter Voraussetzung unterschiedlicher Annahmen über die zukünftige Abwasserbelastung bestimmte Eutrophierungsverläufe resultieren. Die Prüfung der empirischen Gültigkeit des Modells erfolgt über den Vergleich simulierter Daten mit (weiteren) Beobachtungsdaten, aber auch anhand verschiedener Plausibilitätsprüfungen. Wichtige Voraussetzungen für ein valides Modell sind vor allem die Kompetenz des Biologen in Fragen seines Gegenstandsbereichs und der mathematischen Systemanalyse, die Qualität und der Umfang der erhobenen Daten, die in die Modellierung eingehen, und die Differenziertheit, mit der etwa die Dynamik zeitkontinuierlich ablaufender Prozesse im verwendeten mathematischen Kalkül abgebildet werden kann. Allerdings ist die Simulation eines Binnensees mit Hilfe mathematisch-numerischer Modelle nicht selbst ein Binnensee, da das Simulierte und seine Simulation in diesem Fall auf unterschiedlichen Prinzipien beruhen.

Eine Denkpsychologin beobachtet Probanden bei der Bearbeitung des ‚Turm von Hanoi‘-Problems. Sie stellt fest und berichtet dies auch in ihrer Forschungsgruppe, daß die Schwierigkeit der Problemlösung mit der Anzahl der Scheiben zunehme, und zwar insbesondere bei ungetübten Personen. Viele fehlerhafte Züge seien zu beobachten, aber sie habe schon eine Vermutung, in welchen Fällen das Erkennen einer Sackgasse von zentraler Bedeutung für die Aufdeckung des optimalen Vorgehens sei. An ihrem Schreibtisch beginnt die Psychologin, den Problemlöseprozeß, der dem beobachteten Verhalten zugrundeliegen könnte, auf dem Rechner zu rekonstruieren. Sie wählt eine symbolische Repräsentation für das Problem, spezifiziert auf dieser Grundlage das Vorwissen der simulierten Probanden und entwickelt einen Lösungs- und Lernalgorithmus, der in einer geeigneten Programmiersprache als wissensbasiertes System auf dem Rechner implementiert wird. Das Ergebnis ist eine Modellierung, für die die Wissenschaftlerin prüft, ob sie das beobachtete Lösungsverhalten in wichtigen Charakteristika korrekt generiert. Darüber hinaus testet sie weitere, aus der Modellierung abgeleitete Annahmen anhand zusätzlicher empirischer Untersuchungen. Die Modellierung wird um so valider sein, je detaillierter die Information ist, die empirisch über diesen kognitiven Prozeß vorliegt. Wichtig sind die Adäquatheit und Reichhaltigkeit der formalen Sprache, mit deren Hilfe die Modellierung erfolgt, aber auch die Kompetenz der Psychologin im Umgang mit den Elementen dieser Sprache. Natürlich bleibt die Rekonstruktion ein Modell, aber sie ist zugleich ein Beispiel für einen Problemlösungsprozeß.

Können zwischen dem Nachbauen einer Brücke mit den Bestandteilen eines Baukastens, der Modellierung der Eutrophierung eines Sees anhand von Differentialgleichungssystemen und der Rekonstruktion eines Problemlöseprozesses mit Hilfe eines wissensbasierten Systems Analogien gezogen werden?

Allen drei Beispielen ist gemeinsam, daß eine ausschließlich natürlichsprachliche Erfassung des Gegenstandsbereiches, sei es in der Umgangssprache oder einer Fachsprache, zu kurz greift. Erst das Modell aus Bausteinen oder aber die formalisierten Prozeßmodelle erlauben einen aktiven Umgang mit der Modellierung. Bedingungsmanipulationen und Prognosen werden möglich.

Das Brückenbaubeispiel und den Fall der auf dem Rechner simulierten Problemlösungsprozesse verbinden in der Phase der Modellerstellung das ingenieurmäßige, konstruktive Vorgehen in der Modellierung und der geringe Grad an Datenrestriktion (vgl. Tack, in diesem Band). Es handelt sich um ein eher interpretatives Vorgehen. Während keine Parallelen zwischen einem kindlichen Vorgehen und denkpsychologischen Forschungsstrategien gezogen werden sollen, ist der Bezug zwischen ‚Das Modell einer Brücke ist eine Brücke‘ und ‚Die rekonstruktive Modellierung einer Problemlösung ist eine Problemlösung‘ durchaus intendiert. Für die Ökosystemmodellierung und die denkpsychologische Simulation gilt, daß es sich in beiden Fällen um formale Modelle handelt, allerdings im ersten Fall um ein mathematisch-numerisches Modell, im zweiten Fall um ein wissensbasiertes Modell. Beides sind Prozeßmodelle. Eine hohe Datenrestriktion charakterisiert die Phase der Modellgeltungsprüfung.

Halten wir fest: Geistige Prozesse, die uns in der Reflexion bewußt werden, die wir aber auch als Erklärung für das Verhalten anderer heranziehen, lassen sich in unterschiedlicher Weise, umgangssprachlich, fachsprachlich oder formal beschreiben. Wir können Aussagen über ihre Funktion formulieren, über ihr Ziel und über die Bedingungen, unter denen sie auftreten. Darüber hinaus können wir sie aber auch modellieren, indem wir sie formal fassen und die Formalisierung auf einem Rechner in einer geeigneten Programmiersprache implementieren. Man kann dabei Prozesse abbilden, die zu korrekten, aber auch andere, die zu fehlerhaften Ergebnissen führen, man kann eine bestimmte Person in einer definierten Situation vor Augen haben oder aber kognitive Prozesse in prototypischer Weise abzubilden versuchen. Unter dem Blickwinkel der Psychologie bedarf die Modellierung auf dem Rechner der Validierung anhand empirischer Daten. Ziel ist nicht nur eine Rekonstruktion, die bei gegebenen Eingaben zum selben Ergebnis führt, das beim Menschen beobachtet wurde, sondern auch die Nachbildung wesentlicher Ablaufcharakteristika.

## 2. Modellierung als Rekonstruktion

Das Denken und Handeln von Personen beruht auf ihren Deutungen und Auslegungen der sie umgebenden Welt. Personen konstruieren durch Vorgänge der Wahrnehmung und des Sprachverstehens sowie aufgrund von (Vor-) Wissen, Erwartungen und Zielsetzungen eine interne, subjektive Repräsentation von Ausschnitten ihrer tatsächlichen oder einer vorgestellten Umgebung. Diese **Konstruktionen erster Ordnung** oder **mentalen Modelle'** sind als Rekonstruktion der Umgebung wie jede Art der Modellbildung eine Abstraktion. Dies impliziert im Besonderen ihre prinzipielle Unvollständigkeit. Darüber hinaus sind sie nicht unmittelbar beobachtbar. Der Beobachtung zugänglich sind lediglich Verhaltensdaten, aus denen auf das zugrundeliegende mentale Modell geschlossen werden muß. Damit wird die psychologisch valide Rekonstruktion des mentalen Modells einer Person zur zentralen Zielsetzung der psychologischen Theoriebildung. Die in wissenschaftlicher Perspektive zu erarbeitenden **Modelle mentaler Modelle** sind in dieser Hinsicht **Konstruktionen zweiter Ordnung**.

Die Thematik dieses Kapitels ist eine spezielle Methodik zur Rekonstruktion mentaler Modelle, ihre **Modellierung mit Hilfe wissensbasierter Systeme**. Die theoretische Rahmenvorstellung ist das Paradigma der Informationsverarbeitung, das die kognitiv orientierte Psychologie seit Ende der sechziger Jahre dominiert. Grundannahme ist, daß geistige Aktivitäten wie induktive und deduktive Denkvorgänge, Problemlösen, Planungsprozesse, Sprechen und Sprachverstehen, aber auch die Wahrnehmung der Umwelt, als **wissensbasierte Vorgänge der Symbolverarbeitung** aufgefaßt werden können. Diese Betrachtungsweise wird in den weiteren Ausführungen auf der Grundlage des Konzepts wissensbasierter Systeme elaboriert. Ziel ist die Explikation zentraler methodologischer Annahmen und Argumente, die der Modellierung mentaler Modelle durch wissensbasierte Systeme zugrundeliegen.

Damit wird in diesem Kapitel die Auffassung vertreten, daß die Analyse geistiger Aktivitäten als symbolische Berechnungsvorgänge und ihre korrespondierende Rekonstruktion mit technischen Mitteln aufgrund der dazu erforder-

1 Der Begriff ‚mentales Modell‘ wird in der kognitiven Psychologie keineswegs einheitlich verwendet (vgl. auch Herrmann, 1988). In der vorliegenden Arbeit ist damit ein Ausschnitt der internen Repräsentation der äußeren Welt eines Subjekts gemeint, die es ihm ermöglicht, Phänomene zu verstehen, Vorhersagen zu treffen und Inferenzen durchzuführen. Ein wesentliches Charakteristikum ist die Möglichkeit von ‚inneren Simulationen‘, etwa in Form von Schlußfolgerungen und mentaler Probehandlungen. über die An der Repräsentation werden keine Annahmen getroffen. Demgegenüber betrachtet etwa Johnson-Laird (1983) mentale Modelle als eine spezifische Form der Wissensrepräsentation, der sog. analogen Repräsentation (in Abgrenzung von propositionalen und regelbasierten Repräsentationssystemen). Kennzeichen ist bei ihm eine erfahrungsmäßige Komponente, einhergehend mit Vorstellungsbildern, in denen räumliche Relationen zwischen Objekten und/oder zeitliche Relationen zwischen Ereignissen repräsentiert sind,

derlichen lückenlosen Formalisierung und kompromißlosen Explizitheit besonders geeignet sind, zentrale inhaltliche und methodologische Fragen offenzulegen. Zusammenfassend wird eine Position eingenommen, wie sie etwa von Johnson-Laird (1988) formuliert wurde.

Is the mind a computational phenomenon? No one knows. . . . what is clear is that **computability provides an appropriate conceptual apparatus for theories of the mind** (Hervorhebung von den Verfassern). This apparatus takes nothing for granted that is not obvious . . . Yet, . . . any clear and explicit account of, say, how people recognize faces, reason deductively, create new ideas or control skilled actions can always be modelled by a computer program.

(Johnson-Laird, 1988, S. 51f.)

Danach sind Computer nützliche analytische Werkzeuge, um unser Verständnis kognitiver Vorgänge beim Menschen zu verbessern, indem - wie so oft in der Wissenschaftsgeschichte - eine neuartige, aber bekannte Technologie zur Grundlage unseres Verstehens des Unbekannten gemacht wird (vgl. Gigerenzer, 1988, 1991). Computermodelle erlauben die Formulierung und Prüfung theoretischer Vorstellungen in einer Form, die durch die grundlegenden wissenschaftstheoretischen Ideale der **Präzision, Widerspruchsfreiheit** und **Vorhersagbarkeit** gekennzeichnet ist. Die These ist, daß die Psychologie mit den Möglichkeiten der Modellierung geistiger Prozesse am Rechner ein Werkzeug in die Hand bekommen hat, das mächtiger ist als die bisher verfügbaren Ansätze der Modellbildung.

Da in diesem Kapitel methodologische Fragen im Vordergrund der Betrachtung stehen, soll zumindestens über Literaturhinweise auch ein kurzer Überblick über Inhalte kognitiver Modellierungen gegeben werden.

Eine historisch orientierte, breit angelegte Darstellung der Kognitionswissenschaft enthält Gardner (1985). Eine Sammlung wichtiger ‚klassischer‘ Arbeiten findet man in Collins und Smith (1988). Das Buch von Johnson-Laird (1988) stellt ein einführendes, an Inhalte der Allgemeinen Psychologie anknüpfendes Lehrbuch dar, während die von Posner (1989) bzw. Osherson (Osherson & Lasnik, 1990; Osherson, Kosslyn & Hollerbach, 1990; Osherson & Smith, 1990) herausgegebenen Bände handbuchartig in die Kognitionswissenschaft einführen. Ein Überblick zur Rolle der kognitiven Modellierung in der psychologischen Theoriebildung findet sich bei Opwis (1992).

Es folgt eine Auflistung grundlegender klassischer, aber auch beispielhafter moderner Arbeiten zu ausgewählten inhaltlichen Themenbereichen. Zum **Problemlösen und Lernen** sind insbesondere die Arbeiten von Newell und Simon sowie von Anderson zu nennen (etwa: Newell, Shaw & Simon, 1958; Newell & Simon, 1963, 1972; Anzai & Simon, 1979; Anderson, 1982, 1983, 1987). Daneben sei auf die von Holland, Holyoak, Nisbett und Thagard (1986) vorgeschlagene Theorie induktiven Denkens und Lernens sowie auf die Arbeiten

von VanLehn zum Erwerb einfacher prozeduraler Fertigkeiten und zur Rolle des Verstehens von Beispielen für den Erwerb von Problemlöseprozeduren und deren Transfer auf neue Probleme hingewiesen (vgl. VanLehn, 1989; VanLehn, Jones & Chi, 1992). Vielzitierte Autoren zur Wissensrepräsentation sind Quillian (1968) sowie Collins und Loftus (1975) zu semantischen Netzen, Minsky (1975) zu „Frames“ oder Schank und Abelson (1977) zum Skriptkonzept. Eine Sammlung grundlegender Arbeiten findet sich in Brachman und Levesque (1985), ein modernes Lehrbuch ist Davis (1990). Aus der Fülle von einschlägigen Arbeiten zur **Sprache** seien Chomsky (1963), Winograd (1972, 1983) und Allen (1987) herausgegriffen. Berwick (1985) und Pinker (1989) behandeln das Thema Spracherwerb. Informativ orientierte Arbeiten zur Wahrnehmung sind etwa Marr (1982) und Hildreth und Ullman (1989). Schließlich seien noch die Veröffentlichungen von Anderson (1983) und Newell (1990; vgl. auch Laird, Newell & Rosenbloom, 1987) zur Frage der **Architektur des menschlichen kognitiven Apparats** genannt.

Viele wichtige kognitionswissenschaftliche Originalarbeiten werden in der Zeitschrift **Cognitive Science** veröffentlicht. Seit 1991 gibt es auch eine deutschsprachige Zeitschrift **Kognitionswissenschaft**. Daneben sind eine Reihe weiterer Fachzeitschriften zu nennen, wie das **Journal of the Learning Sciences**, **User modeling and User-adapted Interaction**, **Minds and Machines**, sowie mit einer stärkeren KI-Orientierung etwa **Artificial Intelligence**, **Machine Learning** oder **Journal of Artificial Intelligence in Education**. Darüber hinaus veranstaltet die Cognitive Science Society jährliche Konferenzen, deren Tagungsberichte als **Proceedings of the Conference of the Cognitive Science Society** veröffentlicht werden.

Am Anfang der hier angesprochenen Entwicklung der Kognitionswissenschaft, etwa Mitte der fünfziger Jahre, haben Psychologen und Informatiker Hand in Hand gearbeitet. Etwa seit Mitte der siebziger Jahre dominiert zunehmend der Beitrag der Informatik, zumindestens hinsichtlich der Weiterentwicklung der begrifflichen und methodischen Grundlagen. Die Psychologie muß alle Anstrengungen unternehmen, damit sie auch künftig einen gestaltenden Anteil an einer multidisziplinären kognitiven Wissenschaft hat.

Die weiteren Inhalte und der Aufbau des Beitrags lassen sich wie folgt zusammenfassen. Im dritten Abschnitt folgt eine Darstellung ausgewählter theoretischer und formaler Grundlagen. Damit soll eine Verständigung über die Grundbegriffe und die Untersuchungsperspektive erreicht werden. Anhand eines Beispiels wird das Problem der Rekonstruktion menschlichen Wissens mit technischen Mitteln illustriert. Dies führt zu einer Explikation des Konzepts wissensbasierter Systeme und zur **Repräsentationshypothese**. Danach sind wissensbasierte Systeme algorithmische Modelle, die durch die Verwendung und Manipulation formaler Symbolstrukturen gekennzeichnet sind. Es

folgt eine Präzisierung des **Algorithmenbegriffs** und des damit aufs engste verbundenen Konzepts der **Berechenbarkeit**. Danach steht die Verwendung wissensbasierter Systeme als kognitionspsychologische Modelle im Mittelpunkt. Das Verhalten, die Funktionalität und die Struktur technischer, aber auch biologischer Systeme lassen sich auf unterschiedlichen Betrachtungsebenen beschreiben, analysieren und erklären. Dazu wird eine psychologisch motivierte **Mehrebenenbetrachtung** informationsverarbeitender Systeme entwickelt, die einen konstruktiven Umgang mit dem in diesem Zusammenhang zentralen Problem der Vermischung von Ebenen ermöglicht.

Fragen der Bewertung wissensbasierter Systeme als psychologische Modelle werden im vierten Abschnitt behandelt. Dabei wird insbesondere das Kriterium der empirischen Adäquatheit betrachtet und es werden die verschiedenen Funktionen diskutiert, die empirische Befunde im Prozeß der Theorieformulierung und -prüfung erfüllen.

Einige wissenschaftstheoretische Anmerkungen und ein Epilog runden das Kapitel ab.

### 3. Grundlagen wissensbasierter Systeme

Nach der allgemeinen Modelltheorie (Stachowiak, 1973) sind **Modelle** immer Modelle von **etwas**, Abbildungen, Repräsentationen bzw. Rekonstruktionen natürlicher oder künstlicher Originale, die ihrerseits auch wieder Modelle sein können (vgl. hierzu auch den Modellbegriff der (Prädikaten-) Logik; z.B. Genesereth & Nilsson, 1987). Modelle sind dabei ihren Originalen nicht per se zugeordnet, sondern sie erfüllen ihre Funktionen stets für bestimmte Erkenntnissubjekte und/oder Akteure und relativ zu bestimmten Zielen und Zwecken. Hinsichtlich des Mediums, in dem die Modellbildung vorgenommen wird, lassen sich zwei große Klassen von Modellen unterscheiden (vgl. Gigerenzer, 1981): **Symbolische** und **analoge** Modelle.

Bei **symbolischen Modellen** erfolgt die Modellbildung in einem Medium, das - meist per Definition - in eine Menge von elementaren Informationseinheiten (Symbole) zerlegbar ist und über eine Menge von Vorschriften zur Zusammensetzung von komplexen Symbolstrukturen verfügt. Dies impliziert die Eigenschaft der **Grammatikalität** bzw. eine Definition von **syntaktischer Wohlgeformtheit**. Die wichtigsten Unterklassen sind **formale Modelle** numerischer (etwa: probabilistische mathematische Modelle) oder **nicht-numerischer Art** (etwa: logische Kalküle oder mengen- und graphentheoretische Vorstellungen) sowie **natürlich-sprachliche Modelle** (etwa: Theorien auf der Grundlage einer mehr oder weniger präzise definierten **fachwissenschaftlichen Terminologie** oder der **Alltagssprache**).

Hinsichtlich ihrer Brauchbarkeit als wissenschaftliche Modelle weisen formale gegenüber sprachlich formulierten Modellen eine Reihe von Vorteilen auf (Vermeidung von widersprüchlichen Annahmen, Nachweis von lückenhaften Annahmen, Präzisierung der verwendeten Begrifflichkeit). Ausführlich werden die Funktion formaler Modelle, ihre spezifischen Vorteile und Probleme etwa in Deppe (1977) oder Neches (1982) erörtert. Die Unterschiede zwischen numerischen und nicht-numerischen Modellen liegen nicht nur in der Art ihrer Formulierung. Mathematische Modelle auf der Grundlage statistischer oder wahrscheinlichkeitstheoretischer Überlegungen sind Modelle von Daten. Sie zielen zumeist auf ausgewählte und direkt beobachtbare Aspekte psychischer Aktivitäten wie etwa Reaktionszeiten. Sie können in diesem Sinne auch als **„oberflächliche“** Theorien nach Moravcsik (1980) charakterisiert werden. Nicht-numerische Modelle thematisieren dagegen als **„tiefte“** Theorien die dem Verhalten von Personen als zugrundeliegend angenommenen und prinzipiell nicht unmittelbar beobachtbaren psychischen Strukturen und Prozesse selbst. In der Psychologie hatten mathematische Modelle insbesondere in den fünfziger und sechziger Jahren ihre Blütezeit (vgl. das dreibändige Handbuch der mathematischen Psychologie; Luce, Bush & Galanter, 1963), während nicht-numerische Modelle seit den siebziger Jahren dominieren. Allerdings ist in Gestalt des sog. **„Neueren Konnektionismus“** gegenwärtig eine Renaissance einer Klasse numerischer Modelle erkennbar, die Analogien zu neuronalen Prozessen der Informationsverarbeitung aufweisen. Auf die theoretischen Grundlagen dieser, in der Regel als nicht-lineare Differenzen- oder Differentialgleichungssysteme formulierten Modelle kann im Rahmen dieses Beitrags nicht eingegangen werden (vgl. dazu etwa McClelland & Rumelhart, 1986, 1988; Rumelhart & McClelland, 1986; Smolensky, 1988; Strube, 1990).

Die in diesem Beitrag näher ausgeführte Methode zur Rekonstruktion kognitiver Strukturen und Prozesse ist ihre formale, nicht-numerische Modellierung (vgl. Tack, 1987). Das Problem der Modellbildung kann umschrieben werden als das Problem der adäquaten Rekonstruktion „von strukturell-operativen Zusammenhängen menschlichen Wissens in Algorithmen und Datenstrukturen“ (Scheffe, 1986, S.155). Die Besonderheiten und Anforderungen dieser Form der Modellbildung werden im folgenden Abschnitt anhand eines bekannten gedächtnispsychologischen Beispiels, den sog. semantischen Netzwerkmodellen, illustriert. Dabei geht es hier nicht primär um den spezifischen Inhalt dieser Modelle. Vielmehr soll aufgezeigt werden, daß bei der Modellbildung drei unterschiedliche, aber eng aufeinander-bezogene Ebenen relevant sind: Die Ebene der Konzeptualisierung als Formulierung theoretischer Vorstellungen mit Hilfe psychologischer Begriffe, die Ebene der Formalisierung dieser Vorstellungen mit formal-logischen Kalkülen und die Ebene der Implementation als wissensbasierte Systeme.

### 3.1 Konzeptuelle Modelle, formale Kalküle und wissensbasierte Systeme am Beispiel semantischer Netzwerkmodelle

Im Zusammenhang mit der Frage, wie das sprachlich-begriffliche Wissen einer Person im Gedächtnis gespeichert ist, taucht in der psychologischen Literatur immer wieder der Begriff des **semantischen Netzes** auf. Allerdings werden von verschiedenen Autoren sehr unterschiedliche Modellvorstellungen unter diesen Begriff subsumiert (vgl. Opwis & Lüer, in Druck). Gemeinsam ist der theoretische Anspruch, daß sie ein **Modell der Organisation sprachlicher Gedächtnisinhalte** darstellen. In der Mehrzahl der Modelle wird als zugrundeliegende formale Struktur ein **gerichteter Graph** (bzw. allzu oft auch nur eine entsprechende graphische Notation) verwendet. Die Knoten und Kanten eines Graphen stehen für sprachliche Einheiten und Relationen. Die Modelle unterscheiden sich nun insbesondere dadurch, von welchen elementaren sprachlichen Einheiten (insbesondere Begriffe oder Sätze bzw. Propositionen) sie ausgehen, welche klassifikatorischen (etwa Unter-/Oberbegriff), modifikatorischen (Attributzuschreibung), syntaktischen (Nominal-, Verbalphrase usw.) oder semantischen (Handlung, Akteur, Objekt u. ä.) Relationen zugrundegelegt werden und welche Prozesse (Vergleichs-, Inferenzprozesse oder Aktivierungsvorgänge) im Mittelpunkt der Analyse stehen.

#### **3.1.1 Konzeptualisierung: Abstraktionshierarchien als psychologische Modellvorstellung**

Eine der wichtigsten Anwendungen semantischer Netzwerkmodelle ist die Beschreibung von Taxonomien, das sind hierarchisch geordnete Klasseneinteilungen eines Gegenstandsbereichs. Man spricht in diesem Zusammenhang auch von **Abstraktionshierarchien**. Ein bekanntes, in der einschlägigen Literatur seit Anfang der siebziger Jahre immer wieder bemühtes Beispiel ist die folgende, in Abbildung 1 veranschaulichte Abstraktionshierarchie von Tierarten (vgl. Klix, 1988; Möbus, 1988).

Wie man unmittelbar erkennt, ist die Abstraktionshierarchie eine netzartige Struktur, die auf einer Einteilung in Ober- bzw. Unterklassen (bzw. -begriffe) im Sinne einer ‚ist-Teilmenge-von‘ Relation beruht. Beispielsweise gelangt man ausgehend von der Klasse der Lachse in zwei Abstraktionsschritten zur Klasse der Tiere (über Fisch). Die minimale Länge des Weges zwischen zwei Begriffen definiert ihre semantische Distanz. Eine bestimmte Klasse ist nun durch eine Menge von Attributen, Merkmalen oder Eigenschaften definiert. Abstraktion bedeutet Elimination von Merkmalen. Je abstrakter eine Klasse ist, desto weniger charakterisierende Merkmale weist sie auf. Wichtig ist in diesem Zusammenhang aber die Vorstellung **kognitiver (Speicher-) Ökonomie**.



Informationen werden nicht mehrfach abgespeichert, sondern anschaulich ausgedrückt an dem ‚obersten‘ Knoten, für den sie Gültigkeit haben, angehängt und im Bedarfsfall ‚nach unten‘ weitergereicht. Die Abstraktionshierarchie bildet somit zugleich Implikationen und mögliche Inferenzwege ab. Eigenschaften, die auf Oberbegriffe zutreffen, können über diese Wege auf alle ihre Unterbegriffe vererbt werden. Beispielsweise kann eine Frage der Art ‚Hat ein Lachs Kiemen?‘ auf diese Weise positiv beantwortet werden (Da jeder Lachs ein Fisch ist).

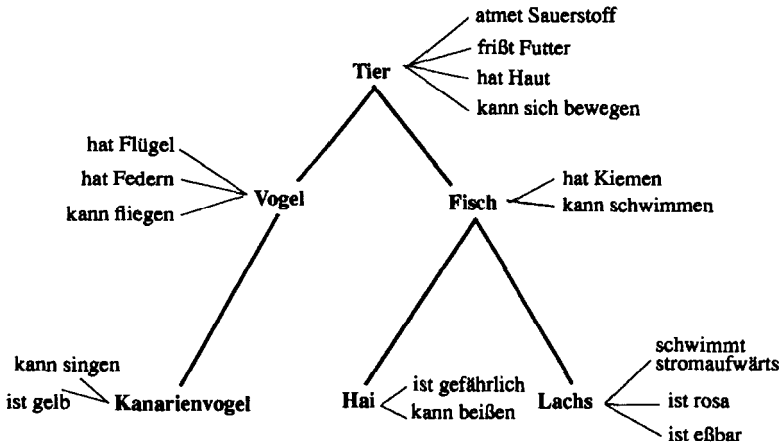


Abb. 1: Taxonomie von Tierarten als Beispiel für eine Abstraktionshierarchie (nach Collins & Quillian, 1969)

Im nächsten Schritt soll die bisher verwendete psychologische Begrifflichkeit präzisiert werden. Dazu ist eine Festlegung der **Objekte** erforderlich, über die Aussagen gemacht werden sollen, sowie die Definition von **Relationen** zwischen den Objekten und von **Operationen** auf den Objekten bzw. Relationen. Dieser Vorgang wird auch als **Konzeptualisierung** bezeichnet (vgl. Genesereth & Nilsson, 1987).

Für dieses Beispiel liegt es nahe, die relevanten Objekte mit den Knoten bzw. mit den in der Abstraktionshierarchie vorkommenden Begriffen zu identifizieren:

{Tier, atmet-Sauerstoff, . . . , Vogel, ...}

Die Menge der in einer Konzeptualisierung betrachteten Objekte definiert zugleich ihren sog. Diskursbereich. Eine Relation ist eine spezielle Beziehung zwischen Objekten des Diskursbereichs. Beispielsweise bilden die Kanten in Abbildung 1 zwei unterschiedlichen Arten von Relationen ab, eine klassifikatorische Relation im Sinne der Unter-/Oberbegriffsrelation und eine modi-

fikatorische Relation, die einem Begriff eine spezielle Eigenschaft zuordnet. Damit korrespondieren die beiden folgenden Mengen von Tupeln:

$$\{ \langle \text{Vogel}, \text{Tier} \rangle, \langle \text{Fisch}, \text{Tier} \rangle, \dots \}$$

und  $\{ \langle \text{Tier}, \text{atmet\_Sauerstoff} \rangle, \langle \text{Tier}, \text{frisst\_Futter} \rangle, \dots \}$

Diese beiden Relationen stellen keineswegs die einzig möglichen Relationen zwischen den betrachteten Begriffen dar. Eine Konzeptualisierung, die aus einer bestimmten Menge von Objekten und einer Menge von betrachteten Relationen besteht, ist prinzipiell unvollständig. Unter Umständen verhindert die gewählte Konzeptualisierung die Darstellung bestimmter Sachverhalte. Ein bekanntes Beispiel hierfür ist die Kontroverse in der Physik über die Natur von Licht. Je nachdem, ob Licht als ein Teilchen- oder als ein Wellenphänomen konzeptualisiert wird, können spezielle Aspekte des Verhaltens von Licht erklärt werden, aber keine der Sichtweisen reicht allein aus. Erst die Entwicklung einer weiteren Konzeptualisierung auf der Grundlage der modernen Quantentheorie hat diese Schwierigkeiten beseitigt, indem sie beide Sichtweisen zusammenführte.

### **3.1.2 Formalisierung: Prädikatenlogische Darstellung von Abstraktionshierarchien**

Haben wir uns für eine Konzeptualisierung entschieden, so können wir die Formalisierung des Wissens vornehmen. Dazu wollen wir im folgenden Abschnitt eine formale Sprache verwenden, das sog. Prädikatenkalkül bzw. die Prädikatenlogik. Die wesentlichen Vorteile sind:

- die Unterscheidung von Variablen und Konstanten, wobei Variablen benötigt werden, um Eigenschaften von Objekten darzustellen, ohne dabei diese Objekte explizit benennen zu müssen;
- die Verwendung logischer Operatoren wie ‚ $\wedge$ ‘ (Konjunktion), ‚ $\vee$ ‘ (Disjunktion) und ‚ $\Rightarrow$ ‘ (Implikation), die es ermöglichen, aus einfachen komplexe Aussagen zu bilden, ohne dabei die Wahrheit oder Falschheit der einfachen Aussagen angeben zu müssen;
- die Verwendung von Quantoren wie ‚ $\forall$ ‘ (Allquantor) und ‚ $\exists$ ‘ (Existenzquantor), die es ermöglichen, über alle Objekte Aussagen zu machen, ohne die Objekte im einzelnen aufzählen zu müssen bzw. die Existenz eines Objektes mit bestimmten Eigenschaften anzunehmen, ohne das Objekt selbst identifizieren zu müssen;
- eine präzise Definition der Bedeutung prädikatenlogischer Ausdrücke (denotative Semantik).

Zur Formalisierung unseres Beispiels benötigen wir nicht das gesamte Sprachrepertoire der Prädikatenlogik, sondern nur ausgewählte Bestandteile. Dabei werden Konstanten bzw. Variablen im folgenden durch eine Folge aus Buchstaben oder Ziffern dargestellt, wobei das erste Zeichen bei einer Konstanten ein Großbuchstabe und bei einer Variablen ein Kleinbuchstabe ist. Die Tatsache, daß der Begriff ‚Vogel‘ (als Objektkonstante) ein Unterbegriff des Begriffs ‚Tier‘ ist, kann dann durch die Einführung eines konstanten Relationensymbols ‚Ist-ein‘ in der folgenden Weise notiert werden:

Ist\_ein(Vogel,Tier)  
Ist\_ein(Fisch,Tier) usw.

Entsprechend kann die zweite Relationsart durch die Einführung eines weiteren Relationensymbols ‚Hat\_Eigenschaft‘ notiert werden:

Hat\_Eigenschaft(Tier,Atmet\_Sauerstoff)  
Hat\_Eigenschaft(Tier,Frißt\_Futter) usw.

Auf diese Weise können die betrachteten Relationen zwischen Begriffen als eine Menge von **Propositionen** im Sinne logischer Axiome formalisiert werden.

Allerdings sind mit dieser Formalisierung noch keine Inferenzvorgänge spezifiziert. Fragen der Art ‚Ist ein Lachs ein Tier?‘ oder ‚Hat ein Lachs Kiemen?‘ können nicht beantwortet werden. Dazu sind Operationen im Sinne logischer Implikationen erforderlich, die Beziehungen zwischen Relationen definieren. Sie legen fest, wie aus gegebenen Fakten (Axiome oder Prämissen) neue Aussagen (Theoreme oder Konklusionen) abgeleitet werden können. Die Ober-/Unterbegriffsrelation zwischen zwei beliebigen Begriffen einer Abstraktionshierarchie läßt sich rekursiv mit Hilfe zweier Regeln **O1** (Terminierungsfall) und **O2** (rekursiver Fall) definieren:

- (O1) begriff1 ist ein Unterbegriff von begriff2, wenn zwischen beiden Begriffen eine ‚Ist-ein‘ Relation besteht.
- (O2) begriff1 ist ein Unterbegriff von begriff2, wenn zwischen begriff1 und einem begriff3 eine ‚Ist-ein‘ Relation besteht und begriff3 ein Unterbegriff von begriff2 ist.

Die prädikatenlogische Formalisierung dieser Regeln ist:

$$\text{Qx Qy} \quad \text{Ist\_ein}(x,y) \Rightarrow \text{Ist\_Unterbegriff}(x,y)$$

$$\text{Qx Qy} \quad \forall z \text{ Ist\_ein}(x,z) \wedge \text{Ist\_Unterbegriff}(z,y) \Rightarrow \text{Ist\_Unterbegriff}(x,y)$$

Zur vollständigen Konzeptualisierung der Abstraktionshierarchie fehlen noch Inferenzregeln zur Vererbung von Eigenschaften. Dazu sind ebenfalls zwei Regeln erforderlich:

- (O3) begriff1 besitzt die Eigenschaft x, wenn zwischen begriff1 und x eine ‚Hat\_Eigenschaft‘ Relation besteht.
- (O4) begriff1 besitzt die Eigenschaft x, wenn begriff1 ein Unterbegriff von begriff2 ist und zwischen begriff2 und x eine ‚Hat\_Eigenschaft‘ Relation besteht.

Die korrespondierende prädikatenlogische Formalisierung dieser Regeln ist:

$$\forall x \forall y \quad \text{Hat\_Eigenschaft}(x,y) \Rightarrow \text{Besitzt\_Eigenschaft}(x,y)$$

$$\forall x \forall y \forall z \quad \text{Ist\_Unterbegriff}(x,z) \wedge \text{Hat\_Eigenschaft}(z,y) \Rightarrow \text{Besitzt\_Eigenschaft}(x,y)$$

Die vorgenommene logikorientierte Konzeptualisierung einer Domäne kann nun im nächsten Schritt als ein wissensbasiertes System implementiert werden. Im Beispiel bietet sich dazu die Programmiersprache PROLOG an (vgl. Sterling & Shapiro, 1986; Malpas, 1987).

### 3.1.3 Implementation: Abstraktionshierarchien als wissensbasierte Systeme

Ein PROLOG-Programm besteht im wesentlichen aus einer Wissensbasis, die alle bekannten Fakten (Axiome) und Regeln (Implikationen) umfaßt. Dabei sind folgende syntaktischen Unterschiede zur logischen Formalisierung im vorigen Abschnitt zu beachten: Regeln und Fakten werden mit einem Punkt abgeschlossen, ihre Namen beginnen ebenso wie konstante Argumente mit einem Kleinbuchstaben, während Variablen mit einem Großbuchstaben beginnen; die konjunktive Verknüpfung von Teilzielen wird durch ein Komma dargestellt. Tabelle 1 enthält die vollständige Implementation einer Abstraktionshierarchie als PROLOG-Programm.

Tabelle 1: Implementation einer Abstraktionshierarchie als PROLOG-Programm

---

<b>% Relationen der Abstraktionshierarchie als PROLOG-Fakten</b>	
ist_ein(vogel, tier).	ist_ein(fisch, tier).
ist_ein(kanarienvogel, vogel).	ist_ein(hai, fisch).
ist_ein(lachs, fisch).	
hat_Eigenschaft(tier, atmet-Sauerstoff).	hat_Eigenschaft(tier, frißt-Futter).
hat_Eigenschaft(tier, hat_Haut).	hat_Eigenschaft(tier, kann_sich-bewegen).
hat_Eigenschaft(vogel, hat_Flügel).	hat_Eigenschaft(vogel, hat_Federn).
hat_Eigenschaft(vogel, kann_fliegen).	hat_Eigenschaft(fisch, hat_Kiemen).
hat_Eigenschaft(fisch, kann_schwimmen).	hat_Eigenschaft(kanarienvogel, ist_gelb).
hat_Eigenschaft(kanarienvogel, kann_singen).	hat_Eigenschaft(hai, ist_gefährlich).
hat_Eigenschaft(hai, kann_beißen).	hat_Eigenschaft(lachs, ist_eßbar).
hat_Eigenschaft(lachs, schwimmt_stromaufwärts)	
hat_Eigenschaft(lachs, ist_rosa).	

---

Tabelle 1 (Fortsetzung)

---

<b>% Regeln zur Definition der Unter-/Oberbegriffsrelation zwischen zwei Begriffen</b>	
ist_Unterbegriff(Unterbegriff, Oberbegriff) :-	ist_ein(Unterbegriff, Oberbegriff)
ist_Unterbegriff(Unterbegriff, Oberbegriff) :-	ist_ein(Unterbegriff, Begriff), ist_Unterbegriff(Begriff, Oberbegriff)
% Regeln zur Vererbung von Eigenschaften	
besitzt_Eigenschaft(Begriff, Eigenschaft) :-	hat_Eigenschaft(Begriff, Eigenschaft).
besitzt_Eigenschaft(Begriff, Eigenschaft) :-	hat_Eigenschaft(Oberbegriff, Eigenschaft).

---

Berechnungen werden in PROLOG als deduktive Berechnungsvorgänge aufgefaßt. Dazu ist neben der Wissensbasis ein domänenunabhängiger, allgemeiner Inferenzmechanismus erforderlich, der auf der Menge der Fakten und Regeln im Sinne eines verallgemeinerten Modus Ponens operiert und durch eine Anfrage angestoßen wird, etwa:

```
?- ist_ein(Objekt,tier).
```

In dieser Anfrage ist ‚Objekt‘ eine (PROLOG-) Variable. Das PROLOG-System prüft, ob die Anfrage aus der Wissensbasis ableitbar ist. Dies ist möglich, wenn die Variable instantiiert werden kann, d.h. wenn in der Wissensbasis eine Konstante gefunden wird, für die sie im Sinne eines Platzhalters stehen kann. Aus Tabelle 1 ist ersichtlich, daß die Variable ‚Objekt‘ durch die Konstante ‚vogel‘ instantiierbar ist. Durch Eingabe eines Semikolons kann eine Anfrage fortgesetzt werden, um evtl. weitere Antworten zu erhalten.

```
?- ist_ein(Objekt,tier).
```

```
Objekt = vogel;  
Objekt = fisch;  
no
```

Beispiele für weitere Anfragen sind etwa für „Ist ein Hai ein Tier?“

```
?- ist_Unterbegriff(hai,tier).
```

oder

```
?- ist_Unterbegriff(Begriff,Klasse).
```

Die letzte Anfrage liefert alle in der Wissensbasis direkt oder indirekt vorhandenen Unter-/Oberbegriffsrelationen. Die nächste Anfrage liefert alle Eigenschaften der Klasse der Kanarienvögel als Antwort, wobei auch die bei den Oberbegriffen nicht redundant gespeicherten Eigenschaften inferiert werden:

```
?- besitzt_Eigenschaft(kanarienvogel,Eigenschaft).
```

Wie das Prolog-Programm in Tabelle 1 illustriert, sind wissensbasierte Systeme im Unterschied zu herkömmlichen Programmen durch eine weitgehende Ex-

plizierung und Modularisierung ihrer Wissensbestände charakterisiert. Wichtig ist dabei, daß prozedurale Wissensbestände hinsichtlich ihrer Aktivierung a priori keinerlei Festlegungen unterworfen sind. Operatoren enthalten typischerweise die Prüfung von Bedingungen, unter denen sie prinzipiell anwendbar sind. Dadurch ergibt sich eine regelartige Struktur: ‚Wenn Bedingung, dann Aktion‘. Einzelne Wissenseinheiten werden aufgrund ihrer spezifischen Inhalte aktiviert, also aufgrund eines assoziativen Zugriffs und nicht entsprechend einer vorgegebenen starren Abarbeitungssequenz.

Eine zentrale Eigenschaft wissensbasierter Systeme im Sinne ablauffähiger Computerprogramme ist ihre **operative Interpretierbarkeit**. Es resultiert ein neuer Typ von Daten: Das beobachtbare Modellverhalten als Verhaltenssimulation, das ausschließlich auf der Grundlage des Modells abgeleitet wird. Mit anderen Worten formuliert: Wissensbasierte Systeme sind **Prozeß- und Performanzmodelle**, insofern sie die **Generierung von Modelldaten** ermöglichen.

Zusammenfassend läßt sich festhalten, daß wissensbasierte Systeme gekennzeichnet sind durch die Verwendung formaler syntaktischer Strukturen, die als repräsentations sprachliche Ausdrücke oder Symbolstrukturen interpretiert werden können. Sie repräsentieren als Propositionen, was das System über eine Anwendungsdomäne ‚weiß‘. Diese Sichtweise hat Smith (1982) in seiner Hypothese der **Wissensrepräsentation** (knowledge representation hypothesis) herausgestellt:

Any mechanically embodied intelligent process will be comprised of structural ingredients that a) we as external observers naturally take to represent a propositional account of the knowledge that the overall process exhibits, and b) independent of such external semantic attribution, play a formal but causal and essential role in engendering the behavior that manifests that knowledge.

(Smith, 1982, S.2)

### 3.2 Algorithmen, Berechenbarkeit und symbolverarbeitende Systeme

Aus der Sicht des Paradigmas der Symbolverarbeitung, das Newell als „the most fundamental contribution“ (1980, p. 136) ansieht, die von seiten der Informatik bzw. KI bisher in die Kognitionswissenschaft eingebracht wurde, können mentale Modelle mit Hilfe wissensbasierter Systeme rekonstruiert werden, wobei im Mittelpunkt der Betrachtung die Manipulation intern repräsentierter Symbolstrukturen steht.

Wo liegen die Ursachen für die Attraktivität dieser Betrachtungsweise? Zwei Gesichtspunkte sind wesentlich: Die Eigenschaften der **Immaterialität** und der **Berechnungsuniversalität** (vgl. auch Pylyshyn, 1989).

Analysieren wir zur Explikation dieser Eigenschaften einige Beispiele:

- Ist die Addition zweier natürlicher Zahlen ‚berechenbar‘?
- Ist die Klimaentwicklung ‚berechenbar‘?
- Sind menschliche kognitive Vorgänge wie Wahrnehmung, Problemlösen, Sprachverarbeitung ‚berechenbar‘?

Offenkundig hängt die Beantwortung dieser und vergleichbarer Fragen einerseits von unserem Verständnis des Begriffs ‚berechenbar‘ und andererseits der angesprochenen Vorgänge ab. Intuitiv unstrittig dürfte auf den ersten Blick allein die positive Beantwortung der ersten Frage sein. Die Addition zweier natürlicher Zahlen ist zweifellos berechenbar, aber warum bzw. in welcher Hinsicht eigentlich? Eine naheliegende, nichts desto trotz zu kurz greifende Antwort ist der Hinweis, daß es schließlich ‚Maschinen‘ und ‚Automaten‘ der unterschiedlichsten Art gibt, die eine Addition zweier natürlicher Zahlen durchführen können (z. B. Abakus, Taschenrechner, Computer). Damit Maschinen funktionieren, müssen sie zuvor entsprechend ‚instruiert‘ oder ‚programmiert‘ worden sein, d.h. entscheidend ist die Existenz (mindestens) eines Algorithmus, der genau angibt, wie eine Addition durchzuführen ist. Ein **Algorithmus** wird üblicherweise definiert als **ein durch eine endliche Vorschrift eindeutig festgelegtes effektives Verfahren zur Lösung einer Klasse von Problemen**. Ein Algorithmus akzeptiert beliebige zulässige Werte als Eingabe, verarbeitet diese wie in der Verfahrensbeschreibung festgelegt, so lange, bis das Ende der Berechnungen erreicht ist und das Ergebnis vorliegt. Effektivität meint dabei, daß ein Algorithmus - angewendet auf zulässige Eingabewerte - nach endlich vielen Schritten anhält und das ‚richtige‘ Ergebnis zurückliefert. Die eindeutige Festlegung durch eine endliche Vorschrift impliziert, daß ein Algorithmus in mechanischer Weise von einem symbolverarbeitenden System, einer Maschine oder einem Programm abgearbeitet werden kann.

Dabei ist ein **symbolverarbeitendes System** durch die folgenden grundlegenden Bestandteile definiert (Newell, 1980):

- eine **Menge primitiver Symbole** oder Zeichen (das ‚Alphabet‘ des Systems);
- eine **Menge von Regeln** zur Generierung komplexer, zusammengesetzter Symbolstrukturen oder Zeichenketten (die ‚Grammatik‘ des Systems);
- ein (oder mehrere) **Speicher**, in dem Symbole/Symbolstrukturen abgelegt und aus dem sie abgerufen werden können (das ‚Gedächtnis‘ des Systems);
- eine **Menge grundlegender Operationen** zur Verarbeitung der Symbole/Symbolstrukturen (etwa: Erzeugung, Zugriff, Veränderung, Löschung);
- ein (oder mehrere) **Prozessor(en)**, der die Arbeitsweise des Systems steuert und kontrolliert.

Nach dieser Definition sind symbolverarbeitende Systeme zunächst in einem abstrakten, theoretischen Sinn logisch-formale Systeme. Sie dürfen nicht mit real existierenden Rechenmaschinen, speziellen Programmiersprachen oder konkret implementierten Computerprogrammen verwechselt oder gleichgesetzt werden.

Die Konzipierung einer speziellen **Architektur des menschlichen kognitiven Apparats** als Spezialfall eines symbolverarbeitenden Systems ist eine zentrale Aufgabe der kognitionspsychologischen Theoriebildung. Eine **kognitive Architektur** ist ein spezifisches System von Speicherstrukturen („Gedächtnisse“) und Operationen („Verarbeitungsvorgänge“), das den Rahmen zur Rekonstruktion kognitiver Leistungen festlegt. Bekannte Beispiele kognitiver Architekturen sind das ACT\*-Modell von Anderson (1983) und das System SOAR (vgl. Laird, Rosenbloom & Newell, 1987; Newell, 1990). Eine allgemeine Diskussion des Architekturproblems, einschließlich einer vergleichenden Gegenüberstellung von ACT\* und SOAR findet sich bei Newell, Rosenbloom und Laird (1989).

Bisher wurde der Begriff der Berechenbarkeit informell verwendet. Ein Vorgang ist berechenbar, wenn es einen Algorithmus gibt, der von einem symbolverarbeitenden System ausgeführt werden kann. Um zu einer präzisen Operationalisierung des Algorithmusbegriffs zu gelangen, können drei Zugangsweisen unterschieden werden (vgl. Krämer, 1988, für eine geschichtlich orientierte Darstellung formaler Kalküle).

**Präzisierung mit Hilfe des mathematischen Funktionsbegriffs.** Die Arbeiten von Gödel (1931), Church (1936), Kleene (1936) und Turing (1936) auf dem Gebiet der mathematischen Logik legten die Grundlagen für diese Art der Präzisierung des Algorithmusbegriffs mit Hilfe des Konzepts der berechenbaren Funktionen. Sie führten zur Definition grundlegender Klassen berechenbarer Funktionen. Die wichtigsten Klassen sind die **elementaren Funktionen** (etwa: Identitäts-, konstante und Nachfolgefunktion), die **primitiv rekursiven Funktionen** und die **p-rekursiven Funktionen**. Umfassende Darstellungen der mathematischen Berechenbarkeits- oder Rekursionstheorie finden sich etwa in den Lehrbüchern von Minsky (1967) oder Rogers (1987).

**Präzisierung mit Hilfe formaler Sprachen/Grammatiken.** Bei dieser Betrachtungsweise, die auf Autoren wie Post (1943), Markov (1954) oder Chomsky (etwa Chomsky, 1963) zurückgeht, werden die Ein- und die Ausgabe von Algorithmen als Worte eines jeweils passenden Alphabets aufgefaßt. Ein Algorithmus ist dabei charakterisiert als eine Menge von Regeln zur schrittweisen Manipulation von Zeichenreihen. Man unterscheidet zwischen **regulären, kontextfreien, kontextsensitiven** und **allgemeinen Sprachen/Grammatiken**.



**Präzisierung mit Hilfe von Klassen von Maschinen.** Die Informatik kennt und studiert eine Reihe wichtiger Klassen von Maschinen oder Automaten als Spezialfälle (abstrakter) symbolverarbeitender Systeme (vgl. Hopcroft & Ullman, 1979; Stetter, 1988). Insbesondere sind dies die Klassen der **endlichen Automaten**, der **Kellerautomaten**, der **linear beschränkte Automaten** und - als die wohl bekannteste Klasse - die **Turingmaschinen**.

Im Grunde handelt es sich bei der Turingmaschine um ein **abstraktes Gedankenmodell**, das die Durchführung theoretischer Analysen zur Beantwortung grundlegender Fragen über die Natur von Rechenprozessen erlaubt. Für praktische Berechnungen ist sie nicht geeignet. Die Bedeutung der Turingmaschine für die Theorie der Berechenbarkeit ergibt sich daraus, daß sie - einen genügend großen, aber endlichen Zeitraum vorausgesetzt - jeden Algorithmus und damit jede Berechnung durchführen kann. Diese **Eigenschaft der Berechnungsuniversalität** und die damit einhergehende Definierbarkeit der algorithmischen Berechenbarkeit, unabhängig von speziellen Klassen von Maschinen, ist von fundamentaler Bedeutung. Einerseits beruht darauf die umfassende Anwendbarkeit programmierbarer Rechenmaschinen zur algorithmischen Lösung der unterschiedlichsten Probleme; andererseits gestattet sie, den Nachweis der algorithmischen Lösbarkeit eines Problems anhand der am besten geeigneten Klasse von Maschinen zu führen.

Anzumerken bleibt, daß von den oben genannten Klassen von Maschinen lediglich die Turingmaschine die Eigenschaft der Berechnungsuniversalität besitzt. Die übrigen Klassen von Maschinen sind weniger leistungsfähig. Für den Zusammenhang von Sprachen/Grammatiken und Maschinentypen gilt: Reguläre Sprachen sind äquivalent zu endlichen Automaten, kontextfreie bzw. kontextsensitive Sprachen sind äquivalent zu Kellerautomaten bzw. zu linear beschränkten Automaten und allgemeine Sprachen sind äquivalent zu Turingmaschinen.

Allerdings bedeutet Berechnungsuniversalität keineswegs, daß alle Vorgänge berechenbar sind. Es gibt Probleme, die **nicht** algorithmisch gelöst werden können. Bekanntestes Beispiel für diese Form der Unvollständigkeit formaler Systeme ist das sog. Halteproblem. Man kann zeigen, daß es keinen Algorithmus gibt, der ‚berechnet‘, ob eine beliebige Turingmaschine anhält, also nach endlich vielen Schritten stehenbleibt oder nicht (vgl. Rogers, 1987, weitere Beispiele nicht berechenbarer Funktionen findet man in Brauer, 1990).

Bei vielen Fragestellungen kommt es für die Frage der Berechenbarkeit darüber hinaus entscheidend auf den **Auflösungsgrad** an, mit dem ein Vorgang betrachtet wird. Berechenbarkeit auf einem geringen Auflösungsgrad impliziert dabei keineswegs Berechenbarkeit auf einem höheren Auflösungsgrad. So bedeutet etwa eine systematische Regelmäßigkeit in aggregierten Daten nicht unbedingt eine Vorhersagbarkeit des Einzelfalls. Je ‚genauer‘ die Vor-

hersage der künftigen Entwicklung sein soll, desto eher wird man den zugrundeliegenden Vorgang als nicht berechenbar, bzw. als beim gegenwärtigen Kenntnisstand nicht berechenbar bezeichnen. Beispielsweise gilt hinsichtlich der Frage nach der Berechenbarkeit langfristiger Klimaentwicklungen: Modelle einer globalen Klimaveränderung existieren, aber sie erlauben keine präzisen Aussagen über lokale Veränderungen. Dabei geht die Meteorologie durchaus von der Annahme aus, daß die Entwicklung des Klimas global und lokal systematischen Regelmäßigkeiten gehorcht, die man zwar im Detail noch nicht kennen mag, aber prinzipiell ‚entdecken‘ und mit Hilfe geeigneter Modelle rekonstruieren könnte. Berechenbarkeit und Vorhersagbarkeit können in diesen und vergleichbaren Fragestellungen auch als Relation zwischen dem vorherzusagenden Geschehen und dem aktuellen Wissen über eben dieses Geschehen aufgefaßt werden. Danach wird ein Geschehen als ‚nicht berechenbar‘ bezeichnet, wenn kein ausreichendes Wissen zu seiner Vorhersagbarkeit vorhanden ist.

Wie sieht es aber nun mit der Berechenbarkeit kognitiver Leistungen des Menschen aus? Zur Beantwortung dieser Frage wird mit Bezug auf Newell (1982) eine Mehrebenenbetrachtung von Theorien der Informationsverarbeitung eingeführt.

### 3.3 Mehrebenenbetrachtung von Theorien der Informationsverarbeitung

Das Verhalten, die Funktionalität und die Struktur vieler technischer, aber auch biologischer Systeme lassen sich auf verschiedenen Ebenen beschreiben, analysieren und erklären. Häufig liegt das Problem gerade ‚darin‘, die richtige Ebene der Beschreibung und Erklärung zu finden. Welche Ebenen sind zu unterscheiden, wenn menschliche kognitive Aktivitäten mit Hilfe wissensverarbeitender Systeme rekonstruiert werden?

Dazu wurden in den letzten Jahren verschiedene Vorschläge gemacht, insbesondere von Marr (1982), Newell (1982) und Pylyshyn (1985), aber auch von Anderson (1990). Wegen ihres nachhaltigen Einflusses auf die gesamte Grundlagendiskussion sollen die Überlegungen von Newell im folgenden kurz vorgestellt werden.

**Newells Mehrebenenbetrachtung.** Ausgangspunkt für Newells Überlegungen war die Frage, welcher ‚Wissensbegriff‘ der KI überhaupt zugrundeliegt und damit insbesondere das Problem, welcher Zusammenhang zwischen Wissen und seiner Rekonstruktion in formalen Strukturen besteht (Newell, 1982; vgl. auch Daiser, 1984). Newell unterscheidet drei grundlegende Ebenen der Betrachtung, die **Wissensebene**, eine **Programm- bzw. Symbolebene** und eine phy-

**sikalische Ebene.** Die einzelnen Ebenen bilden dabei für Newell eine Hierarchie mit zwei wesentlichen Eigenschaften. Sie sind autonom definierbar, aber dennoch gibt es Abhängigkeitsbeziehungen zwischen ihnen in dem Sinne, daß jede Ebene auf der unmittelbar unter ihr angesiedelten Ebene aufsetzt. Dies impliziert, daß jede Ebene - zumindestens im Prinzip - auf die unter ihr liegende Ebene zurückgeführt bzw. ‚reduziert‘ werden kann. Eine derartige, **analytisch-reduktionistische Vorgehensweise** ist eine Voraussetzung für die Bildung von Simulationsmodellen. Wichtig ist in diesem Zusammenhang, daß diese Reduktion keineswegs eindeutig ist. So kann etwa eine theoretische Vorstellung in der Regel auf unterschiedliche Weise formalisiert werden. Ebenso kann ein Algorithmus in verschiedener Weise implementiert werden.

Erklärungen der Funktionalität und des Verhaltens eines Systems sind jeweils innerhalb einer Ebene zu entwickeln. Eine Vermischung von Betrachtungsebenen impliziert keine besseren Erklärungschancen, sondern führt zu begrifflicher Verwirrung! Zu Recht wird beispielsweise immer wieder darauf hingewiesen, daß theoretische Vorstellungen einerseits und Computerprogramme andererseits nicht miteinander verwechselt werden dürfen (vgl. Herrmann, 1982). Darüber hinaus stehen Vorstellungen, die sich auf verschiedene Ebenen eines Systems beziehen, nicht unmittelbar in Konkurrenz zueinander - im Unterschied zu solchen Vorstellungen, die beanspruchen, ein System innerhalb ein- und derselben Ebene zu erklären.

Kehren wir zu den Überlegungen Newells zurück. Oberhalb der Programm- oder Symbolebene liegt als ‚oberste‘ Ebene die Wissensebene. Ihre Existenz wird von Newell als Hypothese formuliert (**knowledge level hypothesis**). In Abhebung von den übrigen Ebenen lassen sich die Besonderheiten der Wissensebene in zwei charakteristischen Eigenschaften zusammenfassen. Erstens ist die Wissensebene nach Newell auf eine radikale Art unvollständig. Dies impliziert, daß Verhalten nicht vollständig beschreibbar und vorhersagbar ist.

Sometimes behavior can be predicted by the knowledge level description; often it cannot. The incompleteness is not just a failure in certain special situations or in some small departures. The term radical is used to indicate that entire ranges of behavior may not be describable at the knowledge level, but only in terms of systems at a lower level (namely, the symbolic level). However, the necessity of accepting this incompleteness is an essential aspect of this level.

(Newell, 1982, S. 105)

Zweitens ist Wissen nach Newell ein Medium, das keinerlei spezifische materielle Zustandsbeschreibung besitzt, sondern lediglich funktional definiert werden kann.

**Knowledge.** Whatever can be ascribed to an agent, such that its behavior can be computed according to principle of rationality.

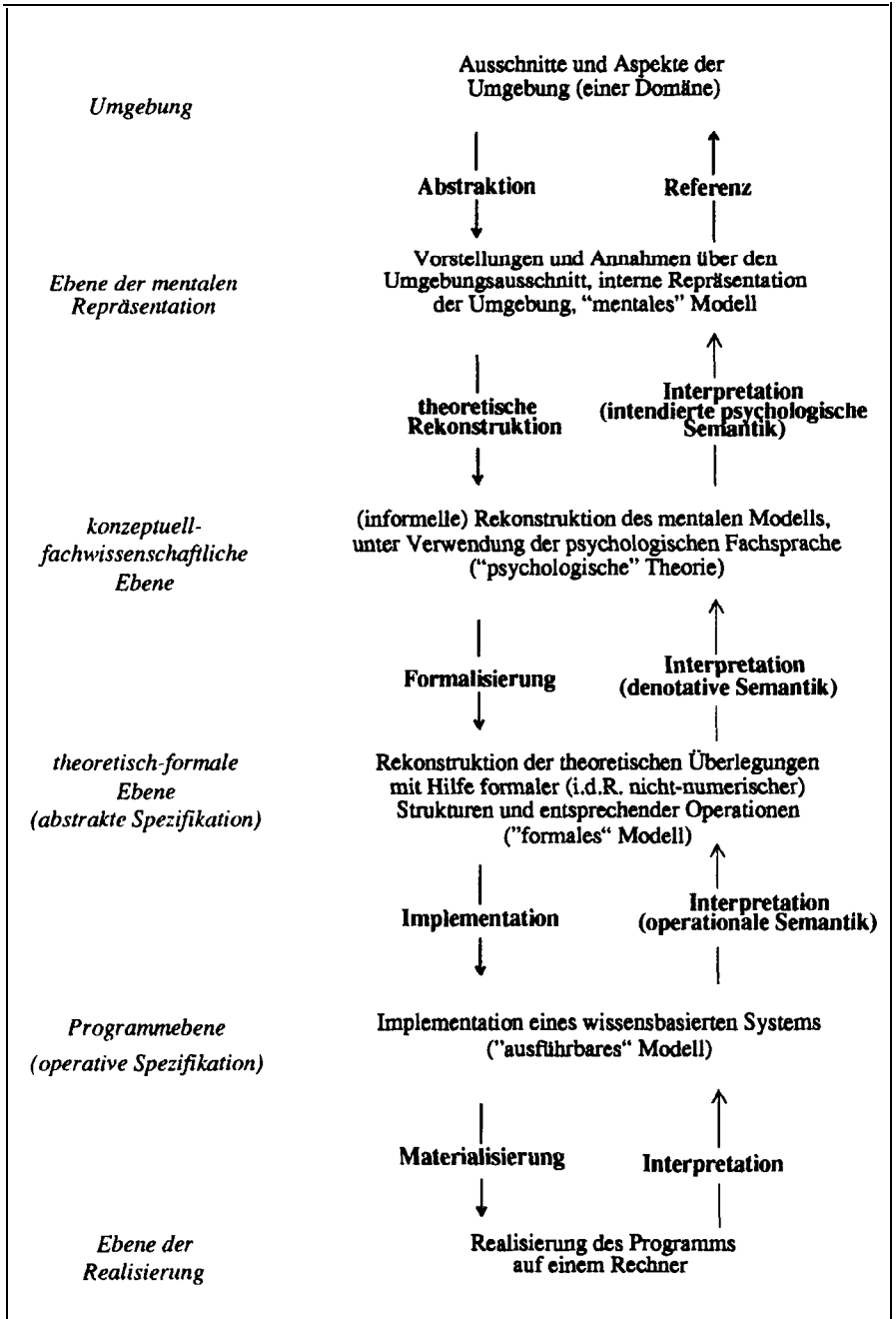


Abb. 2: Mehrebenenbetrachtung von Theorien der Informationsverarbeitung aus psychologischer Sicht

Knowledge is to be characterized entirely **functionally**, in terms of what it does, not **structurally**, in terms of physical objects with particular properties and relations.

(Newell, 1982, S.105)

Wissen wird auf diese Weise zu einem Kompetenzbegriff im Sinne eines Potentials, das verhaltenswirksam werden kann, aber nicht muß und das nicht passiv, sondern nur aktiv im Prozeß der Auswahl und Anwendung realisiert ist. Eine indirekte und approximate materielle Struktur bekommt Wissen erst durch seine Rekonstruktion in Form wissensverarbeitender Systeme auf der Programm- oder Symbolebene.

Abbildung 2 faßt einige Überlegungen aus diesem und den vorangehenden Abschnitten in Form einer Mehrebenenbetrachtung von psychologischen Theorien der Informationsverarbeitung zusammen.

Die **Ebene der mentalen Repräsentation** stellt eine Rekonstruktion von Ausschnitten einer tatsächlichen oder einer vorgestellten **Umgebung** dar. Sie ist eine **Abstraktion**. Umgekehrt bezeichnen die mit dem mentalen Modell angesprochenen Vorstellungen Objekte der Erfahrung. Diese Vorstellungen erlauben das Wiedererkennen von und das gedankliche Operieren mit Objekten der Erfahrung. Diese Beziehung wird **Referenz** genannt.

Die **theoretische Rekonstruktion** des mentalen Modells einer Person unter Verwendung informeller, theoretischer psychologischer Begriffe charakterisiert die **konzeptuell-fachwissenschaftliche Ebene**. Die Interpretation des damit angesprochenen psychologischen Modells erster Art entspricht der **intendierten psychologischen Semantik**.

Die unteren drei Ebenen in Abbildung 2 ergeben sich aus der Verwendung wissensbasierter Systeme zur Rekonstruktion mentaler Modelle. Eine **Formalisierung** der in der Regel natürlich-sprachlichen psychologischen Begriffe führt zur **theoretisch-formalen Ebene**. Ihre **Implementation** entspricht einer Transformation in eine operative Spezifikation auf der **Programmebene**, deren **Materialisierung** schließlich zur **physikalischen Realisierung** des Programms auf einem speziellen Rechner führt. Umgekehrt betrachtet kann eine operative Spezifikation als eine **Interpretation** ihrer jeweiligen Materialisierung aufgefaßt werden. Die Interpretation eines Programms im Sinne seiner **operationalen Semantik** führt zu einer abstrakten Spezifikation, deren **denotative Semantik** im vorliegenden Kontext mit einer Abbildung formalisierter Strukturen und Operationen auf eine psychologische Begrifflichkeit korrespondiert.

Diese Mehrebenenbetrachtung von Theorien der Informationsverarbeitung ist eine idealisierte Betrachtung. Dies wird sichtbar, wenn man die in Abbildung 2 eingeführten Ebenen mit der von Newell entwickelten Betrachtung in Beziehung setzt. Insbesondere die Zuordnung der Wissensebene bereitet Schwierigkeiten. Ihre Funktionalität wird sowohl durch die Ebene der mentalen Re-

Präsentation als auch durch die konzeptuell-fachwissenschaftliche Ebene abgedeckt. Eine saubere Trennung zwischen einer kognitionspsychologischen Begrifflichkeit, die durch sprachliche Ausdrücke wie Ziele, Absichten, Wissen, Pläne und dergleichen gekennzeichnet ist und implementationsorientierten Begriffen wie deklarative und prozedurale Wissensbestände, Kontrollstrategien usw. ist häufig nicht einfach, wenn auch grundsätzlich wünschenswert.

#### **4. Bewertung wissensbasierter Systeme als psychologische Modelle**

Computer science has given psychology a new way of expressing models of cognition that is much more detailed and precise than its predecessors. But unfortunately, the increased detail and precision in stating models has not been accompanied by correspondingly detailed and precise arguments **analyzing and supporting** them.

(VanLehn, Brown & Greeno, 1984, S.237)

In diesem Abschnitt untersuchen wir wissensbasierte Systeme zur Modellierung kognitiver Strukturen und Prozesse anhand der Bewertungskriterien, die üblicherweise an Theorien herangetragen werden. Dazu zählen etwa (vgl. Gadenne, Theoriebewertung, in diesem Band): Die logische Konsistenz und Widerspruchsfreiheit der Theorie, die Einheitlichkeit des Begriffssystems (semantische Einheitlichkeit), der Informationsgehalt und die Tiefe der Theorie, ihre praktische Relevanz und - besonders herausgehoben - ihre empirische Adäquatheit.

##### 4.1 Empirische Adäquatheit

Die empirische Adäquatheit einer Modellierung ist das zentrale Gütemerkmal zur Beurteilung einer psychologischen Theorie. Damit ist auch ein Kriterium angesprochen, das wie kein anderes zwischen einer psychologisch motivierten kognitionswissenschaftlichen Forschung, bei der der Mensch das Objekt der Theoriebildung ist, und dem Forschungsansatz der künstlichen Intelligenz als Teil der Informatik unterscheidet. Aus computerwissenschaftlicher Sicht steht der Entwurf und die Implementation eines intelligenten Systems mit bekannten formalen Eigenschaften und definierten Leistungen im Vordergrund. Die Frage einer psychologisch validen Rekonstruktion menschlicher intelligenter Leistungen ist dabei nicht von primärem Interesse.

###### **4.1.1 Eine theoretische Rahmenvorstellung**

Zur Behandlung der Frage der empirischen Adäquatheit einer kognitiven Modellierung greifen wir auf die theoretische Rahmenvorstellung zurück, die

schon im zweiten Abschnitt skizziert wurde. Sie ist in vereinfachter und zugleich für die hier interessierenden Fragen zugeschnittener Form in Abbildung 3 dargestellt.

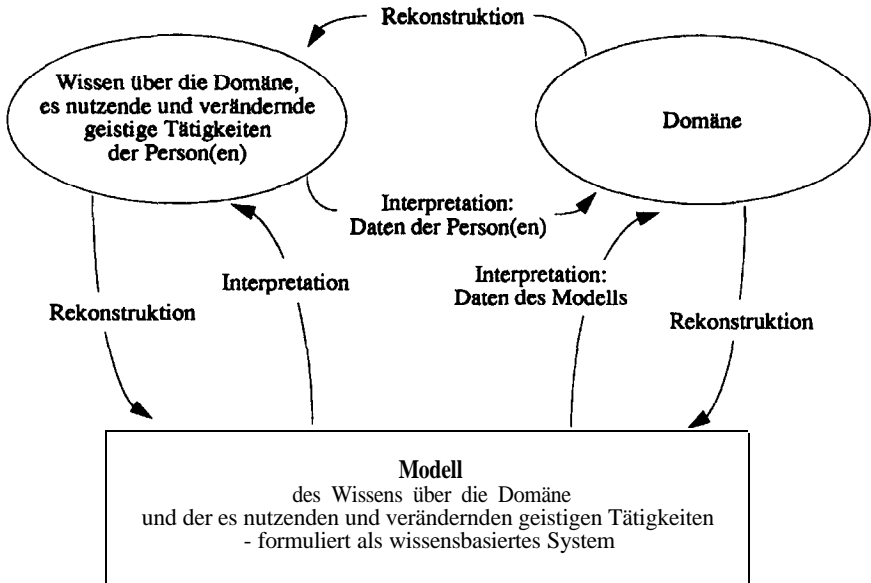


Abb. 3: Bewertung wissensbasierter Systeme: Vergleich von Personen- und Modelldaten zur Überprüfung der empirischen Adäquatheit

Eine Person (oder eine Gruppe von Personen) beschäftigt sich mit einer bestimmten Domäne, einem Ausschnitt der Umgebung. Das Ergebnis der Rekonstruktion ist Wissen über die Domäne. Über die es nutzenden und verändernden geistigen Tätigkeiten resultieren in Interaktion mit der Domäne die Daten der Person(en). Die psychologisch valide Rekonstruktion dieses Wissens und der auf ihm ablaufenden Prozesse ist Ziel der Modellierung, hier seiner Formulierung als wissensbasiertes System. Dieses System soll die Domäne in derselben Weise rekonstruieren, wie dies die betrachteten Personen tun. Überprüfbar ist dies anhand der Interpretation des Modells, d.h. über die Daten, die das Modell in Interaktion mit der Domäne liefert. Modell- und Personendaten stellen die Basis für die Prüfung der empirische Adäquatheit der Modellierung dar.

Wir haben bisher nicht spezifiziert, ob sich die Modellierung nur auf eine Person oder alle Personen einer bestimmten Population beziehen soll. Tatsächlich ist dies ein Punkt, der in der Literatur nicht selten für Verwirrung sorgt. Es ist zu unterscheiden, ob ein idiographisches Modell, das eine bestimmte Person beschreibt, ein für alle Personen einer Population prototypisches Modell oder ein individualisiertes Modell vorliegt, das von allgemeinen, auf alle Personen

**einer Population zutreffenden Annahmen** ausgeht, **darüber hinaus aber interindividuell unterschiedliche Annahmen und Komponenten zuläßt.**

Wissenschaftstheoretisch gesehen sind Modellierungen des letzten Typs am vorteilhaftesten. In ihnen wird klar expliziert, was als allgemein und was als individuell spezifisch angenommen wird. Interindividuelle Varianz wird theoretisch begründet. Individualisierte Modelle erfordern eine Diagnosekomponente, um die individualspezifischen Anteile festlegen zu können. Anhand einer Einzelfallanalyse entwickelte Modellierungen beschreiben zumeist die mentalen Strukturen und Prozesse nur des einen betrachteten Individuums. Allerdings weisen sie oft keine klare Trennung von allgemeinen und personenspezifischen Annahmen auf. Sie stellen unter dem Aspekt der Verallgemeinerbarkeit einen sehr problematischen Fall dar.

In den Tabellen 2 und 3 werden mit den Systemen MULEDS (vgl. Plötzner, Spada, Stumpf & Opwis, 1990) und KAGE (vgl. Plötzner, 1990; Plötzner & Spada, 1992) zwei Beispiele von Modellierungen skizziert, von denen in der einen interindividuelle Unterschiede zugelassen sind, die andere unter Ausklammerung interindividueller Varianz Wissensveränderungen, also ein prototypisches, hier aber ‚dynamisches‘ mentales Modell rekonstruiert. Wir werden in den nachfolgenden Ausführungen zu Fragen der empirischen Prüfung verschiedentlich Bezug auf diese Modellierungen nehmen.<sup>2</sup>

Tabelle 2: Programmsystem MULEDS (Multi-Level-Diagnosis-System)

---

Theoretische Grundlage des mit MULEDS realisierten Diagnosemodells ist die Rahmenvorstellung einer multiplen mentalen Repräsentation von Wissen über physikalische Gegenstandsbereiche entlang der Dimension qualitativ/quantitativ. Methodisch ist MULEDS dem im Zusammenhang mit der Diagnose und Modellierung arithmetischer Fertigkeiten bekannt gewordenen ‚Fehlerbibliotheksansatz‘ zuzuordnen. Zentral sind hierbei zwei Mengen von im Modell vorselektierten Wissens-elementen. Erstens, die Menge korrekter Wissens-elemente, die das zu einer erfolgreichen Problemlösung benötigte bereichsspezifische Wissen repräsentieren. Zweitens, eine Menge fehlerhafter bzw. unvollständiger Wissens-elemente, die typische Denkfehler und/oder Mißkonzepte widerspiegeln.

Es wurden korrekte, fehlerhafte und unvollständige Wissenskomponenten auf jeweils einer qualitativ-relationalen, einer quantitativ-relationalen und einer quantitativen Repräsentationsebene zu den Phänomentypen des zentralen und schiefen Stoßes, eines Teilgebiets der klassischen Mechanik formuliert. Zusätzlich wurde mathematisches/experimentelles Wissen, das der Verknüpfung mehrerer Vorhersagen dient, in Form korrekter Regeln repräsentiert. Zur Implementation wurde der Produktionssysteminterpretierer PRISM verwendet.

Bei Anwendung von MULEDS auf einen Datensatz von Personen, die entsprechende Aufgaben bearbeitet haben, resultieren im positiven Fall ausführbare individuelle Wissensmodelle. Im negativen Fall ist keine Diagnose des Wissens der Person möglich.

---

<sup>2</sup> Diese Arbeiten erfolgten im Rahmen von Forschungsvorhaben, die durch die Deutsche Forschungsgemeinschaft unterstützt wurden (Sp 251/2-x; Sp 251/6).



Tabelle 3: Programmsystem KAGE (Knowledge Acquisition Governed by Experimentation)

---

Basierend auf der Theorie multipler mentaler Repräsentationsebenen (Plötzner et al., 1991) und unter Verwendung analysebasierter Lernmechanismen wird der Erwerb von Wissen über funktionale Beziehungen rekonstruiert. Modelliert wird der Lernprozeß von Schülern, die im Rahmen einer Mikrowelt-Lernumgebung simulierte physikalische Experimente durchführen. Zentrale Annahme ist, daß ein Schüler zunächst mit Hilfe experimentellen und heuristischen Wissens versucht, anhand einzelner Experimente Hypothesen darüber zu formulieren, welche physikalischen Größen wie zusammenhängen. Die anschließende mathematische Elaboration der Art des Zusammenhangs wird dann im Sinne eines Analogieschlusses auf neue experimentelle Situationen übertragen. Solche Analogieschlüsse sind im Modell sowohl innerhalb als auch zwischen Repräsentationsebenen vorgesehen.

Zu den rekonstruierten kognitiven Mechanismen zählen: Identifikation funktional abhängiger Variablen; analoges Schließen unter Heranziehung analoger Fälle, Elaboration des betrachteten Falls, Extrapolation auf den neuen und Evaluation der Inferenz, Abstraktionsbildung und Ableitung neuer experimenteller Vorhersagen. KAGE behandelt insbesondere proportionale und umgekehrt proportionale Beziehungen, und zwar auf verschiedenen Repräsentationsstufen, die sich auf einer Dimension qualitativ/quantitativ anordnen lassen. Im Vordergrund der Modellierung steht die jeweilige Nutzung von Wissen, das auf einer informationsärmeren Repräsentationsebene erworben wurde, zur Formulierung von Hypothesen auf der nächst höheren Ebene.

KAGE liegt implementiert in PROLOG vor.

---

#### **4.1.2 Probleme**

Auf den ersten Blick erscheint die empirische Geltungsprüfung von als wissenschaftsbasierte Systeme realisierten Modellierungen problemlos. Offensichtlich gilt es, die Daten der Personen, also die empirischen Daten, mit den vom Modell generierten Daten zu vergleichen. So könnte man etwa die mit Hilfe des Systems MULEDS verschiedenen Schülern zugeschriebenen Wissensmodelle zur Generierung von Antworten auf weitere Aufgaben heranziehen, um dann diese Modelldaten mit den von den Schülern selbst gezeigten Reaktionen zu vergleichen. Eine Prüfung des Systems KAGE könnte wie folgt ablaufen: es wird eine bestimmte Aufgabensequenz sowohl Schülern wie auch dem System vorgelegt. Die aus der Bearbeitung der Aufgaben resultierende Veränderung des Wissens der Schüler bzw. ihre damit korrespondierenden Antworten werden mit den durch das Modell vorhergesagten Antworten in Beziehung gesetzt. Ähnlich könnte bei anderen Modellierungen, bei denen mit wissenschaftsbasierten Systemen gearbeitet wird, vorgegangen werden. Geltungskriterium ist die Passung von Personen- und Modelldaten.

Dieses Vorgehen ist möglich, da die Arbeit mit wissenschaftsbasierten Systemen zu einem neuen Datentyp, den anhand der Modellierung selbst generierten Daten, führt. Aus der Modellierung resultiert eine bestimmte Form von Perfor-

manz. Sie ergibt sich aus einem besonderen Merkmal dieser Forschungsmethodik, der ‚Lauffähigkeit‘ wissensbasierter Systeme. Ein Computermodell liefert nicht nur eine Beschreibung von Aktivitäten, sondern setzt im Sinne einer Kompetenzmodellierung die Spezifizierung generativer Mechanismen voraus, auf deren Grundlage die in Frage stehenden Verhaltensphänomene generiert werden können. Dies macht ja gerade einen wichtigen Aspekt der Faszination dieses Forschungsansatzes aus. Der Wissenschaftler wird in die Lage versetzt, (Modell-) Verhalten zu generieren und zu manipulieren. Wo liegen aber nun die Probleme?

Ein zentrales theoretische Problem ist die **Unterdeterminiertheit einer Modellierung durch empirische Daten**. Nicht minder gravierend ist das auf den ersten Blick gegensätzliche Problem, daß **Modelldaten und empirische Daten** bei strenger Prüfung fast immer **differieren**.

Was an diesen Problemen ist aber nun der Modellierung in Form von wissensbasierten Systemen anzulasten, und was gilt in gleichem bzw. in noch viel höherem Maße auch für andere Formen der Theorieformulierung?

Modelle - welcher Art auch immer - sind prinzipiell empirisch unterdeterminiert. Eine perfekte Passung von Modell und Daten ist nie zu erwarten. Bei Modellierungen des hier betrachteten Typs treten diese Probleme aber aus verschiedenen Gründen besonders deutlich hervor.

(1) Es werden nicht nur Oberflächenmerkmale geistiger Tätigkeiten (wie etwa Reaktionszeiten) beschrieben, sondern es werden die Strukturen und Prozesse modelliert, die den beobachteten Phänomenen als zugrundeliegend angenommen werden. VanLehn et al. (1984) unterscheiden im Anschluß an Moravcsik (1980) in diesem Zusammenhang zwischen ‚**tiefen**‘ und ‚**oberflächlichen**‘ Theorien. Oberflächliche Theorien nehmen ausschließlich Bezug auf die beobachtbaren Ergebnisse psychischer Vorgänge und setzen diese zumeist korrelativ miteinander in Verbindung. Im Unterschied dazu versuchen tiefe Theorien die ‚black box‘ zwischen Eingangs- und Ausgangsgrößen zu öffnen und jene mentalen Strukturen und Prozesse zu beschreiben, auf die die beobachtbaren Phänomene zurückgeführt werden können.

(2) Computersimulierte Theorien weisen einen sehr hohen Informationsgehalt auf, da sie viele und sehr präzise formulierte Aussagen enthalten. Die Kehrseite tiefer Theorien ist ein ungünstiges Verhältnis zwischen Postulaten und empirischen Daten.

(3) Hinzu kommt, daß zur Gewährleistung der Lauffähigkeit eines wissensbasierten Systems Annahmen in einem Detaillierungsgrad getroffen werden müssen, der den wissenschaftlichen Kenntnisstand zu dem entsprechenden Phänomenbereich häufig weit übersteigt. Im Vergleich zu Theorien, die nur den Effekt bestimmter Bedingungen beschreiben und vorhersagen, erfordert die Modellierung mentaler Prozesse wesentlich differenziertere Annahmen.

(4) Darüber hinaus wird mit Hilfe wissensbasierter Systeme meistens Einzelverhalten generiert. Der Vergleich von Modell- und Personendaten erfolgt dann häufig auch auf dieser Ebene. Es wird nicht mit über alle Personen gemittelten Lösungszeiten, mit relativen Häufigkeiten richtiger Antworten oder ähnlichen aggregierten Daten und damit einem reduzierten Auflösungsgrad gearbeitet. Diese Aussage gilt übrigens auch für Modellierungen, die mit dem Ziel entwickelt werden, prototypisch für alle Personen einer Population zu sein. Hingegen ist die Verwendung aggregierter Daten die Regel bei der statistischen Prüfung verbal formulierter Theorien, aber auch mathematisch-numerischer Modelle. So schreiben Erdfelder und Bredenkamp (in diesem Band):

Empirische Untersuchungen in der Psychologie zeichnen sich fast ausnahmslos dadurch aus, daß substanzwissenschaftliche Fragestellungen mittels statistischer Hypothesentests beantwortet werden. Bemerkenswert ist dieses Vorgehen vor allem deshalb, weil die interessierende psychologische Hypothese und die letztlich getestete statistische Hypothese im Regelfall nicht identisch sind. Psychologische Hypothesen beziehen sich typischerweise auf psychische Vorgänge, die im einzelnen Individuum ablaufen... Statistisch getestet werden aber . . . häufig Populationsaussagen . . .

(5) Ein weiterer Aspekt ist, daß wissensbasierte Systeme zumeist deterministisch formuliert sind. Bei Theorien, die primär Oberflächenmerkmale beschreiben, kann man eine zusätzliche Flexibilität - man spricht auch von einer Robustheit der Theorie gegenüber Falsifikationsbemühungen - dadurch erzielen, daß nur Wahrscheinlichkeitsaussagen formuliert werden. Die Explizitheit der Formulierung wissensbasierter Systeme in Verbindung mit dem Ziel der Generierung von Verhalten läßt probabilistischen Formulierungen wenig Raum. Eine hohe Falsifikationswahrscheinlichkeit ist die negative (oder positive?) Folge.

Tiefe, deterministisch formulierte Theorien in Form wissensbasierter Systeme, die Einzelverhalten generieren, sind empirisch leicht angreifbar. Sind sie damit aber weniger adäquat als die auf einem sehr eingeschränkten Auflösungsgrad formulierten oberflächlichen Theorien?

Um diese Frage beantworten zu können, ist die Prüfung der empirischen Adäquatheit in einen größeren Zusammenhang zu stellen. Wir werden die Rolle von Empirie im Prozeß der Theorieentwicklung aus drei Blickwinkeln betrachten: Empirie als Vorbild für die Modellierung, empirische Geltungsprüfung anhand einer Analyse der Modelldaten selbst, und anhand eines methodisch reflektierten Vergleichs von Modell- und Personendaten.

### **4.1.3 Empirie als Vorbild**

Stellen wir zunächst die Rolle der Empirie als Richterin über die Adäquatheit einer Modellierung zurück. Betrachten wir empirisches **Wissen als Vorausset-**

**zung und Leitfaden für die Modellkonstruktion.** Diese Phase der Theorieentwicklung findet häufig kaum Beachtung in methodischen Erörterungen im Fach Psychologie (vgl. aber Dörner, 1989; in diesem Band).

The function of empirical observations in this research programme is different from its role in traditional experimental psychology. In a traditional study, the purpose of observing a person's performance is to ascertain the effect of some stimulus variable on his responses. In the simulation programme, the function of empirical investigations is, broadly speaking, to describe human Performance in as much detail as possible, in order to constrain model building. . . . knowledge takes the form of **know-how**.

(Ohlsson, 1988, S. 14)

Nach Ohlsson kommt somit empirischen Erhebungen für die Theorieformulierung zur Ableitung von Designentscheidungen bei der Entwicklung wissensbasierter Systeme eine zentrale Rolle zu. Jeder kognitiven Modellierung muß eine intensive empirische Analyse der „Sachverhältnisse“ (Selz, 1913) des betrachteten Phänomenbereichs vorausgehen. Dabei handelt es sich um ein eher interpretatives Vorgehen (vgl. Soeffner & Hitzler, in diesem Band) mit einem geringen Grad an Datenrestriktion. Selbstverständlich werden auch verbal formulierte Theorien und mathematisch-numerische Modelle unter Heranziehung des Vorwissens über die Domäne gestaltet. Merkmale, wie Exaktheit, Detailliertheit und Lauffähigkeit eines wissensbasierten Systems, zwingen aber zu einer wesentlich intensiveren Auseinandersetzung mit dem Gegenstandsbereich und damit zu einer umfassenden Datenerhebung im Vorfeld der Modellierung.

Jeder Schritt in der Modellierung, jede getroffene Annahme ist theoretisch und empirisch zu motivieren und zu begründen. Auf diese Weise wird ein hoher empirischer Gehalt der Modellierung schon in der Phase der Entwicklung des wissensbasierten Systems sichergestellt. Als Beispiel können die Arbeit von Chi, Bassok, Lewis, Reimann und Glaser (1989) und die auf ihr aufbauenden Modellierungen herangezogen werden. Chi et al. (1989) zeigen, daß Schüler insbesondere dann aus vorgegebenen Beispielen lernen, wenn sie die Beispiele reflektierend nachvollziehen und sich selbst erklären. Ansätze einer Modellierung dieses Selbsterklärungseffekts werden von Reimann und Schult (1991) sowie VanLehn, Jones und Chi (1992) diskutiert.

#### **4.1.4 Modellperformanz als Datenquelle**

Ein wichtiger Teil der Geltungsprüfung besteht in einer systematischen Analyse der Modelldaten vor dem Hintergrund des Wissens über den betrachteten Phänomenbereich. Folgende Fragen sind zu prüfen:

- Rekonstruiert die Modellierung die zu erklärenden Phänomene des untersuchten Gegenstandsbereichs (situative Variation) und für den analysierten Personenkreis (individuelle Variation)?
- Generiert die Modellierung auch unsinnige, bisher nicht empirisch beobachtete Phänomene, sog. Sternchen-Daten im Sinne von VanLehn et al., 1984?
- Führen gezielte Eingriffe in das Modell zu den theoretisch erwarteten Änderungen des Modellverhaltens?

Die erste Frage bezieht sich auf die **Suffizienz** einer Modellierung. Es wird geprüft, ob das Modell als hinreichend zur Erzeugung des interessierenden Verhaltens angesehen werden kann. Dabei ist insbesondere zu untersuchen, ob die Phänomene des Gegenstandsbereichs in ihrer gesamten Breite generiert werden können und ob die Modellierung der in der analysierten Population beobachteten interindividuellen Varianz gerecht wird. Diese Prüfung beschränkt sich nicht nur auf die Rekonstruktion korrekten Verhaltens, sondern umfaßt auch fehlerhafte Reaktionen.

Die Prüfung der Suffizienz ist ein strenger Falsifikationstest. Ist ihr Ergebnis negativ, kann auf weitere Schritte der Geltungskontrolle verzichtet werden. Umgekehrt bedeutet ein positives Ergebnis nicht notwendigerweise, daß die Modellierung auch empirisch adäquat ist. In diesem Fall ist ‚nur‘ gesichert, daß die Modellierung hinreichend zur Verhaltensgenerierung ist.

Die Analyse der Frage, ob ein wissenschaftliches System auch empirisch **nicht** beobachtbares Verhalten zeigt, stellt eine weitere Möglichkeit der Falsifizierung dar. Sie wurde von VanLehn (1983, Brown & VanLehn, 1980) im Zusammenhang mit der Entwicklung und Testung ihrer Flickwerktheorie (repair-theory) in die Diskussion eingeführt. Geprüft wurde, ob das zur Modellierung des Erwerbs einfacher arithmetischer Fertigkeiten formulierte wissenschaftliche System neben empirisch auftretenden Fehlern auch unsinnige Fehler generiert. Tatsächlich zeigte sich beim Vergleich verschiedener Modellvarianten, daß mit einer zunehmenden Anzahl erklärter Fehler auch die Häufigkeit bisher nicht beobachteter und teilweise völlig unplausibler Varianten weit überproportional anstieg.

Modellexperimente eröffnen eine weitere Prüfmöglichkeit. Die Modellannahmen werden gezielt modifiziert, etwa indem ausgewählte Wissensbestände entfernt werden. Anschließend wird analysiert, ob sich die theoretisch vorhergesagten Verhaltensänderungen, etwa die Nichtlösbarkeit spezieller Aufgaben in den Modelldaten nachweisen lassen (vgl. Young & O’Shea, 1981).

Insgesamt kommt der Prüfung der Suffizienz bei Modellierungen mittels wissenschaftlicher Systeme eine besondere Rolle zu. Tatsächlich bietet dieser Typ von Modellierungen wie kaum ein anderer die Möglichkeit zu solchen Prü-

fungen. Aus dieser Perspektive wird verständlich, warum Vertreter dieser Forschungsrichtung anderen Ansätzen der Theorieformulierung, bei denen Sufizienz überhaupt nicht gegeben ist, sehr skeptisch gegenüber stehen.

#### **4.1.5 Vergleich von Modell- und Personendaten**

Der Vergleich von Modell- und Personendaten bleibt der wichtigste Weg zur Prüfung der empirischen Adäquatheit einer Modellierung. Wenden wir uns zunächst den Fällen zu, in denen ein direkter Vergleich der Modelldaten über Einzelverhalten mit den Daten einer oder mehrerer Personen erfolgt. Dies ist dann naheliegend, wenn in der Modellierung allgemeine **und** individuelle Komponenten berücksichtigt sind bzw. wenn anhand einer Einzelfallstudie das Verhalten einer einzigen Person rekonstruiert wurde. Zumeist liegt im letztgenannten Fall eine umfassende und dichte Datenerhebung bei einer über einen längeren Zeitraum hinweg beobachteten Person vor (vgl. etwa Simon & Simon, 1978). Dies ermöglicht den Vergleich des Datenprotokolls der Person mit den Aufzeichnungen der Modelldaten. Im Sinne einer Kreuzvalidierung ist es dabei erforderlich, nur einen Teil der Daten zur Modellierung heranzuziehen, um den anderen mit aus dem Modell generierten Prognosedaten vergleichen zu können.

Eine besondere Form einer individualisierten Modellierung stellt die anhand des Systems MULEDS bereits illustrierte kognitionswissenschaftliche qualitative Diagnostik dar (vgl. für ein klassisches Diagnosesystem DEBUGGY, Burton, 1982, und allgemein Spada & Reimann, 1988; Wenger 1987). Bei dieser Form der Diagnostik, für die etwa die Erfassung des Wissens der einzelnen Personen ein typisches Beispiel ist, liegt ein Rahmenmodell vor, das allgemeine, für alle Personen geltende Annahmen beinhaltet. Darüber hinaus werden im Diagnoseprozeß für jede einzelne Person individuelle Anteile festgelegt, beispielsweise die eine Person charakterisierenden korrekten und fehlerhaften Wissenskomponenten. Für jede einzelne Person resultiert ein individualisiertes Modell in Form eines lauffähigen wissensbasierten Systems. Worin besteht nun die empirische Prüfung der einzelnen Modelle und des Diagnosesystems als Ganzem? Anhand der empirischen Daten wird für jede einzelne Person ein Modell spezifiziert. Dabei wird geprüft, ob es möglich ist, anhand dieser Modelle die beobachteten Reaktionen der Personen etwa auf Aufgaben des untersuchten Gegenstandsbereichs zu rekonstruieren. Noch überzeugender ist die Prüfung, wenn sie im Sinne einer Kreuzvalidierung an einem neuen Satz von Daten derselben Personen erfolgt. Das einzelne Modell wird verworfen, wenn die mit seiner Hilfe generierten Daten nicht mit den Antworten der betreffenden Person übereinstimmen. Das gesamte Diagnosesystem ist zurückzuweisen, wenn es bei einer größeren Anzahl der Probanden versagt.

Schwieriger ist die Situation, wenn die Modellierung in einer prototypischen Form vorliegt, die keine individuenbezogene Spezifikationen erlaubt. Dann ist von vornherein damit zu rechnen, daß ein direkter Vergleich von Modell-daten mit den Daten einzelner Personen Abweichungen signalisiert. Wir haben ja bei Modelliermagazinen dieses Typs mit wissensbasierten Systemen die besondere Situation, daß eine allgemeine, auf den prototypischen Probanden bezogene Theorie so formuliert ist, daß sie Einzelverhaltensdaten generiert, die prinzipiell einem Vergleich mit den Daten beliebiger Personen unterzogen werden können. Welcher Weg kann in diesem Fall sinnvollerweise beschritten werden?

Eine empirische Geltungsprüfung ist beispielsweise durch Ableitung von Hypothesen über Daten anhand der Modellierung ganz im Sinne des klassischen, experimentell-statistischen Ansatzes möglich. Ausgehend vom Annahmengengefüge des Modells werden Bedingungen spezifiziert, die theoretisch interessante Effekte auf die Daten erwarten lassen. Unter diesen Bedingungen werden sowohl Modellläufe durchgeführt, als auch Personen getestet. Es wird geprüft, ob sich die in den Modelldaten niederschlagenden Bedingungeffekte auch in statistisch gesicherter Weise anhand der Personendaten feststellen lassen. Charakteristisch für diese Form der empirischen Modellprüfung ist die hohe Datenrestriktion, die ganz im Gegensatz zum Umgang mit Daten in der Phase der Modellentwicklung steht.

Ein Beispiel für eine derartige Modellprüfung wäre, bezogen auf das schon erwähnte Wissenserwerbsmodell KAGE (vgl. Tabelle 3), daß verschiedene Sequenzen von Lernaufgaben dargeboten werden. Sie werden so ausgewählt, daß sie bei Anwendung des Systems KAGE zu einem optimalen, einem suboptimalen und einem Lernergebnis mit fehlerhaftem Wissen führen. An Schülern wird überprüft, ob sich die anhand des Modells vorhergesagten Effekte der Lernbedingungen empirisch bestätigen lassen.

Selbstverständlich prüft ein experimentell-statistisches Vorgehen dieses Typs nicht das gesamte Annahmengengefüge eines wissensbasierten Prozeßmodells. Es stellt sich die Frage nach dem durch die jeweilige Bedingungsvariation erreichten Auflösungsgrad, die bei anders formulierten Theorien allerdings häufig vernachlässigt wird. Und es ist das Problem zu berücksichtigen, auf das Erdfelder und Bredenkamp (in diesem Band) hinweisen, wenn sie die fehlende Passung von Hypothesen über Einzelverhalten und statistischen Populationsaussagen anmerken.

Diese Probleme sind keineswegs spezifisch für die Überprüfung von kognitiven Modellierungen mit Hilfe eines experimentell-statistischen Vorgehens, sondern treffen in gleicher Weise auf die empirische Kontrolle jeder umfassenden Theorie zu. Es hängt vom theoretischen Verständnis und experimen-

tellen Geschick des einzelnen Forschers ab, inwieweit es ihm gelingt, auf experimentellem Wege wichtige Modellaspekte zu prüfen.

#### **4.1.6 Modellvergleichende Argumentation**

VanLehn et al. (1984) setzen sich mit der Frage auseinander, wie nach einer erfolgreichen Suffizienzprüfung einer kognitiven Modellierung Schritte in Richtung auf das Kriterium „Notwendigkeit“ gemacht werden können. Wie lassen sich Hinweise gewinnen, daß die formulierten Annahmen und postulierten Prozesse notwendig sind, um das interessierende Verhalten zu generieren. Sie schlagen dafür ein Verfahren vor, das sie **modellvergleichende Argumentation** (competitive argumentation) nennen. Die Annahmen einer computersimulierten Theorie werden mit denen anderer Theorien oder aber anderer Versionen der Theorie selbst verglichen. Beispielsweise gilt es zu zeigen, daß ein Designprinzip, das der Entwicklung eines wissensbasierten Systems zugrundeliegt, bestimmte Fakten zu erklären vermag, während Variationen dieses Prinzips, obwohl selbst nicht ohne empirischen Wert, in der einen oder anderen Weise mit Problemen belastet sind. Um zu belegen, daß eine bestimmte, in eine Modellierung eingeführte Bedingung notwendig ist, genügt es keineswegs zu zeigen, daß das Suffizienzkriterium erfüllt ist. Wenn nämlich zwei Theorien vorliegen, von denen die eine behauptet, daß Prinzip X hinreichend sei und eine andere, daß dies für ein mit X inkompatibles Prinzip Y gelte, dann versagt das Suffizienzkriterium als Richtschnur zur Entscheidung zwischen den beiden Modellierungen. Es muß vielmehr argumentiert werden, warum Prinzip X aus psychologischer Sicht besser ist als Prinzip Y. In dieser Art einer vergleichenden Argumentation sehen VanLehn et al. (1984) eine realistische Alternative zum letztlich unmöglichen Nachweis, daß eine Annahme notwendig ist. Vergleichende Argumentation sei ein Weg für eine sukzessive Annäherung an psychologisch valide Theorien. Des weiteren verweisen VanLehn et al. (1984) darauf, daß derartige Argumente, etwa empirische Befunde, die eine von mehreren alternativen Theorien belegen, häufig einen längeren Bestand haben, als die Theorien selbst, zu deren Stützung sie vorgelegt wurden.

Eine modellvergleichende Argumentation kann als Verallgemeinerung des bekannten Entscheidungsexperiments, bei dem zwischen zwei Theorien auf der Basis eines empirischen Befundes entschieden wird, angesehen werden. In dieselbe Kategorie fällt auch der allerdings formal wesentlich ausgefeiltere Vergleich von mathematisch-numerischen Modellen, wenn etwa bei probabilistischen Ansätzen ein Likelihood-Vergleich durchgeführt wird, um eine Aussage über die relative Geltung zweier konkurrierender Modelle treffen zu können.



## 4.2 Weitere Bewertungskriterien

Gadenne (in diesem Band) führt weitere Kriterien zur Theoriebewertung an: logische Konsistenz und Widerspruchsfreiheit, semantische Einheitlichkeit, Informationsgehalt und Tiefe von Theorien sowie Einfachheit. Wie sind anhand dieser Bewertungskriterien Modellierungen in Form wissensbasierter Systeme zu beurteilen? Selbstverständlich ist es schwierig, generelle Aussagen zu treffen. Die folgenden Feststellungen beschreiben deshalb den Grad an Kriterienerfüllung, der bei diesem Typ von Theorieformulierung bei einem adäquaten Vorgehen prinzipiell erreichbar ist.

Zur Frage der logischen Konsistenz und Widerspruchsfreiheit bieten lauffähige wissensbasierte Systeme über die Generierung repräsentativer Sätze von Modelldaten eine gute Prüfmöglichkeit. Die Simulation eröffnet einen Weg, Widersprüchen auf die Spur zu kommen. Vergleichbare Prüfmöglichkeiten weisen verbal formulierte Theorien nicht auf.

Die Frage nach der semantischen Einheitlichkeit, also etwa der Einheitlichkeit des verwendeten Begriffssystems ist schwer zu beantworten. Sicherlich erzwingt die Formalisierung durch die jeweilige Festlegung auf einen bestimmten Programmierstil und eine bestimmte Programmiersprache von vornherein eine gewisse Einheitlichkeit. Es ist aber nicht gewährleistet, daß tatsächlich nur semantisch einheitliche Begriffe in die Formalisierung eingehen.

Die Kriterien „Informationsgehalt“ und „Tiefe“ sind bei wissensbasierten Systemen in hohem Maße erfüllt. Die Notwendigkeit der Explizitheit und Detaillierung zwingt zu vielen und präzisen Aussagen. Es resultieren Theorien mit hohem Informationsgehalt und zugleich großer Tiefe. Es ist ja gerade das Anliegen, die „black box“ zwischen Ein- und Ausgabegrößen aufzubrechen und die Prozesse zu spezifizieren, die den Zusammenhang zwischen diesen Größen bedingen.

Einfachheit ist nicht die Stärke von Modellierungen mit Hilfe wissensbasierter Systeme. Das Ziel der Suffizienz und das damit verbundene Merkmal der Lauffähigkeit zwingen dazu, manche Annahmen über Details einzuführen, die nicht im Kern der Theorie stehen und über die möglicherweise keine hinreichende theoretische und empirische Evidenz vorliegt.

Die bisher genannten Theoriebewertungskriterien stammen weitgehend aus einer Zeit, in der der Stellenwert computersimulierter Theorien noch sehr gering war. Müssen sie nicht mit Blick auf diesen neuen Typ der Theorieformulierung modifiziert, präzisiert und vor allem ergänzt werden?

Ohlsson (1988) führt zwei Bewertungskriterien an, die aufs engste mit der Lauffähigkeit wissensbasierter Systeme und mit dem neuen Typ von Perfor-

manz, der Generierung modellbasierter Daten, zu tun haben. Es sind dies die schon angesprochene Suffizienz einer Modellierung und die Anwendbarkeit des wissensbasierten Systems. Das letztgenannte Kriterium hängt eng mit einem klassischen Theorieideal zusammen, nämlich der Frage der Verallgemeinerbarkeit der Ergebnisse. Es gewinnt allerdings im Kontext computersimulierter Theorien einen besonderen Stellenwert, da sich über die Generierung von Modelldaten neue Anwendungsmöglichkeiten ergeben. Beispiele sind etwa:

- die Ableitung präziser Verhaltensvorhersagen unter verschiedenen situativen Bedingungen,
- die Aufdeckung von kognitiven Strukturen und Prozessen, die zu fehlerhaften Reaktionen führen, und
- die Ableitung von Interventionsmaßnahmen mit in der Simulation erprobten Effekten.

Der letzte Punkt läßt sich an dem Konzept der ‚Pseudostudenten‘ von Van-Lehn (1991) verdeutlichen. Er nutzt Computermodellierungen des Wissenserwerbs, um verschiedene Instruktionsstrategien zu erproben. Aus der Simulation wird abgeleitet, welche Strategie den Lernzielen am besten dient. Erst dann erfolgen Felduntersuchungen. Computersimierte Theorien des Wissenserwerbs werden hier zum Testfeld für Instruktionsfragen. Die gezielt eingesetzte Empirie dient der Überprüfung der über die Modelldaten gewonnenen Erwartungen. Im Vergleich zu empirischen Interventionsstudien nach dem Versuch und Irrtum Prinzip ist dieses Vorgehen wesentlich ökonomischer und theoretisch befriedigender.

Tabelle 4: Vergleich formaler Modelle numerischer Art und nicht-numerischer Art

Inwieweit erfüllen formale Modelle numerischer Art (wie mathematische Lernmodelle) und nicht-numerischer Art (wie wissensbasierte Systeme) die bekannten Theoriebewertungskriterien (vgl. Gadenne, in diesem Band)?

Konnektionistische Modelle sind hier ausgeklammert. Auch zur **inhaltlichen** Frage der empirischen Adäquatheit einer Theorie werden keine Aussagen gemacht.

formale Modelle	logische Konsistenz	Informationsgehalt/Tiefe	Einfachheit	Suffizienz	Empirische Prüfbarkeit	Praktische Anwendbarkeit
numerischer Art	ja	eher gering	ja	eher nein	gut	eher gering
nicht-numerischer Art	ja	groß	nein	ja	eher schlecht	oft gegeben

Das zweite, nicht nur von Ohlsson (1988) ins Gespräch gebrachte Theoriebewertungskriterium ist das der Suffizienz. Sie ist Vorbedingung für empirische Adäquatheit. Das Streben nach suffizienter Modellierung hat vielfältige

Effekte auf den Forschungsprozeß. Ein noch nicht angesprochener Effekt ist die Notwendigkeit einer ganzheitlichen Betrachtungsweise im Sinne einer integrierten Behandlung aller Strukturen und Prozesse, die zur Erbringung einer Leistung erforderlich sind. Eine Beschränkung auf eine einzelne Komponente wäre unter diesem Gesichtspunkt nicht hinreichend.

Tabelle 4 faßt die Ausführungen zur Bewertung wissensbasierter Systeme nochmals zusammen, in dem formale Modelle numerischer und nicht-numerischer Art schematisch anhand wichtiger Theoriebewertungskriterien verglichen werden.

## 5. *Wissenschaftstheoretische Anmerkungen*

Ist es wissenschaftstheoretisch gesehen überhaupt begründbar, daß geistige Aktivitäten und intelligente Leistungen des Menschen mit technischen Mitteln untersucht und im Sinne einer maschinellen Rekonstruktion auf mechanische Weise nachgebildet werden?

Zur Versachlichung der Diskussion dieser Frage sollen zunächst zwei Thesen unterschieden werden:

These 1: Wissensbasierte Systeme sind ein nützliches und mächtiges Werkzeug zur Erforschung der menschlichen Kognition. Die Rekonstruktion kognitiver Prozesse in Form von Berechnungsvorgängen ist möglich und theoretisch sinnvoll. Befürworter dieser, nach einer Bezeichnung von Searle (1980) als schwache These **der KI** bekanntgewordenen Position verweisen typischerweise auf die präzisere Formulierung, den größeren Komplexitätsgrad und den höheren Informationsgehalt der zugrundeliegenden Theorien sowie auf die damit einhergehende Eröffnung vielfältiger Anwendungsperspektiven.

These 2: Wissensbasierte Systeme sind nicht nur ein Mittel zur Erforschung der menschlichen Kognition, sondern sie können darüber hinaus selbst als Träger kognitiver Prozesse angesehen werden. Kognition vollzieht sich in Prozessen der Informationsverarbeitung, d.h. kognitive Prozesse sind Berechnungsprozesse. Diese **starke** These **der KI** (Searle, 1980) besagt, daß im Unterschied etwa zur Simulation eines Binnensees mit Hilfe von mathematisch-numerischen Modellen, die natürlich nicht selbst ein Binnensee ist, die Simulation eines kognitiven Prozesses durch ein wissensbasiertes System selbst wiederum ein kognitiver Prozeß ist, da in diesem Fall das Simulierte und seine Simulation nach denselben Prinzipien der Symbolverarbeitung funktionieren. Oder noch provokanter formuliert: In wissensbasierten Systemen laufen geistige Prozesse ab. Solche Systeme besitzen Geist.

Gegenstand der folgenden Ausführungen ist die Frage nach dem wissenschaftstheoretischen Selbstverständnis der in den vorangegangenen Abschnitten dargestellten Methodologie. Ausgangspunkt ist der sog. Turing-Test, der einen pragmatischen, operationalen Zugang zu dem gesamten Problemkreis liefert. Anschließend wird mit der Position des Funktionalismus eine einflußreiche wissenschaftstheoretische Position zur Philosophie des Geistes in ihren Grundannahmen skizziert, bevor dann anhand des Gedankenexperiments von John Searle zum sog. chinesischen Zimmer die Position der starken KI-These kritisch diskutiert wird.

## 5.1 Der Turing-Test

Alan Turing beginnt seine berühmt gewordene, 1950 veröffentlichte Arbeit „Computing machinery and intelligence“ mit den folgenden Worten:

I propose to consider the question, ‘Can machines think?’. This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’. . . .

(Turing, 1950, S. 433)

Im folgenden expliziert er den heute als **Turing-Test** bekannten Gedanken-gang. Dabei geht es um eine Operationalisierung der Frage, wann einem Computer die Fähigkeit zu denken zugeschrieben werden kann. Turing wählt einen pragmatischen Zugang zur Beantwortung: Ein Computer denkt, sofern er ununterscheidbar von einer Person handelt. Dies kann aufgrund eines Imitations-tests entschieden werden. Eine Person soll eine andere Person und einen Computer (bzw. ein Computerprogramm) befragen. Die gesamte Kommunikation (Fragen und Antworten) wird mit technischen Mitteln, ohne jeden Sichtkontakt durchgeführt. Die Aufgabe für die fragende Person besteht darin, die andere Person bzw. den Computer zu identifizieren, indem sie das Frage-Antwort-Spiel entsprechend gestaltet. Falls die fragende Person auch nach einer Reihe von Fragen die andere Person nicht eindeutig zu identifizieren vermag, gesteht man dem Computer zu, den Test erfolgreich bestanden zu haben.

Unabhängig von der Tatsache, daß auch heute noch kein Computersystem existiert, daß diesen Test erfolgreich bestanden hat, führten diese Überlegungen zu einer intensiven Diskussion der Frage, ob Menschen und Computer prinzipiell vergleichbar sind und ob der Turing-Test ein angemessenes Entscheidungsverfahren darstellt. Muß man einem Computer, der den Turing-Test besteht, **notwendigerweise** kognitive Fähigkeiten zuschreiben?

Ein in diesem Zusammenhang immer wieder geäußerter Einwand bezieht sich auf den behavioristisch verengten Vergleichsmaßstab: Nicht das Denken selbst (in Form interner, nicht beobachtbarer mentaler Vorgänge) wird thematisiert,

sondern es werden lediglich die Denkergebnisse in Form beobachtbaren (Antwort-) Verhaltens untersucht. Aber ist es andererseits nicht vernünftig anzunehmen, daß sich jede noch so geschickte Imitation durch zielgerichtetes Prüfen aufdecken läßt, falls ein künstliches System wie ein Computer sein Verhalten auf prinzipiell andere Weise hervorbrächte als ein Mensch? Was ist überhaupt der Status mentaler Vorgänge bzw. Zustände? Handelt es sich um mehr als bloße sprachliche Konventionen? Mit Fragen dieser Art setzt sich die Philosophie des Geistes im Allgemeinen und die Position des Funktionalismus im Besonderen auseinander. Sie sind auch für die Philosophie der Psychologie von grundlegender Bedeutung (vgl. für eine umfassende Darstellung verschiedener Positionen Carrier & Mittelstraß, 1988).

## 5.2 Der Funktionalismus

Der philosophische **Funktionalismus** versucht eine Antwort auf die Frage zu geben, wie mentalistische bzw. intentionale Begriffe wie Wissen (im Sinne von Überzeugungen), Erwartungen, Ziele oder Motive interpretiert werden können (vgl. Carrier & Mittelstraß, 1988; Gadenne, Theorien, in diesem Band). Worauf beziehen sich derartige Begriffe? Welchen Status besitzt eine Aussage der Art „Eine Person befindet sich in einem bestimmten mentalen Zustand, etwa sie hat das Ziel, ein gegebenes Problem zu lösen?“

Der Funktionalismus ist durch die Auffassung gekennzeichnet, daß mentale Zustände ihrer logischen Natur nach von materiellen Zuständen als verschieden angesehen werden müssen. Mentale Zustände können nicht sinnvoll durch Eigenschaften ihrer materiellen Realisierung, also etwa als Gehirnzustände definiert werden, sondern nur funktional. Sie sind charakterisiert durch ihre kausalen Beziehungen zu anderen mentalen Zuständen, zu äußeren Bedingungen und zum Verhalten. Entscheidend ist ihr kausales Potential, durch das sie in der Informationsverarbeitung wirksam werden können. Ein abstrakter funktionaler Zustand ist auf multiple, verschiedenartige Weise materiell realisierbar. Daraus folgt umgekehrt, daß für die Beschreibung der funktionalen Organisation eines informationsverarbeitenden Systems, verstanden als die Gesamtheit seiner funktionalen Zustände und ihrer Relationen, die materielle Realisierung nicht bekannt sein muß.

Angenommen, wir sind Zeugen eines schweren Verkehrsunfalls und beobachten, daß ein anderer Zeuge daraufhin zur nächsten Telefonzelle eilt. Dieses Verhalten kann offensichtlich funktional erklärt werden, wobei die resultierende Erklärung auch zur Verhaltensvorhersage genutzt werden kann. Würde es Sinn machen, dieses Verhalten neurobiologisch beschreiben und erklären zu wollen? Es ist eine der Grundthesen der Kognitionswissenschaft, daß die Erklärung menschlichen Verhaltens durch Rückgriff auf mentalistische bzw.

intentionale Begriffe das erfolgreichste verfügbare Modell für menschliche Verhaltensklärungen darstellt. Damit geht die Auffassung einher, daß mit Hilfe funktionaler Erklärungen Regelförmigkeiten formuliert werden können, die mit anderen Kategorien von Begriffen nicht adäquat faßbar sind.

Welche Implikationen ergeben sich aus einer derartigen Betrachtungsweise? Menschen und in geeigneter Weise programmierte Computer werden als informationsverarbeitende Systeme angesehen, die eine zwar materiell unterschiedlich realisierte, aber dennoch prinzipiell vergleichbare funktionale Organisation aufweisen können. Aus dieser Vergleichbarkeit wird die Gültigkeit der starken These der KI gefolgert. Der Funktionalismus geht davon aus, daß die Prinzipien der funktionalen Organisation eines wissensbasierten Systems im Prinzip denen des menschlichen kognitiven Apparats gleichen: Maschinen sind grundsätzlich zu denselben kognitiven Leistungen wie der Mensch in der Lage (vgl. auch die auf Newell und Simon zurückgehende **Hypothese der Symbolsysteme**; Newell & Simon, 1976).

Für die kognitionspsychologische Theoriebildung hat diese Betrachtung, auch wenn man dem Funktionalismus in seiner extremen Form nicht folgt, zwei Konsequenzen. Erstens, intentionale Begriffe wie Wissen und Ziele sind wissenschaftlich sinnvolle, nicht eliminierbare Konzepte, und es ist methodologisch gerechtfertigt, sie bei der Beschreibung und Erklärung kognitiver Strukturen und Prozesse zu verwenden. Dabei werden derartigen mentalen Inhalten in Analogie zur Arbeitsweise eines wissensbasierten Systems sowohl syntaktisch-kausale als auch semantisch-logische Eigenschaften zugewiesen (vgl. die Hypothese der Wissensrepräsentation in Abschnitt 3.1.3). Zweitens sind alle postulierten Inhalte so zu präzisieren, daß sie als Programm einer Turing-Maschine, also in algorithmischer Weise formulierbar sind.

Zusammenfassend sind mentale Zustände aus der Sicht des Funktionalismus semantisch interpretierbare Größen, die bestimmte Inhalte ausdrücken und insofern semantische Attribute wie Wahrheitswerte oder Referenz aufweisen. Daneben kommen mentalen Zuständen aber auch syntaktische oder formale Eigenschaften zu, die alle diejenigen Merkmale umfassen, die von ihrer Bedeutung unabhängig sind. Damit korrespondiert, daß **kognitive Aktivitäten** als regelhafte Prozesse **über semantisch interpretierten symbolischen Repräsentationen** angesehen werden können, wie dies Pylyshyn (1987) in seinem „**representational metapostulate**“ explizit formuliert hat:

Computation, information processing, and rule-governed behavior all depend on the existence of physically instantiated codes or symbols that refer to or represent things and properties extrinsic to the behaving system. In all these cases the behavior of the systems in questions (be they minds, computers or social systems) are explained not in terms of intrinsic properties of the system itself but in terms of rules and processes that operate on representations of extrinsic things. **Cognition, in other words, is ex-**

**plained in terms of regularities holding over semantically interpreted symbolic representations, . . .** (Hervorhebung von den Verfassern).

(Pylyshyn, 1987, S. 133)

### 5.3 Das chinesische Zimmer

Die starke These der KI bzw. die These einer prinzipiellen Vergleichbarkeit menschlichen und maschinellen Verstehens ist aus verschiedenen Positionen heraus angegriffen worden (vgl. etwa Dreyfus, 1979; Winograd & Flores, 1987; Penrose, 1989). Die Position des vielleicht bekanntesten Kritikers, John Searle, soll in diesem Abschnitt skizziert werden. Dabei wird anknüpfend an die Überlegungen zum Turing Test insbesondere die Frage diskutiert, ob aus dem Verhalten eines Systems dessen kognitive Fähigkeiten erschließbar sind bzw. anders formuliert, ob ‚intelligentes‘ Handeln Verstehen voraussetzt.

Unter dem Begriff **chinesisches** Zimmer ist ein Gedankenexperiment berühmt geworden, das auf Searle (1980) zurückgeht. Ein modernes KI-Sprachverarbeitungssystem besitzt zweifellos eine erstaunliche Leistungsfähigkeit. Es liefert gute Fragen und vernünftige Antworten. Was aber versteht ein „sprachverstehendes“ System wirklich? Ist es nicht so, daß lediglich eine rein syntaktische Verknüpfung gewisser Zeichenreihen, die dem System eingegeben werden, mit gewissen anderen Zeichenreihen, die vom System ausgegeben werden, erfolgt, ohne daß dabei die Bedeutung der Zeichen in irgendeiner Weise ‚verstanden‘ wird? Zur Diskussion dieser Frage entwirft Searle das folgende Szenario:

Eine Person, etwa Searle selbst, befinde sich in einem abgeschlossenen Raum, vor sich einen Stapel in chinesischer Sprache verfaßten Textes, von dem er nicht ein Wort versteht, ja noch nicht einmal weiß, daß es sich um chinesische Schriftzeichen handelt. In diesen Raum wird nun zunächst von außen ein zweiter Stapel chinesischen Textes hereingereicht, zusammen mit einer Menge von in englischer Sprache formulierten Regeln, mit deren Hilfe die Elemente des ersten und des zweiten Stapels auf rein mechanische Weise zueinander in Beziehung gesetzt werden können (etwa ausschließlich aufgrund der Berücksichtigung ihrer äußerlichen Form). Schließlich werde noch ein dritter Stapel chinesischen Textes, zusammen mit einer Menge von englischsprachlichen Anweisungen zur Verfügung gestellt, die es Searle ermöglichen, Elemente des dritten Stapels mit Elementen aus den beiden anderen Stapeln in Beziehung zu bringen und für bestimmte Zeichen aus dem dritten Stapel mechanisch bestimmbare chinesische Zeichen aus den beiden anderen Stapeln zurückzugeben.

Bezugnehmend auf das Skriptkonzept bzw. die hierauf beruhenden Dialogsysteme von Roger Schank (vgl. Schank & Abelson, 1977), die den Anspruch erheben, Sprache „verstehen“ zu können (im Unterschied zum bloßen Erwecken eines entsprechenden Anscheins), führt Searle weiter aus:

Unknown to me, the people who are giving me all of these symbols call the first batch a “script”, they call the second batch a “story”, and they call the third batch “questions”. Furthermore, they call the symbols I gave them back in response to the third batch “answers to the questions” and the set of rules in English that they gave me, they call “the program”. (S.418)

Werden nun Searle in dieser Situation chinesische „Fragen“ hereingereicht, dann können zwar für einen externen Beobachter die chinesischen Zeichen, die Searle als „Antworten“ zurückgibt, durchaus als „sinnvolle“ Antworten erscheinen, vorausgesetzt der externe Beobachter ist der chinesischen Sprache mächtig. Searle aber, der die Fragen mechanisch mit Hilfe der englischen Regeln und Anweisungen und ausschließlich aufgrund äußerlicher („syntaktischer“) Merkmale beantwortet hat, versteht nach wie vor offenbar kein einziges Wort Chinesisch:

It seems to me quite obvious in the example that I do not understand a word of Chinese stories. I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, but I still understand nothing. For the same reasons, Schank's computer understands nothing of any stories, whether in Chinese, English, or whatever, since in the Chinese case the computer is me, and in cases where the computer is not me, the computer has nothing more than I have in the case where I understand nothing. (S.418)

Nehmen wir zur Komplikation der Situation weiter an, daß man dem eingeschlossenen Searle nicht nur chinesische, sondern auch englische Fragen gibt, die Searle als „native speaker“ zweifellos so beantworten könnte, daß seine Antworten für einen externen Beobachter als die Antworten einer die englische Sprache „verstehenden“ Person beurteilt werden. Gibt es zwischen beiden Situationen (Verstehen chinesischer vs. englischer Sprache) einen Unterschied und wenn ja, woran ist er zu erkennen?

Searle ist der Meinung, daß es einen prinzipiellen Unterschied zwischen seinem Verstehen als Mensch und dem Verstehen eines Computers gibt. Menschen verfügen danach über die Fähigkeit, Bedeutung unmittelbar zu erfassen, wofür er den Begriff der **intrinsic Intentionalität** einführt. Menschen wissen, was sie meinen - meint Searle - niemals jedoch ein Computer, da seine Arbeitsweise auf rein syntaktisch-formalen Operationen beruht, die keine Semantik beinhalten. Er begründet seine Argumentation im Kern damit, daß die Eigenschaft der intrinsic Intentionalität an die besondere biologische Ausstattung des Menschen, insbesondere an sein Gehirn, gebunden ist. Diese spezielle materielle, genauer gesagt neurobiologische Bedingtheit mentaler Zu-



stände impliziert für Searle, daß die starke These der KI als unzutreffend zurückzuweisen ist.

Diese Argumentation Searles ist selbst wiederum keineswegs unumstritten. Befürworter des Funktionalismus begegnen ihr etwa mit dem Einwand, daß die Zuschreibung von Verstehen nicht von einer Kenntnis der internen Zustände bzw. Gehirnzustände einer Person abhängen könne. Wovon aber dann? Zu dieser Frage kann Searle keine konstruktive Perspektive aufzeigen. Für ihn steht der Gedanke im Vordergrund, daß ‚richtiges‘ Handeln kein hinreichendes Kriterium für Verstehen ist. Aber welche empirischen Grundlagen für die Zuschreibung kognitiver Fähigkeiten blieben dann für die Psychologie? Damit sind wir aber wieder bei Fragen, die seit Turing im Mittelpunkt der wissenschaftstheoretischen Diskussion stehen.

## 6. Epilog

Was ist aber nun unsere eigene kognitionspsychologische Perspektive auf die zuvor aufgeworfenen wissenschaftstheoretischen Fragen?

Für uns sind wissensbasierte Systeme - entsprechend der schwachen These der KI - ein wichtiges Werkzeug zur Erforschung der menschlichen Kognition. Aber auch die starke These wird von uns nicht als abwegig beurteilt. Dabei ist sorgfältig zwischen den kognitiven Prozessen selbst und ihrem jeweiligen Träger zu unterscheiden. Der Mensch **hat** Geist und er **ist** intelligent, während die Maschine agiert, als **ob** sie Geist hätte, als **ob** sie intelligent sei. Die kognitiven Prozesse selbst können je nach ihren Eigenschaften - aber unabhängig von der Form ihrer materiellen Realisierung und damit von ihrem Träger - als mehr oder minder flexible, intelligente usw. Prozesse bezeichnet werden. Das Modellerte und seine Modellierung funktionieren nach denselben Prinzipien der Informationsverarbeitung. Die Simulation eines Denkprozesses ist ein Denkprozeß, simulierte Wahrnehmung ist Wahrnehmung. Damit ist aber noch nichts darüber ausgesagt, ob menschliches Denken und Wahrnehmen nicht über den informativisch rekonstruierbaren Anteil hinaus durch weitere Anteile geprägt sind.

Diskussionen zur schwachen und starken These der KI übersehen auch häufig, daß man noch weit davon entfernt ist, menschliche Kognition so zu modellieren, daß die Rekonstruktion den Turing-Test bestehen könnte, oder Herrn Searle im chinesischen Zimmer dazu befähigen wurde, einen intelligenten, wenn auch vielleicht unverstandenen Dialog zu führen. Tatsächlich ist dieses Ziel noch fern. Aber wir glauben, daß es eine lohnende Forschungsperspektive aufzeigt.

Weiter oben haben wir darauf hingewiesen, daß für die Beschreibung der funktionalen Organisation eines informationsverarbeitenden Systems die materielle Realisierung nicht bekannt sein muß. Dies darf nicht dahingehend mißverstanden werden, daß die materielle Realisierung unerheblich sei. Sie legt den Raum der Möglichkeiten fest. Bei Ausfall bestimmter Komponenten auf der materiellen Ebene ergeben sich u.U. dramatische Konsequenzen für die funktionale Ebene.

Kehren wir aber zur Frage der Modellierung kognitiver Strukturen und Prozesse mit Hilfe wissensbasierter Systeme zurück. Wir sind der Auffassung, daß wissensbasierte Systeme durch eine Reihe von Eigenschaften charakterisiert sind, die sie zu einer ausgezeichneten Kategorie von Modellen machen, um unser Verständnis kognitiver Vorgänge beim Menschen zu verbessern. Sie erlauben eine Form der Theoriebildung, die durch grundlegende wissenschaftstheoretische Ideale gekennzeichnet ist. Darüber hinaus liegt wissensbasierten Systemen ein äußerst anspruchsvoller Erklärungs begriff zugrunde. Wissensbestände definieren ein Kompetenzmodell im Sinne eines Potentials, das verhaltenswirksam werden kann, aber nicht muß. Ihre Aktivierung im Prozeß der Auswahl und Anwendung setzt die Spezifizierung generativer Mechanismen voraus. Die in Frage stehenden Performanzphänomene werden nicht nur oberflächlich beschrieben, sondern in einer konstruktiven Weise (re-)generiert. Die Eigenschaft der Suffizienz von Computermodellen liefert zugleich eine Art von **hinreichendem Existenzbeweis**, insofern das postulierte Modell sich als hinreichend zur Erzeugung der von ihm zu erklärenden Phänomene erwiesen hat.

Wissensbasierte Systeme führen zu einer **inkrementellen Theorieentwicklung**. Als generative Prozeßmodelle ermöglichen sie eine quasi-experimentelle Methodologie, insofern Hypothesen aufgestellt und ihre Implikationen als Simulationen unmittelbar im Modell überprüft werden können. Durch die Möglichkeit einer differenzierten Analyse der Leistungsfähigkeit, aber auch der Grenzen existierender Computersysteme sowie durch die Verwendung einer präzisen Sprache wird eine systematische Nutzung von bekannten Ergebnissen und ihre gezielte Weiterführung möglich. Neue Methoden wie beispielsweise die modellvergleichende Argumentation wurden entwickelt bzw. sind künftig zu entwickeln, um den damit einhergehenden Anforderungen im Prozeß der Theorieentwicklung und -testung gerecht werden zu können.

Wissensbasierte Systeme sind dann Teil (kognitions-) psychologischer Theoriebildung, wenn ein valides Abbild mentaler Strukturen und Prozesse angestrebt wird. Sind aber nur valide Modelle von Interesse, oder verdienen auch solche Entwicklungen unsere Aufmerksamkeit, bei denen der Mensch nicht explizit als Modell und Maßstab herangezogen wird und der ingenieurmäßige Aspekt der Modellierung im Vordergrund steht? Unsere Antwort auf diese

Frage ist durchaus positiv. Die Unterscheidung zwischen einer ingenieurmäßigen und einer psychologischen Betrachtung intelligenter Leistungen kann im Forschungsalltag kaum durchgehalten werden. Viele Realisierungen wissensbasierter Systeme orientieren sich in der einen oder anderen Form am Menschen und jede psychologisch motivierte Modellierung mit Hilfe wissensbasierter Systeme greift notwendigerweise auf in der Informatik entwickelte Methoden zurück. Ein Beispiel für die psychologische Fruchtbarkeit von Vorstellungen, für die nur selten der Anspruch empirischer Adäquatheit erhoben wird, ist der Bereich des maschinellen Lernens (vgl. etwa Kodratoff, 1988, oder das von Carbonell herausgegebene Sonderheft der Zeitschrift *Artificial Intelligence*, 1989). Diese Arbeiten haben in vielfältiger Weise psychologische Forschungen angeregt, die sich auf der Basis der in diesem Zusammenhang entwickelten Konzepte in neuer Weise mit Fragen menschlichen Lernens beschäftigen.

Diesen Argumenten für eine Modellierung kognitiver Vorgänge mit Hilfe wissensbasierter Systeme stehen aber auch eine Reihe von Problemen gegenüber. An erster Stelle ist das Spannungsfeld zwischen dem Auflösungsgrad, dem Aufwand einer Modellierung und ihrem Geltungsbereich zu nennen: Je feiner der Auflösungsgrad der Betrachtung, desto aufwendiger werden die Modelle, bei gleichzeitig eher eingeschränktem Geltungsbereich. Darüber hinaus hat die Anwendung von Methoden und Modellvorstellungen aus dem Computerbereich in fast schon paradoxer Weise dazu geführt, daß wir zunehmend genauer wissen, in welcher Weise menschliches Denken durch diese Modelle gerade nicht rekonstruiert werden kann. Obwohl wir über die Funktionsweise des informationsverarbeitenden Systems Mensch in vielen Bereichen nur eine begrenzte Anzahl positiver Aussagen machen können, so sind wir doch häufig sicher, daß es so nicht sein kann, wie es in einem speziellen wissensbasierten System postuliert wird, etwa aufgrund unrealistischer Annahmen zur Speicherkapazität.

Weiter haben die Forschungen der letzten vierzig Jahre deutlich gemacht, daß eine ausschließlich logisch-rationale Sicht von Kognition vielen grundlegenden Aspekten menschlichen Denkens und Handelns nicht gerecht wird. Auch ist es kein Zufall, daß gerade das sog. Alltagswissen mit all seinen Lücken, Unschärfen, Inkonsistenzen und Faustregeln sich einer Formalisierung bislang weitgehend entzieht, während für viele sog. ‚höhere‘ kognitive Funktionen gut bewährte Vorstellungen entwickelt werden konnten. In der Forschung zur Künstlichen Intelligenz ist man heute eher in der Lage, Schachprogramme mit der Spielstärke von Großmeistern zu entwickeln als ein halbwegs akzeptables Urlaubsberatungssystem. Kognitionspsychologisch läßt sich eher der über viel quantitatives Formelwissen verfügende physikalische Experte modellieren als ein Physikstudent, der mühsam damit kämpft, sein konzeptuelles Alltagswissen mit dem naturwissenschaftlichen Fachwissen zu integrieren. Und es bedarf

noch eines weiten Weges, einen Roboter zu entwickeln, der wohlbehalten eine Straße überquert, die gegenüberliegende Wirtschaft betritt, dort ein Glas Bier bestellt und eine Unterhaltung beginnt - vielleicht sollten wir letzteres aber auch uns vorbehalten.

## **Literatur**

- Aebli, H. (1980). **Denken: Das Ordnen des Tuns** (Band 1: Kognitive Aspekte der Handlungstheorie). Stuttgart: Klett-Cotta.
- Allen, J. (1987). **Understanding natural language**. Menlo Park, CA: Benjamin/Cummings.
- Anderson, J.R. (1982). Acquisition of cognitive skill. **Psychological Review**, 89, 369-406.
- Anderson, J. R. (1983). **The architecture of cognition**. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. **Psychological Review**, 94, 192-210.
- Anderson, J. R. (1990). **The adaptive** character of thought. Hillsdale, NJ: Erlbaum.
- Anzai, Y. & Simon, H. A. (1979). The theory of learning by doing. **Psychological Review**, 86, 124-140.
- Berwick, R. C. (1985). **The acquisition of syntactic knowledge**. Cambridge, MA: MIT Press.
- Brachman, R. & Levesque, H. (Eds.). (1985). **Readings in knowledge representation**. San Mateo, CA: Morgan Kaufmann.
- Brauer, W. (1990). Grenzen maschineller Berechenbarkeit. **Informatik-Spektrum**, 13, 61-70.
- Brown, J.S. & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. **Cognitive Science**, 4, 379-427.
- Burton, R.R. (1982). Diagnosing bugs in a simple procedural skill. In D. Sieeman & J. S. Brown (Eds.), **Intelligent tutoring systems** (pp. 157-183). New York, NY: Academic Press.
- Carbonell, J. G. (Ed.). (1990). Machine learning. **Artificial Intelligence**, 40.
- Carrier, M. & Mittelstraß, J. (1989). **Geist, Gehirn, Verhalten: Das Leib-Seele-Problem und die Philosophie der Psychologie**. Berlin: De Gruyter.
- Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P. & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. **Cognitive Science**, 13, 145-182.
- Chomsky, N. (1963). Formal properties of grammars. In R.D. Luce, R.R. Bush & E. Galanter (Eds.), **Handbook of mathematical psychology**. Volume II (pp. 323-418). New York, NY: Wiley.
- Church, A. (1936). An unsolvable problem for elementary number theory. **American Journal of Mathematics**, 58, 345-363.

- Collins, A. M. & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407-428.
- Collins, A. M. & Smith, E. E. (Eds.). (1988). *Readings in cognitive science*. San Mateo, CA: Morgan Kaufmann.
- Daiser, W. (1984). **Künstliche Intelligenz** Forschung **und ihre epistemologischen Grundlagen**. Frankfurt/Main: Lang.
- Davis, E. (1990). **Representations of commonsense knowledge**. San Mateo, CA: Morgan Kaufmann.
- Deppe, W. (1977). **Formale Modelle in der Psychologie**. Stuttgart: Kohlhammer.
- Dörner, D. (1989). Die kleinen grünen Schildkröten und die Methoden der experimentellen Psychologie. *Sprache & Kognition*, *8*, 86-97.
- Dreyfus, H. (1979). **What computers can't do: The limits of artificial intelligence**. New York, NY: Harper and Row. (Dt. Was Computer nicht können: Die Grenzen künstlicher Intelligenz. Königstein/Ts.: Athenäum, 1985).
- Gardner, H. (1985). **The mind's new science**. New York, NY: Basic Books (Dt. Dem Denken auf der Spur. Stuttgart: Klett-Cotta, 1989).
- Genesereth, M. R. & Nilsson, N.J. (1987). **Logical foundations of artificial intelligence**. Los Altos, CA: Morgan Kaufmann. (Dt. Logische Grundlagen der Künstlichen Intelligenz. Braunschweig: Vieweg, 1989).
- Gigerenzer, G. (1981). **Messung und Modellbildung in der Psychologie**. München: Reinhardt.
- Gigerenzer, G. (1988). Woher kommen Theorien über kognitive Prozesse? *Psychologische Rundschau*, *39*, 91-100.
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, *98*, 254-267.
- Gödel, K. (1931). über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. *Monatshefte für Mathematik und Physik*, *38*, 173-198.
- Herrmann, T. (1982). Über begriffliche Schwächen kognitivistischer Kognitionstheorien: Begriffsinflation und Akteur-System-Kontamination. *Zeitschrift für Sprache und Kognition*, *1*, 3-14.
- Herrmann, T. (1988). Mentale Repräsentation - ein erläuterungsbedürftiger Begriff. *Sprache & Kognition*, *7*, 162-175.
- Hildreth, E. C. & Ulkman, S. (1989). The computational study of vision. In M. I. Posner (Ed.), **Foundations of cognitive science** (pp. 581-630). Cambridge, MA: MIT Press.
- Holland, J., Holyoak, K., Nisbett, R. E. & Thagard, P. (1986). **Induction: Processes of inference, learning and discovery**. Cambridge, MA: The MIT Press.
- Hopcroft, J. E. & Ullman, J.D. (1979). **Introduction to automata theory, languages, and computation**. Reading, MA: Addison-Wesley.
- Johnson-Laird, P. N. (1983). **Mental models**. Cambridge: Cambridge University Press.
- Johnson-Laird, P.N. (1988). **The Computer and the mind**. Cambridge, MA: Harvard University Press.
- Kleene, S.C. (1936). General recursive functions of natural numbers. *Mathematical Annals*, *112*, 727-742.

- Klix, F. (1988). Gedächtnis und Wissen. In H. Mandl & H. Spada (Hg.), *Wissenspsychologie (S. 19-54)*. München: Psychologie Verlags Union.
- Kodratoff, Y. (1988). *Introduction to machine learning*. London: Pitman.
- Krämer, S. (1988). *Symbolische Maschinen: Die Idee der Formalisierung in geschichtlichem Abriß*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Laird, J. E., Newell, A. & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, **33**, 1-64.
- Luce, R.D., Bush, R.R. & Galanter, E. (Eds.). (1963). *Handbook of mathematical psychology. Volume I-III*. New York, NY: Wiley.
- Malpas, J. (1987). *Prolog: A relational language and its applications*. Englewood Cliffs, NJ: Prentice-Hall.
- Markov, A. A. (1954). *Theory of algorithms*. Moskau: National Academy of Science.
- Marr, D. (1982). *Vision*. New York, NY: Freeman.
- McClelland, J. A. & Rumelhart, D. E. (Eds.). (1986). Parallel *distributed processing: Explorations in the microstructure of cognition. Volume II: Psychological and biological models*. Cambridge, MA: The MIT Press.
- McClelland, J. A. & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs and exercises*. Cambridge, MA: The MIT Press.
- Minsky, M. (1967). *Computation: Finite and infinite machines*. Englewood Cliffs, NJ: Prentice-Hall.
- Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York, NY: McGraw-Hill.
- Möbus, C. (1988). Zur Modellierung kognitiver Prozesse mit daten- bzw. zielorientierten Regelsystemen. In H. Mandl & H. Spada (Hrsg.), *Wissenspsychologie (S. 423-468)*. München: Psychologie Verlags Union.
- Moravcsik, J. M. (1980). Chomsky's radical break with modern tradition (A commentary on Chomsky's Rules and Representations). *The Behavioral and Brain Sciences*, **3**, 28-29.
- Neches, R. (1982). Simulation systems for cognitive psychology. *Behavior Research Methods & Instrumentation*, **14**, 77-91.
- Newell, A. (1980). Physical symbol systems. *Cognitive Science*, **4**, 135-183.
- Newell, A. (1982). The knowledge level. *Artificial Intelligence*, **18**, 87-127.
- Newell, A.** (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Newell, A., Rosenbloom, P.S. & Laird, J.E. (1989). Symbolic architectures for cognition. In M.I. Posner (Ed.), *Foundations of cognitive science* (pp.93-132). Cambridge, MA: The MIT Press.
- Newell, A., Shaw, J. C. & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, **65**, 151-166.
- Newell, A. & Simon, H.A. (1963). GPS, a program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 279-293). New York, NY Mc Graw-Hill.

- Newell, A. & Simon, H.A. (1972). **Human problem solving**. Englewood Cliffs, NJ: Prentice-Hall.
- Newell, A. & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. **Communications of the ACM**, 19, 113-126.
- Ohlsson, S. (1988). Computer Simulation and its impact on educational research and practice. **International Journal of Educational Research**, 12, 5-34.
- Opwis, K. (1992). **Kognitive Modellierung: Zur Verwendung wissensbasierter Systeme in der psychologischen Theoriebildung**. Bern: Huber.
- Opwis, K. & Lüer, G. (in Druck). Modelle der Repräsentation von Wissen. In D. Albert & K.-H. Stapf (Hrsg.), **Enzyklopädie der Psychologie. Gedächtnispsychologie: Erwerb, Nutzung und Speicherung von Information**. Göttingen: Hogrefe.
- Osherson, D. N. & Lasniak, H. (Eds.). (1990). **An invitation to cognitive science. Volume 1: Language**. Cambridge, MA: MIT Press.
- Osherson, D. N. & Kosslyn, S.M. & Hollerbach, J.M. (Eds.). (1990). **An invitation to cognitive science. Volume 2: Visual cognition and action**. Cambridge, MA: MIT Press.
- Osherson, D. N. & Smith, E.E. (Eds.). (1990). **An invitation to cognitive science. Volume 3: Thinking**. Cambridge, MA: MIT Press.
- Penrose, R. (1989). **The emperor's new mind: Concerning computers, minds, and the laws of physics**. Oxford: Oxford University Press. (Dt. Computerdenken: Die Debatte um künstliche Intelligenz, Bewußtsein und die Gesetze der Physik. Heidelberg: Spektrum der Wissenschaft, 1991).
- Pinker, S. (1989). **Learnability and cognition: The acquisition of argument structure**. Cambridge, MA: MIT Press.
- Plötzner, R. & Spada, H. (1992). Analysis-based learning on multiple levels of mental domain representation. In E. De Corte, M. Linn, H. Mandl, & L. Verschaffel (Hrsg.), **Computer-based learning environments and problem solving** (pp. 103-127). Berlin: Springer.
- Plötzner, R., Spada, H., Stumpf, M. & Opwis, K. (1991). Learning qualitative and quantitative reasoning in a microworld for elastic impacts. **European Journal of Psychology and Education**, 4, 501-516.
- Posner, M.I. (Ed.). (1989). **Foundations of cognitive science**. Cambridge, MA: MIT Press.
- Post, E. L. (1943). Formal reductions of the general combinatorial decision problem. **American Journal of Mathematics**, 65, 197-215.
- Polyshyn, Z. W. (1985). **Computation and cognition**. Cambridge, MA: The MIT Press.
- Polyshyn, Z. W. (1987). Cognitive Science. In S. C. Shapiro (Ed.), **Encyclopedia of artificial intelligence, Volume 1** (pp. 120-124). New York, NY: Wiley.
- Polyshyn, Z. W. (1989). Computing in cognitive science. In M. I. Posner (Ed.), **Foundations of cognitive science** (pp.51-91). Cambridge, MA: The MIT Press.
- Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), **Semantic information processing** (pp.227-270). Cambridge, MA: The MIT Press.

- Reimann, P. & Schult, T. J. (1991). Modeling example elaboration strategies. In ***Proceedings of the First International Conference on the Learning Sciences, Evanston, IL***.
- Rogers, H. (1987). ***Theory of recursive functions and effective computability***. Cambridge, MA: The MIT Press.
- Rumelhart, D. E. & McClelland, J. A. (Eds.). (1986). ***Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations***. Cambridge, MA: The MIT Press.
- Schank, R. C. & Abelson, R.P. (1977). ***Scripts, plans, goals and understanding***. Hillsdale, NJ: Erlbaum.
- Scheffe, P. (1986). ***Künstliche Intelligenz - überblick und Grundlagen***. Mannheim: B. I. Wissenschaftsverlag.
- Searle, J. (1980). Minds, brains and programs. ***The Behavioral and Brain Sciences*, 3**, 417-457.
- Selz, O. (1913). über die Gesetze des geordneten Denkverlaufs**. Stuttgart: Speman.
- Simon, D.P. & Simon, (1978). Individual differences in solving physics problems. In R. Siegler (Ed.), ***Children's thinking: What develops*** (pp.325-348)? Hillsdale, NJ: Lawrence Erlbaum.
- Smith, B. C. (1982). ***Reflection and semantics in a procedural language*** (Technical Report). Cambridge, MA: The Massachusetts Institute of Technology.
- Smolensky, P. (1988). On the proper treatment of connectionism. ***Behavioral and Brain Sciences*, 11**, 1-74.
- Spada, H. & Reimann, P. (1988). Wissensdiagnostik auf kognitionswissenschaftlicher Basis. ***Zeitschrift für Differentielle und Diagnostische Psychologie*, 9**, 183-192.
- Stachowiak, H. (1973). ***Allgemeine Modelltheorie***. Berlin: Springer.
- Sterling, L. & Shapiro, E. (1986). ***The art of Prolog: Advanced programming techniques***. Cambridge, MA: MIT Press. (Dt. Prolog: Fortgeschrittene Programmier-techniken. Bonn: Addison-Wesley, 1988).
- Stetter, F. (1988). ***Grundbegriffe der theoretischen Informatik***. Berlin: Springer.
- Strube, (1990). Neokonnektionismus: Eine neue Basis für die Theorie und Modellierung menschlicher Kognition? ***Psychologische Rundschau*, 41**, 129-143.
- Tack, W. (1987). Ziele und Methoden der Wissensrepräsentation. ***Sprache & Kognition*, 3**, 150-163.
- Turing, A. (1936). On computable numbers with an application to the Entscheidungsproblem. ***Proceedings London Math. Society*, 42**, 230-265. A correction, *ibid.*, 43, 544-546.
- Turing, A. (1950). Computing machinery and intelligence. ***Mind*, 59**, 433-460. (Dt. Kann eine Maschine denken?, Kursbuch, 8 (1967), 106-138).
- VanLehn, K. (1983). On the representation of procedures in repair theory. In M.P. Ginsburg (Ed.), ***The development of mathematical thinking*** (pp. 197-252). New York, NY: Academic Press.
- VanLehn, K. (1989). ***Mind bugs***. Cambridge, MA: MIT Press.



- VanLehn, K. (1991). Two pseudo-students: Applications of machine learning to formative evaluation. In R. Lewis & S. Otoriki (Eds.), **Advanced research on computers in education** (p. 17-26). Elsevier: North Holland.
- VanLehn, K., Brown, J.S. & Greeno, J. (1984). Competitive argumentation in computational theories of cognition. In W. Kintsch, J. R. Miller & P. G. Polson (Eds.), **Methods and tactics in cognitive science** (pp.235-262). Hillsdale, NJ: Erlbaum.
- VanLehn, K., Jones, R. M. & Chi, M.T. H. (1992). A model of the self-explanation effect. **The Journal of the Learning Sciences, 2**, 1-60.
- Wenger, E. (1987). **Artificial intelligence and tutoring systems**. Los Altos, CA: Morgan Kaufmann.
- Winograd, T. (1972). Understanding natural language. **Cognitive Psychology, 3**, 1-191.
- Winograd, T. (1983). **Language as a cognitive process. Volume I: Syntax**. Reading, MA: Addison-Wesley.
- Winograd, T. & Flores, F. (1986). **Understanding computers and cognition**. Norwood, NJ: Ablex Publishing Corporation. (Dt. **Erkenntnis - Maschinen - Verstehen**. Berlin: Rotbuch, 1989).

Teil C  
Theorien und Forschungsprogramme

## **Kapitel 6**

### **Forschungsprogramme**

von Theo Herrmann

## **Kapitel 7**

### **Theorien**

von Volker Gadenne

## **Kapitel 8**

### **Heuristik der Theorienbildung**

von Dietrich Dörner

## **Kapitel 9**

### **Theoriebewertung**

von Volker Gadenne

## **Kapitel 10**

### **Induktion**

von Rainer Westermann und Peter Gerjets

Die Psychologie setzt sich wie jede Erfahrungswissenschaft das Ziel, Forschungsgegenstände möglichst strikt zu bestimmen und empirische Ergebnisse zu systematisieren. Das wichtigste Mittel zur Erreichung dieser Ziele sind Theorien. Im Teil C dieses Bandes wird die Stellung von Theorien im Wissenschaftsprozess erörtert. Dieser Prozeß wird als ein dynamisches Netzwerk von Forschungsprogrammen expliziert (Kapitel 6), und es werden (in Kapitel 7) Theorien zunächst für sich betrachtet: Wie sind sie aufgebaut? Was heißt Formalisierung und Mathematisierung von Theorien? Wie verhalten sie sich zur „Realität“? Die Kapitel 8 und 9 handeln dann von der wissenschaftstheoretischen Grundlegung und von methodologischen Regeln der Theorienbildung und Theoriebewertung. Welchen Status haben heuristische Regeln der Theoriebildung? Nach welchen methodologischen Prinzipien werden Theorien evaluiert? Wie lassen sich solche Prinzipien normativ begründen?

Neben deduktive Begründungsstrategien treten Extrapolationen, Generalisierungen und Abduktionen. Solche nicht unproblematischen induktiven Strategien dienen unter anderem der Gewinnung und Rechtfertigung von Verallgemeinerungen und der Stützung von Annahmen durch empirische Befunde (Kapitel 10).

## 6. Kapitel

# Forschungsprogramme

*Theo Herrmann*

Die Methodologie hat die doppelte Aufgabe, dem Wissenschaftler Handlungsregeln anzubieten und ihm geeignete Vorstellungen darüber zu vermitteln, wie seine Wissenschaft „funktioniert“ (vgl. auch Kap. 1). Dieses Kapitel enthält einen Beitrag zur Behandlung der zuletzt genannten Aufgabe. Auch die wissenschaftliche Tätigkeit von Psychologen kann dadurch gefördert werden, daß sie theoretische Modelle bzw. Rekonstruktionen der psychologischen Forschung und ihrer Voraussetzungen kennenlernen. Theoretische Rekonstruktionen der psychologischen Wissenschaft basieren in erheblichem Maße auf Ergebnissen und Vorstellungsweisen der allgemeinen Wissenschaftslehre und der speziellen Wissenschaftstheorie der Psychologie. (Vgl. zur allgemeinen Wissenschaftslehre Essler, 1970, 1971, 1973, 1979; zur Wissenschaftslehre der Psychologie Breuer, <sup>4</sup>1989; Gadanne, 1984; Groeben & Westmeyer, 1975.)

Der Gegenstand des gegenwärtigen Kapitels ist eine theoretische Rekonstruktion der psychologischen Wissenschaft, bei welcher die methodologische Funktion von psychologischen Theorien im Mittelpunkt steht. Wissenschaftliche Forschung wird dabei als ein komplexes Gefüge von Problemlösungsprozessen aufgefaßt. Jeder dieser wissenschaftlichen Problemlösungsprozesse läßt sich als ein Forschungsprogramm interpretieren. Theorien sind Mittel zur Lösung der in Forschungsprogrammen behandelten Probleme. In einem Forschungsprogramm werden Theorien als Mittel zur Lösung des eigenen Problems entwickelt, oder sie werden zu diesem Zweck von anderen Forschungsprogrammen übernommen. So entsteht ein Austausch von Theorien zwischen Forschungsprogrammen. Dieser Theorienaustausch ist Teil der dynamischen Wechselbeziehungen zwischen psychologischen Problemlösungsprozessen bzw. Forschungsprogrammen. Die Problemlösungsdynamik der wissenschaftlichen Psychologie steht auch in erheblichem Ausmaß unter dem Einfluß externaler Bedingungen von der Art der Wissenschaftsmoden, der „Paradigmen“, aber auch der Organisation staatlicher Forschungsförderung, des Publikationswesens, usf. Solche externalen Einflüsse auf Problemlösungsprozesse in psychologischen Forschungsprogrammen sind ebenfalls Gegenstand dieses Kapitels.

Nach allem werden hier psychologische Theorien in ihrer Funktion im Wissenschaftsprozess, konkret: als Mittel zur Lösung von psychologischen Problemen in Forschungsprogrammen behandelt. Die theoretische Rekonstruktion solcher Forschungsprogramme, ihrer internen Struktur, ihrer Verschiedenartigkeit, ihrer Determinanten wie auch ihrer vielfältigen externen Beziehungen soll einen Einblick in Voraussetzungen vermitteln, unter denen konkrete psychologische Forschungsarbeit erfolgt.

## 1. Was sind Forschungsprogramme?

### 1.1 Begriffsexplikation

Man kann den psychologischen Forschungsprozeß als die gemeinsame, interaktive Tätigkeit von Forschern und ihren Mitarbeitern in bestimmten Zeiträumen und an bestimmten Orten, als Teil bestimmter Institutionen, beim Vorliegen bestimmter Ressourcen auffassen. Dabei gilt diese Tätigkeit der Bearbeitung bestimmter fachspezifischer Probleme. (Vgl. auch Herrmann, 1979, S. 28ff.; Stachowiak, 1973, S. 132f.) Ausgangsbegriff der im gegenwärtigen Kapitel dargestellten theoretischen Rekonstruktion dieser Problemlösungstätigkeit ist das Forschungsprogramm (Herrmann, 1976, S. 42f.). Forschungsprogramme lassen sich für den gegenwärtigen Zweck in der folgenden Weise explizieren. (Dabei ist zu beachten, daß sich die hier verwendete Begriffsfassung von derjenigen von Lakatos (1974) unterscheidet. Lakatos versteht als Forschungsprogramme spezifische „Theorieserien“.)

- (1) Ein Forschungsprogramm liegt vor, wenn ein soziales System besteht, für das die folgenden Bestimmungen gelten:
- (2) Das soziale System besteht aus einer Gruppe von wissenschaftlichen Akteuren (z. B. einer Forschungsgruppe), die (2.1) ein gemeinsames Problem P bearbeiten und die (2.2) bei der Bearbeitung des Problems P nach einvernehmlichen (sozial definierten) Handlungsregeln agieren und interagieren.
- (3) Das Problem P ist eine im sozialen System akzeptierte Struktur von Annahmen über einen problematisierten Realitätsbereich. Diese Annahmenstruktur zu P nennen wir den Annahmenkern des Forschungsprogramms. Der Annahmenkern enthält (3.1) in variierender Weise quasidefinitische und existenzkonstatierende sowie implikative (Wenn-dann-) Annahmen. Er enthält (3.2) Annahmen auf verschiedenen semantischen Ebenen.
- (4) Der Annahmenkern enthält (in seinen Annahmen höherer semantischer Ebene) mindestens die folgenden inhaltlichen Komponenten: Annahmen darüber, daß der Istzustand des Realitätsbereichs oder des Wissens über denselben unbefriedigend und verbesserungsbedürftig (= „problema-

tisch“) ist, d.h. daß der Istzustand keinem Sollzustand entspricht; pauschale Annahmen über „Barrieren“, die den Ist- und den/einen Sollzustand trennen; unspezifische Charakterisierungen des/eines Sollzustands und generelle Annahmen über Mittel zur Transformation des Ist- in den/einen Sollzustand.

- (5) Durch den Annahmenkern zu P ist ein Forschungsprogramm während seiner gesamten Existenz wesentlich bestimmt. Insofern sind Annahmenkerne zu P forschungsprogrammspezifisch indisponibel.

## 1.2 Erläuterungen

Aus der Begriffsexplikation von „Forschungsprogramm“ ergibt sich eine Reihe von Folgerungen. Auch sind einige begriffliche Klärungen und Veranschaulichungen erforderlich.

### 1.2.1 *Raum-zeitliche Spezifikation*

Ein Forschungsprogramm existiert in einem bestimmten Zeitraum, insofern die Mitglieder einer Gruppe von Wissenschaftlern das Problem P zu bestimmten Zeiten an bestimmten Orten bearbeiten (vgl. auch Stachowiak, 1973, S. 82ff.). Forschungsprogramme sind also raum-zeitlich spezifiziert. Dabei muß es sich freilich um keine Face-to-face-Gruppe handeln; entscheidend ist, daß die Mitglieder dasselbe Problem P haben und es interaktiv behandeln. (Zur Interaktion vgl. allgemein u. a. Irle, 1975, S. 398ff.) Der Terminus „interaktiv“ ist hier in liberalisierter Weise zu verstehen: Wissenschaftliche Interaktionen können im Grenzfall auch einseitig erfolgen; das Mitglied B der ein Problem P bearbeitenden Gruppe kann in seinem Handeln von Mitglied A beeinflußt werden - nicht aber umgekehrt. Und das „Medium“ der Interaktion kann allenfalls nur aus Texten bestehen: B kann dadurch beeinflußt sein, daß B einen von A verfaßten Text liest. Zur Klarstellung sei aber betont, daß dies nicht-prototypische Formen des forschungsprogrammspezifischen Interagierens sind. Die „kanonische“ Interaktion ist die normale, in wechselseitiger Handlungssteuerung konstituierte wissenschaftliche Kooperation und Kommunikation (z. B. gemeinsame Untersuchungen, laufender Informationsaustausch, Diskussionen, gemeinsame Publikationen usw.).

### 1.2.2 *Probleme und Annahmenkerne*

Die wissenschaftliche Tätigkeit wird, vor allem in der wissenschaftstheoretischen Tradition des Kritischen Rationalismus, als das Lösen von Problemen

aufgefaßt (Popper, 1972 u.a.O.; vgl. auch Albert, 1982). Shapere (1974) hat die wissenschaftlichen Problemlösungsvorgänge theoretisch systematisiert. Er nennt den Realitätsbereich, auf den sich ein Problem bezieht (= „domain“) ein „body of information“, das für den Problemlöser ein „Object for investigation“ ist.

Die wesentliche identitätsverleihende Komponente eines Forschungsprogramms ist das Problem P. Was hier als Problem generell und als wissenschaftliches Problem speziell verstanden werden soll, kann u. a. mit Hilfe der beiden folgenden Bestimmungsstücke erläutert werden: Zum einen gehört zu einem Problem notwendigerweise dasjenige, was vom Problemlöser problematisiert wird, wofür er eine Lösung sucht. Der Problemlöser hat also immer Annahmen über den problematisierten Realitätsbereich. Zum anderen kogniziert der Problemlöser diesen Realitätsbereich notwendigerweise als problematisch (d.h. als irgendwie defizitär, veränderungsbedürftig, lösungsbedürftig o. dgl.). Anders formuliert (vgl. dazu Hussy, 1984 (Bd. 1), S. 114ff.), befindet sich der Realitätsbereich, so wie ihn der Problemlöser auffaßt, in einem defizienten Istzustand. Angezielt ist vom Problemlöser ein Sollzustand des Realitätsbereichs, der aber vom Istzustand zunächst durch eine „Barriere“ getrennt ist. Die Lösung des jeweiligen Problems besteht dann in der Transformation des Istzustands in den Sollzustand, also in der Überwindung der „Barriere“. Die Überwindung der „Barriere“ ist im Falle der Problemlösung nicht durch die einfache Anwendung von erlernten Routineoperationen möglich; Problemlösungen verlangen vom Problemlöser kreative Leistungen (s. unten S. 256).

Das Problem P eines Forschungsprogramms ist durch seinen Annahmehkern bestimmt (s.oben). Der Annahmehkern von P umfaßt die beiden genannten Bestimmungsstücke: Annahmen über einen Realitätsbereich und Annahmen darüber, daß und inwiefern er problematisch ist.

Unter Bezugnahme auf Dörner (1979, S. 11ff.) kann man wissenschaftliche Probleme wie folgt spezifizieren: Unbeschadet der Tatsache, daß Wissenschaftler in Forschungsprogrammen immer auch Routineaufgaben und solche Teilprobleme lösen müssen, bei denen Kriterien für die Erreichung des jeweiligen Sollzustands (= Zielkriterien) bekannt sind, handelt es sich bei den als ganze betrachteten Problemen P doch um Probleme ohne klare Zielkriterien: Zufolge des Annahmehkerns befindet sich der Realitätsbereich - wie bei jedem Problem - zunächst in einem defizienten Istzustand. Die Annahmen zu diesem Istzustand sind relativ klar und bestimmt. Die Akteure haben aber keine hinreichende Information darüber, wie der erstrebte Sollzustand beschaffen sein soll. Der Sollzustand schwebt den Problemlösern allenfalls vage vor und kann zum Beispiel zunächst nur in der Art beschrieben werden, etwas solle besser erkannt, besser beschreibbar, besser erklärbar oder besser beherrschbar

sein. Häufig verfügen die Proponenten von Forschungsprogrammen erst dann über adäquate Annahmen zu Zielkriterien, wenn sie ihr Ziel erreicht haben. Vor allem grundlagenwissenschaftliche Forschungsprogramme (s. unten S.270) arbeiten kaum einmal unter der Maxime: „Gegeben ist das wohldefinierte Ziel Z; gesucht ist der Weg W dahin.“

Wir erläutern das bisher Ausgeführte an einem Beispiel aus der psychologischen Grundlagenforschung (vgl. Herrmann, 1976, S. 118): In einem langjährigen Forschungsprogramm ist das menschliche Leistungsstreben bzw. die Leistungsmotivation behandelt worden. Prominente Mitglieder des sozialen Systems, in dem die Leistungsmotivation bearbeitet wurde, sind u. a. McClelland (1955), Atkinson (1958), Heckhausen (1967), Weiner (1972) und Meyer (1973). (Vgl. dazu Heckhausen, 1989, S. 231ff.)

Der Annahmenkern dieses Programms enthält - ohne Anspruch auf Vollständigkeit - die folgenden Annahmen:

Menschen handeln. Es gibt eine Klasse menschlicher Handlungsweisen, die unter dem Gesichtspunkt des Leistens bewertet werden. Menschen setzen für ihr Handeln Normen. Sie erleben angesichts ihrer Normsetzungen Erfolge und Mißerfolge. Sie hoffen auf Erfolge und fürchten Mißerfolge. Sie suchen Erfolgssituationen auf und meiden Mißerfolgssituationen. Die Hoffnungen und Befürchtungen hängen von früheren individuellen Erfahrungen ab. Das Leistungsstreben ist etwas anderes als das Streben nach Anerkennung oder nach Macht. Menschen unterscheiden sich danach, welche Leistungsnormen sie sich setzen, wie stark sie zur Erfolgshoffnung oder Mißerfolgsschreck neigen, wieweit sie Leistungssituationen suchen oder meiden. (Usf.) - Man weiß zuwenig über die Gründe für diese Unterschiede. Man kann die Unterschiede empirisch nur schlecht erfassen. Die wissenschaftliche Erfassung von Unterschieden des Leistungsstrebens und die Aufklärung der Gründe ihres Zustandekommens erscheinen möglich, und sie sind erwünscht. (Usf.) - (Für eine abweichende Rekonstruktion vgl. Stephan, 1990, S. 140ff.)

Annahmenkerne sind kognitive Strukturen, nicht Strukturen von bereits formulierten Aussagen; diese kognitiven Strukturen können (zunächst) sehr vage, inexplizit und nur schwer verbalisierbar sein. Die Mitglieder des sozialen Systems, in dem P bearbeitet wird, können allenfalls, vor allem zu Beginn eines Programms, die Kernannahmen nur schlecht explizieren, sie können sie weitgehend stillschweigend präsupponieren, usf. Die Problemexplikation, die begriffliche und theoretische Präzisierung des Problems, seine sprachliche Fixierung usf., die Shapere (1974) die Bearbeitung von „Domainproblemen“ nennt (s. auch unten S.270), sind bereits ein Teil der Problembearbeitung. - Die sprachliche Fassung der kognitiven Struktur eines Annahmenkerns zu P



ist demzufolge eine mehr oder minder irrumsgefährdete Rekonstruktion (vgl. dazu auch Herrmann, 1976).

Man kann am Beispiel der Leistungsmotivation erkennen, daß der Annahmenkern quasi-definitivische (und damit „analytische“, empirisch gehaltfreie, unwiderlegbare) Annahmen enthält. Wenn zum Beispiel eine Annahme lautet: „Menschen handeln.“, so wird hier das mit „Mensch“ Gemeinte durch das mit „handeln“ Gemeinte erläutert; auf dieses Attribut des Menschen wird hingewiesen. Man kann vielleicht auch annehmen, daß der Mensch (auch) durch sein Handeln definiert ist. (Was „Mensch“ und „handeln“ jeweils bedeuten, gehört dabei bereits zum Basiswissen des Menschen (unserer Kultur) und wird bereits konsensuell vorausgesetzt: Jede Problemkonstitution setzt dieses alltägliche Basiswissen bzw. die Alltagssprache voraus. Vgl. auch v. Savigny, 1975.)

Einige Annahmen des Programms sind existenzkonstatierend („es gibt ...“). Andere beziehen sich auf Implikationen oder ähnliche Zusammenhänge: Wenn sich Menschen Leistungsnormen setzen, so erleben sie Erfolge oder Mißerfolge. Oder Erfolgshoffnung und Mißerfolgsschmerz hängen von früheren Erfahrungen ab. Solche implikativen Annahmen sind mögliche Vorgänger für spätere nomologische, empirisch prüfbare Behauptungen bzw. Hypothesen. - Durch die implikativen Annahmen werden in besonders deutlicher Weise denkbare Vorstellungen zu einem Realitätsbereich ausgeschlossen: Der Annahmenkern des Forschungsprogramms zur Leistungsmotivation enthält zum Beispiel eben nicht die Annahme, daß Menschen Rückenschmerzen bekommen oder daß sie aggressiv werden, wenn sie sich Leistungsnormen setzen. Vielmehr erleben sie Erfolgshoffnung oder Mißerfolgsschmerz. Oder: Erfolgshoffnung und Mißerfolgsschmerz hängen nicht vom Einfluß böser Geister oder vom Geburtsgewicht ab. Vielmehr hängen sie von früheren individuellen Erfahrungen ab.

Annahmenkerne sind semantisch mehrstufig aufgebaut: Einige Annahmen beziehen sich auf andere Annahmen. So gehört es zum Beispiel zum Annahmenkern der Leistungsmotivation, daß man über die Gründe des Zustandekommens von Unterschieden des Leistungsstrebens zuwenig weiß, daß man darüber gern mehr wissen möchte.

### **1.2.3 Zur Methodenwahl**

Generelle Annahmen höherer semantischer Stufe beziehen sich auch auf Wege und Mittel zur Transformation des Ist- in einen Sollzustand des programm-eigenen Problems P bzw. zur Überwindung der jeweiligen „Barriere“ (s. oben S. 254). Insofern Forschungsprogramme ihre „Identität“ wesentlich durch ih-

ren Annahmehkern zum Problem P erhalten, sind sie auch durch diejenigen zum Kern gehörenden Annahmen charakterisiert, welche sich auf mögliche Transformationsmittel beziehen. Zwar ist die Schaffung konkreter Methoden oder Verfahren oder ihr Import aus anderen Forschungsprogrammen bereits ein Teil des Problemlösungsprozesses, also der versuchten Transformation des Ist- in einen Sollzustand, doch findet man im überdauernden Annahmehkern selbst schon generelle Maßgaben zur Methodenwahl; d.h. zentrale Annahmen zum Realitätsbereich (zum Beispiel implizierte „Menschenbildvorstellungen“) präjudizieren die Wahl von Methoden. Da zum Beispiel Erfolgshoffnungen und Mißerfolgsschmerz eben nicht vom Einfluß böser Geister oder vom Geburtsgewicht abhängen (s. oben), kommen im Forschungsprogramm zur Leistungsmotivation von vornherein keine Methoden der Geisterbeschwörung, aber auch nicht der Einsatz einer Säuglingswaage infrage. Das „Zu-Erforschende“, so wie es die Mitglieder eines Forschungsprogramms auffassen, kodeterminiert die im Forschungsprogramm verwendbaren Forschungsmittel.

Ein anderes Beispiel: Das Ebbinghaus'sche Forschungsprogramm zum Gedächtnis (Ebbinghaus, 1885; vgl. auch Reese, 1973) basiert auf einem Annahmehkern, der die Methodik des Erlernens „sinnloser Silben“ geradezu fordert. Das Problem des Gedächtnisses war dort so konstituiert, daß zum Beispiel eine Abfragemethode zum Behalten des „wesentlichen Inhalts“ von Romanen oder die Methode des Nacherzählens von exotischen Märchen apriori ausscheiden. Methodik und Problembestimmung waren hier besonders eng verschwistert, wodurch das „Ebbinghaus-Programm“ eben auch durch seine besondere Methodik der „sinnlosen Silben“ wesentlich charakterisiert ist. Diese so enge und unwandelbare Verknüpfung von Methodik und Problemkonstitution gilt aber offensichtlich bei weitem nicht für alle Forschungsprogramme. Sie gilt zum Beispiel nicht für das Forschungsprogramm zur Leistungsmotivation.

Wie weiter unten gezeigt werden soll (S.285), leiten sich die methodischen Mittel zur forschungsprogrammspezifischen Problembearbeitung auch zum guten Teil von forschungsprogrammübergreifenden, oft schnell wechselnden Methodenmoden und von epochalen („paradigmatischen“) Handlungsvorgaben her. Das wissenschaftliche „Klima“, in dem das Forschungsprogramm zur Leistungsmotivation entstand, erlaubte den Einsatz laborexperimenteller Methodik. (Die Frage, ob man Menschen „messen“ darf oder soll, stand damals offensichtlich nicht zur Debatte. Vgl. dazu auch Kap.4.) Man war zugleich so wenig auf einen orthodoxen methodologischen Behaviorismus festgelegt (vgl. Briskman, 1972), daß bei der Erforschung der Leistungsmotivation sofort auch textanalytische Verfahren mit einem erheblichen interpretativen Anteil Verwendung fanden (Heckhausen, 1989, S. 233ff.). Später brachte es eine Wissenschaftsmode der Sechziger- und Siebzigerjahre, nämlich die Neigung, die Attributionstheorie in der Psychologie fast ubiquitär anzuwenden, mit sich,

daß diese Theorie auch in das Programm zur Leistungsmotivation importiert wurde (vgl. Weiner, 1972). Und mit dieser seinerzeit modischen Theorie importierte man zugleich die Standardmethodik zur Erfassung der Kausalattribution. - Solche Determinanten der Methodenwahl lassen sich nicht primär auf die Beschaffenheit des Annahmehkerns eines Forschungsprogramms zurückführen.

Es sei hinzugefügt, daß die Methodenwahl in Forschungsprogrammen überdies nicht selten unter dem folgenden externalen Einfluß steht (s. auch unten S.285): Die persönliche Kompetenz und die materiellen Möglichkeiten, bestimmte Methoden und Verfahren (bereits) verwenden zu können, verleihen - innerhalb von Forschungsprogrammen sowie im Vergleich zwischen Forschungsprogrammen - den Akteuren soziale Statusvorteile mit allen ihren Konsequenzen (Kariervorteile, Vorteile bei der Einwerbung von Förderungsmitteln usf.). Das war zum Beispiel seinerzeit beim Kennen und Anwendenkönnen der mathematischen Faktorenanalyse der Fall (vgl. Orlik, 1967). Der Experte für diese Methode (nicht so sehr für die ihr zugrundeliegende Theorie) galt weithin als hochqualifizierter Forscher, eben weil er bereits die Faktorenanalyse irgendwo anzuwenden gelernt hatte, was für die meisten Psychologen noch nicht zutraf. Eine solche Sachlage (ver-) führt zur prestigemotivierten Verwendung von Methoden, wobei die programmspezifische (problemspezifische) Eignung der Methode als Auswahlkriterium zumindest zeitweilig zurücktritt. (Für das Forschungsprogramm zur Leistungsmotivation scheint der prestigemotivierte Einsatz von Methoden nicht charakteristisch zu sein.)

Es folgt generell für die Methodenwahl in Forschungsprogrammen: Die Methodenwahl ist multipel determiniert. Zum Teil resultiert sie aus dem Annahmehkern des Programms; zentrale Annahmen zum problematisierten Realitätsbereich schließen bestimmte Methoden bzw. Methodenklassen apriori aus und legen den Einsatz anderer nahe. Insofern forschungsprogrammspezifische Arbeit im Kontext genereller „paradigmatischer“ Vorstellungen und kürzerfristiger Wissenschaftsmoden erfolgt, wirken sich auch diese programmexternen Einflüsse auf die Methodenwahl aus. Dies so, daß programmexterne Einflüsse entweder zentrale Annahmen zum problematisierten Realitätsbereich selbst determinieren, wodurch, wie soeben erörtert, bestimmte Methoden ausgeschlossen und andere nahegelegt sind. Oder diese epochalen oder kürzerfristigen Einflüsse führen zeitweilig zur programmübergreifenden, innerhalb der Psychologie weithin homogenen Verwendung bestimmter Methodenklassen und zur Vermeidung anderer. (Im „Zeitalter“ des Behaviorismus war zum Beispiel kein Platz für introspektive Methoden.) Ein besonderer Gesichtspunkt liegt in der beschriebenen prestigemotivierten Methodenverwendung. Diese kann auch mit der „Selbsterhaltungstendenz“ von Forschungsprogrammen (s. unten S.260f.) in Zusammenhang gebracht werden.

### **1.2.4 Programmspezifische Indisponibilität**

Innerhalb eines Forschungsprogramms kann kein Bearbeitungsergebnis dazu zwingen, den Annahmehauptbestandteil zu P zur Disposition zu stellen (Herrmann, 1979, S. 32ff.). Für die Psychologie als eine Erfahrungswissenschaft gilt insbesondere: Keine auf Beobachtungen, Experimenten o. dgl. beruhende Erfahrung, die im Wege der gemeinsamen, der Problembewältigung geltenden Arbeit innerhalb eines Forschungsprogramms gewonnen wird, kann (der Explikation auf S.253ff. zufolge) dazu zwingen, das Problem P qua Problem zu ersetzen. Eine Problemersetzung bedeutet vielmehr die Ersetzung eines Forschungsprogramms durch ein anderes. (Zu programmeeigenen Problem-Derivaten vgl. unten S.270 und 272.)

Selbstverständlich können sich Probleme P nicht als in einer Weise falsch herausstellen, wie sich empirische Hypothesen allenfalls als falsch herausstellen können. Sie können u.U. aber als uninteressant, unfruchtbar, nutzlos, nicht adäquat bearbeitbar o. dgl. gelten. Auch Forschungsprogramme können - als Arbeit an solchen Problemen P - nicht strikt an der Erfahrung scheitern; sie können als solche nicht „falsifiziert“ werden. Auch sie können aber als uninteressant, unfruchtbar, nutzlos, nicht adäquat bearbeitbar o. dgl. aufgegeben, eingestellt werden, „versanden“ o. dgl.; Forschungsprogramme können als Teilprogramme in andere Forschungsprogramme aufgehen (s. unten S. 264). (Vgl. zum generellen Problem der Nichtfalsifizierbarkeit auch den Kritischen Rationalisten Andersson, 1988.) Das das Problem P bearbeitende soziale System kann sich auflösen. Ein Forschungsprogramm kommt ersichtlich auch zu seinem Ende, wenn zwar das Problem P nicht als uninteressant, unfruchtbar, nutzlos, nicht adäquat bearbeitbar o. dgl. gilt, wenn aber die Gruppe von wissenschaftlichen Akteuren aus wissenschaftsinternen oder wissenschaftsexternen Gründen aufhört zu existieren: Ein wissenschaftsinterner Grund kann zum Beispiel liegen, daß eine attraktive Wissenschaftsmode entsteht, der sich viele Mitglieder der Akteursgruppe opportunistisch oder wie auch immer anschließen, so daß sie für die interaktive Bearbeitung des Problems P keine Zeit mehr haben. Wissenschaftsexterne Gründe können politischer oder ökonomischer Natur sein. Zum Beispiel kann psychologische Forschungsarbeit zur Humanisierung der Arbeit oder zur Hochbegabtenförderung in bestimmten Staaten aus politischen Gründen durch Mittelverweigerung verhindert werden.

### **1.2.5 Soziales System und Regulation**

Das soziale System, in welchem das jeweilige Forschungsproblem bearbeitet wird, ist in Anlehnung an Hejl(1984, **1989**) wie folgt zu verstehen: Individuen

sind nur soweit Mitglieder des Systems (der Forschungsgruppe o. dgl.), als sie das Problem P bearbeiten und dabei interaktiv handeln. (Insofern und nur insofern auch kognizieren sich die Individuen - auch gegenseitig - als Mitglieder einer und derselben Gruppe.) Die beteiligten Individuen sind normalerweise Mitglieder in mehreren sozialen Systemen (z.B. Familie, Fakultät, Herausgebergremium, Gutachtergruppe, politische Partei, andere Forschungsgruppe). Soziale Systeme können dadurch miteinander verbunden sein, daß sie Individuen als Mitglieder gemeinsam haben. (Beispiel: Eine ein Forschungsproblem bearbeitende Forschungsgruppe und eine wissenschaftliche Fachgesellschaft haben viele Individuen gemeinsam.)

Es ist zu beachten, daß die Mitglieder einer ein Forschungsproblem bearbeitenden Gruppe selbstverständlich nicht nur als Mitglieder dieser Gruppe existieren. Sie alle haben zusätzliche Merkmale, sind zusätzlichen Einflüssen ausgesetzt, usf. über die systemexternen Merkmale der Gruppenmitglieder und über systemexterne Einflüsse, denen sie ausgesetzt sind, ist das jeweilige soziale System und damit das bearbeitete Forschungsproblem derart kodeterminiert, daß der Verlauf der forschungsprogrammspezifischen Problembearbeitung nicht allein aus den Voraussetzungen des Problems P selbst heraus verstanden werden kann (s. auch unten S.266ff.). Vielleicht wird zum Beispiel bei der Bearbeitung eines Problems ein dysfunktionales Problemlösungsmittel trotz vieler Mißerfolge deshalb beibehalten, weil die Gruppenmitglieder inzwischen älter und etablierter geworden sind und sich nur noch schlecht auf neue Methoden, Theorien o. dgl. umstellen können. Oder es wird in einem Forschungsprogramm plötzlich der „ökologische Kontext“ von Sachverhalten untersucht, weil die Individuen, welche Mitglieder der Forschungsgruppe sind, von einer epochalen politisch-ökologischen Erweckungsbewegung erfaßt wurden.

Man kann Forschungsprogrammen, insofern sie soziale Systeme sind, die Eigenschaft zuschreiben, regulativ zu sein (Hejl, 1989): Soziale Systeme halten danach ihre Grundparameter (d.h. ihre Existenz und den Bestand ihrer Organisationsform) im ständigen Prozeß der Variation vieler anderer Parameter konstant. (Vgl. allgemein Maturana, 1982, S.86f.; Stadler & Kruse, 1986.) Äußere Einwirkungen auf das System bzw. auf Systemteile können dann als „Störungen“ betrachtet werden, die im System zu kompensatorischen Regulationsprozessen führen. Falls sich die wissenschaftliche Tätigkeit in Forschungsprogrammen zumindest partiell in diesem Sinne regulationstheoretisch interpretieren läßt, kann man fragen: Wieweit dienen wissenschaftliche Tätigkeiten in Forschungsprogrammen der Aufrechterhaltung des Programms und seiner Organisation oder dem Bestand von Teilgruppen, die am Programm mitarbeiten (vgl. u.a. Sahner, 1979)? Wieweit läßt sich u.a. die prestigemotivierte Methodenwahl (s. oben S. 258) in dieser Weise interpretieren?

Stegmüller (1969, S. 138ff.) und andere haben mit ihrer Konzeption eines pragmatischen Erklärungsbegriffs die Annahme oder Ablehnung einer wissenschaftlichen Erklärung vom gesamten „geistigen Hintergrund“ der um Erklärung bemühten wissenschaftlichen Akteure abhängig gemacht. Es muß beachtet werden, daß dieser „geistige Hintergrund“ seinerseits u. a. erfahrungs- und interessendeterminiert, durch soziale Normen bzw. Konventionen gesteuert und durch organisatorische, institutionelle und andere soziale Rahmenbedingungen begrenzt ist. Die Akzeptierung oder Ablehnung von Erklärungen seitens der Mitglieder eines Forschungsprogramms kann auf der Basis ihres „geistigen Hintergrunds“ regulationstheoretisch interpretiert werden: Man akzeptiert dann genau die gruppenexistenz-konservierenden Erklärungen. Und da die Akzeptanz von Theorien nicht zuletzt auf der ihr zugeschriebenen Erklärungskraft beruht, entscheidet die „Selbsterhaltungstendenz“ von Forschungsprogrammen über das Schicksal psychologischer Theorien mit. - Gruppenexistenzkonservierende Strategien bei der Publikation empirischer Forschungsergebnisse sind mehrmals nachgewiesen worden (Sahner, 1982).

## *2. Probleme P und strukturalistische Rekonstruktionen*

### **2.1 Fragestellung**

Probleme P sind, wie ausgeführt, durch Annahmenkerne expliziert. In welchem begrifflichen und funktionalen Verhältnis stehen diese Annahmenkerne zu den Theorien T des metatheoretischen Strukturalismus (s. Kap.7.5 und Kap. 10.6)? Diese Frage ist bisher innerhalb der Psychologie nicht eingehend diskutiert worden. Einige Ansätze findet man bei Westermann (1987), Breuer (1989) und Stephan (1990). (Zum allgemeinen vgl. auch Stegmüller, 1973, 1986.) In heutiger Sicht liegen für die Psychologie die nachfolgenden Vorstellungen zum Verhältnis beider Rekonstruktionsweisen nahe (s. Herrmann, 1976, 1987; vgl. auch Balzer, Moulines & Sneed, 1987; Westmeyer (Ed.), 1989).

### **2.2 Einige Grundannahmen des metatheoretischen Strukturalismus**

Eine stark vereinfachte und unvollständige Skizze des strukturalistischen Theorienverständnisses sieht wie folgt aus (vgl. Kap. 7.5): Theorien T (genauer: Theorieelemente T) lassen sich als aus einem Theoriekern K(T) und einem (offenen) intendierten Anwendungsbereich I(T) bestehend rekonstruieren. Theoriekerne K(T) sind durch diejenigen Realitätsbereiche bzw. Entitäten bestimmt, die sich in der Begrifflichkeit von T beschreiben lassen und die Annahmen von T erfüllen: Die Begrifflichkeit von T enthält im wesentlichen Ausdrücke (Terme) für Mengen, Relationen und Funktionen. Die Ausdrücke

sind bei psychologischen Theorien in der Regel semantisch interpretiert; wir sprechen in der Folge von Begriffen. Die Annahmen betreffen (zum Teil implikative) Zusammenhänge der Begriffe. Die Begrifflichkeit von  $T$  enthält zum einen Begriffe (Relationen u. dgl.), die nur unter Rückgriff auf die (erfolgreich angewandte) Theorie  $T$  selbst bestimmt bzw. angewandt werden können. (Zum Beispiel können dann Größen von Funktionen nur unter Rückgriff auf  $T$ , zu denen die Funktionen gehören, berechnet werden.) Diese Teilmenge der Begriffe von  $T$  nennt man  $T$ -theoretische Begriffe. Die Theorie  $T$  enthält zum anderen unabhängig von  $T$  bestimmbare Begriffe: die nicht- $T$ -theoretischen Begriffe. (Vgl. auch Balzer et al., 1987, S. 55.)

Realitätsbereiche bzw. Entitäten sind potentielle Modelle  $M_p(T)$  von  $T$ , wenn sie sich in der Begrifflichkeit von  $T$  beschreiben lassen. Partielle potentielle Modelle bzw. Partialmodelle  $M_{pp}(T)$  von  $T$  sind solche Entitäten, bei deren Beschreibung die nicht- $T$ -theoretischen Begriffe von  $T$  verwendet werden. Modelle  $M(T)$  von  $T$  sind potentielle Modelle von  $T$ , die darüber hinaus auch die Annahmen von  $T$  erfüllen.

Die Menge intendierter Anwendungen  $I(T)$  von  $T$  kann als Teilmenge der Menge partieller potentieller Modelle  $M_{pp}(T)$ , also als Teilmenge der Menge derjenigen Realitätsbereiche verstanden werden, die mittels der nicht- $T$ -theoretischen Begriffe von  $T$  beschreibbar sind. Sie ist von ihrer komplementären Teilmenge von  $M_{pp}(T)$  in pragmatischer Weise unterschieden: Proponenten der Theorie  $T$  seligieren bzw. deklarieren einen (offenen) Teil von  $M_{pp}(T)$  als Menge intendierter Anwendungen  $I(T)$ . Nicht alles, was (gegeben  $T$ ) mit deren nicht- $T$ -theoretischen Begriffen beschrieben werden kann ( $= M_{pp}(T)$ ), wird de facto als eine Anwendung von  $T$  behandelt. (Vgl. auch Sneed, 1971, S. 183f.) Stellt sich heraus, daß eine intendierte Anwendung durch  $T$ -theoretische Begriffe ergänzt werden kann und daß die Annahmen von  $T$  für diese intendierte Anwendung gelten, so ist die intendierte Anwendung zu einem Modell  $M(T)$  von  $T$  geworden: Die intendierte Anwendung ist dann eine erfolgreiche Anwendung von  $T$ .

Eine Teilmenge der Menge intendierter Anwendungen  $I(T)$  bildet die Menge der paradigmatischen Anwendungen  $I_o(T)$ . Solche intendierten Anwendungen sind bereits erfolgreich angewendet worden. Durch sie ist  $T$  wesentlich gekennzeichnet. Durch die Menge der Realitätsbereiche bzw. Entitäten  $I_o(T)$ , die für den Anwendungsbereich von  $T$  sozusagen beispielhaft ist, läßt sich die Natur einer Theorie  $T$  besonders gut charakterisieren.

Aus Raumangel wird hier auf die Erläuterung zweier weiterer Komponenten von Theorien  $T$  verzichtet; auf ‚Querverbindungen‘ und auf ‚intertheoretische Relationen‘ (vgl. dazu Kap. 7.5).

## 2.3 Annahmenkerne und Theorien im Lichte des Strukturalismus

Annahmenkerne zu Problemen P lassen sich in strukturalistischer Sicht tentativ selbst als Theorien bzw. Theorieelemente T beschreiben. Nach der Explikation von Forschungsprogrammen (s. oben S. 253ff.) bleiben die Annahmenkerne zu P innerhalb eines Forschungsprogramms indisponibel.

Aus dem jeweiligen Annahmenkern werden jedoch innerhalb des Forschungsprogramms in variabler Weise Derivate gewonnen: Das Problem P wird präzisiert, erweitert, eingeschränkt, für die Problembearbeitung in Teilprobleme zerlegt, formalisiert, metrisiert usf. Unter gewissen Vorannahmen kann man P-Derivate als das Resultat aus Kernannahmen und Sekundärannahmen auffassen (Herrmann, 1976, S.50ff.). Über die Zufügung von im Programm erarbeiteten Sekundärannahmen können aus Problemen P empirische Erwartungen gewonnen werden. Treffen die empirischen Erwartungen angesichts empirischer Ergebnisse nicht zu, so können die empirischen Erwartungen und allenfalls die Sekundärannahmen als falsch bzw. ungeeignet zurückgewiesen werden, nicht aber die Kernannahmen. (Vgl. dazu auch Lakatos, 1974.) Auch die Derivate sind mit den Mitteln des metatheoretischen Strukturalismus als Theorieelemente T rekonstruierbar. Insofern können Annahmenkerne zu P zusammen mit ihren im Forschungsprogramm vorliegenden Derivaten - strukturalistisch betrachtet - als sich über die Zeit hinweg ändernde Theorienetze oder (genauer:) als zu verschiedenen Zeiten unterschiedliche Theorienetze verstanden werden. (Vgl. u. a. Westermann, 1987, S. 34ff.; Stephan, 1990, s. 55ff.)

Wir betrachten im Augenblick nur den forschungsprogrammspezifisch indisponiblen Annahmenkern zu P (und nicht die Derivate). Der Annahmenkern kann also in strukturalistischer Weise tentativ als Theorie  $T_p$  rekonstruiert werden. Wie etwa das informelle Beispiel der Leistungsmotivation deutlich macht (s. oben), ist die theoretische Struktur  $T_p$ , die den Annahmenkern zu P bildet, sehr vage (d.h. schlecht rekonstruierbar) und, zumindest zu Beginn des Programms, zumeist mit einem erheblichen Anwendungsdefizit behaftet: Soweit es paradigmatische Anwendungen  $I_o(T_p)$  gibt, sind diese in der Regel (zum Beginn des Forschungsprogramms) keine Anwendungen, die aus wissenschaftlicher Tätigkeit resultieren; die beispielhaften Anwendungen ( $I_o(T_p)$ ) des Annahmenkerns  $T_p$  stammen häufig aus Alltagserfahrungen oder sind nur „ausgedacht“. Das Forschungsprogramm selbst dient ja u.a. der mit wissenschaftlichen Methoden realisierten Anwendung des Annahmenkerns zu P auf einen Realitätsbereich. So stammten zum Beispiel zu Beginn der wissenschaftlichen Angstforschung die Erfahrungen der Forscher mit Angstphänomenen aus dem vorwissenschaftlichen Alltag und aus praktischen (zum Beispiel psychiatrischen) Berufserfahrungen. (Vgl. auch Krohne, 1976.) Immerhin kann



zumindest im Grundsatz unterstellt werden, daß Annahmenkerne zu P, die als  $T_p$  rekonstruiert werden, über  $M_p(T_p)$ ,  $M_{pp}(T_p)$  und (angenähert) über  $M(T_p)$  sowie über  $I(T_p)$  explizierbar and. (Man dürfte auch  $T_p$ -theoretische Begriffe spezifizieren können.)

Man kann nach allem Annahmenkerne zu P strukturalistisch als Theorien bzw. Theorieelemente  $T_p$  auffassen. Dies jedoch mit folgender Einschränkung (- weshalb mehrmals von einer tentativen Rekonstruierbarkeit die Rede war): Annahmenkerne zu P sind semantisch mehrstufig (s. oben); sie enthalten Annahmen über andere Annahmen. Zum Beispiel die „pragmatische“ Annahme, daß die empirische Erfassung individueller Unterschiede des Leistungsstrebens noch nicht (hinreichend) realisiert, aber erwünscht ist, ist zwar Teil des Annahmenkerns zur Leistungsmotivationsproblematik P, aber sie ist mutmaßlich kein Teil einer Theorie  $T_p$ . (Daß der Theorieproponent x wünscht, gilt im allgemeinen nicht als Teil von Theorien, die x enthalten.) Streng betrachtet, kann der Annahmenkern zu P also nur partiell als  $T_p$  rekonstruiert werden: soweit er eine Beschreibungsbegrifflichkeit für einen Realitätsbereich und quasi-theoretische Annahmen zum Zusammenhang von Begriffen enthält. Soweit der Annahmenkern aber „pragmatische“ Komponenten enthält, mit denen eine auf einen Wirklichkeitsausschnitt bezogene kognitive Struktur erst zum Problem (für jemanden) wird, hat der Annahmenkern zu P einen Bedeutungsüberschuß<sup>3</sup> über seine bloße Theoretizität ( $T_p$ ) hinaus.

In einem Forschungsprogramm kann der Versuch gemacht werden, Theorien  $T'$  zu erfinden bzw. aus fremden Programmen zu importieren (s. unten S.267f.);  $T'$  soll geeignet sein, ein Derivat des programmeigenen Problems P zu einer erfolgreichen Anwendung von  $T'$  zu machen. Man sucht also eine Theorie  $T'$  und entwickelt oder modifiziert ein Derivat aus P derart, daß das Derivat zu einem Partialmodell  $M_{pp}(T')$  von  $T'$  werden kann. Im günstigen Fall kann dann durch die Ergänzung durch T-theoretische Begriffe und durch den Nachweis der Geltung von Annahmen von  $T'$  das Derivat von P zu einem Modell  $M(T')$  von  $T'$  werden. Und damit hat nicht nur  $T'$  eine neue erfolgreiche Anwendung erfahren - das interessiert die Akteure, die das Problem P bearbeiten, kaum. Vielmehr ist das Problem P (via Derivat) mittels  $T'$  so rekonstruiert worden, daß man eben das Problem P besser versteht bzw. daß man zum Beispiel im Lichte des  $T'$ -theoretischen Rekonstrukts von P Komponenten von P in ihrem gesetzmäßigen Zusammenhang besser begreift. Bisweilen kann man den Zusammenhang von P-Komponenten erst empirisch prüfen, nachdem man die P-Komponenten T'-theoretisch rekonstruiert (d.h. zu einem Anwendungsfall der Theorie  $T'$ , also zu einem Element von  $I(T')$  gemacht) hat. Zum Beispiel sind Fragen des Zusammenhangs von Furcht und individueller Erfahrung wesentlich verbessert behandelbar geworden, nachdem Proponenten der Angstforschung das Angstproblem P mittels der Theorie  $T'$  des operanten Konditionierens rekonstruiert haben. (Vgl. Krohne, 1976.)

### 3. Programmnetze und Austauschbeziehungen

Wissenschaften wie die Psychologie können als in steter historischer Wandlung begriffene Netzwerke aufgefaßt werden (Herrmann, 1976, 1987): Die Netzwerkknoten sind die Forschungsprogramme. Forschungsprogramme entstehen und vergehen; sie können auseinander hervorgehen; sie können sich teilen, wobei P in Teilstrukturen zerfällt; sie können auch verschmelzen, usf. Die Netzwerkkanten stellen sich als Austauschbeziehungen zwischen den Netzwerkknoten, d.h. den einzelnen Forschungsprogrammen, dar. Die psychologische Forschung kann als ein Ausschnitt aus einem größeren Wissenschaftsnetzwerk interpretiert werden. Dieser Netzwerkausschnitt ist nicht randscharf. Man kann den Netzwerkausschnitt als das Wissenschaftssystem der Psychologie interpretieren. Zu diesem Wissenschaftssystem verhalten sich die einzelnen Forschungsprogramme wie Subsysteme eines Systems (s. unten S. 266). Der Netzwerkausschnitt, den das Wissenschaftssystem der Psychologie bildet, kann sowohl über die Gemeinsamkeit von Problemkonstitutionen P als auch über Gemeinsamkeiten und Interaktionen der Mitglieder der Forschungsprogramme, die ihre Subsysteme sind, bestimmt werden (s. unten S.266).

Es gibt Forschungsprogramme, für die es fraglich ist, ob sie schon oder noch zur Psychologie gehören. Man denke zum Beispiel an einige wahrnehmungspsychologische, sprachpsychologische oder organisationspsychologische Programme (vgl. z.B. Szagun, 1986). Typische psychologische Forschungsprogramme sind zum Beispiel die Erforschung der Leistungsmotivation (s. oben), des Modell-Lernens und der kognitiven Entwicklung des Kindes (vgl. z.B. Oerter & Montada, 1982, S. 633ff.). Es gibt psychologische Forschungsprogramme, bei denen Probleme, die auch außerhalb der Psychologie und der Wissenschaft überhaupt bestehen, zumeist nach entsprechender Rekonstruktion, wissenschaftlich bearbeitet werden (z. B. Angstforschung, vgl. Krohne, 1976; psychologische Verhandlungsforschung, vgl. Crott, 1979). Bei anderen Forschungsprogrammen handelt es sich um Probleme, die erst im Wissenschaftsprozeß selbst entstanden sind (z.B. Erforschung des Ultrakurzzeitpeichers, vgl. Lachman, Lachman & Butterfield, 1979; Erforschung von Repressors und Sensitizers, vgl. Herrmann, 1991).

Für das soziale System (die Forschungsgruppe o. dgl.), in dem ein forschungsprogrammspezifisches Problem P (z.B. die Angst) erforscht wird, gehören andere soziale Systeme, die andere forschungsprogrammspezifische Probleme P' bearbeiten (z. B. die Leistungsmotivation), zur Systemumgebung. Das schließt nicht aus, daß mehrere psychologische Forschungsprogramme - unter entsprechender systemtheoretischer Perspektive - als Subsysteme eines intermediären Systems höherer Ordnung (vgl. psychologische Teildisziplinen wie Persönlichkeitsforschung, Arbeits- und Organisationspsychologie) verstanden

werden können. Dann enthalten die Probleme P und P' der beteiligten psychologischen Forschungsprogramme viele gemeinsame Komponenten von Kernannahmen. Und zwischen den Mitgliedern dieser Forschungsprogramme bestehen starke Gemeinsamkeiten und relativ intensive Interaktionen. Wie vermerkt, nennen wir hingegen das System, das alle psychologischen Forschungsprogramme als Subsysteme umschließt, das Wissenschaftssystem der Psychologie.

Nach der zuvor dargestellten Netzwerkvorstellung kann das Wissenschaftssystem der Psychologie als Teilnetz des generellen Wissenschaftsnetzwerks verstanden werden. Die Forschungsprogramme sind die Knoten im Netzwerk. Nachdem nun, wie soeben vermerkt, einige psychologische Forschungsprogramme besonders ähnliche Annahmenkerne haben und besonders eng miteinander interagieren können (Beispiele: Persönlichkeitsforschung, Arbeits- und Organisationspsychologie), bildet eine solche Gruppe von psychologischen Forschungsprogrammen ein Teilnetz des Netzwerks der Psychologie, welches seinerseits als Teilnetz des Wissenschaftsnetzwerks aufgefaßt werden kann. Die Vernetzung der Forschungsprogramme ist in dieser Weise hierarchisiert zu denken.

Die Auffassung, daß Forschungsprogramme spezifische soziale Subsysteme sind, impliziert nach der zu Beginn erfolgten Erläuterung zum Terminus „Forschungsprogramm“, daß alle beteiligten Subsysteme, die intermediären Systeme und das übergeordnete System der Psychologie (= Wissenschaftssystem) jeweils über Annahmenkerne zu Problemen P bestimmt sind (s. auch unten S. 280). Nur insofern Wissenschaftlergruppen Probleme P bearbeiten, sind sie Subsysteme eines übergeordneten (Wissenschafts-) Systems. Wenn nun eine Forschungsgruppe, die P bearbeitet, zu einer Fakultät einer Universität gehört, so sind die Fakultät und die Universität, die nicht durch die Bearbeitung von Problemen P gekennzeichnet sind, bezüglich der Forschungsgruppe keine übergeordneten sozialen Systeme; die Forschungsgruppe ist kein Subsystem der Fakultät oder der Universität. Vielmehr sind die Fakultät und die Universität nach der gegenwärtigen Festlegung systemfremde Systeme, die mit der Forschungsgruppe lediglich Individuen teilen und zur Systemumgebung der Forschungsgruppe gehören. Es gilt generell: Die interne Struktur der Wissenschaften kommt mit der institutionellen Struktur staatlich-administrativer Wissenschaftsorganisation nicht überein.

#### *4. Zur Bearbeitung von Problemen in Forschungsprogrammen*

Wie bereits erläutert (S.254), handelt es sich bei der Bearbeitung von Problemen P in Forschungsprogrammen im wesentlichen um die Lösung von Problemen ohne klare Zielkriterien. Die Klarheit der Zielkriterien fehlt zumal

fast stets bei Versuchen, wissenschaftliche Innovationen zu erreichen: neue Theorien zu entwickeln, Theorien auf neue Realitätsbereiche anzuwenden, Begriffe (Konstrukte) mit geeigneten Beobachtungs- und Meßvorschriften (= „Operationalisierungen“, „Empirisierungen“) zu versehen, neue Verfahrensweisen zur Datenerhebung oder Datenverarbeitung zu gewinnen, usf.: Hier erkennt man, wie schon erwähnt, in der Regel die Kriterien dafür, wann das Ziel erreicht ist, erst dann, wenn man das Ziel erreicht hat.

In Forschungsprogrammen ist bei der Bearbeitung des Problems P innovatives Verhalten gefordert. Klages (1967, S. 94ff.) hat vor längerem zwei extreme Stile dieses Verhaltens unterschieden: das „Jagd-Verhalten“ und das „Entdeckungsverhalten“. (Daneben bespricht er als Mischform das „Aussiebungsverhalten“.) Entdeckungsförmige Innovation ist weitgehend hypothesenfrei und überraschungsoffen. Hier findet man zunächst kaum ein systematisches, sondern ein noch ungehemmt phantasievolles Vorgehen, das im Laufe der Zeit methodischer, planvoller und kombinatorischer wird, bis es endlich zur Bildung expliziter Annahmenstrukturen von der Art der Theorien führt. - Das jagdförmige Innovationsverhalten arbeitet mit vorfixierten Ergebniserwartungen. Hier prüft man zum Beispiel erwartete Wenn-dann-Beziehungen in systematischer Form nacheinander ab.

Bei Forschungsprogrammen dürfte zumindest in ihren frühen Phasen entdeckungsförmige Innovation vorherrschen. Shapere (1974) spricht hier von der Lösung von („vortheoretischen“) Domainproblemen. Bei diesen (Teil-) Bearbeitungen von Problemen P in Forschungsprogrammen handelt es sich etwa darum herauszufinden, was genau zum problematisierten Realitätsbereich gehört, wie man den problematisierten Realitätsbereich unterteilen kann, wie Komponenten des Realitätsbereichs zu erfassen und begrifflich zu präzisieren sind; man möchte neue, mit dem Problem verbundene „Fakten“, „weitere Beispiele“ usf. finden („fact finding“). In dieser Phase sucht man zunächst noch nicht nach einer (oder der!) Theorie, mit deren Findung das Problem P oder Komponenten desselben explizit rekonstruiert, also „auf den Begriff gebracht“, und theoriegeleiteter Forschung zugänglich ist. Interpretationen, Deutungen stehen im Vordergrund (vgl. dazu auch Kap. 3). Erst im Laufe der Bearbeitung von „Domainproblemen“ wird der Suchraum für im engeren Sinne theoretische Entwicklungen generiert und sukzessiv verkleinert; das innovative Vorgehen wird planvoller und nähert sich immer mehr dem Jagd-Verhalten nach Klages.

Theoretische Entwicklungen in einem Forschungsprogramm sind grundsätzlich in zweierlei Weise möglich: Theorien werden entweder innerhalb des Programms generiert, oder sie werden (als Realisierung einer Austauschbeziehung) aus einem anderen Forschungsprogramm importiert. Im gegenwärtigen Zusammenhang interessiert insbesondere der Import von Theorien: Im Sinne

strukturalistischer Auffassung (s.oben S. 264) führt der Import einer Theorie dazu, daß der Annahmenkern zu P oder ein Derivat des Annahmenkerns des importierenden Programms zur intendierten Anwendung der importierten Theorie T' wird. Die importierte Theorie kann auch, nach Dörner (1979), als ein Entwurf des Zielzustands des eigenen Problems aufgefaßt werden, der (bei Problemlösungen ohne klare Zielkriterien) versuchsweise verwendet und auf interne und externe Widersprüche überprüft wird. So hat man zum Beispiel versucht, das Problem der Entstehung von Unterschieden der Leistungsmotivation (s. oben S. 255) so zu spezifizieren, daß Derivate entstehen, von denen eines als intendierte Anwendung von Pawlovs klassischer Konditionierung und ein anderes als intendierte Anwendung von Hebbs Theorie der Konsequenzen variabler Diskrepanzen zwischen Ist- und Sollzuständen dienen. Es wurde - strukturalistisch formuliert - angestrebt, die beiden importierten Theorien zur (auch jeweils T'-theoretischen) Beschreibung des Motivationsentstehungsproblems zu verwenden. Und es wurde geprüft, ob die beiden Theorien (ihre Annahmen) auf die derart beschriebenen Komponenten der Motivationsentstehung zutreffen. (Vgl. dazu McClelland (Ed.), 1955.) - Später wurden andere Derivate desselben Problems mit Hilfe eines anderen Theorieimports zu rekonstruieren versucht: der Theorie der Kausalattribution. (Vgl. dazu Meyer, 1973.)

Beim Versuch, eine importierte Theorie als Entwurf für einen erwünschten Zustand (Zielzustand) des Problems P bzw. eines entsprechenden Derivats des Annahmenkerns zu P zu nutzen, können (nach Dörner, 1979) externe Widersprüche auftreten. Das kann - in Anlehnung an den Strukturalismus - zweierlei bedeuten: (i) Aus P können keine Derivate erzeugt werden, die in der Begrifflichkeit der importierten Theorie T' (bzw. mindestens eines ihrer Theorieelemente) mit Einschluß der T'-theoretischen Begriffe beschrieben werden können. Daraus folgt: Das Problem P (als Explicandum) kann dann nicht durch die Theorie T' (als Explicans) theoriespezifisch rekonstruiert werden. Die Theorie T' ergänzt (verbessert o.dgl.) nicht die Konzeptualisierung von P mittels ihrer ganz spezifischen T'-Begrifflichkeit. Hier kann man von einem theoretisch-begrifflichen Scheitern der importierten Theorie T' bei der Lösung des Problems P sprechen. - (ii) Theorien T' können auch empirisch scheitern: Aus P können dann zwar Derivate erzeugt werden, die (nach (i)) auch in der T'-theoretischen Begrifflichkeit von T' beschrieben werden können. Doch ergeben empirische Prüfungen auf Dauer nicht, daß die Annahmen von T' in zufriedenstellender Weise auf das Problem P bzw. auf sein Derivat zutreffen. Anders formuliert: Beim empirischen Scheitern findet man auf Dauer zum Annahmenkern zu P kein Derivat, das folgende Bedingungen erfüllt: Es wird in der Begrifflichkeit von T' beschrieben; es enthält Aussagen über bestehende empirische Sachverhalte; den Aussagen über bestehende empirische Sachverhalte widersprechen auf Dauer keine empirischen Behauptungen, die mit T' verbunden sind. (Vgl. auch Stegmüller, 1986, S. 64f.)

Wir lassen hier offen, wie es (nach (ii)) zu Aussagen über empirische Sachverhalte kommt. Zum Beispiel mag es sich bei solchen Aussagen um bei den Akteuren des Forschungsprogramms konsensuell für „unwiderleglich richtig“ gehaltene Ergebnisse von experimentellen Untersuchungen handeln (vgl. auch Kap. 9). Das empirische Scheitern von Theorien T' bei der Bearbeitung von Problemen P ist immer ein De-facto-Mißerfolg, der u.a. auch auf Täuschungen und Fehlern der Akteure eines Forschungsprogramms bezüglich des Zutreffens von Aussagen über empirische Sachverhalte beruhen kann. Es ist hier - wie entsprechend beim theoretisch-begrifflichen Scheitern - nicht an ein Scheitern „an der Realität“ schlechthin zu denken (vgl. auch Andersson, 1988).

Es leuchtet zudem unmittelbar ein, daß eine importierte Theorie T' beim Scheitern bei der Lösung des Problems P nicht in irgendeinem Sinne „falsifiziert“ ist. (Vgl. auch Stegmüller, 1986.) Die Theorie T' hat lediglich einen Anwendungsfall eingebüßt. Die Akteure des importierenden Forschungsprogramms haben dagegen einen Rückschlag bei der Lösung ihres Problems P erlitten. (Sie können jetzt zum Beispiel versuchen, eine andere Theorie T'' zu importieren.) Im Mißerfolg haben sich das Problem P und die Theorie T' funktional entkoppelt; sie gehen sozusagen getrennt ihrer Wege. T' ist damit nicht „falsch“, „widerlegt“ o.dgl.; es besteht per se kein Grund, T' als zielführende Konstruktion aufzugeben, ebenso wie per se kein Grund besteht, die Bearbeitung von P (und damit das importierende Forschungsprogramm) aufzugeben. (Nur wenn es für T' auf Dauer gar keine erfolgreichen Anwendungen gibt oder wenn man für P auf Dauer keinerlei taugliche Problemlösungsmittel (theoretische Entwürfe usf.) findet, werden solche Konsequenzen plausibel.)

## 5. Programm-Typen

### 5.1 Übersicht

Forschungsprogramme können nach der unterschiedlichen Art ihrer Probleme P bzw. ihrer Annahmenkerne (und nach daraus resultierenden Unterschieden der sozialen Systeme, in denen die Probleme P bearbeitet werden) akzentuierend und ohne scharfe Grenzziehung wie folgt eingeteilt werden: Grundlagenwissenschaftliche Forschungsprogramme der Psychologie stehen (psycho-) technologischen Forschungsprogrammen gegenüber. Die grundlagenwissenschaftlichen Forschungsprogramme lassen sich in Sachproblem-Programme und Theorie-Programme einteilen. Für die technologischen Programme bietet sich eine Unterscheidung in Techniken-Programme und Programme zur Entwicklung operativen Wissens („Wissensprogramme“) an. (Vgl. auch Bunge, 1967; Klages, 1967; Stachowiak, 1973.) - In früheren Publikationen von Herr-

mann (1976, S. 29ff. u.a.O.) wurden Sachproblem-Programme als „Domain-Programme“ und Theorie-Programme als „Quasi-paradigmatische Programme“ bezeichnet.

## 5.2 Grundlagenwissenschaftliche und technologische Programme

Grundlagenwissenschaftliche und technologische Forschungsprogramme unterscheiden sich primär in der Art und Weise, wie Realitätsbereiche problematisiert werden, nicht aber so sehr dadurch, welche Realitätsbereiche problematisiert werden. So kann man zum Beispiel den Realitätsbereich der menschlichen Arbeit (vgl. dazu Hacker, 1978; Leontjew, 1977) sowohl grundlagenwissenschaftlich als auch technologisch problematisieren. Die Art der Problematisierung aber ist unter den Gesichtspunkten, die anschließend erörtert werden sollen, verschieden. (Nach der Explikation von Forschungsprogrammen (s. oben S. 253ff.) sind es also vor allem die Annahmen höherer semantischer Ebene, die die Annahmenkerne grundlagenwissenschaftlicher und technologischer Forschungsprogramme unterscheiden.)

### 5.2.1 *Grundlagenwissenschaftliche Programme*

Zu diesem Typ gehört u.a. das als Beispiel herangezogene Leistungsmotivations-Forschungsprogramm. Grundlagenwissenschaftliche Programme setzen sich die Aufgabe, den im jeweiligen Annahmenkern zu P angezielten Realitätsbereich besser zu durchschauen (vgl. auch Shapere, 1974).

Mit der Abarbeitung von „vorthoretischen“ Domain-Problemen (s. oben), durch Deutungen bzw. Interpretationen oder mittels der Generierung oder des Imports von expliziten Theorien (s. oben) soll erstens der Realitätsbereich, auf den sich der Annahmenkern bezieht, unter Verwendung präziserer Begriffe und Aussagen und unter Herbeiziehung aller für relevant gehaltenen Informationen interpretierend rekonstruiert (= expliziert) werden. Damit wird sozusagen eine Was-Frage beantwortet. So entstehen auch Derivate von P (s. oben S.259). Man mag zum Beispiel die Frage danach, was Angstentstehung ist, verbessert dadurch beantworten wollen, daß man - bei Import der Theorie des operanten Konditionierens (s. auch unten S. 277) - die Angstentstehung als einen Prozeß operanten Konditionierens rekonstruiert und analysiert.

Das bessere Durchschauen des Realitätsbereichs bedeutet zweitens, daß einerseits (1) regelhafte bzw. gesetzmäßige Zusammenhänge zwischen mehreren Komponenten des in bestimmter Art explizierten Realitätsbereichs festgestellt und allenfalls für Erklärungen und Vorhersagen von Ereignissen genutzt wer-

den. Es werden also Warum-Fragen beantwortet: „Warum ist dieses Kind ängstlich?“ - „Strenge Mütter haben ängstliche Kinder, und seine Mutter ist streng.“ Und es bedeutet andererseits (2), mit Hilfe von theoretischem „Gesetzeswissen“ (z. B. einer Erziehungsstil-Theorie) die unterstellten Regelmäßigkeiten oder Gesetzmäßigkeiten (nach (1) ihrerseits zu erklären bzw. zu begründen. Dabei geht es um die Beantwortung von Inwiefern-Fragen: Inwiefern treten Komponenten des explizierten Realitätsbereichs (nach (1)) in gesetzmäßiger Weise simultan oder sukzessiv auf? (Vgl. auch Stegmüller, 1969, s. 90ff.)

Im Sinne des metatheoretischen Strukturalismus (s. oben S. 261) handelt es sich beim Beantworten von Inwiefern-Fragen (im hier unterstellten Sinne) darum, daß ein zur theoretischen Erklärung anstehendes Derivat des programmeigenen Problems P zu einem Partialmodell ( $M_{pp}(T')$ ) einer generierten oder importierten Theorie T' wird und daß das Derivat zu einem Modell (M(T')) dieser Theorie T' ergänzt werden kann. Es sind also besonders die theoretischen Annahmen von T', die dann durch die im P-Derivat enthaltenen Behauptungen zu Ereigniszusammenhängen im Realitätsbereich erfüllt sind und die damit die (nach (1)) konstatierten Ereigniszusammenhänge theoretisch erklären. (Vgl. dazu auch Westermann, 1987, S. 75f.) Angstforscher mögen derart (ad (1)) gesetzmäßige Zusammenhänge zwischen der gemessenen Ängstlichkeit von Kindern und dem gemessenen Erziehungsstil ihrer Eltern feststellen und sie für Ereigniserklärungen und Vorhersagen nutzen. Und sie mögen (ad (2)) den gefundenen Zusammenhang zwischen Erziehungsstil und ängstlichem Verhalten aufgrund einer Erziehungsstiltheorie T' theoretisch erklären. (Vgl. dazu Schneewind & Herrmann (Hrsg.), 1980.)

Forschungsprogramme, bei denen sich die Bearbeitung des Problems P auf die Beantwortung der Was-Frage beschränkt, können unter gewissen Vorbehalten als nicht-nomologische („qualitative“, „verstehende“, „hermeneutische“ o.dgl.) Forschungsprogramme bezeichnet werden (vgl. Kapitel 3).

Das Durchschauen eines Problems bedeutet auch, von der vorwissenschaftlich gegeben Fülle und Komplexität eines Realitätsbereiches abzusehen und die erfahrbare Wirklichkeit zu dekomponieren (Klages, 1967). Und es bedeutet, im dargestellten Sinne Entwürfe für die theoretische Rekonstruktion des im Annahmehern repräsentierten Realitätsbereichs zu machen und sie unter möglicher Vermeidung interner oder externer Widersprüche (s. o. S. 268f.) anzuwenden. Diese Entwürfe sind stark „verkürzte“ Abbildungen („Modelle“) des zum Forschungsproblem gemachten Realitätsausschnitts (Stachowiak, 1973). Wer etwa die Angstentstehung als einen Prozeß des operanten Konditionierens rekonstruiert, sieht mit der Wahl eben dieses „Modells der Angstentstehung“ notwendigerweise von manchen Merkmalen der Angstentstehung ab, die auch zu unseren kultureigenen Erfahrungen mit der Angstentstehung



gehören mögen. Eine Theorie zur Angstenstehung (vgl. Krohne, 1976) kann aber nur dann wissenschaftlich präzise und empirisch prüfbar sein, wenn sie nicht alles und jedes berücksichtigt, was jemals im Zusammenhang mit der Angstenstehung geschehen ist oder was darüber gedacht werden kann. - Man kann zusammenfassen, daß psychologische Grundlagenforschung stets dekomponierend und thematisch-reduktiv ist.

### **5.2.2 Technologische Programme**

Diese Programme stehen unter der dominanten Maxime, die Effizienz der nicht-forschenden Praxis von Psychologen und „Psychologie-Importeuren“ zu erhöhen, d.h. operatives Hintergrundwissen und standardisierte Techniken (als normierte Handlungsanweisungen) bereitzustellen. Angezielt sind Problemlösungsvorgänge, die „in der Praxis funktionieren“, also die auch verlässlich, nebenwirkungsfrei, routinisierbar, wirtschaftlich sowie rechtlich und faktisch akzeptierbar sind. So mag die in einem Forschungsprogramm erfolgende Entwicklung eines lernzielorientierten Tests zwar den neuesten Erkenntnissen einer avancierten Testtheorie entsprechen; wenn dieser Test aber nicht zur Verbesserung der Erkennung bestimmter lernzielbezogener Leistungen beiträgt, wenn er zu schwierig anzuwenden, wenn er zeitökonomisch untragbar oder wenn er aus rechtlichen Gründen unakzeptabel ist, so stellt er eine mißglückte technologische Problemlösung dar. (Vgl. Alich & Rössner, 1983; Bunge 1967; Brocke, 1978, Westmeyer, 1977).

In der Regel verbietet es sich bei technologischen Problemlösungen, in starkem Maße von der konkreten Beschaffenheit des jeweiligen „praktischen Problems“ zu abstrahieren. So mag es z.B. in einem Forschungsprogramm nicht um die „abstrakte Natur“ des menschlichen Lernens überhaupt, sondern um das ganz konkrete Problem des Lernens von Schülern in bestimmten Sonderschulen und um Verbesserungen des Lernerfolgs gehen. Fast immer sind es auch nicht die neuesten, originellsten, kühnsten und riskantesten Ideen, die für die nicht-forschende Praxis hilfreich sind. Bewährtheit, Nebenwirkungsfreiheit, Verlässlichkeit, Akzeptanz und ähnliche Merkmale sind hier von höherem Wert. Sind technologische Lösungen zwar effizient, so können aber dabei die theoretischen „Inwiefern-Fragen“ (s. oben) doch ungeklärt bleiben; die theoretische Begründung auch von effizienten technologischen Lösungen ist nicht selten mangelhaft.

Ersichtlich wird in technologischen Programmen auf unterschiedliche Weise versucht, geeignetes Handlungswissen zu generieren (Stachowiak, 1973, S.99ff.). Das kann u.a. durch die Instrumentalisierung von auf Derivate des Annahmehkerns zu P erfolgreich angewandten theoretischen Annahmen einer Theorie T' (s.o. S.259) erfolgen: Von gut bewährten Annahmen über einen

Realitätsbereich kann man zu technologischen Regeln gelangen, die das praktische Handeln steuern. Man kann auch durch geeignete Forschungsmaßnahmen das Hintergrundwissen für praktisches Handeln verbessern. Daneben tritt die Gewinnung von Handlungswissen aus praxisbezogenen Ad hoc-Erfahrungen der wissenschaftlichen Akteure, aus methodisch unkontrollierten Erfahrungskumulationen der Berufspraxis, auch aus Expertenüberzeugungen usf. Das für die nichtforschende Praxis der Bewertung oder Änderung konkreter „Systeme“ (z.B. Klient, Betrieb, Fan-Club) benötigte Handlungswissen gewinnt man häufig dadurch, daß man Erfahrungen, die man (bei eklektischer Verwendung aller dieser Wissensquellen) im Umgang mit möglichst vielen konkreten „Systemen“ ähnlicher Art kumuliert hat, in flexibler Weise auf den jeweils neuen Fall anzuwenden lernt. Diese Variante des Erwerbs von häufig völlig neuartigem Handlungswissen ist ersichtlich von derjenigen verschieden, bei der dieses Wissen primär durch die Anwendung einer vorliegenden Theorie T' auf praktisch relevante Realitätsbereiche gewonnen wird. - Der Austausch technologischer Programme mit sozialen Systemen, in denen nicht-forschende Praxisarbeit geleistet wird, ist naturgemäß viel intensiver als bei grundlagenwissenschaftlichen Programmen. (Vgl. auch Bunge, 1967, S. 132ff.; Westmeyer, 1977.)

Die technologische Forschung ist in keiner Weise der Grundlagenforschung nachgeordnet (Herrmann, 1979, S. 164ff.). Grundlagenwissenschaftliche Forschungsprogramme können von den Ergebnissen technologischer Forschungsprogramme profitieren. Ein Beispiel ist die Entwicklung allgemeiner Handlungstheorien, die von arbeitspsychologischen Systematisierungen stark befruchtet worden ist (Hacker, 1978). Umgekehrt sind diese maßgeblichen arbeitspsychologischen Entwicklungen ihrerseits in starkem Maße in grundlagenwissenschaftlichen Theoriebildungen zur menschlichen Tätigkeit fundiert (vgl. u. a. Leontjew, 1977). - Im übrigen bedeutet die Unterscheidung von grundlagenwissenschaftlichen und technologischen Forschungsprogrammen selbstverständlich nicht, einzelne Forscher könnten nicht Mitglieder in beiderlei Forschungsprogrammen sein. So mag ein Organisationspsychologe sowohl an einem sozialpsychologischen Grundlagenproblem als auch an einem technologischen Optimierungsproblem arbeiten.

### 5.3 Technologische Techniken-Programme und „Wissensprogramme“

Technologische Programme können in erster Linie das Ziel verfolgen, standardisierte Techniken (z. B. Tests, standardisierte Therapietechniken o. dgl.) zu erarbeiten. Man kann dann von Techniken-Programmen sprechen. Eine solche Technik anzuwenden heißt, vorgegebene Handlungsregeln zu befolgen, also

z. B. nach einem Testmanual vorzugehen. Man kann Bewertungstechniken (Tests usw.) und Veränderungstechniken (Therapietechniken usw.) unterscheiden.

Technologische Programme können auch - als „Wissensprogramme“ - primär ein für die nicht-forschende Praxis unmittelbar nutzbares operatives (Hintergrund-) Wissen erarbeiten, also die Arbeit des nicht-forschend Handelnden z. B. dadurch effizienter zu machen versuchen, daß instrumentalisierbares Wissen über die Wirksamkeit von Werbetexten oder über die Lernmotivation in Grundkursen der gymnasialen Oberstufe oder über Bettnässen erarbeitet wird.

Technologische Forschungsprogramme innerhalb der Psychologie verfolgen in der Regel das Ziel, sowohl operatives Hintergrundwissen als auch Standardtechniken zu generieren. Die Einteilung in Techniken-Programme und „Wissensprogramme“ verweist eher auf „Mischungsverhältnisse“ und läßt sich zur Angabe jeweils dominierender Zielrichtungen bzw. Problemstrukturierungen verwenden.

Auch die Techniken- und Wissensprogramme erfüllen die Voraussetzungen von psychologischen Forschungsprogrammen überhaupt (s.o. S. 252). Technologische Programme haben spezifische Probleme P: z.B. das Lernen von Schülern in Sonderschulklassen, das Stottern, die hochbegabten Grundschüler, Qualitätszirkel in Industriebetrieben, die Akzeptanz neuer Technologien in Institutionen, das Erlernen von Programmiersprachen, den Vandalismus in Städten, psychologische Aspekte der Drogensucht u. dgl., aber auch die diagnostische Erfassung von Aphasien, die Prüfung von Studienbewerbern, die optimale Gestaltung forensischer Explorationen u.dgl. Die Probleme P beziehen sich hier fast stets auf schon in außerwissenschaftlichen Lebensbereichen relativ gut formulierbare, zirkumskripte, in relevanten Gruppen der Bevölkerung konsensuell interpretierte Realitätsbereiche. Die Annahmenkerne zu diesen Realitätsbereichen werden im technologischen Programm - wie jedes Problem P - expliziert, präzisiert, durch „Faktenbeschaffung“ angereichert und unter Beiziehung von Theorien rekonstruiert. Beim Import von Theorien geht man in technologischen Programmen betont effizienzgesteuert und eklektisch vor (vgl. auch Klages, 1967).

Grundlagenwissenschaftliche und technologische Theorien unterscheiden sich nach Bunge (1967, S. 121ff.) wie folgt: Zwar kulminiert in beiderlei Theorien die bisherige Forschung, und beide Arten sind zugleich Handhaben für die weitere Forschung. Doch geben technologische Theorien in spezifischer Weise die Basis her für die Gewinnung von Handhaben, die den Verlauf optimalen praktischen Handelns leiten. Diejenigen Akteure, die nichtforschende, technisch-praktische Tätigkeiten ausüben, handeln zumeist in Beachtung von Entscheidungen, die im Lichte der jeweils besten technologischen, nicht aber un-

bedingt der besten grundlagenwissenschaftlichen Erkenntnis getroffen werden. Die meisten grundlagenwissenschaftlichen Erkenntnisse sind nämlich für das technisch-praktische Handeln zu abgelegen oder schlicht irrelevant. Technologische Theorien beziehen sich, Bunge zufolge, nicht, wie grundlagenwissenschaftliche Theorien, primär darauf, was man als ohnedies geschehend registrieren kann, sondern darauf, was jeweils getan werden sollte, um etwas hervorzubringen, zu vermeiden, zu ändern oder zu verbessern (1967, S. 123). Und das bedeutet eben, daß technologische Theorien vor allem nützlich, effizient, verläßlich und routinisierbar sein müssen.

Bunge (1967, S. 121ff.) unterscheidet inhaltlich-technologische Theorien von operativ-technologischen Theorien. Erstere richten sich auf die „Objekte“ des technischen Handelns. Sie haben oft grundlagenwissenschaftliche Vorgängertheorien, deren Spezifikationen sie sind, und generieren technologisches Hintergrundwissen. (So wird beim Aufbau pädagogisch-technologischer Theorien des Schulkindes auf theoretische Entwicklungen der Kinderpsychologie qua Grundlagenforschung zurückgegriffen.) Die operativ-technologischen Theorien beziehen sich auf das praktisch-technische Handeln selbst (z. B. Testtheorien). Auch sie pflegen von vornherein auf technische Zwecksetzungen hin angelegt zu sein und generieren technologische Regeln.

## 5.4 Grundlagenwissenschaftliche Sachproblem- und Theorie-Programme

Die bisher verwendeten Beispiele für Forschungsprogramme (u. a. Leistungsmotivationsprogramm, Angstprogramm) beziehen sich überwiegend auf Sachprobleme. Das Leistungsstreben, die Angst, auch das Tiefensehen, die optischen Täuschungen, die visuelle Suche, das Kurzzeitgedächtnis, das komplexe Problemlösen, die Sprachproduktion, die Einstellungsänderung, die Moralentwicklung beim Kinde usf. usf. gehören zum großen und offenen Bereich unterschiedlich umfangreicher psychologischer Sachprobleme. Die grundlagenwissenschaftliche Erforschung solcher Sachprobleme erfolgt (in der bisher in diesem Kapitel erläuterten Weise) im Rahmen von Sachproblem-Programmen. (Vgl. Herrmann, 1976.)

Probleme P von Forschungsprogrammen können daneben auch cum grano salis auf Theorien (d.h. auf theoretische Ideen, Auffassungsmuster, allgemeine Konzeptualisierungen, ideologische oder quasi-mythologische Weltstrukturierungen, Vorstellungen über Relationen und Funktionen in der Wirklichkeit usf.) bezogen sein. (Vgl. dazu auch Shapere, 1974.) Solche quasi-theoretischen Strukturen sind hier die problematisierten Realitätsbereiche, die im Wege der Konstitution eines Annahenkerns zum Problem P eines Programms werden.

Wir sprechen dann von Theorie-Programmen. Quasi-theoretische Strukturen können zu den „sozialen Sachverhalten“ gerechnet und sie können derart als Teile einer (sozialen, historischen) Realität verstanden werden. Und diese Realitätsbereiche können im Grundsatz ebenso problematisiert werden wie andere Realitätsbereiche auch.

Zu den problematisierten theoretischen Ideen, die Gegenstand der Bearbeitung in psychologischen Theorie-Programmen geworden sind, gehört z.B. die generelle hedonistische Idee, daß Menschen gern glücklich, erfolgreich und zufrieden sind und gern Tätigkeiten ausüben, die zum Erfolg und zur Zufriedenheit führen. Hieraus wurde das Forschungsprogramm des operanten Konditionierens. (Vgl. dazu Hilgard & Bower, 1971.) - Oder ein kompliziertes Gefüge zum Teil aus Alltagsüberzeugungen und zum Teil aus Ergebnissen anderer Wissenschaften (z.B. der Physik) wurde zum Annahmestamm des Forschungsprogramms der „Berliner Gestalttheorie“. (Zum Annahmestamm vgl. Metzger, 1954.) Eine der im gestalttheoretischen Programm bearbeiteten Annahmen ist bekanntlich diejenige, daß das Ganze mehr sei als die Summe seiner Teile. Eine andere besteht darin, daß für das menschliche Erleben und für zentrale Prozesse des Gehirns identische Funktionsgesetze gelten (= leibseelischer „Isomorphismus“). Man kann sich letzteres am Beispiel des für das menschliche Erleben charakteristischen „Prägnanzprinzips“ verdeutlichen: Menschen erleben ihre Umgebung einfacher, geschlossener, prägnanter, als sie „wirklich“ ist; ein Winkel von  $89^\circ$  erscheint uns als ein rechter Winkel - nie aber umgekehrt. Das Hirngeschehen, das mit diesem Erleben einhergeht, folgt nun (nach gestalttheoretischer Auffassung) felddynamischen Prinzipien äquivalenter Art: Auch bei der physiologischen Reizverarbeitung entstehen vereinfachte und „präzisierte“ zerebrale Repräsentationen der Umgebungsreize. Das Erleben und die Hirnprozesse folgen beide dem „Prägnanzprinzip“ und sind insofern funktionsgleich. Grundsätzlicher formuliert: Das Erleben und die höheren Hirnvorgänge sind, der Gestalttheorie zufolge, lediglich zwei verschiedene Aspekte derselben Realität; sie sind identisch strukturiert („isomorph“).

Psychologische Theorie-Programme rekonstruieren auch quasi-theoretische Strukturen, die zu den fundamentalen Auffassungen anderer Wissenschaften gehören. So kann man nach der allgemeinen Erfahrung der Geologie geologische Schichten unterscheiden: Die tiefsten Schichten sind am ältesten; man kann höhere Schichten manipulieren, ohne damit die tieferen zu tangieren, nicht aber umgekehrt. (Ufs.) Diese Schichten-Konzeption wurde zu einem problematisierten und forschungsprogrammspezifisch bearbeiteten Annahmestamm der älteren psychologischen Persönlichkeitsforschung. (Vgl. dazu Rothacker, 1952.) Oder man kann die Informationsverarbeitung im Computer - generell bei Turing-Maschinen - oder auch die Informationsverarbeitung in neuronalen Netzwerken als Grundidee verschiedener Theorie-Programme nutzen. (Vgl. Gigerenzer, 1988; Lachman et al., 1979; Levelt, 1991.)

Theorie-Programme haben nach allem theorieartige Strukturen zu ihrem Thema und Problem. Diese Strukturen, die in Theorie-Programmen expliziert und bearbeitet werden, bilden meist bereits so etwas wie Vorformen von gesetzesförmigen theoretischen Annahmen. So ist bereits in der hedonistischen Idee, daß man dasjenige gern tut, was Zufriedenheit verspricht, nahegelegt, daß man etwas um so wahrscheinlicher tut, je erfolgreicher es ist. (Das aber ist schon eine quasi-nomologische Annahme.) Mit der Explikation und Elaborierung des Annahmengkerns von Theorieprogrammen bilden sich relativ geschlossene Einzeltheorien. So entstanden als spezifische Ausformungen der allgemeinen „hedonistischen“ Konzeptualisierung, daß Menschen dasjenige gern tun, was zufrieden macht, mehrere Einzeltheorien (Lerntheorien), u.a. von Thorndike, Skinner und Hull (vgl. Hilgard & Bower, 1971). Solche Einzeltheorien können als Derivate (s.o. S.272) des indisponiblen Annahmengkerns dieses Forschungsprogramms aufgefaßt werden. Bei der Entwicklung einer Einzeltheorie aus dem Annahmengkern eines Theorie-Programms spielt die Spezifizierung der empirischen Erfassung wesentlicher Komponenten der für das Programm charakteristischen Ideen bzw. Begriffe (Konstrukte) eine entscheidende Rolle. So ist Skinners Einzeltheorie ohne sein spezifisches Meß- und Erhebungsverfahren zur Häufigkeit von Reaktionen (Skinner, 1938) nicht vorstellbar. (Vgl. auch Shapere, 1974, S. 531f.)

Derivate des Annahmengkerns von P in einem Theorie-Programm (= Einzeltheorien) können zu intendierten Anwendungen von anderen im Programm generierten Einzeltheorien T' werden. So ist zum Beispiel Thorndikes Konditionierungstheorie als ein Derivat (= Einzeltheorie) im genannten hedonistischen Theorie-Programm interpretierbar. Und man kann diese Theorie als Anwendungsfeld für Skinners später entstandene Verstärkungstheorie (= T') verstehen, die ebenfalls ein Derivat im hedonistischen Programm ist. (Vgl. dazu Skinner, 1938.)

Charakteristisch für Theorie-Programme ist der Sachverhalt, daß für die im Programm per Präzisierung und Elaboration des Annahmengkerns entstandenen Einzeltheorien möglichst viele programmexterne Anwendungen gesucht und ausprobiert werden. Skinners Theorie wurde oder wird auf die Angstentstehung, auf die Aggressionsgenese, auf Tierdressuren verschiedener Art, auf die Erziehungsstilforschung, auf Fragen der pädagogischen Psychologie, auf den Spracherwerb und viele andere Sachprobleme angewandt. Anders formuliert: Die in einem Theorie-Programm entstandene Einzeltheorie von Skinner wurde in vielen grundlagenwissenschaftlichen Sachproblem-Programmen (und auch in technologischen Programmen) als (importierte) Theorie T' genutzt. Die Bearbeitung von Problemen P in Theorie-Programmen erfolgt generell in starkem Ausmaß über den Export in Sachproblem-Programme. Über den Erfolg von Theorie-Programmen wird zum Teil durch die Arbeit der (importierenden) Akteure von Sachproblem-Programmen an ihrem Problem P ent-

schieden. Zwischen den sozialen Systemen eines Sachproblem- und eines Theorie-Programms kann so die Austauschbeziehung des „Zulieferns von Anwendungen“ entstehen.

Die Proponenten einer anwendungsgerechten Theorie T' können die erfolgreichen Anwendungsversuche in einer Reihe von gleichzeitig oder sukzessiv arbeitenden Sachproblem-Programmen als wachsenden Erfolg bei der Problemlösung in ihrem Theorie-Programm verbuchen. Als Folge davon kann sich eine zeitweilig dominierende Stellung des betreffenden Theorie-Programms im Wissenschaftssystem ergeben (vgl. „Gestalttheoretische Ära“, Siegeszug Skinners, Dominanz der Computer-Metapher usw.). - Selbstverständlich werden daneben auch in hohem Maße innerhalb der sozialen Systeme, die Theorie-Probleme bearbeiten, Anwendungsversuche für die eigenen Einzeltheorien unternommen. (Auch hier sind die Gestalttheoretiker ein gutes Beispiel. Vgl. Metzger, 1954.)

Die Sachproblem-Programme und die Theorie-Programme sind Teilmengen der Forschungsprogramme, wie sie zu Beginn dieses Kapitels erläutert wurden. Vielleicht sind die Sachproblem-Programme etwas eher „prototypisch“ für psychologische Forschungsprogramme als Theorie-Programme. Dennoch erfüllen die Theorie-Programme ebenso die Definition psychologischer Forschungsprogramme.

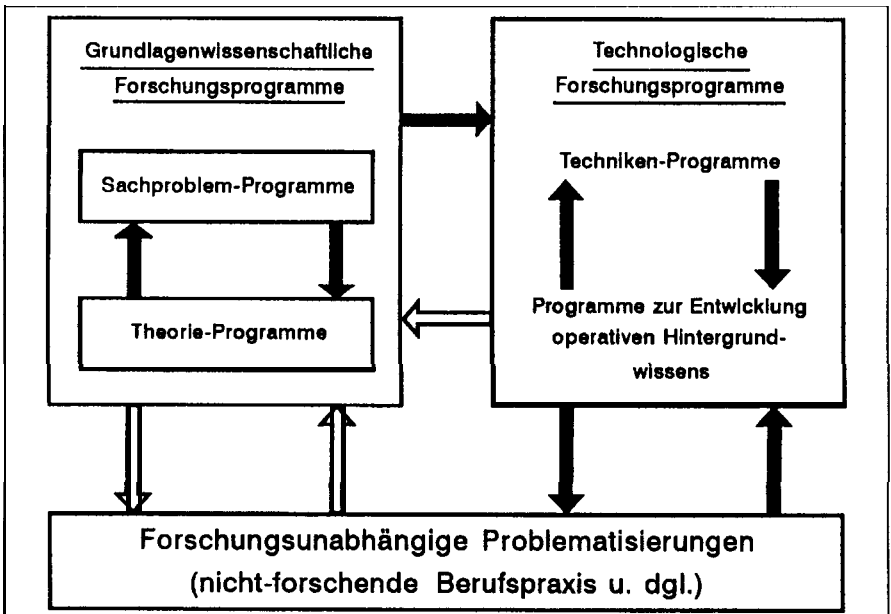


Abb. 1: Programmtypen und einige Beeinflussungsrichtungen (s. Text)

Sachproblem- und Theorie-Programme unterscheiden sich fundamental in der strategischen Zielrichtung, die in Kernannahmen höherer semantischer Ordnung konstituiert ist: In Sachproblem-Programmen folgt die Bearbeitung der zentralen Problemrichtung: „Wie kann man das Problematisierte P erläutern, erklären, begründen?“ In Theorie-Programmen arbeitet man nach der Fragestellung: „Was kann man mittels des Problematisierten P erläutern, erklären, begründen?“ Sachproblem-Programme haben - pauschal formuliert - ein Explicandum und Explanandum (= ein Sachproblem) und suchen (u.a.) Explicantien und Explanantien (= Theorien). Theorie-Programme haben umgekehrt potentielle und zunächst nur wenig angewandte Explicantien und Explanantien (= Theorien) und suchen (als Anwendungsfälle) Explicanda und Explananda (= Sachprobleme).

Abbildung 1 gibt einen Überblick über die wichtigsten in diesem Abschnitt besprochenen Sachverhalte. (Die gefüllten Pfeile geben Hauptbeeinflussungsrichtungen, die umrandeten Pfeile Nebenrichtungen der Beeinflussung an.)

## 6. Einige externale Einflüsse auf Forschungsprogramme

### 6.1 Internale und externale Einflüsse

Der zu einem bestimmten Zeitpunkt existierende Zustand eines Forschungsprogramms, vor allem der Status der Problembearbeitung, ist, wie sich aus den Ausführungen dieses Kapitels ergibt, u. a. von der Beschaffenheit des Problems P, von der Problemlösungsdynamik innerhalb des Programms, also von den aus der wissenschaftlichen Arbeit selbst erwachsenden Fortschritten, Hindernissen, Verzögerungen, innovativen Durchbrüchen u. dgl. und damit auch von den Leistungen der wissenschaftlichen Akteure und ähnlichen programm-spezifischen Bedingungen, und nicht zuletzt von der Zeitstrecke abhängig, die bereits für die Problembearbeitung zur Verfügung stand. Die am Beginn eines Forschungsprogramms vorliegende Problemkonstitution kann entweder vage und schlecht definiert sein oder bereits in außerwissenschaftlich gut vorstrukturierter Form vorliegen; sie kann ein sehr weites Feld betreffen oder zirkumskript sein; sie kann in unterschiedlichem Ausmaß mit (heuristischen) Hinweisen auf Lösungsmöglichkeiten versehen sein. Man kann mehr oder minder viel Zeit für die Problembearbeitung gehabt haben. Es können bezüglich der Problembearbeitung in unterschiedlichem Maße beschleunigende oder hemmende Ereignisse aufgetreten sein, die aus der Problembearbeitung selbst resultieren. Alle diese Bedingungen führen zu einer forschungsprogrammspezifischen, internalen Problemlösungsdynamik, die für jedes Forschungsprogramm separat zu analysieren ist. (Insofern bei der Problemlösungsdynamik



Theorien eine Rolle spielen (s.o. S.276), handelt es sich zugleich auch um eine interne Theoriedynamik.)

Der internalen Dynamik sind externale Einflüsse auf Forschungsprogramme - und damit auf den Zustand eines Forschungsprogramms zu einem bestimmten Zeitpunkt - aufgelagert, von denen einige erläutert werden sollen. Externale Einflüsse entstehen nicht problemspezifisch, im Zuge der Bearbeitung genau des Problems P innerhalb des sozialen Systems des betreffenden Forschungsprogramms, sondern sie beeinflussen die programminterne Arbeit in anderer Weise:

- Forschungsprogramme als soziale Systeme können, wie dargestellt, als Subsysteme von übergeordneten Systemen, vor allem von Wissenschaftssystemen (z. B. der Psychologie), verstanden werden (s. oben S. 266). Temporäre oder dauernde Eigenschaften dieses übergeordneten Systems stellen sich als Kodeterminanten der Problembearbeitung in den als Subsysteme figurierenden Forschungsprogrammen dar.
- Wie ausgeführt (s.o. S.266), sind wissenschaftliche Akteure nur soweit Mitglieder des sozialen Systems (einer Forschungsgruppe o.dgl.), wie sie sich an der Bearbeitung des Problems P gemeinsam und interaktiv beteiligen. Die Akteure haben daneben systemexterne Eigenschaften und sind systemexternen Einflüssen ausgesetzt. Die systemexterne Beschaffenheit der Mitglieder eines Forschungsprogramms kann ebenfalls als Kodeterminante für die Problembearbeitung in einem Forschungsprogramm aufgefaßt werden.
- Man kann weitere externale Einflußgrößen auf Forschungsprogramme ausgrenzen: die räumlichen, personellen, administrativen, medienbezogenen und ähnliche Bedingungen, wie sie zum Beispiel durch politische Entscheidungsprozesse verschiedener Art, durch die jeweilige materielle Situation von Hochschulen, die Organisation der finanziellen Wissenschaftsförderung, das Verlagswesen, hochschul- und arbeitsrechtliche Regelungen in Staaten, die Organisation wissenschaftlicher Fachgesellschaften u. dgl. mitbestimmt werden.

Im Wissenschaftssystem werden (per Berufssozialisation u. dgl.) generelle Normen und Rationalitätsmaximen und auch „handwerkliche“ Verfahrensregeln an die Mitglieder derjenigen Forschungsprogramme vermittelt, die zu diesem Wissenschaftssystem gehören. So besteht zum Beispiel für Psychologen ein allgemeiner Konsens, daß man nicht plagiiert, daß Theorien keine logischen Widersprüche enthalten dürfen oder daß man den inferenzstatistischen Chi-Quadrat-Test nicht auf Maßzahlen anwenden darf. Im Wissenschaftssystem gibt es generelle Vorgaben für Problemkonstitutionen und für Strategien der Problembearbeitung. Übertretungen werden sanktioniert. So kann aber der Zustand eines Wissenschaftssystems im Extrem auch zur

Verunmöglichung der Bearbeitung von Problemen P in Forschungsprogrammen führen. Die Durchführung von Problembearbeitungen kann so auch wesentlich erschwert oder erleichtert sein. Beispiele: In der Zeit nach dem 11. Weltkrieg wäre eine forschungsprogrammsspezifische Bearbeitung einer Rasenseelenlehre (vgl. dazu Graumann (Hrsg.), 1985) schon aus Gründen wissenschaftlicher Tabuisierung und Sanktionierung nicht möglich gewesen. - Nach der programmübergreifenden Überzeugungslage der Psychologie der Sechzigerjahre wäre eine Bearbeitung des Problems des Spracherwerbs mittels des Imports einer behavioristischen Lerntheorie sanktioniert und somit zumindest stark erschwert worden (vgl. Hörmann, 1977).

Die systemexterne Beschaffenheit der Forschungsprogramm-Mitglieder kann ebenfalls die Bearbeitung von Problemen P fördern, erschweren oder im Extremfall zum Erliegen bringen. Man denke dabei z.B. an politische Überzeugungsänderungen und geänderte Handlungspräferenzen bei Individuen, die wissenschaftliche Akteure in Forschungsprogrammen sind. (Beispiel: Verzicht auf die Mitarbeit an laborexperimentellen, sog. „positivistischen“ Forschungsprogrammen während und nach der Studentenbewegung um 1968.) - Die Entwicklung einer Theorie in einem technologischen, zum Beispiel einem arbeitspsychologischen Forschungsprogramm kann dadurch gefördert werden, daß Mitglieder dieses Programms zugleich an Theorieentwicklungen in einem anderen - grundlagenwissenschaftlichen - Programm mitarbeiten. Das technologische Programm profitiert so durch Kompetenzerweiterungen seiner Mitglieder, die diese im anderen Programm erhalten. Dies unbeschadet der Sachlage, daß bei technologischen und grundlagenwissenschaftlichen Forschungsprogrammen höchst unterschiedliche Problematisierungsweisen und Zielsetzungen vorliegen (s. oben S. 269). (Ersichtlich kann - umgekehrt - in dieser Weise auch die Problemlösung im grundlagenwissenschaftlichen Programm gefördert werden.)

Daß sehr unterschiedliche politische und administrativ-organisatorische Einflüsse das Schicksal von Forschungsprogrammen immer wieder beeinflussen, ist offensichtlich. In der ehemaligen DDR durften bestimmte Forschungsprobleme nicht bearbeitet werden, andere wurden gefördert (vgl. Herrmann, 1976, S. 132ff.). Die Gesamtalimentierung deutscher Universitäten, politische Verteilungsentscheidungen innerhalb der Universitäten, Studien- und Prüfungsordnungen u. dgl. wirken sich auf die Beteiligung deutscher Wissenschaftler an Forschungsprogrammen aus. Oder man denke an die variable Förderungspolitik in bestimmten Staaten, so z.B. an die temporäre Förderung der Kreativitätsforschung in den USA (sog. „Sputnik-Schock“, vgl. auch Ullmann, 1968) oder an die erhöhte Bereitschaft zur Förderung von „Umweltthemen“ in der Bundesrepublik Deutschland der Achtzigerjahre. Die Erforschung der Sprachproduktion und Sprachrezeption ist zur Zeit u.a. dadurch erschwert, daß die Sprachpsychologie nicht zum Curriculum der Diplom-Psy-

chologen in Deutschland und anderen Ländern gehört und daß (demzufolge) eine Spezialisierung in der Sprachpsychologie für die wissenschaftliche Karriere jüngerer Psychologen kontraindiziert erscheint. Alle diese „lokalen“ Bedingungen können den Fortgang auch überregional (international) betriebener Forschungsprogramme beeinflussen.

## 6.2 Wissenschaftsparadigmen

Auch in psychologischen Veröffentlichungen findet man häufig die mit vielen unterschiedlichen Bedeutungen versehenen Ausdrücke „Paradigma“ und „Wissenschaftsparadigma“.

Im gegenwärtigen Zusammenhang verstehen wir als Paradigma eine wissenschaftstheoretische Konstruktion, die auf Th.S. Kuhn zurückgeht (Kuhn, 1962 (dt. 1967), 1970 u.a.O.; vgl. auch Andersson, 1988; Carrier, 1986). Kuhn und seine Nachfolger haben den Ausdruck „Paradigma“ sehr unterschiedlich (und unterschiedlich präzise) expliziert (vgl. Masterman, 1970). Hier soll damit folgendes verstanden werden (Kuhn, 1967, S.123ff.): Prototypische Paradigmen sind das Ptolemäische Weltbild und die Newtonsche Physik. Paradigmen sind grundlegende, in einer großen Wissenschaftlergemeinschaft akzeptierte „Theorien“: Die Wissenschaftler, die einem Paradigma zugerechnet werden, „sehen“ die Welt in bestimmter Weise, ja sie „leben“ in einer bestimmten Welt. Wissenschaftler, die verschiedenen Paradigmen zugehören, reden über Verschiedenes in verschiedener „Sprache“, sie stellen unterschiedliche Fragen, sie fühlen sich verschiedenen sozialen Systemen zugehörig, sie verwenden verschiedene Metaphern und Analogien, sie benutzen unterschiedliche Forschungsmethoden, usf. (Vgl. auch Lachman, Lachman & Butterfield, 1979, S28ff.). Bei Paradigmen denkt man häufig an „paradigmatische“ Experimente (z.B. Galileis Untersuchungen zum freien Fall) oder den Einsatz „paradigmatischer“ Apparaturen und Instrumente (z.B. Fernrohr, Wilsonsche Nebelkammer). Da alle Erfahrungen „theorieabhängig“ sind, kann man innerhalb eines Paradigmas nur bestimmte Erfahrungen machen und andere nicht. Das bedeutet auch, daß verschiedene Paradigmen nicht direkt vergleichbar, d.h. daß sie inkommensurabel sind und nicht anhand „der Realität“ verglichen werden können. Die grundlegenden Auffassungsmuster, die man als Paradigmen bezeichnet, sind ersichtlich nicht falsifizierbar. Der Übergang einer Wissenschaftlergemeinschaft von einem zu einem anderen Paradigma ist eine „wissenschaftliche Revolution“; er gleicht für den einzelnen Akteur einer Bekehrung oder Konversion. Man kann, Kuhn zufolge, nur feststellen, daß nach einiger Zeit, falls ein Paradigma ein anderes Paradigma besiegt hat, die Mehrzahl der Wissenschaftler zu Anhängern des neuen Paradigmas geworden sind; irgendwelche strikten und explizierbaren Widerlegungen des Alten sind hin-

gegen nicht feststellbar. Oft sterben die Vertreter des alten Paradigmas einfach aus. (Vgl. dazu auch Andersson, 1988, S.37.)

Um die Kuhnsche Paradigmen-Konzeption ist innerhalb der Wissenschaftslehre jahrzehntelang gestritten worden. (Vgl. u.a. Agazzi, 1987; Bayertz, 1987; Carrier, 1986; Gholson & Barker, 1986; Kordig, 1971; Lachman et al., 1979; Lakatos, 1974; Laudan, 1977; Musgrave, 1971; Stegmüller, 1973; Suppe (Ed.), 1977; Toulmin, 1972 u.v.a.) Hier interessiert lediglich, ob und wieweit der Paradigmenbegriff geeignet ist, externe Einflüsse auf Forschungsprogramme verständlich zu machen.

Zweifellos sind Paradigmen als sehr grundlegende, die Wissenschaften fundamental beeinflussende, nicht leicht und schnell austauschbare Globalkonzeptionen zu verstehen. Es handelt sich nicht nur um flüchtige Denkmoden u. dgl. (s. unten). Man kann die vorliegende Psychologiegeschichte daraufhin untersuchen, ob sich in ihr Paradigmen und „revolutionäre“ Paradigmenwechsel abgespielt haben. (Vgl. u.a. Briskman, 1972; Kriz, Lück & Heidbrink, 1987; Lachman et al., 1979, S.20ff.; Palermo, 1971; Weimer, 1974.) So hat man den Übergang von der „Wundtschen“ Erlebnispsychologie zur Gestalttheorie, auch denjenigen von der Erlebnispsychologie der Wundtschen Ära zur Psychologie als „objektiver Verhaltenswissenschaft“ im Behaviorismus und Funktionalismus und auch den Übergang vom Behaviorismus zur Kognitiven Psychologie und generell zum Informationsverarbeitungsansatz als veritable Paradigmenwechsel verstanden. Man unterstellt auch, daß innerhalb einer Wissenschaft wie der Psychologie simultan mehrere Paradigmen bestehen können (Lachman et al., 1979, S.33).

Unter dem Aspekt der im Mittelpunkt dieses Kapitels stehenden Forschungsprogramme stellt sich die Sachlage wie folgt dar: Die Unterstellung wissenschaftlicher Paradigmen konfligiert nicht mit dem Forschungsprogramm-Ansatz. Falls man für die Psychologiegeschichte die Existenz Kuhnscher Paradigmen annehmen will (z. B. Erlebnispsychologie, Behaviorismus, Psychoanalyse, Informationsverarbeitungsansatz), so handelt es sich um Strukturen von Annahmen über große Realitätsbereiche, die von einer Wissenschaftlergemeinschaft in einem Zeitabschnitt vertreten und zugleich problematisiert werden. Diese Annahmenstruktur enthält auch globale Ideen über Problemlösungsrichtungen sowie präskriptive Komponenten, in denen in globaler Weise (methodologisch) festgelegt ist, was in der psychologischen Forschung erlaubt ist und was nicht. Man kann diese Annahmenstruktur (= Paradigma) als Meta- $P_M$  betrachten, welches für ein Wissenschaftssystem, zum Beispiel die Psychologie, konstitutiv ist. Dieses System bildet insofern ein „Meta-Forschungsprogramm“. Es ist nicht auszuschließen, daß im gleichen Zeitraum mehrere „Mets-Forschungsprogramme“ bzw. Wissenschaftssysteme bestehen, an denen Akteure beteiligt sind, die allesamt als Psychologen bezeichnet werden. (Dann existieren gleichzeitig mehrere psychologische Paradigmen.)

Ein Wissenschaftssystem ist, wie mehrmals erwähnt, bezüglich derjenigen sozialen Systeme (Forschungsgruppen usw.), in denen forschungsprogrammspezifische Probleme  $P$  bearbeitet werden, ein übergeordnetes System. Da die hier besprochenen sozialen Systeme durch die Bearbeitung von Problemen  $P$  definiert sind, ergibt sich aus dieser Betrachtung, daß mehrere Forschungsprogramme als „nach einem und demselben Paradigma arbeitend“ verstanden werden können, falls die Annahmenkerne ihrer jeweiligen Probleme  $P_1, P_2 \dots P_n$  eine Kernannahmenmenge bilden, die Annahmen des Metaproblems  $P_M$  des Wissenschaftssystems wesentlich enthält. Anders formuliert: Mehrere Forschungsprogramme sind im Falle der gemeinsamen Zugehörigkeit zu einem Wissenschaftsparadigma durch Annahmenkerne konstituiert, die einander insofern ähnlich sind, als sie alle einige wesentliche Merkmale desjenigen Metaproblems  $P_M$  teilen, durch das die Wissenschaftlergemeinschaft gekennzeichnet ist, deren Subsysteme die Forschungsprogramme sind.

Bei alledem ist, wie ausgeführt, jedes an einem Paradigma beteiligte Forschungsprogramm sowohl internal kodeterminiert als external nicht nur durch seine Teilhabe am Paradigma bestimmt. Immerhin ist aber die Bearbeitung des forschungsprogrammspezifischen Problems  $P$  sowohl in seiner Konstitution als auch bezüglich der Strategien zu seiner Bearbeitung paradigmenspezifisch beeinflusst. Beispiel: Wenn man das Behalten und Erinnern (in einem Sachproblem-Programm) von vornherein über Annahmen zu Gedächtnisspeichern, zur Suche im Speicher, zur Ordnung im Speicher usw. konzeptualisiert und bearbeitet, so teilt das fragliche Forschungsprogramm wesentliche fundamentale Annahmen mit dem globalen Informationsverarbeitungsparadigma. Oder wenn das Theorie-Programm des operanten Konditionierens (s.o. S.277) keine Annahmen zum Erleben von Erfolgen, Mißerfolgen usw. oder zur Erfassung solcher Erlebnisse per Introspektion enthält, so teilt es damit Grundvoraussetzungen des behavioristischen Paradigmas.

### 6.3 Moden und Metaphern

Wissenschaftsmoden (Herrmann, 1987) findet man auch in der Psychologie. Es handelt sich dabei oft um den Einsatz von Methoden, Verfahren o. dgl., die jeweils soeben bekannt geworden sind, aber noch nicht ubiquitär - sozusagen von jedermann - auf der Basis erworbener Kompetenzen verwendet werden können (Beispiele: Faktorenanalyse (s. oben S. 258), Multidimensionales Skalieren, PDP-Modelle). Bisweilen besteht, wenn die fragliche Methode o. dgl. dann überall begriffen worden ist und verwendet wird, eine nachfolgende, neue Mode darin, daß das zuvor modische Verfahren kritisiert und als inadäquat erwiesen wird (Beispiel: Faktorenanalyse; Orlik, 1967). Soweit es sich bei alledem nicht um eine aus der internalen Problemlösungsdynamik

erwachsende Mittelwahl, sondern um einen Mitteleinsatz als Profilierungsstrategie von Wissenschaftlern und Wissenschaftlergruppen handelt, lassen sich diese Moden als Einflüsse der systemexternen Beschaffenheit von Forschungsprogramm-Akteuren auffassen (s.o. S. 258).

Werden Methoden, Verfahren, aber auch Theorien temporär in großen Wissenschaftlergemeinschaften (forschungsprogramm-übergreifend) als gut, attraktiv, unverzichtbar o.dgl. betrachtet, so kann es sich dabei auch um Implikationen eines generellen Paradigmas (s. oben S.257) handeln, das in die einzelnen Forschungsprogramme „diffundiert“. Man kann Wissenschaftsmoden auch als höherfrequente Wellen verstehen, die die „Grundwellen“ des jeweiligen Wissenschaftsparadigmas sozusagen überlagern. Dabei dürften diese relativ kurzweiligen Moden wohl so beschaffen sein, daß sie den Grundannahmen des jeweiligen Paradigmas nicht widersprechen. - Insofern wirken sich Wissenschaftsmoden auf Forschungsprozesse nicht nur im Wege systemexterner Merkmale von Forschungsprogramm-Akteuren aus, sondern auch als Wirkungen des übergeordneten Systems, dessen Subsystem das Forschungsprogramm ist. (Beide Effekte schließen sich ersichtlich nicht aus.)

Ein für die Psychologie nicht hinreichend untersuchter, möglicher Tatbestand könnte darin bestehen, daß eine Wissenschaftsmode, die ihrerseits vielleicht aus der plötzlichen Verfügbarkeit eines Forschungswerkzeugs hervorgeht, zur wissenschaftlichen Revolution führt und damit zum Beginn eines neuen, lange durchgehaltenen psychologischen Paradigmas werden kann. Dies wahrscheinlich dann, wenn das bestehende (alte) Paradigma - wie auch immer - krisenhaft degeneriert ist und somit zur Ablösung ansteht (vgl. Kuhn, 1967). In dieses Vakuum stößt dann gleichsam die Modeerscheinung hinein. Eine zunächst allenfalls als kurzdauernde Mode diagnostizierbare Innovation fixiert sich zu einem über längere Zeit quasi-stationären Zustand. Zum Beispiel konnte für die zunächst in einigen wenigen Konferenzen in den USA rasch etablierte Psycholinguistik sprachstrukturalistischer Provenienz (Chomsky, 1957, 1959) zunächst wohl kaum vorausgesehen werden, daß sie die zuvor beherrschende behavioristische Psychologie sprachlichen Verhaltens (Skinner, 1957) geradezu in der Art einer wissenschaftlichen Revolution in wenigen Jahren völlig „besiegte“ und für Jahrzehnte beherrschend blieb. Die Generative Transformationsgrammatik als damals neuartiges Werkzeug der Sprachanalyse hat dabei eine wichtige Rolle gespielt. (Vgl. auch Hörmann, 1981.)

Gigerenzer (1988) hat gezeigt, daß Werkzeuge des Forschers, zum Beispiel die Statistik oder der Computer, auf dem Wege der Metaphernbildung zur Erzeugung bestimmter Theorien führen können: So ergibt die Beschäftigung des Forschers mit dem Computer eine Theorie, derzufolge kognitive Prozesse ihrer Natur nach Rechengvorgänge sind (so Pylyshyn, 1980, S. 114). Oder die Arbeit mit statistischen Methoden resultiert in der Annahme, die menschliche

Kognition sei ihrer Natur nach „intuitive Statistik“ (so Rucci & Tweney, 1980). In diesen Fällen „internalisiert“ der Forscher, so Gigerenzer (1988, S.95), sein Werkzeug und macht daraus eine kognitive Theorie; ein Forschungsproblem wird hier in Analogie des Werkzeuggebrauchs des Forschers rekonstruiert.

Die Entstehung von Paradigmen darf auch im Zusammenhang mit der technischen Verfügbarkeit von Forschungsressourcen gesehen werden (vgl. Carrier, 1986; s. auch Andersson, 1988). Ohne die Erfindung elektronischer Rechner (und ohne die Entwicklung der Theorie dieser Rechner) wäre der Informationsverarbeitungsansatz wohl gar nicht oder zumindest nicht in der vorliegenden Form zum (mutmaßlichen) Wissenschaftsparadigma der Psychologie geworden (vgl. auch Lachman et al., 1979). (Allenfalls hätte sich vielleicht ein informationstheoretisch-kybernetisches Paradigma herausgebildet. Vgl. Klix, 1971.)

Zur Funktion von Metaphern bei der Theoriegenese in der Psychologie sei auf Beiträge von Farrell (1986), Gholson & Barker (1986), Reese (1973), Estes (1978), Gigerenzer (1988) und Herrmann (1983) hingewiesen. Eine empirische Untersuchung der psychologischen Metaphorik im Zeitraum von 1894 bis 1975 bieten Gentner & Grudin (1985): Danach sind die häufigsten Metaphern für seelische Vorgänge räumlicher oder systemischer Natur. Farrell (1986) identifiziert die Entstehung von Wissenschaftsparadigmen mit der Requirierung einer jeweils neuen Metapher für globale, vom Psychologen problematisierte Realitätsbereiche. So verstehen, Farrell zufolge, Newell & Simon (1972) im Paradigma der Informationsverarbeitung die „menschliche Seele“ als Computer; G.H. Mead (1973) versteht aus der Sicht des psychologischen Interaktionismus die „Seele“ hingegen als ein Feld, das sich zwischen zwei Individuen oder zwischen einem Individuum und einem Gegenstand ausbreitet. Aus diesen verschiedenen „Bildern“ von der menschlichen „Seele“ ergeben sich u.a. völlig verschiedene Vorstellungen zur Forschungsmethodik. Informationstheoretiker bevorzugen u.a. „lautes Denken“ (und das Reaktionszeitparadigma), Interaktionisten untersuchen soziale Interaktionen (Farrell, 1986, S.719).

Es zeigt sich hier: Programmübergreifende (quasitheoretische) Metaphern beeinflussen auch die Methodenwahl im einzelnen Programm. Und der programmübergreifende (oft modische) „Werkzeuggebrauch“ kann auch die programmspezifische Theorieverwendung beeinflussen. Zwischen Methodik und Theorieverwendung bestehen komplizierte Wechselwirkungen.

## 6.4 Weitere externale Einflußquellen

Der Einfluß der Politik, der Administration, aber auch wissenschaftlicher Fachgesellschaften (vgl. Heckhausen, 1987), des Publikationswesens u.dgl. auf die Entwicklung der Wissenschaften ist Gegenstand der Wissenschaftssozio-

logie (vgl. u.a. Crane, 1967; Gaston (Ed.), 1978; Merton, 1973; Weingart (Hrsg.), 1972) und kann hier nicht im einzelnen dargelegt werden. Daß es sich dabei um zum Teil sehr wirksame externale Einflüsse auf die Bearbeitung von Problemen P in Forschungsprogrammen handelt, wurde schon betont (s. oben S. 279ff.). Hier sei nur noch darauf hingewiesen, daß sich die Bearbeitung von Problemen P nicht nur in der kommunikativen Interaktion innerhalb des jeweiligen Forschungsprogramms realisiert, sondern daß Kommunikation mit wesentlichen Konsequenzen auch zwischen dem Programm und seiner Systemumgebung erfolgt. So gehört zum Austausch zwischen den Netzwerkknoten des Wissenschaftsnetzwerks (s. oben S. 265f.) auch die Publikation von Problemstellungen, Problemlösungsmitteln, Problemlösungsergebnissen. Solche Publikationen können die Problemlösungsprozesse in anderen Forschungsprogrammen beeinflussen und allenfalls Merkmale des gesamten Wissenschaftssystems modifizieren helfen. Sie können auch in vielfältiger Weise die Verfügung über Forschungsressourcen beeinflussen, indem sie z.B. Mitgliedern des publizierenden Forschungsprogramms soziales Ansehen verleihen, wie auch umgekehrt das bereits vorhandene Ansehen von Mitgliedern die Publikationschancen erhöht (vgl. Merton, 1968). Das bedeutet aber, daß die Verfügung über Publikationsentscheidungen durch Zeitschriftenherausgeber, wissenschaftliche Verlagsberater usf. (= „Türhüterfunktion“) eine nicht zu unterschätzende externale Einflußgröße des Verlaufs von Problemlösungsprozessen in Forschungsprogrammen ist. Bei diesen „Türhütern“ handelt es sich weitgehend um „nicht-kontrollierte Kontrolleure“, deren Verhaltenssteuerung wissenschaftspsychologisch noch nicht hinreichend untersucht worden ist (vgl. aber Crane, 1967; Merton, 1973). Ein ähnliches Problem stellen die Wissenschaftler dar, die andere Wissenschaftler im Zusammenhang mit der finanziellen Wissenschaftsförderung prüfen.

Eine weitere externale Einflußquelle ist die Etablierung von tradierten Zitiergewohnheiten oder sogar von explizit vereinbarten Zitierkartellen, also von Konventionen darüber, wen man - im Austausch - zitiert und wen nicht. (Vgl. dazu u.a. Garfield, 1979; Lück, 1988; Metraux, 1985; Rice, Borgman & Reeves, 1988; Sahner, 1979; 1982; Scheerer, 1988; van Strien, 1988.)

## 6.5 Zusammenfassung

Zusammenfassend lassen sich nach den in diesem Kapitel dargestellten Gesichtspunkten wesentliche Bedingungen für den Verlauf von Problemlösungsvorgängen in Forschungsprogrammen wie folgt unterteilen:

Ein Forschungsprogramm wird primär durch die Beschaffenheit des in ihm bearbeiteten (forschungsprogrammspezifischen) Problems P determiniert (= internaler Einfluß).



Ein Forschungsprogramm empfängt daneben im wesentlichen die folgenden externalen Einflüsse:

- Zwischen Forschungsprogrammen bestehen Austauschbeziehungen. So sind in Theorie-Programmen generierte Theorien T' Explikations- bzw. Explanationsmittel, die in Sachproblem-Programmen genutzt werden; zugleich sind dann in Sachproblem-Programmen entwickelte Derivate des Problems P Explicanda bzw. Explananda, die für die Theorie-Programme intendierte Anwendungen der Theorien T' darstellen.
- Es gibt vielfältige Einflüsse aus dem Wissenschaftssystem, dessen Subsystem das Forschungsprogramm ist: Die Forschungsprogramme eines Wissenschaftssystems haben generelle wissenschaftsethische Normen und Rationalitätsmaximen wie auch „handwerkliche“ Verfahrensregeln gemeinsam, die im Wissenschaftssystem (per Berufssozialisation der Mitglieder u. dgl.) an die Mitglieder der Subsysteme vermittelt werden. Aus dem jeweiligen Annahenkern zum Metaproblem  $P_M$  (= Wissenschaftsparadigma) des Wissenschaftssystems gelangen zudem Annahenkomponenten in den jeweiligen Annahenkern zum Problem P des Forschungsprogramms. (Die wissenschaftsethischen Normen, Rationalitätsmaximen und „handwerklichen“ Verfahrensregeln können dabei als partiell paradigmenspezifisch gelten.) Weitere Einflüsse auf das Forschungsprogramm ergeben sich aus kurzweiligen, im Wissenschaftssystem grassierenden Moden, deren Verbreitung allerdings auch über systemexterne Attribute von Forschungsprogramm-Mitgliedern interpretierbar ist (s. oben).
- Wesentliche externalen Bedingungen für die Entstehung und den Verlauf von Forschungsprogrammen sind die politischen, administrativen u. dgl. Verhältnisse in Staaten und andere „lokale“ Sachverhalte (s. oben). Insofern das soziale System, in dem ein programmspezifisches Problem P bearbeitet wird, immer auch räumlich situiert ist (s. oben S.281), wird es durch diese externalen Verhältnisse (je nach der Lokalisierung von Mitgliedern oder von Teilorganisationen des Programms in heterogener Weise) tangiert.
- Insofern die Mitglieder eines Forschungsprogramms auch systemexterne Eigenschaften haben und systemexternen Einflüssen ausgesetzt sind, wird das Forschungsprogramm durch mitgliederspezifische Bedingungen kodeterminiert. Als Beispiele dienen die Übernahme von Modeverfahren als Profilierungsmittel für einzelne Mitglieder und die Überalterung von führenden Mitgliedern eines Programms, durch welche der Import neuer und erfolgversprechender Theorien o. dgl. behindert wird.

Abbildung 2 gibt abschließend eine Skizze der wichtigsten Zusammenhänge.

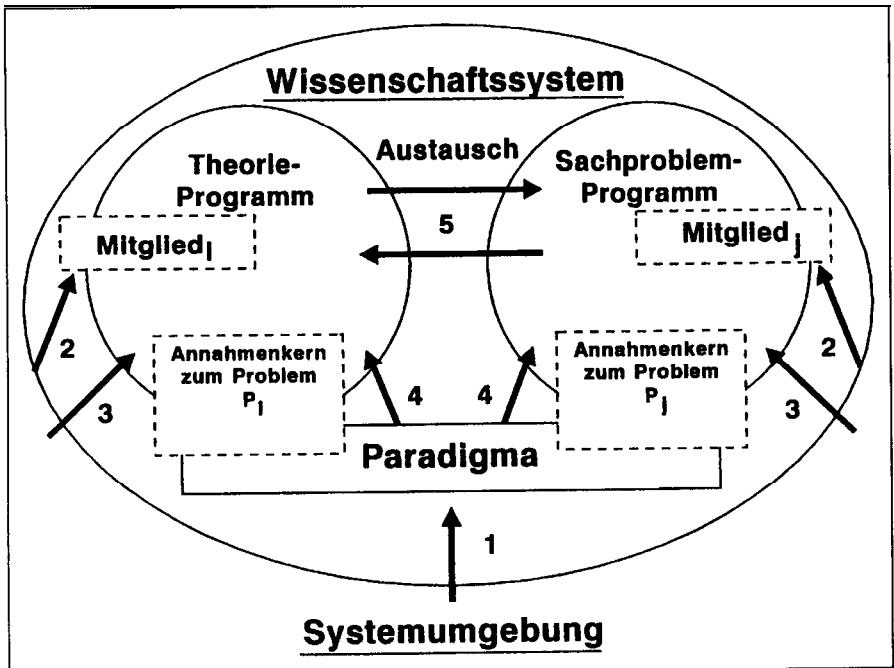


Abb. 2: Wichtige Determinanten des Verlaufs von Problemlösungsprozessen in Forschungsprogrammen (s. Text)

- (1) Generelle Einflüsse auf das Wissenschaftssystem
- (2) Einflüsse auf systemexterne Merkmale von Forschungsprogramm-Mitgliedern
- (3) Lokale Einflüsse auf einzelne Forschungsprogramme
- (4) Wissenschaftsinterne Einflüsse auf einzelne Forschungsprogramme („Paradigmen-Einflüsse“)
- (5) Wechselseitige Forschungsprogramm-Einflüsse (Austausch)

## Literatur

- Agazzi, E. (1987). A system-theoretic approach to the problem of the responsibility of science. *Zeitschrift für allgemeine Wissenschaftstheorie*, **18**, (1-2), 30-49.
- Albert, H. (1982). *Die Wissenschaft und die Fehlbarkeit der Vernunft*. Tübingen: Mohr.
- Alish, L.-M. & Rössner, L. (1983). Operative Modelle als Technologische Theorien. In H. Stachowiak (Hrsg.), *Modelle - Konstruktionen der Wirklichkeit* (S. 147-170). München: Fink.
- Andersson, G. (1988). *Kritik und Wissenschaftsgeschichte*. Tübingen: Mohr.
- Atkinson, J. W. (Ed.) (1958). *Motives in fantasy, action, and society*. Princeton, N.J.: D. van Nostrand Company.

- Balzer, W., Moulines, C.-U. & Sneed, J. (1987). **An architectonic for science. The structuralist program.** Dordrecht: Reidel.
- Bayertz, K. (1987). Wissenschaftsentwicklung als Evolution? **Zeitschrift für allgemeine Wissenschaftstheorie**, **18 (1-2)**, 61-91.
- Breuer, F. (1989). **Wissenschaftstheorie für Psychologen: eine Einführung** (4. Aufl.). Münster: Aschendorff.
- Briskman, L. B. (1972). Is a Kuhnian analysis applicable to psychology? **Science Studies**, **2**, 87-97.
- Brocke, B. (1978). **Technologische Prognosen: Elemente einer Methodologie der angewandten Sozialwissenschaften.** Freiburg: Alber.
- Bunge, M. (1967). **Scientific research.** Vol. I, II. New York: Springer.
- Carrier, M. (1986). Wissenschaftsgeschichte, rationale Rekonstruktion und die Begründung von Methodologien. **Zeitschrift für allgemeine Wissenschaftstheorie**, **17 (2)**, 201-228.
- Chomsky, N. (1957). **Syntactic structures.** The Hague: Mouton.
- Chomsky, N. (1959). Review: Verbal behavior. By B. F. Skinner. **Language**, **35**, 26-58.
- Crane, D. (1967). Gatekeepers of science: Some factors affecting the selection of articles for scientific publication. **American Psychologist**, **2**, 195-201.
- Crott, H. (1979). **Soziale Interaktion und Gruppenprozesse.** Stuttgart: Kohlhammer.
- Dörner, D. (1979). **Problemlösen als Informationsverarbeitung.** Stuttgart: Kohlhammer.
- Ebbinghaus, H. (1885). **über das Gedächtnis.** Leipzig: Duncker & Humblot.
- Essler, W. K. (1970). **Wissenschaftstheorie I. Definition und Reduktion.** Freiburg: Alber.
- Essler, W. K. (1971). **Wissenschaftstheorie II. Theorie und Erfahrung.** Freiburg: Alber.
- Essler, W. K. (1973). **Wissenschaftstheorie III. Wahrscheinlichkeit und Induktion.** Freiburg: Alber.
- Essler, W. K. (1979). **Wissenschaftstheorie IV; Erklärung und Kausalität.** Freiburg: Alber.
- Estes, W. K. (1978). The information-processing approach to cognition: A confluence of metaphors and methods. In W. K. Estes (Ed.), **Human information processing** (= Handbook of Learning and Cognitive Processes, Volume 5), (pp. 1-18). Hillsdale: Erlbaum.
- Farreh, E. (1986). Metaphor and psychology: A reply to Gholson and Barker. **American Psychologist**, **41 (6)**, 719-720.
- Gadenne, V. (1984). **Theorie und Erfahrung in der psychologischen** Forschung. Tübingen: Mohr.
- Garfield, E. (1979). **Citation indexing - its theory and application in science, technology, and humanities.** New York: John Wiley.
- Gaston, J. (Ed.). (1978). **Sociology of science.** San Francisco: Jossey-Bass.
- Gentner D. & Grudin, J. (1985). The evolution of mental metaphors in psychology: A 90-year retrospective. **American Psychologist**, **40 (2)**, 181-192.

- Gholson, B. & Barker, I? (1986). On metaphor in psychology and physics: Reply to Farrell. **American Psychologist**, 41 (6), 720-721.
- Gigerenzer, G. (1988). Woher kommen Theorien über kognitive Prozesse? **Psychologische Rundschau**, 39, 91-100.
- Graumann, C.F. (Hrsg.). (1985). **Psychologie im Nationalsozialismus**. Berlin, Heidelberg: Springer.
- Groeben, N. & Westmeyer, H. (1975). **Kriterien psychologischer Forschung**. München: Juventa.
- Hacker, W. (1978). **Allgemeine Arbeits- und Ingenieurpsychologie**. Bern: Huber.
- Heckhausen, H. (1967). **The anatomy of achievement**. New York: Academic Press.
- Heckhausen, H. (1987). Zur Rolle und Bedeutung wissenschaftlicher Fachgesellschaften. **Beiträge zur Hochschulforschung**, 9 (4), 325-358.
- Heckhausen, H. (1989). **Motivation und Handeln** (2. Aufl.). Berlin, Heidelberg: Springer.
- Hejl, P.M. (1984). Towards a theory of social systems: Self-organization and self-maintenance, self-preference and syn-reference. In H. Ulrich & G. J. B. Probst (Eds.), **Self-organization and management of social systems. Insights, promises, doubts, and questions** (pp. 60-78). Berlin, Heidelberg, New York: Springer.
- Hejl, P.M. (1989). **Self-regulation in social systems: explaining the process of research**. LUMIS Schriften Nr.21. Universität GM Siegen.
- Herrmann, Th. (1976). **Die Psychologie und ihre Forschungsprogramme**. Göttingen: Hogrefe.
- Herrmann, Th. (1979). **Psychologie als Problem**. Stuttgart: Klett-Cotta.
- Herrmann, Th. (1983). Nützliche Fiktionen. Anmerkungen zur Funktion kognitionspsychologischer Theoriebildungen. **Sprache und Kognition**, 2, 85-99.
- Herrmann, Th. (1987). Theoriendynamik in psychologischen Forschungsprogrammen. In J. Brandstädter (Hrsg.), **Struktur und Erfahrung in der psychologischen Forschung** (S.71-89). Berlin: De Gruyter.
- Herrmann, Th. (1991). **Lehrbuch der empirischen Persönlichkeitsforschung** (6. Aufl.). Göttingen: Hogrefe.
- Hilgard, E.R. & Bower, G.H. (1971). **Theorien des Lernens**. Stuttgart: Klett.
- Hörmann, H. (1977). **Psychologie der Sprache**. Berlin, Heidelberg: Springer.
- Hörmann, H. (1981). **Einführung in die Psycholinguistik**. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Hussy, W. (1984). **Denkpsychologie**. (2 Bände) Stuttgart: Kohlhammer.
- Irle, M. (1975). **Lehrbuch der Sozialpsychologie**. Göttingen: Hogrefe.
- Klages, H. (1967). **Rationalität und Spontaneität**. Gütersloh: Bertelsmann.
- Khx, F. (1971). **Information und Verbalten**. Bern: Huber.
- Kordig, C. R. (1971). **The justification of scientific change**. Dordrecht: Reidel.
- Kriz, J., Lück, H.E. & Heidbrink, H. (1987). **Wissenschafts- und Erkenntnistheorie**. Opladen: Leske & Budrich.
- Krohne, H. W. (1976). **Theorien zur Angst**. Stuttgart: Kohlhammer.

- Kuhn, T. S. (1967). **Die Struktur wissenschaftlicher Revolutionen**. Frankfurt a.M.: Suhrkamp.
- Kuhn, T. S. (1970). Logic of discovery or psychology of research? In I. Lakatos & A. Musgrave (Eds.), **Criticism and the growth of knowledge** (pp. 1-23). Cambridge: Cambridge University Press.
- Lachman, R., Lachman, J. W. & Butterfield, E. C. (1979). **Cognitive psychology and information processing**. Hillsdale, N.J.: Erlbaum.
- Lakatos, I. (1974). Falsifikation und Methodologie wissenschaftlicher Forschungsprogramme. In I. Lakatos & A. Musgrave (Hrsg.), **Kritik und Erkenntnisfortschritt** (S. 89-189). Braunschweig: Vieweg.
- Laudan, L. (1977). **Progress and its problems**. Berkeley: University of California Press.
- Leontjew, A. N. (1977). **Tätigkeit, Bewußtsein, Persönlichkeit**. Stuttgart: Klett.
- Levelt, W. J. M. (1991). Die konnektionistische Mode. **Sprache & Kognition, 10**, 61-72.
- Lück, H. E. (1988). Ein paar Anmerkungen und Ergänzungen zu Pieter J. van Striens Aufsatz „De Nederlandse psychologie in het internationale krachtenveld“. **Geschichte der Psychologie, 5 (3)**, 5-13.
- Masterman, M. (1970). The nature of a paradigm. In I. Lakatos & A. Musgrave (Eds.), **Criticism and the growth of knowledge** (pp. 59-89). Cambridge: Cambridge University Press.
- Maturana, H. R. (1982). **Erkennen: die Organisation und Verkörperung von Wirklichkeit**. Braunschweig: Vieweg.
- McClelland, D.C. (Ed.). (1955). **Studies in motivation**. New York: Appleton Century Crofts.
- Mead, G. H. (1973). **Geist, Identität und Gesellschaft aus der Sicht des Sozialbehaviorismus**. Frankfurt a.M.: Suhrkamp.
- Merton, R.K. (1968). Science and democratic social structure. In R.K. Metton (Ed.), **Social theory and social structure** (pp.604-615). New York: The Free Press.
- Merton, R. K. (1973). **The sociology of science. Theoretical and empirical investigations**. Chicago, London: The University of Chicago Press.
- Metraux, A. (1985). Die angewandte Psychologie vor und nach 1933 in Deutschland. In C. F. Graumann (Hrsg.), **Psychologie im Nationalsozialismus (S. 221-262)**. Berlin: Springer.
- Metzger, W. (1954). **Psychologie**. Darmstadt: Steinkopff.
- Meyer, W. U. (1973). **Leistungsmotiv und Ursachenerklärung von Erfolg und Mißerfolg**. Stuttgart: Klett.
- Musgrave, A. E. (1971). Kuhn's second thoughts. **British Journal of the Philosophy of Science, 22, 287-306**.
- Newell, A. & Simon, H. A. (1972). **Human problem solving**. Englewood Cliffs: Prentice Hall.
- Oerter, R. & Montada, L. (1982). **Entwicklungspsychologie**. München: Urban & Schwarzenberg.
- Orlik, P. (1967). Das Dilemma der Faktorenanalyse - Zeichen einer Aufbaukrise in der modernen Psychologie. **Psychologische Beiträge, 10**, 87-98.

- Palermo, D.S. (1971). Is a scientific revolution taking place in psychology? *Science Studies*, **1**, 135-155.
- Popper, K. (1972). *Objective knowledge*. Oxford: University Press.
- Pylyshyn, Z. (1980). Cognitive representations and the process architecture distinction. *The Behavioral and Brain Sciences*, **3**, 154-169.
- Keese, H. W. (1973). Models of memory and models of development. *Human Development*, **16**, 397-416.
- Rice, R. E., Borgmann, C. L. & Reeves, B. (1988). Citation networks of communication journals 1977 - 1985. Cliques and positions, citations made and citations received. *Human Communication Research*, **15** (2), 256-283.
- Rothacker, E. (1952). *Die Schichten der Persönlichkeit* (5.Aufl.). Bonn: Bouvier.
- Rucci, A. J. & Tweney, R. D. (1980). Analysis of variance and the „second discipline“ of scientific psychology: A historical account. *Psychological Bulletin*, **87**, 166-184.
- Sahner, H. (1979). Veröffentlichte empirische Sozialforschung: Eine Kumulation von Artefakten? Eine Analyse von Periodika. *Zeitschrift für Soziologie*, **8** (3), 267-278.
- Sahner, H. (1982). Zur Selektivität von Herausgebern: Eine Input-Output Analyse der „Zeitschrift für Soziologie“. *Zeitschrift für Soziologie*, **11** (1), 82-98.
- Savigny, v., E. (1975). Inwiefern ist die Umgangssprache grundlegend für die Fachsprachen? In J.S. Petöfi, A. Podlech & E. v. Savigny (Hrsg.), *Fachsprache - Umgangssprache* (S. 1-32). Kronberg/Ts.: Scriptor.
- Scheerer, E. (1988). Fifty volumes of Psychological Research/Psychologische Forschung. *Psychological Research*, **50**, 71-82.
- Schneewind, K. A. & Herrmann, Th. (Hrsg.). (1980). *Erziehungstilforschung. Theorien, Methoden und Anwendung der Psychologie elterlichen Erziehungsverhaltens*. Bern: Huber.
- Shapere, D. (1974). Scientific theories and their domains. In E Suppe (Ed.), *The structure of scientific theories* (pp.518-565). Urbana, Ill.: University of Illinois Press.
- Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton Century Crofts.
- Skinner, B.F. (1957). *Verbal behavior*. New York: Appleton Century Crofts.
- Sneed, J. (1971). *The logical structure of mathematical physics*. Dordrecht: Reidel.
- Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Berlin, New York: Springer.
- Stadler, M. & Kruse, P. (1986). Gestalttheorie und Theorie der Selbstorganisation. *Gestalt Theory*, **8**, 75-98.
- Stegmüller, W. (1969). *Wissenschaftliche Erklärung und Begründung*. (Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band I) Berlin, Heidelberg: Springer.
- Stegmüller, W. (1973). *Theorie und Erfahrung. Zweiter Teilband. Theorienstrukturen und Theoriendynamik*. (Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band II). Berlin, Heidelberg: Springer.
- Stegmüller, W. (1986). *Theorie und Erfahrung: Dritter Teilband. Die Entwicklung des neuen Strukturalismus seit 1973*. (Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band II) Berlin, Heidelberg: Springer.

- Stephan, E. (1990). Zur **logischen** Struktur **psychologischer Theorien**. Berlin, Heidelberg: Springer.
- Strien, v. I? (1988). De Nederlandse psychologie in het internationale krachtenveld. **De Psycholoog**, Oktober, 575-585.
- Suppe, F. (Ed.). (1977). **The structure of scientific theories**. Urbana: University of Illinois Press.
- Szagan, G. (1986). **Spruchentwicklung beim Kind** (3.Aufl.). München: PVU.
- Toulmin, S. E. (1972). Human **understanding**. Vol. 1. Princeton: Princeton University Press.
- Ulmann, G. (1968). **Kreativität**. Weinheim: Beltz.
- Weimer, W.B. (Ed.). (1974). **Cognition and the symbolic** processes. Hillsdale, N.J.: Erlbaum.
- Weiner, B. (1972). **Theories of motivation**. Chicago: Markharn.
- Weingart, P. (Hrsg.). (1972). **Wissenschaftssoziologie, Bd. 1. Wissenschaftliche Entwicklung als sozialer Prozeß**. Frankfurt: Athenäum Verlag.
- Westermann, R. (1987). **Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie**. Berlin, Heidelberg: Springer.
- Westmeyer, H. (1977). Verhaltenstherapie: Anwendung von Verhaltenstheorien oder kontrollierte Praxis? In H. Westmeyer & N. Hoffmann (Hrsg.), **Verhaltenstherapie: Grundlegende Texte (S. 187-203)**. Hamburg: Hoffmann & Campe.
- Westmeyer, H. (Ed.). (1989). **Psychological theories from a structuralist point of view**. Berlin, Heidelberg: Springer.

## 7. Kapitel

# Theorien

### **Volker Gadenne**

Es gehört zur Zielsetzung der Erfahrungswissenschaften, ihre Ergebnisse zu *systematisieren*. Eine bestimmte Art der Systematisierung ist bereits die **Erklärung** einzelner beobachteter Tatsachen durch allgemeine Hypothesen. Versucht man nun, mehrere solcher Hypothesen, die sich auf denselben Gegenstandsbereich beziehen, wiederum durch noch grundlegendere Annahmen zu erklären, so gelingt es unter Umständen, ein **deduktives System** von Hypothesen zu konstruieren, eine **Theorie**. Dabei wird angestrebt, möglichst viele empirische Befunde (bestätigte allgemeine Hypothesen) eines bestimmten Forschungsgebietes aus möglichst wenigen **Grundannahmen (Axiome, Postulate)** abzuleiten. Die Grundannahmen sind eine Art hochkomprimierter Zusammenfassung des Wissens auf einem bestimmten Gebiet. Zusätzlich sollte eine gute Theorie aber auch **neuartige Vorhersagen** über bislang nicht untersuchte Sachverhalte ermöglichen. Sie tragen in besonderem Maße dazu bei, die Theorie einem kritischen Test auszusetzen.

Die Forschung in den fortgeschritteneren Erfahrungswissenschaften ist in hohem Maße theoriegeleitet. Das isolierte Sammeln von Daten, deren Klassifikation, die Suche nach relevanten Variablen, das Experimentieren zum Auffinden interessanter Ideen - solche wenig theoriegeleiteten Aktivitäten werden mit dem Erkenntnisfortschritt zunehmend abgelöst durch Tätigkeiten, die auf Theorien bezogen sind: die Ausarbeitung der Theorie, die Ableitung von Folgerungen für verschiedene Ausgangsbedingungen, die empirische Prüfung von Vorhersagen. Beispiele für besonders umfassende Theorien, die, nachdem sie sich durchgesetzt hatten, die Forschungstätigkeit sehr stark zu determinieren begannen, sind Newtons Mechanik oder Darwins Evolutionstheorie. In der Psychologie ist eine solche Dominanz einer einzelnen Theorie nicht so deutlich auszumachen wie in der Physik oder Biologie. Doch wird zweifellos auch hier das Ziel der Theorienkonstruktion im beschriebenen Sinne verfolgt. Zwar gab es immer auch Denkrichtungen, die das **nomologische** Vorgehen, die Suche nach Gesetzen und die Erklärung auf der Grundlage von Gesetzen, in der Psychologie für verfehlt hielten. Betrachtet man aber Systeme von Annahmen



wie diejenigen von Tolman (1932), Lewin (1936), Hull (1943), Festinger (1957), Atkinson (1964), Norman und Rumelhart (1975), Anderson (1983), so wird deutlich, daß die Konstruktion erklärungskräftiger Aussagensysteme einen wesentlichen Anteil psychologischer Forschung ausmacht.

Theorien in diesem Sinne stehen nicht am Anfang der Erforschung eines Gebietes, sondern entstehen, sobald die gewonnenen Einzelerkenntnisse auf diesem Gebiet zur Systematisierung und Erklärung auffordern. Auf der anderen Seite geht „Theorie“ in einem erweiterten Sinne des Wortes aber jeglicher Erfahrung voraus, nämlich in Form von Erwartungen, einzelnen Hypothesen, Alltagsauffassungen, Weltanschauungen, die nachweislich einen Einfluß darauf haben, welche Informationen eine Person beachtet und was sie als Beobachtungsergebnis festhält. Dies ist gemeint, wenn in der Wissenschaftstheorie (in Einklang mit der Sicht der Kognitionspsychologie) betont wird, daß Erfahrung **theorieabhängig** sei.

## **1. Aufbau und Struktur von Theorien**

Der Begriff „Theorie“ wird in sehr unterschiedlicher Weise verwendet. Das Spektrum reicht von formalisierten und axiomatisierten deduktiven Systemen, die vielen Wissenschaftstheoretikern als Ideal vorschweben, bis hin zu Gedankengebäuden, die in Form eines fortlaufend geschriebenen Prosatextes vorliegen, eventuell unterstützt durch Abbildungen, aus dem manchmal nur schwerlich eindeutige Aussagen herausgearbeitet werden können. Es gibt in der Wissenschaftstheorie mehrere Metatheorien darüber, was eine wissenschaftliche Theorie ist, die von der „Standardkonzeption“ (received view) über die „Weltanschauungs-Auffassung“ bis zu „semantischen Ansätzen“ (unter anderem der „Non-Statement-View“) reichen (vgl. Suppe, 1977). Im gegebenen Rahmen ist es nicht möglich, diese Konzeptionen einzeln vorzustellen, da dadurch die besonderen Aspekte psychologischer Theorienbildung kaum mehr zur Sprache kommen könnten. Im folgenden liegt das Schwergewicht auf diesen Besonderheiten, wobei ein Verständnis von „Theorie“ zugrunde gelegt wird, das sich an der herkömmlichen „Aussagenkonzeption“ orientiert und weitgehend dem intuitiven Theorienverständnis in der Psychologie entspricht. („Aussagenkonzeption“ darf nicht mit dem „received view“ gleichgesetzt werden, der für die neopositivistischen Konzeptionen, vor allem diejenige Carnaps, steht. Im folgenden werden weder dessen formalsprachliche Ideale noch eine eindeutige Trennung zwischen theoretischer und Beobachtungssprache akzeptiert.) - Die Charakteristika psychologischer Theorienbildung sollen nun zunächst an einem Beispiel erläutert und anschließend allgemein aufgezeigt werden. Es empfiehlt sich, eine weithin bekannte Theorie heranzuziehen, die dem Ideal einer guten Theorienbildung im obenstehenden

Sinne entspricht und an der möglichst viele Aspekte diskutiert werden können. Die Theorie der Leistungsmotivation von Atkinson (1957, 1964) ist hierzu gut geeignet (besser als die spätere Theorie von Atkinson & Birch, 1970). Sie enthält klar formulierte Annahmen und erlaubt es unter anderem, eine einfache Art der mathematischen Modellbildung zu erläutern. Wie Lewin ist Atkinson ein Vertreter des „galileischen“ Denkens, wonach es geboten ist, viele typische wie atypische Befunde durch dieselben Grundannahmen zu erklären und so einen möglichst umfassenden Gegenstandsbereich einheitlich zu systematisieren (vgl. zur Kritik an dieser Vorgehensweise Bischof, 1981).

## 1.1 Ein Beispiel: Die Theorie der Leistungsmotivation

Gegenstandsbereich dieser Theorie sind Personen, den Anwendungsbereich bilden alle Fälle, in denen eine Person unter Leistungsgesichtspunkten handelt, in denen sie weiß, daß ihr Verhalten mit Hilfe eines Leistungsstandards (durch sie selbst oder durch andere) als Erfolg oder Mißerfolg bewertet wird. Die Theorie wurde im Rahmen eines Forschungsprogramms über Leistungsmotivation entworfen, als dessen Begründer McClelland (1951) gilt. Die empirischen Befunde, die Atkinson erklären wollte, entstammen zu einem großen Teil (jedoch nicht völlig) den Arbeiten der Forschergruppe um McClelland. Aus methodologischer Sicht kann man die Fakten, von denen Atkinson sich leiten ließ, grob in zwei Klassen unterteilen. Es ging ihm zum einen um die „typischen“ Befunde, die Fakten, die mit den damals verfügbaren theoretischen Konzeptionen einigermaßen zufriedenstellend erklärt werden konnten. Hierzu gehören eine Reihe von „Selbstverständlichkeiten“ des Alltagswissens, etwa, daß Personen sich Leistungsziele setzen, daß sie diese mit Ausdauer verfolgen, daß sie sich jedoch in Anspruch und Ausdauer interindividuell unterscheiden usw. Solche Beobachtungen hatten zuvor Murray (1938) dazu veranlaßt, neben anderen Bedürfnissen ein „need for achievement“ zu postulieren. McClelland interessierte sich ebenso wie Murray für das Leistungsmotiv. Er übernahm das von Murray konstruierte Instrument zur Messung von Bedürfnissen, den TAT, und entwickelte ihn zur Messung des Leistungsmotivs weiter. - Zu den typischen Befunden gehören weiterhin eine Reihe von wissenschaftlichen Untersuchungsergebnissen, die ebenfalls auf die Existenz eines Leistungsmotivs im Sinne einer allgemeinen und stabilen Disposition verweisen. Aus der großen Zahl von Befunden, die von Atkinson (1964) angeführt werden, seien folgende genannt: Personen erinnern unerledigte Aufgaben besser als erledigte (Zeigarnik-Effekt) und nehmen unerledigte Aufgaben nach Unterbrechung wieder auf. - Personen heben ihr Anspruchsniveau bei Erfolg an. - Bei experimenteller Herstellung leistungsbezogener Situationen zeigen Personen im TAT leistungsbezogenes Phantasieverhalten. - Personen mit relativ höheren (Leistungsmotiv-) Test-

werten in leistungsbetonten Situationen haben auch die relativ höheren Testwerte in entspannten Situationen. - Personen mit höheren Leistungsmotiv-Testwerten wurden in ihrer Kindheit mehr zu Leistung und Selbständigkeit erzogen. - Personen mit höherem Leistungsmotiv erbringen bessere Leistungen in arithmetischen und sprachlichen Aufgaben.

Von besonderer Bedeutung sind aber die „atypischen“ Befunde, die beim damaligen Stand des Forschungsprogramms „Leistungsmotivation“ als schwer erklärlich erschienen. Beispielsweise wurde auch ein dem Zeigarnik-Effekt entgegengesetzter Effekt gefunden. - Manche Personen senken nach Erfolg ihr Anspruchsniveau. - Personen mit relativ hoher Testangst zeigen nach Mißerfolg einen Leistungsabfall, Personen mit relativ geringer Testangst dagegen einen Leistungsanstieg. - Die Leistung einer Person nimmt mit dem Anreiz (Gewinn) und mit der Gewinnerwartung bei einer Tätigkeit zunächst zu, bei sehr hoher Gewinnerwartung sinkt sie jedoch wieder. - Kinder mit hoher Leistungsmotivation wählen Aufgaben von mittlerer Schwierigkeit häufiger als Kinder mit niedriger Leistungsmotivation.

Die von Atkinson entwickelte Theorie ist stark in Ideen Lewins und Tolmans verwurzelt. Die Grundidee war durch die Theorie des Anspruchsniveaus von Escalona (1940) und Festinger (1942) bereits formuliert worden. Atkinson nimmt an, daß jede Person zwei leistungsbezogene Motive hat: ein Erfolgsmotiv  $M_e$  und eine Furcht vor Mißerfolg  $M_m$ .  $M_e$  ist die Disposition, auf eine erbrachte Leistung stolz zu sein,  $M_m$  die Disposition, wegen eines Mißerfolges Scham zu empfinden. Beide werden als relativ überdauernde und situationsunabhängige Persönlichkeitseigenschaften aufgefaßt, die (unabhängig voneinander) mehr oder weniger stark ausgeprägt sein können. - Vier weitere Variablen sind nach der Theorie für das leistungsorientierte Handeln relevant: die subjektive Wahrscheinlichkeit  $W_e$ , mit der eine Person als Resultat ihres Handelns einen Erfolg erwartet; die subjektive Wahrscheinlichkeit  $W_m$  eines Mißerfolges; der Anreiz  $A_e$ , d.h. der erwartete Stolz im Falle eines Erfolges; und der negative Anreiz  $A_m$ , die erwartete Beschämung im Falle eines Mißerfolges. - Wenn nun eine Person die Möglichkeit hat, durch eine bestimmte Handlung, z.B. durch Wahl einer Aufgabe von einer bestimmten Schwierigkeit, Erfolg zu suchen (mit  $W_e$ ) oder durch das Unterlassen dieser Handlung Mißerfolg zu vermeiden, so haben ihre Tendenz  $T_e$ , Erfolg zu suchen und ihre Tendenz  $T_m$ , Mißerfolg zu vermeiden, folgende Stärken:

$$A_1) \quad T_e = M_e \times W_e \times A_e$$

$$A_2) \quad T_m = M_m \times W_m \times A_m$$

Die resultierende Tendenz, von der die Richtung und Stärke des Verhaltens abhängt, beträgt:

$$A_3) \quad T_r = T_e + T_m$$

Für die beiden subjektiven Wahrscheinlichkeiten gilt:

$$A_4) \quad W_e + W_m = 1$$

Bei Erhöhung der Erfolgswahrscheinlichkeit sinkt der Anreiz eines Erfolges, und mit sinkender Mißerfolgswahrscheinlichkeit steigt der negative Anreiz eines Mißerfolges (Beschämung):

$$A_5) \quad A_e = 1 - W_e$$

$$A_6) \quad A_m = - (1 - W_m)$$

Diese sechs Annahmen werden in der Darstellung von Atkinson (1964) und auch in der Rekonstruktion von Weiner (1972, 1980) als grundlegende Annahmen der Theorie angeführt (vgl. auch Heckhausen, 1980). Da keine aus einer anderen logisch ableitbar ist, jedoch alle zur Ableitung weiterer Annahmen benötigt werden, haben sie den Status von **Axiomen**. - Eine erste Aussage, die Atkinson ableitet, folgt aus  $A_1$  und  $A_5$  durch einfaches Ersetzen:

$$T_1) \quad T_e = M_e \times W_e \times (1 - W_e)$$

Und daraus ergibt sich eine weitere wichtige Folgerung:

$T_1')$  Bei konstantem  $M_e$  ist  $T_e$  dann maximal, wenn  $W_e$  den Wert 0,5 hat (also z.B. bei Aufgaben von mittlerer Schwierigkeit).

Entsprechend kann man aus  $A_2$ ,  $A_4$ ,  $A_5$  und  $A_6$  ableiten:

$$T_2) \quad T_m = M_m \times (1 - W_e) \times (- W_e) \\ = - (M_m \times W_e \times (1 - W_e))$$

Und daraus folgt:

$T_2')$  Bei konstantem  $M_m$  hat der negative Wert  $T_m$  seinen maximalen Betrag bei  $W_e = 0,5$ .

Weiterhin sieht man unmittelbar, daß aus  $A_3$  zusammen mit den beiden gerade abgeleiteten Theoremen  $T_1$  und  $T_2$  folgt:

$$T_3) \quad T_r = M_e \times W_e \times (1 - W_e) - M_m \times W_e \times (1 - W_e) \\ = (M_e - M_m) \times (W_e \times (1 - W_e))$$

Alle diese Aussagen sind, da sie aus den Axiomen folgen, Theoreme. Mit  $T_3$  werden die bekannten und interessanten Folgerungen aus der Theorie ersichtlich: Je mehr das Motiv  $M_e$  relativ zu  $M_m$  überwiegt, desto stärker wird die betreffende Person Aktivitäten mit  $W_e = 0,5$  suchen. Je mehr dagegen das Motiv  $M_m$  relativ zu  $M_e$  überwiegt, desto stärker wird die betreffende Person Aktivitäten mit  $W_e = 0,5$  vermeiden. Bezieht man diesen Zusammenhang auf eine Situation, in der unter verschiedenen Aufgaben eine zur Bearbeitung ausgewählt werden muß, so wird daraus ein bekanntes Theorem:

T(Wahl): Erfolgsmotivierte ( $M_e > M_m$ ) wählen bevorzugt mittelschwere Aufgaben; Mißerfolgsmotivierte ( $M_m > M_e$ ) vermeiden mittelschwere Aufgaben am meisten.

T(Wahl) war Gegenstand einiger wichtiger empirischer Untersuchungen, die zur Prüfung der Theorie durchgeführt wurden (Atkinson & Litwin, 1960). T(Wahl) drückt nach Atkinsons Auffassung genau die beiden letzten zitierten „atypischen“ Befunde aus. Diese Befunde (als allgemeine Hypothesen formuliert) wurden damit zu einem Theorem der neuen Theorie. Allerdings ist zu beachten, daß Atkinsons Befund nicht ohne eine **Zusatzannahme** zu T(Wahl) in Beziehung gesetzt werden kann. Diese Zusatzannahme  $Z_1$  besagt, daß die Leistung einer Person mit dem Wert  $T_r$  ansteigt. Aus den erwähnten Axiomen allein ist nichts über Leistungen selbst ableitbar.

Bei der Herleitung von T(Wahl) wurde eine weitere Annahme ( $Z_2$ ) implizit vorausgesetzt: In einer Situation mit verschiedenen Handlungsmöglichkeiten führt eine Person diejenige Handlung aus, bei der  $T_r$  am stärksten ist. Diese Annahme, die der Lewinschen Tradition entstammt, wird in der Theorie nicht explizit als Axiom angeführt, jedoch bei Folgerungen verwendet.

Aus  $T_3$  ergibt sich auch eine problematische Konsequenz: Personen, bei denen  $M_m$  überwiegt, müßten leistungsbezogene Aktivitäten immer vermeiden, denn nach  $T_3$  ist die resultierende Tendenz bei solchen Personen immer negativ, wenn auch in unterschiedlichem Maße. Daß auch solche Personen gelegentlich leistungsbezogene Tätigkeiten ausüben, führt Atkinson auf extrinsische Motive zurück, die von der Leistungsmotivation im hier verstandenen Sinne zu unterscheiden sind: Erwartung äußerer Belohnungen, Furcht vor Strafe usw.  $A_3$  ist demnach genau genommen so zu interpretieren:  $T_r = T_e + T_m + T_{\text{extr}}$

Es wird also ausdrücklich eingeräumt, daß der wichtigste von der Theorie zu erklärende Sachverhalt, nämlich  $T_r$  (in der Auswirkung als Aufgabenwahl, Anspruchsniveau, Ausdauer, Leistung), auch von Faktoren abhängt, die nicht zum Gegenstand der Theorie gehören und über die aus der Theorie deshalb auch nichts gefolgert werden kann. Eine Anwendung der Theorie erfordert Zusatzannahmen ( $Z_3$ ) darüber, ob in einer bestimmten Anwendungssituation extrinsische Motive vorliegen und in welcher Richtung sie wirken.

Zwei weitere Theoreme betreffen die Ausdauer und das Anspruchsniveau von Personen. Feather (1961) leitete aus der Theorie eine differenzierte Aussage über Ausdauer ab. „Ausdauer“ bedeutet hierbei, daß eine Person trotz fortgesetzten Mißerfolgs bei einer Aufgabe erneute Versuche macht, sie zu lösen.

T(Ausd): Bei relativ leichten Aufgaben ( $W_e > 0,5$ ) sind Erfolgsmotivierte ausdauernder als Mißerfolgsmotivierte; bei relativ schwierigen Aufgaben ( $W_e < 0,5$ ) sind dagegen Mißerfolgsmotivierte ausdauernder als Erfolgsmotivierte.

Zur Ableitung dieses Theorems muß die Zusatzannahme  $Z_4$  gemacht werden, daß bei Mißerfolg die wahrgenommene Schwierigkeit der betreffenden Aufgabe ansteigt. Dadurch rückt diese Aufgabe, wenn sie zunächst als relativ leicht empfunden wurde, in Richtung auf  $W_e = 0,5$ , was ihre Attraktivität für Erfolgsmotivierte vergrößert und für Mißerfolgsmotivierte verringert. Bei schwierigen Aufgaben, die nach Mißerfolg noch schwieriger erscheinen, vergrößert sich umgekehrt die Attraktivität für Mißerfolgsmotivierte.

Moulton (1965) zeigte, daß die Theorie folgende Aussage über das Anspruchsniveau von Personen impliziert:

$T(\text{Anspr})$ : Erfolgsmotivierte verändern ihr Anspruchsniveau in typischer, Mißerfolgsmotivierte in untypischer Weise.

Ein typisches Anspruchsniveau bedeutet hierbei, daß eine Person, die bei einer Aufgabe von  $W_e = 0,5$  Erfolg hat, als nächstes eine objektiv schwierigere Aufgabe wählt (die jetzt nämlich subjektiv in die Nähe von  $W_e = 0,5$  rückt) und im Falle eines Mißerfolges zu einer (objektiv) leichteren Aufgabe übergeht. Mißerfolgsmotivierte wählen manchmal bei Erfolg als nächstes eine leichtere und bei Mißerfolg eine schwierigere Aufgabe, nämlich dann, wenn sie durch diese Wahl Aufgaben erhalten, die möglichst weit von einer  $W_e = 0,5$  entfernt sind. Daß dies als „untypisch“ gilt, drückt natürlich nichts anderes aus, als daß die Theorie in diesem Punkt einen **neuartigen** Sachverhalt vorhersagt, d.h. einen Sachverhalt, der den bisherigen theoretischen Erwartungen nicht entspricht. - Auch dieses Theorem beruht zum einen auf der Aussage, daß Erfolgsmotivierte eine

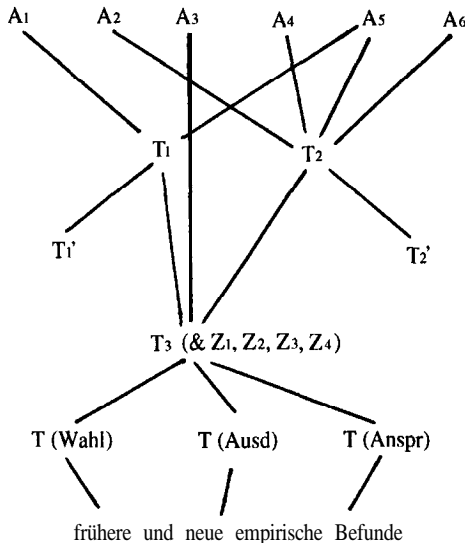


Abb. 1

We = 0,5 suchen, während Mißerfolgsmotivierte dies gerade vermeiden, zum anderen auf der Annahme, daß Erfolg bzw. Mißerfolg die wahrgenommenen Schwierigkeiten von Aufgaben in systematischer Weise verändern (Z4).

Die angeführten Axiome und Theoreme ergeben die in Abbildung 1 dargestellte **deduktive Struktur**.

Die Verbindungslinien in dieser Abbildung stehen für die Beziehung der logischen Ableitbarkeit: Eine Aussage ist aus der Gesamtheit der darüber stehenden Aussagen ableitbar, mit denen sie durch eine Linie verbunden ist.

## 1.2 Theorien als deduktive Systeme von Aussagen

Nach der herkömmlichen Auffassung, die heute manchmal als Aussagenkonzeption (statement view) bezeichnet wird, ist eine erfahrungswissenschaftliche Theorie ein deduktives System von gesetzesartigen Aussagen, die sich auf einen bestimmten Gegenstandsbereich beziehen. Die Elemente des deduktiven Systems zerfallen in zwei Klassen: Die Grundannahmen (Axiome, Postulate) sind im Rahmen des betreffenden Systems nicht ableitbar; die (unendlich vielen) aus ihnen ableitbaren Aussagen heißen Theoreme. Daß eine Aussage ein Axiom ist, besagt nur etwas über ihre Position in einem logischen Ableitungssystem und nichts über Wahrheit oder Evidenz. Axiome werden in den Erfahrungswissenschaften nach dem Gesichtspunkt gewählt, daß möglichst wenige Grundannahmen zusammen einen möglichst hohen Informationsgehalt haben, d.h. daß möglichst viel aus ihnen abgeleitet werden kann, darunter viele bestätigte Hypothesen des betreffenden Forschungsgebietes. Um dies zu leisten, beziehen sich die Axiome einer Theorie gewöhnlich auf abstrakte, der Beobachtung nicht direkt zugängliche Variablen und sind meist alles andere als evident, ganz abgesehen davon, daß ein Erlebnis der Evidenz nichts zu beweisen vermag.

Da die Theoreme aus den Axiomen logisch folgen, haben sie weniger Informationsgehalt als diese. In einem gewissen Sinne ist mit den Axiomen „alles gesagt“. Dennoch kann es sein, daß eine Folgerung aus einer Theorie für den Betrachter einen „subjektiven Überraschungswert“ hat. So sagen etwa im obestehenden Beispiel T1' und T2' viel weniger aus als die Axiome, doch kann man letzteren nicht unmittelbar ansehen, daß sie diese Implikationen haben. Die gefolgerten Aussagen beziehen sich andererseits mehr auf Sachverhalte, die man erfahren und mit empirisch-psychologischen Methoden überprüfen kann. Die empirische Prüfung, Bestätigung oder Entkräftung der Axiome erfolgt indirekt auf dem Weg über die Prüfung der aus ihnen gefolgerten Aussagen.

Nicht alle Aussagen, die in den Erfahrungswissenschaften gewöhnlich als Folgerungen aus Theorien ausgegeben werden, folgen wirklich allein aus den Axiomen im engeren Sinne. Meist werden zahlreiche Zusatzannahmen heran-

gezogen. Einige Annahmen dieser Art wurden im Beispiel aufgezeigt, weitere kommen in Abschnitt 2.3 zur Sprache. Ob man solche Annahmen mit zur Theorie zählt oder als zusätzliches (Hintergrund-) Wissen von der Theorie abtrennt, ist letztlich nur eine Frage des Sprachgebrauchs.

Ist es gerechtfertigt zu sagen, daß psychologische Theorien deduktive Systeme sind? Die Elemente eines deduktiven Systems lassen sich bei jeder psychologischen Theorienbildung nachweisen: Es werden einzelne Annahmen gemacht, die als Hypothesen oder Gesetze bezeichnet werden. Es werden Schlußfolgerungen durchgeführt, indem aus einzelnen oder mehreren Annahmen andere abgeleitet werden. Und zweifellos werden auch immer gewisse Annahmen als die grundlegenden betrachtet. Allerdings sind die meisten psychologischen „Theorien“ bisher nicht so dargestellt worden, daß man leicht erkennen könnte, was Grundannahmen und abgeleitete Annahmen sind und wie die Ableitungsschritte aussehen. Kaum ein Urheber einer bekannten psychologischen Theorie hat sein Werk in systematischer Weise zu axiomatisieren versucht. Hull (1943, 1952) dürfte die einzige Ausnahme sein. Manche Theorien wurden, nachdem sie bekannt geworden waren, von anderen Wissenschaftlern systematisch rekonstruiert. Zum Beispiel haben McCorquodale und Meehl (1954) Tolmans Theorie als ein System mit zwölf Postulaten dargestellt (nach Hulls Vorbild). Eine solche Rekonstruktion ist bei vielen psychologischen Theorien nicht einfach. Allerdings hat Rapaport (1959) gezeigt, daß auch im Falle der psychoanalytischen Theorie, die von Freud bekanntlich in wenig systematischer Form dargestellt worden ist, einige wenige, voneinander trennbare Grundannahmen herausgearbeitet werden können.

Was in der Psychologie als Theorie bezeichnet wird, ist nur annäherungsweise ein axiomatisch-deduktives System, das zudem meist in noch nicht systematisierter Darstellung gegeben ist. Das „annäherungsweise“ soll zum Ausdruck bringen, daß der Versuch der Systematisierung in den meisten Fällen auf gewisse Unklarheiten und Unvollständigkeiten stoßen wird. Manche psychologischen „Theorien“ sind noch besser als „Vorformen« aufzufassen, aus denen man mit mehr oder weniger großer Mühe und Eindeutigkeit ein deduktives System entwickeln könnte.

Eine systematische Rekonstruktion, die Grundannahmen und Ableitungsbeziehungen deutlich macht, hat folgenden Nutzen: Erstens wird auf diese Weise klar, was die Theorie aussagt und was nicht. Dies ist eine wichtige Voraussetzung dafür, die gesamte Theorie zu verstehen und trägt auch dazu bei, sie effektiver anwenden zu können. Wer das Gefüge der vielen Einzelaussagen überblickt, kann in optimaler Weise neue Ableitungen und Anwendungen vornehmen. Auf diese Weise wird zweitens auch die Prüfbarkeit der Theorie verbessert. Sowohl die Prüfung auf Widerspruchsfreiheit als auch empirische Tests setzen klare Ableitungsbeziehungen voraus. Nur dann kann im An-



schluß an empirische Untersuchungen eindeutig beurteilt werden, welche Teile der Theorie von den Ergebnissen betroffen sind.

### 1.3 Theorie und Gesetz

Die Grundannahmen einer Theorie sind **gesetzesartige** Aussagen. Falls sie bereits eine gewisse empirische Bestätigung besitzen, werden sie als **Gesetze** bezeichnet, andernfalls als **Gesetzeshypothesen**. Eine **Gesetzmäßigkeit** ist der objektive Zusammenhang zwischen Variablen, der durch ein Gesetz dargestellt wird. Gesetzesartige Aussagen sind **universelle** Aussagen, sie beziehen sich auf eine offene (potentiell unendliche) Klasse von Fällen. In der Psychologie kann dies eine offene Klasse von Personen sein. Ein psychologisches Gesetz kann aber auch auf nur eine Person und eine offene Klasse von Situationen und Zeitpunkten bezogen sein (ein „Individualgesetz“; vgl. Groeben & Westmeyer, 1975). Eine offene Klasse ist dadurch charakterisiert, daß ihre Elemente nur durch Allgemeinbegriffe und nicht durch Bezugnahme auf einmalige Gegebenheiten definiert werden. Offene Klassen sind z.B.: alle Menschen, alle Frauen, alle Menschen ab dem 60. Lebensjahr. Geschlossene Klassen wären dagegen: alle lebenden Einwohner der USA, alle Studenten europäischer Universitäten. Bei geschlossenen Klassen ist eine Auflistung der Mitglieder im Prinzip vorstellbar, bei offenen nicht.

Nach einer konsequent empiristischen Auffassung behauptet eine Gesetzesaussage nicht mehr, als daß ein bestimmter Zusammenhang universell besteht. Dem steht die Auffassung entgegen, daß eine Gesetzesaussage nicht nur einen Zusammenhang behauptet, der faktisch universell ist, sondern der darüber hinaus (natur-) **notwendig** ist. Ersteres bedeutet nur, daß immer und überall, wenn A faktisch eingetreten ist oder eintreten wird, B folgt. Ein naturgesetzlich notwendiger Zusammenhang impliziert darüber hinaus: Wenn A der Fall gewesen wäre, so wäre auch B eingetreten; dies kann wahr oder falsch sein unabhängig davon, ob A jemals stattfindet. Für Vertreter der Auffassung, daß Theorien mit dem Ziel erstellt werden, Ausschnitte der Realität zu beschreiben (vgl. 3.1), drückt ein Gesetz eine **strukturelle Eigenschaft der Realität** aus (vgl. zur Problematik der Gesetzesartigkeit Armstrong, 1983).

Aus abstrakter Sicht ist eine Gesetzesaussage eine **Wenn-dann-Aussage** über eine (anzugebende) offene Klasse von Individuen (bzw. Situationen, Zeitpunkten). Die Wenn-Komponente gibt den **Anwendungsbereich** der Aussage an. Sie enthält Bedingungen, durch die zusätzlich zur Festlegung des **Gegenstandsbereiches** (Bezugsklasse) noch genauer festgelegt wird, wann man die eigentliche Aussage (die in der Dann-Komponente folgt) als anwendbar betrachtet, z.B., wenn es sich um normale Personen handelt, die ein Leistungsmotiv besitzen und in der gegebenen Handlungssituation einen Gütemaßstab

für verbindlich halten. Solche Bedingungen werden ebenso wie die Bezugs-klasse in der Praxis der Wissenschaften gewöhnlich nicht explizit erwähnt, sondern als selbstverständlich vorausgesetzt. Erwähnt wird lediglich die Dann-Komponente, die aussagt, daß zwischen bestimmten (qualitativen oder quantitativen) Variablen eine bestimmte Beziehung besteht. Manchmal ist diese Dann-Komponente selbst wieder eine Wenn-dann-Aussage, etwa, wenn eine kausale Beziehung zwischen zwei Ereignissen behauptet wird (ein klassisches Beispiel: Frustration führt zu Aggression). Häufig findet man auch Gesetzesaussagen der Form: Je größer (kleiner) X, desto größer (kleiner) Y. Solche **Je-desto-Aussagen** kommen oft auch in Verbindung mit kausalen Wenn-dann-Aussagen vor. (Je größer die Frustration, desto größer die Aggression.) Psychologische Gesetzhypothesen müssen allerdings nicht von dieser einfachen Art sein (wie am Beispiel der Leistungsmotivation schon aufgezeigt wurde), sondern können durchaus auch Angaben über spezielle funktionale Beziehungen (Linearität, Exponentialfunktion) enthalten. Die Aussagenkonzeption legt nichts darüber fest, wie viele Variablen in die Aussage einbezogen werden und wie komplex die angenommenen Relationen zwischen ihnen sein sollen. Insbesondere folgt aus dieser Theorienkonzeption in keiner Weise, daß Gesetzhypothesen sich jeweils auf zwei Variablen bei einseitig ausgerichteter Kausalität beziehen müßten.

Je nachdem wie die behauptete Beziehung zwischen den Variablen beschaffen ist, unterscheidet man verschiedene Arten von Gesetzen. Bei **deterministischen** Gesetzen wird eine strikte Beziehung behauptet, bei **probabilistischen** Gesetzen eine Wahrscheinlichkeitsbeziehung. **Sukzessionsgesetze** besagen etwas über einen zeitlichen Ablauf von Ereignissen, **Koexistenzgesetze** über das gleichzeitige Bestehen von Variablenwerten.

Aus der Struktur von Gesetzesaussagen ergibt sich, daß Theorien nicht Wissen über Einzel-tatsachen, sondern Wissen über Zusammenhänge bereitstellen. Wenn nach einem Gesetz G auf A immer B folgt, so sagt dies nichts darüber aus, wann und wo A und B tatsächlich eintreten, ja nicht einmal, ob sie überhaupt jemals eintreten. G erlaubt es lediglich, für jeden Ort und Zeitpunkt auszuschließen, daß A ohne B geschieht. Insofern haben Naturgesetze logisch den Charakter von „Verboten“. Zusammen mit A erlaubt G allerdings den Schluß auf B.

#### 1.4 Zur Präzisierung der formalen Struktur einer Theorie: Formalisierung und Mathematisierung

Wenn von einer Theorie als einem deduktiven System die Rede ist, so wird oft von der Idealvorstellung ausgegangen, daß die Theorie in einer **formalen Sprache** zu axiomatisieren ist. Wie so etwas durchgeführt werden kann, soll

hier nur angedeutet werden (vgl. Carnap, 1939; Bunge, 1967). Zunächst muß eine formale Sprache gewählt werden, die reichhaltig genug ist, um die Aussagen der Theorie und die Beziehungen zwischen ihnen auszudrücken. Allgemeine Aussagen der Form „Wenn A, dann B“ lassen sich in der Prädikatenlogik erster Ordnung ausdrücken:  $(x)(Ax \rightarrow Bx)$ . Wenn in einer Theorie aber funktionale Beziehungen zwischen quantitativen Variablen vorkommen, wie es bei der Theorie der Leistungsmotivation der Fall ist, benötigt man eine Prädikatenlogik höherer Ordnung und einen Kalkül der Arithmetik. Zum Aufbau der Theorie wäre nun anzugeben, welche **Zeichen** zur Theorie gehören, z.B. T, M, W, A, p, a usw. Logische Zeichen sind durch die gewählte formale Sprache bereits vorgegeben, ebenso **Formregeln**, nach denen aus Zeichen Ausdrücke gebildet werden dürfen, z.B.  $M_e(p) > M_m(p)$ . Vorgegeben sind weiterhin Ableitungsregeln, nach denen aus bestimmten Ausdrücken andere deduzierbar sind, z.B. darf von „P“ und „ $P \rightarrow Q$ “ auf „Q“ geschlossen werden.

Das erste Axiom aus dem obenstehenden Beispiel könnte dann etwa so formuliert werden:

$$(p)(a)(C_{pa} \rightarrow T_e(p,a) = M_e(p) \times W_e(p,a) \times A_e(p,a))$$

Solchen rein **syntaktischen** Gebilden wäre nun noch eine **Interpretation** hinzuzufügen, aus der hervorgeht, daß sich die Variablen p und a auf Personen bzw. Aufgaben beziehen, C auf die Anwendungsbedingungen,  $M_e$ ,  $W_e$ ,  $A_e$  und  $T_e$  auf die entsprechenden Merkmalsdimensionen und  $M_e(p)$ ,  $A_e(p,a)$  usw. auf numerische Werte in diesen Dimensionen. Gemäß dieser Interpretation besagt der formale Ausdruck: Für alle Personen p und alle Aufgaben a gilt: Wenn die Bedingungen C erfüllt sind (zu denen unter anderem gehört, daß p mit der Aufgabe a konfrontiert ist und daß p einen Leistungsstandard für verbindlich hält), dann besteht zwischen den Merkmalswerten  $M_e$ ,  $W_e$ ,  $A_e$  und  $T_e$  die angegebene funktionale Beziehung.

Die logischen Ableitungsbeziehungen zwischen Axiomen und Theoremen bestehen allerdings völlig unabhängig von einer gewählten Interpretation. Sie können auf rein syntaktischer Ebene analysiert werden. Die gesamte Theorie setzt sich also zusammen aus einem formalen Teil, der wegen der Möglichkeit, Ableitungen aus Axiomen durchzuführen, auch **formaler Kalkül** genannt wird, und einer Interpretation, die den formalen Teil zu realen Sachverhalten in Beziehung setzt.

Bei einer Diskussion der Formalisierung psychologischer Theorien empfiehlt es sich, zwei Arten des Gebrauchs formaler Mittel zu unterscheiden: erstens die Verwendung von Teilen der Mathematik (z.B. Algebra, Analysis, Wahrscheinlichkeitstheorie) zur präzisen Formulierung einzelner Gesetzesaussagen; zweitens die Formalisierung einer gesamten Theorie in einer formalen Sprache

(zu der die Mathematik gehören kann). Im folgenden sei die erste Vorgehensweise kurz als **Mathematisierung**, die zweite als Formalisierung bezeichnet. Die erste findet man in der mathematischen Psychologie und in allen Naturwissenschaften. Die dargestellte Theorie der Leistungsmotivation ist ein relativ einfaches Beispiel für die Mathematisierung psychologischer Zusammenhänge. Ziel einer Mathematisierung ist es, differenzierte Zusammenhänge auf möglichst präzise und dennoch kurzgefaßte Weise auszudrücken. Eine entsprechende Darstellung in verbaler Sprache wäre, wenn sie überhaupt möglich ist, wegen ihrer Länge ziemlich unübersichtlich. Durch Mathematisierung kann der Informationsgehalt einer Theorie, die zunächst nur in Wenn-Dann- und Je-Desto-Aussagen vorliegt, erhöht werden. Beispielsweise hat die Aussage „ $Y = a X + b$ , mit  $a > 0$ “ einen höheren Informationsgehalt als „Je größer  $X$  desto größer  $Y$ “. Eine Mathematisierung erlaubt es weiterhin, die logischen Beziehungen zwischen verschiedenen Aussagen mit Hilfe eindeutiger Regeln zu analysieren. In den Sozialwissenschaften werden mathematisierte Theorien meist als „mathematische Modelle“ bezeichnet (vgl. zum Begriff des Modells auch Abschnitt 5; weiterhin Kap. 1, Abschnitt 2.1 dieses Bandes sowie Kap. 5). Versuche mathematischer Modellbildung lohnen sich um so eher, je mehr empirisches Wissen über ein Forschungsgebiet bereits vorliegt; Voraussetzung ist weiterhin eine hinreichend präzise Meßbarkeit der Variablen, um die es geht. Allerdings wäre es nicht zweckmäßig, mit der Theorienbildung bis zur Konstruktion präziser Meßinstrumente zu warten, da letztere wiederum die Lösung inhaltlich-theoretischer Probleme erfordern. Die Entwicklung präziser inhaltlicher Theorien und die Entwicklung von Meßtheorien bzw. Meßinstrumenten sollte in enger Wechselbeziehung zueinander vorgenommen werden.

Die Formalisierung einer Theorie, wie sie oben in sehr vereinfachter Weise beschrieben wurde, ist weder in der Psychologie noch in anderen Erfahrungswissenschaften üblich. Die theoretische Physik beispielsweise macht von dieser Möglichkeit keinen Gebrauch, obwohl sie mehr als jede andere Erfahrungswissenschaft ihre Gesetze mathematisiert. Eine gesamte physikalische Theorie wird üblicherweise dargestellt mit Hilfe mathematischer Ausdrücke und der Umgangssprache (angereichert durch wissenschaftliche Termini). Auf diese Weise wurde oben auch die Theorie der Leistungsmotivation formuliert. Es wäre nun allerdings nicht gerechtfertigt, daraus zu schließen, daß Formalisierungen in der Psychologie keinen Nutzen haben könnten. Je weniger nämlich die einzelnen Annahmen einer Theorie mathematisiert sind, desto mehr vermag eine Formalisierung zur Klarheit beizutragen. Dies ist leicht einzusehen: Wenn die einzelnen Gesetze bereits in Gestalt von mathematischen Formeln und ihre Beziehungen untereinander als mathematische Ableitungen gegeben sind, so ist bereits relativ klar, was die einzelnen Gesetze besagen und was die Struktur der gesamten Theorie ist. Anders verhält es sich, wenn eine Theorie in verbaler Formulierung, ohne klare Trennung zwischen verschiede-

nen Aussagen und ohne eindeutige Ableitungsbeziehungen vorliegt. In diesem Fall zwingt eine Formalisierung dazu, die einzelnen Aussagen zu präzisieren und die deduktive Gesamtstruktur offenzulegen. Hierzu kann manchmal die Prädikatenlogik ausreichend sein. Die auf diese Weise erreichbare Präzisierung trägt allerdings im Unterschied zur Mathematisierung nicht dazu bei, den Informationsgehalt der zuvor verbal formulierten Theorie zu erhöhen. Wenn z. B. verbale Aussagen der Form „Wenn A dann B“ in prädikatenlogische wie „ $(x)(Ax \rightarrow Bx)$ “ überführt werden, so wird dadurch zwar ein Zusammenhang unmißverständlicher ausgedrückt, es wird aber in einem gewissen Sinne nicht mehr gesagt als zuvor. Ersetzt man dagegen eine verbale Aussage wie „Leistungsmotivierte wählen mittlere Risiken“ durch ein mathematisiertes Gesetz wie  $T_r = (M_e - M_m) \times (W_e \times (1 - W_e))$ , so bedeutet dies nicht nur, daß eine präzisere Sprache verwendet wird, sondern daß zusätzlich über den Variablenzusammenhang selbst sehr viel genaueres ausgesagt wird. Die Mathematisierung der Gesetze schließt natürlich eine zusätzliche Formalisierung der gesamten Theorie nicht aus.

Der Wert der Formalisierung gesamter Theorien wird sehr unterschiedlich eingeschätzt. In der Psychologie ist besonders Westmeyer (1973) für den vermehrten Gebrauch formaler Methoden eingetreten. In den letzten Jahren sind einige psychologische Theorien auf der Grundlage des Non-Statement-View formalisiert worden (vgl. 5). Insgesamt ist die Zahl formalisierter Theorien bisher aber relativ klein. - Auch in der Wissenschaftstheorie besteht über den Stellenwert der Formalisierung keine einheitliche Auffassung. Im Neopositivismus wurde ihr größte Bedeutung beigemessen (Carnap, 1939). Dagegen wird z. B. im kritischen Rationalismus das rationale Element der Wissenschaft sehr viel mehr in der Art der Anwendung bzw. Prüfung von Theorien als in deren formaler Darstellung gesehen.

## ***2. Zur Beziehung zwischen Theorie und Empirie***

Es ist charakteristisch für erfahrungswissenschaftliche Theorien, daß sie Termini enthalten, die in der Umgangssprache entweder überhaupt nicht vorkommen (Elektron, Gen, semantisches Netz) oder die innerhalb der Theorie eine Bedeutung erhalten sollen, die sich mit der herkömmlichen nicht deckt (Kraft, Bedürfnis, Erwartung, Schema). Wie kann die Bedeutung solcher theoretischer Begriffe bestimmt werden? Und in welcher Beziehung stehen theoretische Begriffe zu erfahrbaren Sachverhalten? Ein großer Teil der Tätigkeit von Erfahrungswissenschaftlern ist der Zuordnung **von theoretischen zu Beobachtungsbegriffen** gewidmet. Die Zuordnungsproblematik wird dort meist als „Operationalisierung“ oder „Wahl von Indikatoren“ bezeichnet.

## 2.1 Operationale Definition versus Operationalisierung

Es besteht heute eine weitgehende Übereinstimmung darüber, daß die in den Wissenschaften vorgenommenen Zuordnungen von theoretischen zu beobachtbaren Sachverhalten, oft als „Operationalisierung“ bezeichnet, nicht den Charakter von **operationalen Definitionen** im eigentlichen Sinne haben. Theoretische Begriffe sind nicht, wie Bridgman (1927, S. 5) meinte, „synonym mit der entsprechenden Menge von Operationen“, und sie lassen sich nicht vollständig auf Beobachtungsbegriffe (blau, rund, hart, größer als) zurückführen. Die Forderung nach vollständigen operationalen Definitionen stand in Einklang mit dem Sinnkriterium des logischen Empirismus, wonach nur solche Sätze als sinnvoll gelten sollten, über deren Wahrheit oder Falschheit mit Hilfe der Beobachtung definitiv entschieden werden kann. Weder der Operationalismus noch das Sinnkriterium konnten aufrechterhalten werden. Man erkannte bald, daß gerade in den erfolgreichen Erfahrungswissenschaften theoretische Begriffe verwendet werden, die gegenüber den mit ihnen verknüpften Beobachtungsbegriffen einen **Bedeutungsüberschuß** haben. Auch konnte gezeigt werden, daß wichtige Funktionen von Theorien an diesen Bedeutungsüberschuß gebunden sind, der Theorien offen für neue Anwendungen macht. Die verschiedenen Phänomene in einem Forschungsgebiet können nur dann durch wenige Grundannahmen erklärt werden, wenn die Begriffe in diesen Grundannahmen auf vielfältige Weise zu empirischen Variablen in Beziehung gesetzt werden. Würde man (wie Bridgman vorschlug) davon ausgehen, daß z. B. jede operationale Definition von Kraft einen eigenständigen Kraftbegriff einführt, so wurde aus den entsprechenden physikalischen Theorien eine Anhäufung von vielen einzelnen Korrelationen. Theorien wurden dadurch auch ihre Erklärungskraft und ihre heuristische Funktion zur Entdeckung neuer Gesetzmäßigkeiten verlieren (Hempel, 1958). Neue Anwendungen einer Theorie erfordern, daß ein Forscher seine Phantasie betätigt, um Ideen darüber zu entwickeln, in welchen anderen Situationen als den bisher untersuchten sich die von der Theorie postulierten Zusammenhänge oder Prozesse noch manifestieren könnten, z.B. Prozesse der Anregung des Leistungsmotivs und der Risikowahl. Würde man das Leistungsmotiv mit einem Testwert im TAT gleichsetzen, so dürfte sich das Denken des Forschers konsequenterweise nur auf TAT-Situationen richten. Auch könnte mit einer solchen Theorie nichts erklärt werden, da ein Testwert im Gegensatz zu einem Motiv oder einer Erwartung keinen Erklärungsgrund für das Handeln einer Person darstellt.

Die Schwierigkeiten des Operationalismus führten dazu, daß die Neopositivisten ihre Forderung nach vollständiger operativer Definition abschwächten und nur noch eine **partielle Interpretation** der theoretischen Begriffe durch Beobachtungsbegriffe verlangten. Beide sind durch **Zuordnungsregeln** (auch: interpretative Sätze, Brückenprinzipien) miteinander zu verknüpfen, mit deren

Hilfe aus den Grundannahmen einer Theorie Aussagen über Beobachtbares abgeleitet werden können. Nicht jeder theoretische Begriff muß in einer Zuordnungsregel vorkommen, da er durch seine Verknüpfung mit anderen theoretischen Begriffen indirekt zur Erfahrung in Beziehung stehen kann. Die theoretischen Begriffe sind durch die Zuordnungsregeln in ihrer Bedeutung nur partiell bestimmt und stets offen für eine nähere Bestimmung durch neue Zuordnungsregeln.

## 2.2 Zur Abgrenzung zwischen theoretischen und Beobachtungsbegriffen

Als eine Schwierigkeit der zuletzt wiedergegebenen Auffassung stellte sich heraus, daß es zwischen theoretischen Begriffen und Beobachtungsbegriffen (und entsprechend zwischen theoretischen postulierten und beobachtbaren Sachverhalten) keine eindeutige Grenze gibt. Dies ist ein großes Problem für den Empirismus, denn er benötigt eine eindeutige Abgrenzung, nach der das Beobachtbare als das unproblematische, gesicherte Fundament der Erkenntnis erwiesen werden kann, das theoretischen Aussagen ihre Rechtfertigung verleiht. Als beobachtbar sollte gelten, was in der Wahrnehmung unmittelbar gegeben ist und noch keiner Interpretation oder Schlußfolgerung unterzogen wurde. Nun werden zwar Wahrnehmungsinhalte in der Tat als unmittelbare „Gegebenheiten“ erlebt. Wie jedoch immer wieder aufgezeigt wurde, sind diese Wahrnehmungserlebnisse nur das Endprodukt eines Prozesses, der durch zahlreiche Faktoren beeinflusst wird. Von seiten der Philosophie wurde der Einfluß der verfügbaren Denk- und Sprachkategorien betont. Die Wahrnehmungsforschung demonstrierte experimentell den Einfluß von Erfahrungen bzw. Erwartungen auf die Wahrnehmung. Aus der in der heutigen Psychologie überwiegend vertretenen Sicht ist Wahrnehmung das Ergebnis eines komplexen Prozesses der Informationsverarbeitung, dessen Verlauf nur zum Teil durch die sensorisch empfangenen Informationen und zusätzlich durch bereits gespeicherte Informationen (Erfahrung, Wissen) sowie durch die Verarbeitungsprozeduren selbst bestimmt wird. Die resultierende und erlebte Wahrnehmung ist bereits eine „Interpretation von Zeichen“, die über die mit den Sinnen empfangene Information hinausgeht. - Wenn Wahrnehmungen dann sprachlich dargestellt werden, so wird nochmals über das Gegebene hinausgegangen, weil die zur Beschreibung benutzten Begriffe meist mehr ausdrücken als Inhalte von Wahrnehmungserlebnissen des Beobachters. In diesem Sinne sind Beobachtungen niemals „reine Beobachtungen“, sondern erfolgen, wie Popper sagt, „im Lichte von Theorien“ (wobei „Theorie“ hier allerdings nicht allein auf Aussagensysteme bezogen ist, sondern auch Einstellungen, Erwartungen und Begriffskategorien meint).

Ein weiterer Aspekt des Abgrenzungsproblems besteht darin, daß es ziemlich willkürlich wäre, zur Beobachtung nur die menschlichen Sinnesorgane zuzulassen, da diese keineswegs zuverlässiger sind als andere Instrumente und durchaus auch nicht besser verstanden werden. Sind z.B. folgende Sachverhalte beobachtbar: ein dünner Draht, der nicht mit bloßem Auge, jedoch mit einer Brille deutlich sichtbar ist; das Verhalten einer Person, gesehen durch eine Einwegscheibe? Wird die Frage bejaht, so ist es unvermeidlich, auch Viren, Gene und Moleküle als beobachtbar zu bezeichnen, denn die jeweils erforderlichen Instrumente sind ähnlich zuverlässig und werden mindestens so gut verstanden wie das Funktionieren des Auges selbst. Und die entsprechenden Beobachtungen können heute weit zuverlässiger durchgeführt werden als frühere Beobachtungen der Gestirne mit bloßem Auge, deren Ungenauigkeit erhebliche Fehlschlüsse über Theorien zur Folge hatte (vgl. Chalmers, 1986, Kap.3).

In gewisser Hinsicht sind also der Beobachtbarkeit keine vorherbestimmten Grenzen gesetzt. Auf der anderen Seite sind alle Beobachtungen nur in dem Maße als wissenschaftliche Ergebnisse begründet, in dem gewisse Voraussetzungen zutreffend sind, die von den impliziten Annahmen des Beobachters bis zu den Theorien über das benutzte Meßinstrument reichen. In diesem Sinne sind Beobachtungsbeschreibungen **theorieabhängig** („theoriegeladen“), und dies hat zur Folge, daß eine prinzipielle Trennung zwischen beobachtbaren und theoretischen Sachverhalten nicht möglich ist.

Indessen läßt sich nicht bestreiten, daß in einem gegebenen Kontext bestimmte Sachverhalte leichter und zuverlässiger beobachtbar sind als andere. Die Zeigerstellung auf einem Ziffernblatt und die Markierung in einem Fragebogen sind ohne Zweifel besser beobachtbar als ein Gedanke einer anderen Person oder eine Persönlichkeitseigenschaft. Daher werden im Rahmen von empirischen Untersuchungen stets die besser beobachtbaren Variablen als die zu beobachtenden ausgewählt, um dadurch Aussagen über andere Variablen zu testen. Dies ist vernünftig, erfordert aber weder eine Zwei-Sprachen-Theorie (theoretische versus Beobachtungssprache) noch die erkenntnistheoretische Annahme, daß die Beobachtungssätze ein gesichertes, der Kritik prinzipiell entzogenes Fundament darstellen. Die Unterscheidung zwischen beobachtbaren und theoretischen Sachverhalten ist an den jeweiligen Kontext gebunden. Was als beobachtbar gilt, hängt von pragmatischen Umständen ab, z.B. von den Fähigkeiten und der Schulung des Beobachters sowie den technischen Möglichkeiten.

### 2.3 Zum Status von Zuordnungsannahmen

Zusammenhänge zwischen gut beobachtbaren und weniger gut beobachtbaren Sachverhalten werden in den Erfahrungswissenschaften zwar angenommen, doch selten explizit als Zuordnungsregeln oder interpretative Sätze ausgespro-



chen. In der Psychologie sind solche Zuordnungsannahmen in den Beschreibungen der „Operationalisierung“ von Variablen enthalten bzw. in der gewählten Methode zur Beobachtung, Befragung oder Messung. Etwas vereinfacht kann man psychologische Zuordnungsannahmen in der Form darstellen: Wenn die Untersuchungs- oder Testbedingungen U erfüllt sind, gilt: Der (weniger gut beobachtbare) Zustand T liegt genau dann vor, wenn der (gut beobachtbare) Zustand B vorliegt. Gemäß einer solchen Annahme kann man versuchen, T durch experimentelle Manipulation von B eintreten zu lassen oder T anhand von B zu diagnostizieren. B kann eine Reaktionszeit sein, das Drücken einer bestimmten Taste, eine sprachliche Äußerung, ein Testwert; T steht für Sachverhalte wie ein kognitiver Prozeß, eine Emotion, eine Einstellung oder Persönlichkeitseigenschaft. Jede einzelne Zuordnung oder „Operationalisierung“ ist dabei in Wirklichkeit ein oft recht komplexes Gefüge von mehreren Hypothesen und semantischen Beziehungen. So beruht z.B. die „Zuordnung“ zwischen einer Ausprägung des Leistungsmotivs und einem Testwert im TAT auf einer Theorie darüber, wie Motive die Phantasie einer Person beeinflussen. Sie ist also weit davon entfernt, eine bloße sprachliche Konvention oder gar Definition zu sein. - In Experimenten, in denen etwa versucht wurde, Frustration oder kognitive Dissonanz herzustellen, setzen die Untersucher voraus, daß die Versuchspersonen bestimmte Interessen, Ziele oder Normen haben; nur unter diesen Voraussetzungen ist nämlich zu erwarten, daß die experimentelle Manipulation tatsächlich eine Frustration oder kognitive Dissonanz zur Folge hat. - Bei Operationalisierungen kognitiver Prozesse anhand der Variable „Reaktionszeit“ wird vorausgesetzt, daß die Theorie der verwendeten physikalischen Zeitmessung zutreffend ist, daß die Meßinstrumente in der vorgesehenen Weise funktionieren, daß die Versuchspersonen die Instruktion verstanden haben und daß sie ausreichend motiviert sind, sie zu befolgen. Nur dann kann eine systematische Beziehung zwischen der physikalisch gemessenen Zeitdauer und der Zeitdauer des hypothetischen kognitiven Prozesses erwartet werden. Solche Voraussetzungen zählen zu den **Hilfshypothesen** einer Untersuchung bzw. **zu dem Hintergrundwissen**, das man zum Zweck dieser Untersuchung unhinterfragt akzeptiert (in einem anderen Kontext jedoch hinterfragen kann). Der Vorgang der Operationalisierung besteht wesentlich in der Wahl einer Reihe von Hilfhypothesen. Es kann nicht gefordert werden, das Hintergrundwissen jeder einzelnen Untersuchung vollständig zu explizieren; dies wäre ein unnötiger und letztlich gar nicht durchführbarer Aufwand. Das Hintergrundwissen umfaßt zum Teil Annahmen, die im jeweiligen Fachgebiet allgemein bekannt sind und in anderen Zusammenhängen problematisiert und erforscht werden, etwa Annahmen über verwendete Skalen oder psychometrische Tests. Wünschenswert ist allerdings eine bessere Explikation derjenigen Hilfsannahmen, die für eine empirische Untersuchung (insbesondere eine neuartige) spezifisch sind und manchmal noch nicht entdeckte Irrtümer enthalten (vgl. Gadenne, 1984, S. 31ff.).

Operationalisierungen können in sehr unterschiedlichem Maße begründet sein. Eine große Zahl von psychologischen Untersuchungen wurde mit ad hoc konstruierten Schätzskalen durchgeführt, ohne daß eine Kontrolle der Reliabilität und Validität der Messung erfolgte, geschweige denn eine Überprüfung des Skalenniveaus. Es dürfte weitaus mehr Untersuchungen dieser Art geben, als Operationalisierungen auf der Grundlage eines Meßmodells, dessen Test eine Beantwortung des Repräsentations- und Eindeutigkeitsproblems ermöglicht (vgl. dazu Kap. 13). Die zur Illustration herangezogene Forschung über Leistungsmotivation liegt zwischen diesen Extremen. McClelland und Atkinson haben zur Validierung ihrer Meßinstrumente ein Verfahren gewählt, das weitgehend einer **Konstruktvalidierung** im Sinne von Cronbach und Meehl (1955) entspricht. Verglichen mit dem üblichen Vorgehen haben sie sehr viel Aufwand in die Erstellung guter Operationalisierungen investiert. Nach den Kriterien der axiomatischen Meßtheorie sind dennoch die Meßinstrumente für die beiden Motive  $M_e$  und  $M_m$  keineswegs zufriedenstellend. Stephan (1990) argumentiert, daß ein meßtheoretisch adäquates Vorgehen im Falle von Atkinsons Theorie erfordern würde, eine polynomial verbundene Metrisierung zu entwickeln. - Damit soll allerdings nicht die gelegentlich vertretene Auffassung übernommen werden, daß ohne axiomatisch fundierte Messung überhaupt kein empirischer Test einer Theorie möglich sei. Nur ist eben ein Test einer Theorie um so eindeutiger möglich, je mehr die Hilfsannahmen als geprüft und bestätigt gelten können (vgl. Kap.9, 2.1).

### **3. Zur Beziehung zwischen Theorie und Realität**

#### 3.1 Realismus und Instrumentalismus

Die **Darstellung** von Sachverhalten durch Aussagen wird im allgemeinen als eine wichtige Funktion der menschlichen Sprache angesehen. Zu der Frage, ob auch theoretische Aussagen dazu geeignet sind, Tatsachen darzustellen, gibt es jedoch keine einheitliche Auffassung. Der **Realismus** sieht es als eine wichtige Funktion von Theorien an, über Ausschnitte der Wirklichkeit hypothetische Aussagen zu machen, während der **Instrumentalismus** Theorien als nützliche Instrumente betrachtet, die jedoch nicht die Funktion haben, Erkenntnisse über die Wirklichkeit auszudrücken. Nach der realistischen Auffassung dienen Theorien dem Zweck, Erkenntnisse über die Zusammenhänge und Prozesse auszudrücken, die den beobachteten Phänomenen zugrunde liegen. Theorien sollen diese Zusammenhänge möglichst **zutreffend darstellen**, sie sollen mit den Tatsachen möglichst gut **übereinstimmen** und in diesem Sinne der **Wahrheit** möglichst nahekommen. Dabei gilt es als selbstverständlich, daß Darstellungen der Wirklichkeit grundsätzlich **selektiv** sind: Sie beziehen sich immer nur auf Teilbereiche der Wirklichkeit und auf einzelne

Aspekte dieser Bereiche, die zudem oft noch in vereinfachter Weise wiedergegeben werden. Der Realismus umfaßt zusätzlich die Annahme, daß Erkenntnis der Wirklichkeit möglich ist. Die meisten heutigen Versionen des Realismus meinen hierbei nicht sichere Erkenntnis, sondern **bewährtes** (aber nichtsdestoweniger fehlbares und deshalb immer vorläufiges) Wissen: Wenn man Theorien kritisch überprüft und korrigiert, so hat man Grund zu der Erwartung, auf längere Sicht zu Theorien zu gelangen, die mit den Tatsachen besser übereinstimmen als ihre Vorgänger. Der Realismus darf also nicht mit der Annahme verwechselt werden, „im Besitz der Wahrheit“ zu sein; die meisten aktuellen Versionen des Realismus vertreten vielmehr das Gegenteil dieser Annahme: Wahrheit dient als **regulative Idee**, ohne daß jemals Gewißheit darüber erlangt werden kann, ob eine einzelne Aussage wahr ist.

Der Realismus ist zentraler Bestandteil des **kritischen Rationalismus** (Popper, 1983; Albert, 1987), wird jedoch auch unabhängig von diesem vertreten. In der Psychologie wurde er als „kritischer Realismus“ bereits von Oswald Külpe (1912, 1920, 1923) ausführlich dargelegt (vgl. auch Gadenne, 1984). Eine realistische Interpretation psychologischer Aussagen, die z.B. von der Anregung eines Motivs oder von einem Wahrnehmungsvorgang handeln, bedeutet, daß man mit solchen Aussagen auf ein wirkliches Geschehen Bezug nehmen will. Eine solche Absicht wird von Wissenschaftlern meist nicht explizit mit Hilfe von Begriffen wie „Wahrheit“ und „Wirklichkeit“ ausgesprochen, oft aber auf andere Weise, etwa durch die erklärte Absicht, herausfinden zu wollen, wie bestimmte Vorgänge (Entscheidungsfindung, Erinnerung) **tatsächlich** beschaffen sind. Viele Argumente, die in der Forschungspraxis ständig vorkommen, ergeben nur auf der Grundlage eines Realismus einen Sinn, zumindest dann, wenn man sie beim Wort nimmt: Ist es in diesem Experiment wirklich gelungen, das Leistungsmotiv der Vpn anzuregen? Vermag der TAT einen solchen Motivationszustand wirklich zu erfassen? - Wer Operationalisierungsfragen (und auch das Problem der Konstruktvalidität) als Fragen nach der Beziehung zwischen einem gut beobachtbaren (manifesten) Indikator und einem weniger gut beobachtbaren (latenten) Sachverhalt auffaßt, setzt implizit auch letzteren als real voraus, da nur etwas Reales sich in seinen Wirkungen manifestieren kann (gemäß der Aussage Lewins, daß dasjenige real ist, das Effekte hervorruft). Ob man selbst zu einer realistischen Sichtweise neigt, kann man durch einfache Fragen testen: Wer z.B. als Experimentator überlegt, ob die Vpn die Instruktion verstanden haben und es dabei als selbstverständlich betrachtet, daß es das Verstanden-Haben nicht nur im Denken des Forschers, sondern im Denken der Vp wirklich gibt, hat in bezug auf diesen Sachverhalt eine realistische Position.

McCorquodale und Meehl (1948) bezeichneten solche theoretischen Begriffe, die sich auf reale, nicht beobachtbare Sachverhalte beziehen sollen und daher nicht operational definierbar sind, als **hypothetische Konstrukte** (z.B. „bio-

physical trait“ bei Allport, „anxiety“ bei Dollard und Miller), im Unterschied **zu** den **intervenierenden Variablen**, die nur die Funktion einer abgekürzten Beschreibung einer Vielzahl von Reiz-Reaktions-Beziehungen erfüllen und vollständig durch letztere definierbar sind (z. B. „reflex reserve“ bei Skinner). Die allgemeine Verwendung der Ausdrücke „hypothetisches Konstrukt“ und „intervenierende Variable“ in der Psychologie orientiert sich jedoch nicht immer an dem Vorschlag von McCorquodale und Meehl; oft werden beide synonym mit „theoretischer Begriff“ gebraucht.

Der Instrumentalismus bestreitet nun, daß die Darstellung realer Sachverhalte ein mögliches und sinnvolles Ziel der Theorienbildung sei. Er betont stattdessen die **nützlichen Funktionen** von Theorien, zu denen insbesondere die **Vorhersage beobachtbarer Sachverhalte** gehört. Theorien sind nicht wahr oder falsch, sondern mehr oder weniger nützlich. Da allerdings der Realismus Theorien auch als nützliche Instrumente ansieht, liegt der Unterschied im zuerst genannten Punkt: Nach instrumentalistischer Ansicht sind Theorien Instrumente zu anderen Zwecken als der Darstellung realer Sachverhalte. (Schlick, 1938; Toulmin, 1953). Eine spezielle Form des Instrumentalismus, die in jüngster Zeit viel Beachtung gefunden hat, ist der von van Fraassen (1980) entwickelte „konstruktive Empirismus“. Danach werden Theorien zwar als darstellende Aussagen entworfen, das Ziel der empirischen Forschung richtet sich dann aber nicht darauf, die Richtigkeit der Darstellung zu prüfen, sondern nur den Vorhersageerfolg (empirische Adäquatheit). In der Psychologie vertritt Herrmann (1979, 1983) die Auffassung, daß Theorien im Rahmen von Forschungsprogrammen als Problemlösungsmittel dienen und nach ihrer diesbezüglichen **Tauglichkeit** beurteilt werden. Theorien sind eher als „nützliche Fiktionen“ anzusehen denn als Versuche, die Realität darzustellen oder sich der Wahrheit anzunähern.

Der Instrumentalismus wendet sich im allgemeinen nicht gegen die Idee der Darstellung und Wahrheit schlechthin: Beide sind angemessen im Bereich des Beobachtbaren. Beobachtungsaussagen können wahr oder falsch sein; es macht nur keinen Sinn, den Wahrheitsbegriff auf theoretische Aussagen anzuwenden, die keine erfahrbaren Sachverhalte zum Gegenstand haben. (Wenn auch die Beobachtungsaussagen keine Darstellungsfunktion hätten, so wäre kaum einsichtig zu machen, wozu es nützlich sein sollte, sie aus Theorien ableiten zu können.) Theorien selbst können mehr oder weniger **empirisch adäquat** sein, je nachdem, wie zutreffend die aus ihnen ableitbaren empirischen Vorhersagen sind. Es ist nicht zulässig, empirische Adäquatheit als methodischen Anhaltspunkt für Wahrheit zu verwenden. Die empirische Adäquatheit sagt nichts darüber aus, wie das Innere der Black Box beschaffen ist. Zu jedem beobachtbaren Verhalten eines Systems gibt es Unbegrenzt viele mögliche Theorien über die Mechanismen, die es erzeugt haben könnten. Theorien sind durch die Empirie grundsätzlich „unterbestimmt“, Über die Existenz von Sa-

chen wie Moleküle, Gene, Motive, kognitive Prozesse macht der Instrumentalist deshalb keine Annahme. Manchmal findet sich sogar die weitergehende These, daß es keine theoretischen Sachverhalte gibt. So argumentiert z.B. Heckhausen (1980, S.28), daß es Motive „in Wirklichkeit“ nicht gebe, daß sie „keine Tatsachen“ seien, weil sie „nicht unmittelbar beobachtet“ werden können. Für die heutigen instrumentalistischen Positionen ist es jedoch eher charakteristisch, offen zu lassen, ob theoretische Begriffe reale Gegenstücke haben.

Wie Musgrave (1981) gezeigt hat, kann ein Instrumentalismus „lokal“ vertreten werden. Man kann allgemein die Zielsetzung haben, Theorien in Richtung auf Übereinstimmung mit den Tatsachen hin zu korrigieren und dennoch eine bestimmte Theorie instrumentalistisch interpretieren. Dies ist dann angemessen, wenn diese Theorie eine gewisse empirische Adäquatheit aufweist, gleichzeitig jedoch Anhaltspunkte dafür vorliegen, daß sie die tatsächlichen Vorgänge in grundlegenden Punkten falsch darstellt. Als Beispiel sei eine Kritik an einer Klasse von kognitiven Theorien angeführt. Diese Theorien enthalten die Annahme, daß kognitive Operationen in einem Arbeitsspeicher ausgeführt werden, während gleichzeitig im übrigen Teil des kognitiven Systems keine Aktivität stattfindet. Gegen solche theoretischen Ansätze ist eingewendet worden, daß sie insofern einen Irrweg beschreiten wurden, als nach heutigem Wissen über das Zentralnervensystem jede Nervenzelle als ein ständig aktiver Prozessor interpretiert werden kann (vgl. dazu Jackendoff, 1987, S. 29ff.). Wer diese Kritik teilt, zugleich aber der Meinung ist, daß entsprechende Theorien in gewissem Maß empirisch adäquat sind, kann sie in instrumentalistischer Deutung weiterhin verwenden.

Bei der Beurteilung der Frage, ob gewisse theoretische Entitäten fiktiver Natur sind, muß allerdings konsequent zwischen Theorien und den anschaulichen Modellen unterschieden werden, wie man sie zur Illustration von Theorien in allen Erfahrungswissenschaften verwendet (z. B. Moleküle als Billardkugeln). Wer sich etwa Gedächtnisspeicher als rechteckige Kästen vorstellt, kann zu dem voreiligen Ergebnis kommen, daß es solche Speicher „in den Köpfen von Personen nicht wirklich gibt.“ Wenn man sich dagegen klar macht, daß eine Aussage von der Art „Eine Information befindet sich im Arbeitsspeicher“ als eine Aussage über einen Zustand oder Prozeß mit bestimmten Eigenschaften (Repräsentationsformat, Zeitdauer) und Beziehungen zu anderen Zuständen verstanden werden kann, so ist es eher vorstellbar, daß auf diese Weise eine zutreffende Beschreibung von abstrakten Eigenschaften eines realen Systems geliefert werden kann.

Ein zentraler Punkt in der Kontroverse zwischen Realismus und Instrumentalismus ist die schon erwähnte Frage nach der Möglichkeit von Erkenntnis. Die Erkenntnistheorie und die Wissenschaftsgeschichte lehren die grundsätzliche **Fehlbarkeit** aller Erkenntnisversuche. Dies hat viele Wissenschaftler zu

einer skeptischen oder relativistischen Einstellung gegenüber der Möglichkeit der Erkenntnis geführt (Kuhn, 1962; Feyerabend, 1976). Im modernen Instrumentalismus findet diese Skepsis ihren Ausdruck in der Aufgabe des Erkenntnisanspruchs in bezug auf die nicht beobachtbare Wirklichkeit. In extremer Weise wird die Zurücknahme eines Erkenntnisanspruchs neuerdings vom **radikalen Konstruktivismus** vorgenommen (vgl. Glasersfeld, 1987), der sich auf Kant beruft und auf Erkenntnisse der Neurobiologie zu stützen versucht. Als **autopoietische Systeme** (Systeme, die so organisiert sind, daß sie in einem zyklischen Prozeß ihre Bestandteile und deren Organisation reproduzieren; vgl. Maturana, 1985; Roth, 1986) seien wir „selbstreferentiell“ und „operativ geschlossen“: Jeder Gehirnzustand sei das Resultat einer Interaktion früherer Gehirnzustände; die Reaktion der Sinnesorgane lasse keinen Schluß auf die Beschaffenheit ihrer Ursache zu. Die Vorstellung von einer äußeren Realität sei eine vom Gehirn konstruierte, überlebensdienliche Fiktion. Theorien sind als nützliche Instrumente aufzufassen. - Gegen diese Art der Argumentation kann freilich eingewendet werden, daß sie ihre eigenen Ausgangsannahmen widerlegt: Wenn wir nichts über die Realität wissen, dann auch nicht, daß wir autopoietische Systeme sind, ein Gehirn haben usw. (Wendel, 1990). Auch bereits die vom radikalen Konstruktivismus vorgenommene Interpretation neurobiologischer Erkenntnisse ist angezweifelt worden (Nüse, Groeben, Freitag & Schreier, 1991). Wie viele skeptische Richtungen unterscheidet auch der radikale Konstruktivismus nicht konsequent zwischen sicherer Erkenntnis, die er zu Recht als nicht erreichbar erklärt, und Erkenntnis überhaupt. Die Vertreter des Realismus nehmen die Fehlbarkeitsannahme, den **Fallibilismus**, zwar zum Anlaß, das klassische Ideal der Erkenntnisgewißheit aufzugeben (Albert, 1968, 1987). Dennoch wird **es** als ein **rationales** Unternehmen angesehen, durch kritische Prüfung von Theorien sich dem Ziel einer zutreffenden Darstellung von Tatsachen anzunähern. Empirische Adäquatheit dient hierbei als Hilfsmittel zur Orientierung bei der Verbesserung von Theorien.

Bereits bei den Formulierungen des Realismus bzw. Instrumentalismus setzt die Kritik der jeweiligen Gegenseite an. Was bedeutet es genau, sich mit aufeinanderfolgenden Theorien der Wahrheit „anzunähern“? Es kann im allgemeinen nicht bedeuten, daß frühere Theorien falsch und heutige Theorien wahr sind, denn es muß damit gerechnet werden, daß nahezu alle heutigen Theorien mit hohem Informationsgehalt auch irgendwelche falschen Konsequenzen enthalten. Es ist versucht worden, die Idee der besseren Übereinstimmung mit den Tatsachen mit Hilfe des Begriffs der **Wahrheitsnähe** zu explizieren, doch sind die bisherigen Konzeptionen der Wahrheitsnähe mit schwierigen Problemen behaftet (vgl. die Beiträge in Kuipers, 1987). Musgrave (1979) argumentiert, daß es in einem gewissen Sinne heute doch mehr wahre Theorien gibt als früher, nämlich insofern, als spätere Theorien korrigierte und wahre Versionen ihrer Vorgänger enthalten.

Der Instrumentalismus wiederum benötigt, um überhaupt formuliert werden zu können, eine eindeutige Trennung zwischen dem Bereich des Beobachtbaren und des nicht Beobachtbaren. Die Trennung muß eindeutig sein, da auf ihr die Begründung aufgebaut werden soll, daß im zweiten Bereich eine radikale Zurücknahme des Erkenntnisanspruchs notwendig ist. Warum aber soll z.B. eine Fingerbewegung als real gelten, nicht jedoch eine Nervenzelle, die an der Steuerung der Bewegung beteiligt ist? Wie schon ausgeführt wurde, konnte bisher kein zufriedenstellendes Kriterium zur Abgrenzung des Beobachtbaren gefunden werden (vgl. dazu Musgrave, 1982).

In der Psychologie stößt die Formulierung des Instrumentalismus auf ein besonderes Problem: Der Instrumentalist enthält sich einer Existenzannahme in bezug auf theoretische Entitäten wie Atome, Gene usw. Doch nimmt er immerhin an, daß die Wissenschaftler, die über solche Sachverhalte reden, entsprechende theoretische **Begriffe, Hypothesen, Theorien** haben. Nun müssen aber gemäß der instrumentalistischen Position, wenn man sie auf die kognitive Psychologie anwendet, auch Begriffe, Propositionen und Hypothesen von Personen als **theoretische** Sachverhalte eingestuft werden. Nur die Laute und Schriftzeichen, die andere Personen von sich geben, gehören dem Bereich des Beobachtbaren an, nicht aber das, was damit ausgedrückt wird. Eine kognitiv-psychologische Theorie darf also nicht behaupten, daß Personen die Begriffe, Propositionen, Hypothesen usw., die sie ihnen zuschreibt und über die sie als psychologische Theorie etwas aussagt, wirklich hätten. Nun muß man aber fragen, welchen Sinn diese skeptische Enthaltensamkeit in bezug auf die Psychologie haben kann, wenn die instrumentalistische Position bereits in ihrer Formulierung voraussetzt, daß Personen (nämlich als Wissenschaftler) **Begriffe, Propositionen und Hypothesen über theoretische Entitäten haben**, also eben das, was eine psychologische Theorie nicht als real annehmen darf. Der Instrumentalismus scheint auf solche Theorien nicht sinnvoll anwendbar zu sein, die nicht nur selbst aus Begriffen und Propositionen bestehen, sondern sich auch noch auf solche beziehen.

### 3.2 Der Gegenstand psychologischer Theorien

Es ist nicht charakteristisch für die Erfahrungswissenschaften und wäre auch nicht zu empfehlen, zunächst ihren jeweiligen Gegenstand allgemein definieren zu wollen, um ihn dann zu erforschen. Vielmehr werden mit der Entwicklung neuer Theorien auch neue „Gegenstände“ (z. B. Moleküle, Elektronen) erschlossen, und alte Gegenstände werden möglicherweise aufgegeben. Dabei ist es keineswegs selbstverständlich, daß die verschiedenen Gegenstände verschiedener Theorien (z.B. Leistungsmotivation, Aufmerksamkeit, Sprachverstehen) alle etwas gemeinsam haben, das sich dann als der Gegenstand der

Psychologie schlechthin ausgeben ließe. Die Psychologie ist ebensowenig die Wissenschaft (allein) vom Bewußtsein oder gar von der Seele, wie die heutige Physik die Wissenschaft vom Physischen im alltäglichen Sinne. - Die Frage nach dem Gegenstand wird hier aus zwei Gründen aufgegriffen: Zum einen spielt sie in der Philosophie des Geistes und in der Philosophischen Psychologie eine zentrale Rolle, meist in Gestalt der Fragen, wie psychologische Begriffe zu interpretieren sind, was ihr Bezug ist, falls sie einen haben. Es gibt hierzu verschiedene Auffassungen, die jeweils mit bestimmten methodologischen Vorstellungen verbunden sind. Zweitens gibt es in der Psychologie selbst seit der kognitiven Revolution eine bestimmte Interpretationsproblematik: Während der Dominanz des Behaviorismus und Neobehaviorismus war es üblich, Psychologie als Wissenschaft vom Verhalten zu definieren. Eine Psychologie vom **Bewußtsein**, etwa im Sinne von Wilhelm Wundt (1908), galt gemäß der deklarierten Methodologie des Behaviorismus als nicht möglich. Mit den „kognitiven“ Theorien wurde nun das traditionelle **mentalistische** Vokabular wieder eingeführt. Es gilt seither als akzeptabel, Sachverhalte wie Wahrnehmung, Vorstellung, Aufmerksamkeit, Verstehen, Wissen, Denken zum Gegenstand der Theorienbildung zu machen, oder auch nicht-kognitive mentale Sachverhalte, wie z.B. Erlebnisse von Erfolg oder Mißerfolg. In entsprechenden Theorien findet sich das mentalistische Vokabular neben neuen Begriffen, die zu einem großen Teil aus den Informationswissenschaften stammen. Damit stellt sich die Frage, inwieweit sich die mentalistischen Ausdrücke auf subjektive Erlebnisse, auf Bewußtseinszustände beziehen sollen, und wenn nicht, worauf sonst.

In Hinblick auf diese Thematik lassen sich verschiedene Arten der psychologischen Theorienbildung unterscheiden. Erstens gibt es Ansätze, die überhaupt keine mentalistischen Annahmen enthalten und sie teilweise entschieden ablehnen. Hierzu gehören unter anderem die Ansätze des Behaviorismus und Neobehaviorismus, die Verhalten durch theoretische Konstruktionen über interne Reiz-Reaktionsverbindungen zu erklären versuchten (z.B. die Hull-Spence-Schule, nicht dagegen Theoretiker, die, wie Tolman, Erwartungen und Ziele postulierten). Weiterhin zählen hierzu alle Versuche, Verhalten allein durch Bezugnahme auf physiologische Vorgänge zu erklären. Eine Methodologie, die allein solche Ansätze befürwortet und in gewisser Hinsicht das Erbe des radikalen (metaphysischen) Behaviorismus übernommen hat, ist der **eliminative Materialismus** (Rorty, 1965; Churchland, 1981, 1984). Danach sind mentale Ereignisse wie Schmerzen, Vorstellungen, Gedanken usw. nichts weiter als **theoretische Entitäten** innerhalb von Theorien, die sich inzwischen angeblich als völlig inadäquat erwiesen haben und die deshalb ebenso ad acta gelegt werden sollten wie „Wärmesubstanz“ und „Phlogiston“, oder auch wie „Dämonen“ und „Hexen“, die man als theoretische Entitäten erfand, als wissenschaftliche Erklärungen mancher Naturphänomene noch nicht verfügbar



waren. Allein die Neuropsychologie verspricht Fortschritte; die kognitive Psychologie wird als ebenso inadäquat angesehen wie die ältere mentalistische Psychologie. - Es ist zu Recht eingewendet worden, daß diese Auffassung eine Methodologie einer „science fiction“ ist und sich kaum für die zeitgenössische Psychologie eignet. Ohne die enormen Fortschritte auf dem Gebiet der Neuropsychologie zu leugnen, ist es eine unbestreitbare Tatsache, daß sich mit rein physiologischen Annahmen die meisten empirischen Befunde der Psychologie nicht erklären lassen. Das absichtsvolle, zielgerichtete Verhalten, für das sich die Psychologie interessiert, läßt sich auf diese Weise nicht erklären und vorhersagen (Fodor, 1968, 1987; Pylyshyn, 1984). Die meisten Handlungen außerhalb einfacher experimenteller Situationen können ohne die Annahme von Absichten oder Zielen nicht einmal beschrieben werden, da sie nicht auf eine einzige Klasse von Bewegungsformen reduzierbar sind. Die als „mentalistisch“ abgelehnte Psychologie vermag solche Erklärungen zu geben und ist insofern vorläufig nicht ersetzbar. - Diese Kritik richtet sich natürlich in keiner Weise gegen physiologische Psychologie selbst, die nicht mit dem Anspruch des eliminativen Materialismus verbunden sein muß.

Dominierend ist aber in den letzten zwanzig Jahren die Gruppe derjenigen Theorien, die sich der mentalistischen Begriffe bedienen. Dabei besteht nun erstens die Möglichkeit, letztere als **theoretische Begriffe** zu interpretieren, mit denen jedoch nicht auf den Erlebnisaspekt Bezug genommen werden soll (Oswald, 1980; Carrier & Mittelstraß, 1989). Wenn etwa gesagt wird, daß eine Person einen Schmerz empfindet oder eine Erwartung hat, so wird ihr damit ein Zustand hypothetisch zugeschrieben, um ihr Verhalten zu erklären. Doch wird damit nicht behauptet, daß die Person einen entsprechenden Bewußtseinszustand, ein Erlebnis von bestimmter Qualität hat, wie man ihn aus der „inneren Erfahrung“ kennt. - Auch die als **Funktionalismus** oder **funktionaler Materialismus** bezeichnete Auffassung sieht von dem Erlebnisaspekt ab: Ein mentaler Zustand ist weder durch seine Erlebnisqualität noch durch physikalische bzw. physiologische Eigenschaften definiert, sondern durch die kausalen **Beziehungen** zu anderen internen Zuständen, zu äußeren Bedingungen und zum Verhalten, kurz durch die **funktionale Organisation** des betreffenden Systems (Putnam, 1960; Fodor, 1968, 1975, 1987), die, wie das Programm eines Computers, durch Sätze der Form beschrieben werden kann: Wenn das System im Zustand Z ist, so führt Input I zum Output O und zum internen Zustand Z'. Personen können dementsprechend als **informationsverarbeitende Systeme** aufgefaßt werden (Newell & Simon, 1972). - Eine funktionalistische Methodologie kann mit der in der Künstlichen Intelligenz zum Teil verfolgten Zielsetzung kombiniert werden, intelligente Prozesse unabhängig von ihrer physischen Realisierung (letztlich also auch unabhängig von Lebewesen) zu verstehen (Goldstein & Papert, 1977). Funktionalismus läßt sich jedoch auch mit der herkömmlichen Zielsetzung der Psychologie verbinden, das Verhalten

von Personen zu erklären. (Auf die Unterschiede und Verschränkungen zwischen beiden Vorhaben und auf entsprechende methodologische Probleme kann hier nicht näher eingegangen werden.) Im zweiten Fall sind kognitive Theorien jedenfalls Theorien über Personen (bzw. höherentwickelte Tiere). Wenn es gelingt, Verhalten durch Theorien der Informationsverarbeitung zu erklären, kann man in einem zweiten Schritt versuchen, die physiologischen Mechanismen herauszufinden, die Personen dazu in die Lage versetzen, ihre Verhaltensleistungen zu erbringen (s. dazu auch 4.3). Funktionalisten sehen diesen Weg als zweckmäßiger an als den Versuch, die komplexen Leistungen des Zentralnervensystems durch unmittelbare Analyse physiologischer Prozesse verstehen zu wollen. - Gegen den Funktionalismus wurde eingewendet, daß eine rein funktionalistische Beschreibungsweise das spezifische mentaler Zustände, ihre **Intentionalität** und **Erlebnisqualität**, nicht zu erfassen vermag (Searle, 1980; Block, 1980; vgl. auch Kap. 5, Abschnitt 5.3 dieses Bandes).

Zweitens gibt es die Möglichkeit, mit (einigen) mentalistischen Termen in herkömmlicher Weise **Bewußtseinszustände** zu bezeichnen. Das im Verlauf der fünfziger Jahre zunehmende Interesse an kognitiven Vorgängen war zunächst auch ein Interesse am Aspekt der **Subjektivität**, an einer Verbindung von objektivistischer und phänomenologischer Psychologie (Miller, Galanter & Pribram, 1960; Wann, 1964, Graumann, 1965). Wenn auch die heutige kognitive Psychologie keineswegs als Wissenschaft vom Bewußtsein angesehen werden kann, wird doch nicht selten auf das Bewußtsein Bezug genommen. Es gibt sogar ausführlich ausgearbeitete (kognitive) Theorien des Bewußtseins (Mandler, 1985; Jackendoff, 1987; Baars, 1988; Gadenne & Oswald, 1991), die bewußte Zustände etwa mit den Inhalten des Arbeitsgedächtnisses, der fokalen Aufmerksamkeit oder mit bestimmten Repräsentationsformen in Beziehung setzen. Die meisten entsprechenden Theorien nehmen an, daß die bewußten Ereignisse nur einen Teil des gesamten mentalen Geschehens ausmachen. (Nicht bewußt sind z.B. die Analyse elementarer Merkmale in der Wahrnehmung oder die Verwendung grammatikalischer Tiefenstrukturen bei der Sprachproduktion.) Die Beurteilung von Aussagen, die sich (teilweise) auf Bewußtseinszustände beziehen, wirft erneut die Frage auf, inwieweit und in bezug auf welche mentalen Zustände **Introspektion** und **Erlebnisbeschreibung** zuverlässig möglich sind (vgl. Lyons, 1986; Gadenne & Oswald, 1991).

### 3.3 Idealisierung und Unvollständigkeit

Es ist einsichtig, daß Theorien als Darstellungsmittel hochgradig selektiv sein müssen. Das Ziel, mit wenigen Annahmen viele Phänomene zu erklären, erfordert eine Konzentration auf wenige Variablen, die man für erklärungsrelevant hält. Darüber hinaus werden bei der Theorienbildung aber auch Verein-

fachungen und gedankliche **Idealisierungen** in Kauf genommen, über deren fiktiven Charakter man sich völlig im klaren ist. Idealisierungen bzw. **ideale Modelle** finden sich gerade im Zusammenhang mit Theorien von hoher Erklärungskraft. Es gibt strenggenommen keinen freien Fall, keine Massenpunkte, homogenen Körper, idealen Gase, reibungslosen Flüssigkeiten. Physiker nehmen solche Idealisierungen vor, weil es auf diese Weise eher möglich ist, Naturgesetze zu finden bzw. sie mathematisch zu formulieren und entsprechende Berechnungen durchzuführen. (Beispielsweise erfordert die Infinitesimalrechnung Kontinuitätsannahmen, von denen man oft annimmt, daß sie in Wirklichkeit nicht zutreffen.) Aus ähnlichen Gründen findet man in den Sozialwissenschaften theoretische Konstruktionen wie vollkommen freie Märkte, vollkommen informierte Individuen, Personen, die ihre Sprache vollkommen beherrschen, deren Handeln nur durch das Leistungsmotiv oder nur durch bestimmte kognitive Vorgänge bestimmt ist usw. Bei der Formulierung von Gesetzhypothesen und der Durchführung von Ableitungen wird dann zunächst davon ausgegangen, daß man es mit idealen Individuen bzw. Situationen zu tun hätte. Später kann man die Theorie so weiterentwickeln, daß zusätzliche relevante Variablen Berücksichtigung finden. Auch können zum Zweck der Vorhersage weitere relevante Variablen (Reibungskräfte, zusätzliche Motive) als **Korrekturfaktoren** einbezogen werden (Barr, 1971, 1974). Insofern sind Idealisierungen durchaus mit einer realistischen Interpretation von Theorien vereinbar (Bunge, 1967). Die realistische Zielsetzung zeigt sich daran, daß es als wichtig erachtet wird, wie stark der Idealfall vom realen abweicht und wie man die Abweichung eventuell verringern bzw. korrigieren kann.

Ein mit der Idealisierung verwandter Sachverhalt ist die unvollständige Angabe relevanter Variablen. Wenn eine Theorie zu dem Zweck konstruiert wurde, gewisse empirische Phänomene zu erklären, so weiß man in der Regel bereits im voraus, daß diese Phänomene auch noch durch andere Faktoren als die von der Theorie genannten kausal beeinflußt werden. Bei einzelnen Annahmen, etwa der Hypothese, daß Erwartungen die Wahrnehmung beeinflussen oder daß „*vididness*“ die Bereitschaft zu Schlußfolgerungen erhöht, ist es auf den ersten Blick ersichtlich, daß nur **eine** kausal relevante Variable genannt werden soll und weitere Einflüsse auf die abhängige Variable als selbstverständlich angenommen werden. Entsprechendes gilt aber auch für ausgearbeitete Theorien, wie bereits am Beispiel der Leistungsmotivation gezeigt wurde: Zu der Hauptaussage dieser Theorie ist hinzuzufügen, daß auch „*extrinsische Motive*“ eine Rolle spielen können, die nichts mit der eigentlichen, „*intrinsischen*“ Leistungsmotivation zu tun haben. Extrinsische Motive können das Verhalten in der Tat entscheidend bestimmen. Die Theorie sagt jedoch weder, welche extrinsischen Motive es im einzelnen gibt noch wie der jeweilige Wirkungsmechanismus beschaffen ist. In Anlehnung an die Ausführungen über Idealisierung kann man sagen, daß sich die Theorie auf ein ideales **Modell**

der Wirklichkeit bezieht, nämlich auf ideale Personen, deren Handeln ausschließlich durch die intrinsische Leistungsmotivation bestimmt ist. Stattdessen kann man aber auch sagen (und dies ist vielleicht eindeutiger), daß sich die Theorie auf reale Personen bezieht, die zur Erklärung relevanten psychologischen Faktoren jedoch nur unvollständig angibt.

**Unvollständigkeit** im hier verwendeten Sinne ist eine Beziehung zwischen einer Hypothese oder Theorie und den mit ihr zu erklärenden Sachverhalten (s. auch 4.2). Eine Theorie und entsprechend auch eine Erklärung können mehr oder weniger vollständig sein. Im allgemeinen wird man vollständigere Theorien (Erklärungen) bevorzugen und Theorien entsprechend konstruieren bzw. entwickeln, doch sind diesem Bemühen grundsätzliche Grenzen gesetzt. Das absichtsvolle Verhalten, für dessen Erklärung sich Psychologen interessieren, ist immer eine Funktion der **gesamten** psychologischen Sachverhalte, die von einzelnen Forschungsrichtungen als Persönlichkeitseigenschaften, Einstellungen, Motive, Emotionen, kognitive Prozesse usw. bezeichnet werden. Kein Verhalten ist nur durch visuelle Wahrnehmung, Langzeitgedächtnis, Urteilsfehler, Introversion usw. determiniert. Es gibt aber keine psychologische Theorie, die mit einheitlichen Begriffen alle relevanten Faktoren benennen und die Gesetzmäßigkeiten angeben könnte, nach denen diese in Wechselwirkung das Verhalten einer Person hervorbringen. Jede psychologische Theorie, über die derzeit geforscht wird, läßt gewisse Faktoren außer acht, von denen man mit guten Gründen behaupten kann, daß sie auf die untersuchten abhängigen Variablen ebenfalls eine Wirkung haben. Eine psychologische „Gesamttheorie“ wäre zwar wünschenswert, entsprechende Versuche sind jedoch, wie sich in der Geschichte der Psychologie gezeigt hat, entweder gescheitert, oder sie sind bloße Forschungsperspektiven geblieben, deren Realisierung im einzelnen auf Probleme stieß, die nicht gelöst werden konnten.

Es gibt noch einen anderen Grund dafür, daß psychologische Erklärungen eines äußeren Verhaltens nicht vollständig sein können. Jedes äußere Verhalten ist unter anderem von physikalischen Bedingungen abhängig, die mit einer psychologischen Theorie nicht erfaßt werden können. Wenn z.B. alle psychologisch beschreibbaren Bedingungen dafür erfüllt sind, daß eine Versuchsperson eine Armbewegung macht oder ein bestimmtes Wort ausspricht, könnten störende Bedingungen auf physikalischer Ebene eintreten, die das vorhergesagte Verhalten verhindern. Die Vp könnte z.B. in diesem Augenblick einen Gehirnschlag erleiden. Auf psychologischer und auch auf biologischer Ebene gibt es folglich keine ausnahmslos gültigen Naturgesetze zur Erklärung des Verhaltens von Systemen. Die formulierten Gesetzhypothesen sind immer für den Fall gedacht, daß das analysierte System (Person, Lebewesen) nicht zerstört und in seinem Funktionieren nicht zu sehr gestört wird. Insbesondere die letzte Bedingung läßt sich kaum als präzise Anfangsbedingung in eine Theorie aufnehmen, da es unabgrenzbar viele Möglichkeiten der „Störung“ gibt.

Die methodologische Frage, welches Ausmaß an Idealisierung und Unvollständigkeit angemessen ist, kann nicht allgemein beantwortet werden, da hier die Fragestellung und die Komplexität des Gegenstandes eine entscheidende Rolle spielen. Wie bei Problemen allgemein kann es auch in der Theorienbildung eine nützliche Strategie sein, zunächst grobe Problemlösungen zu versuchen, die später im Detail ausgearbeitet werden. Zu viel an Vereinfachung kann aber auch daran hindern, überhaupt adäquate Lösungswege zu finden. Vereinfachung kann z.B. bedeuten, in den theoretischen Annahmen mit möglichst wenig Variablen auszukommen und möglichst wenige und einfache Beziehungen zwischen ihnen zu postulieren.

#### **4. Theorie und Erklärung**

Die Möglichkeit, Erklärungen auf der Grundlage von Gesetzen vorzunehmen, gilt als das Hauptziel nomologischer Wissenschaft. Im Rahmen dieses Kapitels ist es leider nicht möglich, die umfangreiche Problematik des Erklärens in allen Aspekten abzuhandeln. Hauptzweck der folgenden Ausführungen ist es daher, auf psychologisch wichtige Aspekte zu verweisen.

##### 4.1 Deduktiv-nomologische Erklärung

Eine Erklärung gibt eine Antwort auf eine **Warum-Frage**: Warum erinnern Personen die Items, die am Anfang oder Ende einer Lernliste plaziert waren, besser als diejenigen in mittlerer Position (serieller Positionseffekt)? Warum erhöhen Personen bei Erfolg ihr Anspruchsniveau? Warum ist Peter an diesem Nachmittag ungewöhnlich aggressiv? - Offensichtlich kann eine Erklärungsfrage sowohl in bezug auf einen **allgemeinen Zusammenhang** als auch in bezug auf ein **singuläres Ereignis** gestellt werden. Die Erklärung von allgemeinen Hypothesen oder Gesetzen wurde in der Einleitung zu diesem Kap. bereits angesprochen: Zielsetzung bei der Theorienbildung ist es, möglichst viele verallgemeinerte Befunde eines Forschungsgebietes durch eine Theorie zu erklären. Im weiteren kann versucht werden, die Gesetze einer Theorie oder die Gesetze mehrerer Theorien durch eine noch grundlegendere Theorie zu erklären. Im Zusammenhang mit der Leistungsmotivation haben wir eine Vielzahl von empirischen Befunden (allgemeinen Hypothesen) genannt, die durch die Theorie Atkinsons erklärt werden konnten. Die Erklärung besteht aus der **logischen Ableitung** der zu erklärenden Aussage (**Explanandum**) aus den zur Erklärung dienenden Aussagen (**Explanans**). Wie oben aufgezeigt wurde, muß das Explanans häufig nicht nur Gesetze im engeren Sinne, sondern auch Zusatzannahmen enthalten.

Genaugenommen ist es allerdings eine Vereinfachung, zu sagen, daß die zu erklärenden Hypothesen oder Gesetze aus grundlegenden Gesetzen „logisch abgeleitet“ wurden; sie werden meist auch korrigiert (vgl. Popper, 1972, Kap. V). Betrachten wir z.B. den einstmals als typisch geltenden Befund (bestätigte Hypothese), daß Personen nach Erfolg ihr Anspruchsniveau erhöhen. Atkinsons Theorie erklärte, warum dieser Befund unter vielen Bedingungen zu erwarten ist, aber sie zeigte zugleich auf, unter welchen Umständen ein „atypischer“ Befund auftreten wird. Zahlreiche korrigierende Erklärungen sozialwissenschaftlicher Befunde hat Malewski (1967) gegeben.

Eine wichtige Anwendung von Theorien besteht in der Erklärung **singulärer Ereignisse**. Angenommen, eine Person *a* hat unter mehreren Aufgaben die Aufgabe *i* ausgewählt. Dies sei das Explanandum-Ereignis. *E* sei die Beschreibung dieses Sachverhalts. Es kann davon ausgegangen werden, daß *a* in der betreffenden Situation einen Leistungsmaßstab für verbindlich hielt ( $A_1$ ); weiterhin ist nachweisbar, daß bei *a* das Motiv  $M_e$  das Motiv  $M_m$  überwiegt und daß für *a* die Aufgabe *i* unter allen verfügbaren Aufgaben am nächsten bei  $W_e = 0.5$  lag ( $A_2$ ). In diesem Falle läßt sich das Ereignis *E* erklären, indem man aus dem Gesetz *T*(Wahl) zusammen mit den Annahmen  $A_1$  und  $A_2$  die Aussage *E* deduziert.  $A_1$  und  $A_2$  werden als **Randbedingungen** oder **Anfangsbedingungen** bezeichnet. (Man kann auch die mehr kontextuelle Bedingung  $A_1$  als Rand- oder Rahmenbedingung und nur  $A_2$  als Anfangsbedingung bezeichnen). Allgemein läßt sich die Idee einer solchen **deduktiv-nomologischen** (DN-) Erklärung so fassen, daß eine Explanandum-Aussage *E* aus Anfangsbedingungen und Gesetzen logisch abgeleitet wird (vgl. Hempel, 1977):

$$\frac{A_1, A_2, \dots, A_n}{G_1, G_2, \dots, G_r}$$

*E*

Dies ist die Präzisierung der alten Erklärungsidee, wonach bei einer Erklärung eine Ereignisabfolge (*E* folgt auf die Anfangsbedingungen) unter ein allgemeines Gesetz (bzw. mehrere Gesetze) **subsumiert** wird (daher auch „**covering law-model**“ genannt). Die weitere Präzisierung dieser Erklärungskonzeption wirft eine Reihe von schwierigen Problemen auf, die jedoch hier nicht behandelt werden können (vgl. etwa Stegmüller, 1983). Auf ein für die Psychologie besonders wichtiges Kriterium einer adäquaten Erklärung sei jedoch hingewiesen: die **unabhängige Prüfbarkeit des Explanans** (vgl. Popper, 1972, Kap. 5). Damit ist gemeint, daß die erklärenden Annahmen prüfbar sein müssen, wobei der zu erklärende Sachverhalt nicht als Prüfungsinstanz anerkannt wird. Wenn man z.B. die Tatsache, daß jemand seinen Vater umgebracht hat, dadurch erklären wollte, daß er ihn unbewußt gehaßt hat, zur Prüfung dieser

Annahme jedoch nur auf die Tat selbst verweisen kann, so liegt keine unabhängige Prüfbarkeit vor.

## 4.2 Statistische und unvollständige Erklärung

Statistische Erklärungen sind dann angemessen, wenn man es anstatt mit deterministischen mit statistischen Zusammenhängen zu tun hat. Auch im statistischen Fall ist es möglich, Gesetze aus grundlegenden Gesetzen zu deduzieren. Es handelt sich dann um eine **deduktiv-statistische** Erklärung (Hempel, 1977). Man findet diesen Erklärungstyp vor allem in der mathematischen Psychologie. So wird z.B. durch probabilistische Lernmodelle eine Hypothese darüber formuliert, mit welchen Wahrscheinlichkeiten ein Individuum in einer definierten Situation verschiedene Reaktionen zeigt; es wird weiterhin ausgesagt, wie jede dieser Reaktionen die Wahrscheinlichkeitsverteilung für den nachfolgenden Versuch ändert (vgl. Coombs, Dawes & Tversky, 1975, Kap. 9). Daraus läßt sich die Wahrscheinlichkeitsverteilung nach n Versuchen berechnen. Es kann also der Lernerfolg (ausgedrückt als Wahrscheinlichkeitshypothese über die Reaktionen) nach einer gegebenen Zahl von Versuchen durch Wahrscheinlichkeitsannahmen über elementare Prozesse erklärt werden.

Viel größere Probleme bereitet die statistische Erklärung eines **Einzelfalles** in Analogie zur DN-Erklärung. Eine DN-Erklärung kann im einfachsten Fall als ein deduktiv-logischer Schluß von A und dem Gesetz „Immer wenn A, dann B“ auf das Explanandum B dargestellt werden. Dementsprechend könnte man eine statistische Erklärung als eine Art „abgeschwächten Schluß“ konzipieren, indem man das deterministische Gesetz durch ein statistisches ersetzt:  $p(B/A) = r$ . Das deduktive Argument wurde durch ein induktives Argument ersetzt, das dem Explanandum zwar keine Sicherheit, jedoch eine hohe **induktive Wahrscheinlichkeit** (nahe 1) verleiht (Hempel, 1977). In Ermangelung einer entsprechenden induktiven Logik erwies sich dieser Weg jedoch bisher als nicht erfolgreich und ist fast gänzlich aufgegeben worden. Statistische Erklärungen werden nicht mehr als Argumente aufgefaßt. Auch gegen die von Hempel ursprünglich erhobene Forderung nach einer hohen Wahrscheinlichkeit wurden überzeugende Einwände vorgebracht (vgl. Salmon, 1971). - Unter den verschiedenen statistischen Erklärungsmodellen, die nach Hempels Ansatz vorgeschlagen worden sind, hat dasjenige von Salmon (1971, 1984) besondere Beachtung erfahren. Westmeyer (1974) sowie Groeben und Westmeyer (1975) bewerten dieses Modell (in Stegmüllers Version) als grundlegend für die Psychologie, nicht zuletzt für die psychologische Diagnostik (vgl. auch Gadenne, 1988). Hier kann nur die Grundidee des Modells dargestellt werden, ohne technische Details. Angenommen, es sei zu erklären, daß Albert ein Auto gestohlen hat (B). Albert gehört der Grundgesamtheit G der amerikanischen

Jugendlichen an.  $p(B/G)$  sei relativ klein. Man sucht nun nach Merkmalen, die für B statistisch relevant sind, z.B. Geschlecht, Familienstand der Eltern, soziale Schicht.  $A_1, A_2$  usw. seien verschiedene Wertekombinationen in diesen Merkmalen, etwa „männlich/Eltern geschieden/Mittelschicht“. Für jede Kombination sei ein statistisches Gesetz der Form  $p(B/A_i) = r_i$  bekannt. Dabei wird vorausgesetzt, daß die verschiedenen Teilklassen von G, die dadurch definiert sind, daß jeweils eine der Bedingungskonstellationen  $A_1, A_2$  usw. auf sie zutrifft, in bezug auf B **homogen** sind. Dies bedeutet, daß keine weitere Bedingung C existiert, derart, daß  $p(B/A_i \text{ und } C)$  von  $p(B/A_i)$  verschieden wäre. (Mit anderen Worten: Die statistischen Gesetze sind **vollständige** Aussagen; es gibt für B keine weiteren, statistisch relevanten Bedingungen.) - Die eigentliche Erklärung besteht nun einfach in der Angabe aller  $p(B/A_i)$  sowie in der Angabe der im konkreten Fall vorliegenden Bedingungen  $A_i$ . Die Erklärung informiert darüber, wie wahrscheinlich das Explanandum unter den gegebenen Umständen war und wie wahrscheinlich es unter anderen Umständen gewesen wäre. (Sie behauptet also keineswegs, daß B sehr wahrscheinlich war, ja nicht einmal, daß B eher zu erwarten als nicht zu erwarten war.)

Zur Anwendung dieses Modells in der Psychologie muß betont werden, daß es trotz der geringen Anforderungen bezüglich der Höhe der Wahrscheinlichkeiten in anderer Hinsicht sehr hohe Anforderungen stellt: Gemäß der Homogenitätsbedingung muß es sich um „naturgesetzliche“ Wahrscheinlichkeiten handeln. Was dies bedeutet, läßt sich vielleicht am deutlichsten anhand des radioaktiven Zerfalls illustrieren. So hat z.B. ein Radon-Atom eine Zerfallszeit von 3.82 Tagen; es zerfällt innerhalb dieser Zeitspanne mit der Wahrscheinlichkeit 0.5. Diese Zerfallszeit bleibt dieselbe, wenn man eine Radonprobe erhitzt, unter Druck setzt oder weitere Bedingungen variiert. Nun muß jedoch ernstlich bezweifelt werden, ob es außerhalb der Physik der Elementarteilchen Gesetze mit solchen Wahrscheinlichkeiten gibt. Im Beispiel des Autodiebstahls verhält es sich sicherlich nicht so. Kriminalitätsraten bleiben nicht völlig konstant, wenn sich ökonomische, soziologische, ja meteorologische Bedingungen ändern, ganz abgesehen von der Schwierigkeit, die entsprechenden Wahrscheinlichkeiten aufgrund von Statistiken zu ermitteln.

Für viele psychologische Erklärungsfragen sind nur solche Aussagen verfügbar, die oben als **unvollständige Hypothesen** bezeichnet wurden, etwa von der Art: Eine Erhöhung von X bewirkt **ceteris paribus** eine Erhöhung von Y. - Es kann z.B. als hoch bestätigt gelten, daß ein Medikament eine Senkung der Konzentrationsfähigkeit bewirkt. Dies impliziert noch kein Wissen darüber, wie wahrscheinlich die Wirkung ist (im Sinne eines statistischen Gesetzes, falls ein solches existiert) bzw. wie stark der durchschnittliche Effekt ist, bezogen auf eine große Grundgesamtheit von Personen. Was sagt unter solchen Umständen ein Erklärungsvorschlag überhaupt aus? Man kann dies so formulieren und als **unvollständige Erklärung** bezeichnen (Gadanne, 1990): Es wird



**erstens** davon ausgegangen, daß es eine bestätigte unvollständige Hypothese gibt (eventuell als Teil einer Theorie), wonach A ein Kausalfaktor für B ist. **Zweitens** wird gesagt, daß A in der gegebenen Situation aufgetreten ist. Zusammen implizieren die beiden Annahmen, daß A **zu B kausal beigetragen** hat: Was auch immer sonst noch mitgewirkt haben mag, A war daran beteiligt.

### 4.3 Weitere Erklärungsarten

Von den genannten Erklärungsarten sind einige andere unterschieden worden, die zum Teil von speziellem psychologischem Interesse sind. Durch eine **dispositionelle** Erklärung wird versucht, ein Verhalten durch Verweis auf eine Disposition (z. B. Introversión, Durchsetzungsfähigkeit) zu erklären. Eine **genetische** Erklärung soll das Endprodukt einer Entwicklung erklären, indem sie auf die Aufeinanderfolge der einzelnen Entwicklungsstadien und deren immanente Systematik verweist. Eine **teleologische** Erklärung liegt vor, wenn das Explanans auf zukünftige Ereignisse bzw. Ziele Bezug nimmt. - Es konnte nun gezeigt werden, daß es sich hierbei nicht um eigenständige Erklärungsarten handelt, sondern um Unterarten oder Anwendungen der bereits behandelten Erklärungsarten. Die dispositionelle Erklärung verwendet (und benötigt) solche speziellen Gesetze, in denen Dispositionen (etwa Persönlichkeitseigenschaften) vorkommen; oder sie interpretiert Dispositionen selbst als gesetzesartige „Annahmengenüge“ (vgl. Herrmann, 1973). Die genetische Erklärung läßt sich als Reihe von mehreren einzelnen (DN- oder statistischen) Erklärungen darstellen, die sich jeweils auf den Übergang von einem Entwicklungsstadium zum nächsten beziehen. Und eine teleologische Erklärung sagt, daß mentale Zustände, die sich inhaltlich auf die Zukunft beziehen (Absichten, Zielsetzungen), als Ursachen gewirkt und dadurch das zu erklärende Verhalten hervorgebracht haben. - Es sei hierzu auf die klassischen Arbeiten von Hempel verwiesen (enthalten in Hempel, 1977).

Aus der Sicht der neueren kognitiven Psychologie verdient die **Instantiierungserklärung** besondere Aufmerksamkeit (Haugeland, 1978; Cummins, 1983; Eimer, 1990). Es geht hierbei nicht um die kausale Erklärung eines Einzelereignisses, sondern um die Erklärung einer Eigenschaft: Was liegt der Eigenschaft (Disposition, Fähigkeit) D zugrunde? Welche Mechanismen instantiieren D, d.h. setzen das (z.B. kognitive) System dazu in die Lage, das für D charakteristische Verhalten zu zeigen? D ist durch gewisse Input-Output-Gesetzmäßigkeiten charakterisiert; insofern handelt es sich hier um eine spezielle Art der Erklärung von Gesetzen durch grundlegendere Gesetze. Eine Instantiierungserklärung erfordert es, D detailliert zu beschreiben und eventuell in Komponenten zu zerlegen. Sie erfordert weiterhin eine Analyse des betreffenden Systems in Subsysteme und ihre Eigenschaften, soweit sie für D

relevant sind. Die eigentliche Erklärung ist vollzogen, wenn gezeigt werden kann, daß ein aus diesen Subsystemen in bestimmter Weise aufgebautes System die Eigenschaft D haben muß. Wenn D eine mentale Eigenschaft ist, so nehmen die heutigen psychologischen Erklärungen etwa Bezug auf Subsysteme wie Gedächtnisspeicher und damit verbundene Prozessoren, aus deren Zusammenwirken sich z.B. die Gesetzmäßigkeiten des Behaltens und Vergessens ableiten lassen. Das weitere Ziel einer psychologischen Instantiierungserklärung müßte darin bestehen, die Eigenschaften solcher Subsysteme durch Bezugnahme auf physiologische Subsysteme zu erklären. Letzteres wirft bekanntlich erhebliche Probleme auf und ist bisher vielleicht am ehesten in Teilbereichen der visuellen Wahrnehmung gelungen (vgl. Eimer, 1990).

Ein wichtiges methodologisches Prinzip bei einer Instantiierungserklärung (gegen das in der kognitiven Psychologie nicht selten verstoßen wird) besagt, daß die zu erklärende Eigenschaft nicht auch als Eigenschaft der Subsysteme auftauchen darf. Wenn z.B. Aufmerksamkeit oder Sprachverstehen erklärt werden sollen, so dürfen den Prozessoren des Arbeitsgedächtnisses nicht die Fähigkeiten zugeschrieben werden, „Aufmerksamkeit auszuüben“ oder eine Proposition zu „verstehen“ (vgl. auch Herrmann, 1982); andernfalls wird das Ziel verfehlt, das Explanandum durch Verweis auf Mechanismen bzw. elementare Prozesse wirklich zu erklären. Ein Subsystem in psychologischen Theorien, das häufig mit zu vielen Fähigkeiten ausgestattet wird (relativ zum Explanandum), ist die „Exekutive“ bzw. der „zentrale Prozessor“ (vgl. Gadenne & Oswald, 1991). Auf die hierin liegende Gefahr der **Pseudoerklärung** ist in der Psychologie seit langem hingewiesen worden, häufig im Zusammenhang mit der Psychoanalyse (Skinner, 1956): Obwohl Freud nomologische Erklärungen anstrebte, hat er die von ihm angenommenen Subsysteme (Ich, Es, über-Ich) oft sehr anthropomorph beschrieben.

#### 4.4 Theorie, Erklärung und Kausalität

Manche theoretischen Aussagen werden als **Kausalaussagen** interpretiert. Was sagt ein **Kausalgesetz** aus, was eine **singuläre Kausalaussage**? Antworten hierauf sind Voraussetzung für die weitere Frage, wie man eine Kausalaussage **überprüfen** kann (vgl. dazu Kap. 9, 12 und 14). Eine ausführliche Abhandlung der Kausalitätsprobleme findet man bei Stegmüller (1983).

Der Begriff der Kausalität ist im psychologischen Denken ebenso fest verwurzelt wie in vielen anderen Wissenschaften; Ingenieure, Mediziner und auch Juristen scheinen auf ihn nicht verzichten zu können. Mit dem Aufkommen des logischen Empirismus wurde zeitweise die Auffassung vertreten, daß der Kausalbegriff hoffnungslos metaphysisch sei und in den Wissenschaften keinen Platz habe. Mach und Russell argumentierten, daß man ihn sinnvoller-

weise durch den der „funktionalen Abhängigkeit“ ersetzen sollte. Unter dem Einfluß des Behaviorismus schienen viele Psychologen dem Vorschlag zugeeignet, nicht mehr von „Ursachen“, sondern, auf der Grundlage empirischer Gesetze, von „Bedingungen“ des Verhaltens zu sprechen. Inzwischen hat sich die Situation wieder gewandelt, und es ist in der Psychologie weithin selbstverständlich geworden, Kausalbeziehungen zu postulieren und zu untersuchen. Die Bemühungen um **Kontrolle** in der Versuchsplanung und ein großer Teil der Entwicklungen in der Korrelationsstatistik dienen dem erklärten Ziel, Kausalbeziehungen nachzuweisen bzw. Kausalmodelle zu testen (vgl. Kap. 12 und 15). Hierbei liegt allerdings keine einheitliche und zum Teil überhaupt keine geklärte Auffassung von Kausalität zugrunde.

Die Aussage, daß das konkrete Ereignis A das konkrete Ereignis B verursacht habe, behauptet zumindest, daß A und B stattgefunden haben und daß A vorausgegangen ist. Was sagt sie darüber hinaus? Die einflußreiche Antwort Humes, daß zusätzlich nichts weiter behauptet werde als die **regelmäßige Aueinanderfolge** der Ereignisse (Ereignistypen) A und B, gilt heute im allgemeinen als unzureichend. Danach müßte man nämlich behaupten, daß der Schlag einer Uhr den Schlag einer anderen Uhr verursacht, sofern nur beide regelmäßig hintereinander schlagen, solange sie existieren. Die Regelmäßigkeit enthält nicht das Hervorbringen von B durch A und die dabei unterstellte kausale Notwendigkeit. (Hume meinte freilich, daß diese Notwendigkeit nur in unserem Geist existiere, ein Resultat unserer Erwartung aufgrund von Gewohnheit.) Als adäquater müssen jedenfalls jene Lösungsvorschläge gelten, die an die Stelle einer Regelmäßigkeit die Idee der **naturgesetzlichen Notwendigkeit** setzen (die allerdings selbst nicht leicht zu fassen ist; vgl. 1.3). Danach impliziert eine singuläre Kausalaussage, daß es eine wahre Gesetzhypothese G gibt, die A mit B verknüpft. Hierbei wird A oft nur einen (unentbehrlichen) Teil eines Bedingungskomplexes nennen, der insgesamt für B hinreichend ist (vgl. dazu Mackie, 1974). Weiterhin wird G oft eine unvollständige Hypothese sein.

Wie schon angedeutet wurde, faßt man Ursachen und Wirkungen gewöhnlich nicht als Zustände, sondern als **Ereignisse** auf, wobei die Ursache der Wirkung vorausgeht. Danach sind **Kausalgesetze Sukzessionsgesetze** (im Unterschied zu Koexistenzgesetzen). Für psychologische Zusammenhänge erscheint dies akzeptabel, da hier (zumindest praktisch gesehen) bei Sukzessionsgesetzen davon ausgegangen werden kann, daß entweder A oder B zuerst eingesetzt hat. (Aus grundsätzlicherer Perspektive treten hier nämlich Probleme auf: Nach der Relativitätstheorie ist die Frage, ob A auf B folgt oder umgekehrt, vom Bezugssystem des Betrachters abhängig, wogegen die Kausalrelation für jedes Bezugssystem dieselbe bleibt; die Kausalrelation ist somit prinzipiell nicht auf die Relation der zeitlichen Abfolge zurückführbar.)

Ebenso wie mit der Gesetzesartigkeit ist die Kausalität eng mit der Idee des Möglichen verknüpft. Nach Lewis (1973) hängt B genau dann kausal von A ab, wenn folgende **kontrafaktische** Aussage wahr ist: Wenn A nicht stattgefunden hätte, so hätte auch B nicht stattgefunden. Die Explikation kontrafaktischer Aussagen erfolgt mit Hilfe der Begriffe der **möglichen Welt** und der Ähnlichkeit zwischen möglichen Welten. - So wichtig allerdings die Beziehung zwischen Kausalität und Möglichkeit sein mag, wird doch eine Rückführung des Kausalbegriffs auf grundlegend einfachere Begriffe so nicht erreicht: Man ist nun mit der Frage konfrontiert, was eine mögliche Welt ist; logische Möglichkeit ist in diesem Zusammenhang zu weit, und faktische Möglichkeit verweist zurück auf das, was mit den Naturgesetzen (darunter die Kausalgesetze) in Einklang steht.

Eine weitere Kausalitätsauffassung ist aus psychologischer Sicht besonders interessant: A verursacht B genau dann, wenn wir B dadurch herstellen oder **verhindern** können, daß wir A herstellen oder verhindern (Collingwood, 1940). Kausalität wird hier mit der Möglichkeit zur **Manipulation** gleichgesetzt (wobei allerdings zu fragen ist, ob die Idee der Manipulation (herstellen, verhindern) nicht schon die Idee der Verursachung voraussetzt). In der Tat spielt die Erfahrung der Effekte des eigenen Handelns eine wichtige Rolle beim individuellen Erwerb des Kausalitätsbegriffs. Auch ist der Sprachgebrauch durch gewisse handlungsbezogene Aspekte geprägt. So erscheint es z.B. noch eher akzeptabel, zu sagen, daß die Pendellänge (die wir direkt manipulieren können) die Schwingungsdauer verursacht (die wir nur indirekt über die Länge beeinflussen können) als das Umgekehrte, obwohl die Variablenwerte immer gleichzeitig bestehen. Jedoch ist es fraglich, ob man diese pragmatischen Aspekte zur Grundlage einer metatheoretischen Konzeption machen sollte. Man gerät in Schwierigkeiten, sobald es darum geht, die Bedeutung von Kausalassagen über nicht-manipulierbare Sachverhalte zu erklären.

Vermutlich läßt sich die Idee der Verursachung, die man mit den Worten ausdrückt: A **bringt** B hervor, nicht vollständig durch wesentlich einfachere Begriffe und extensionale Logik definieren. Dies wäre kein zwingender Grund, den Kausalitätsbegriff nicht zu verwenden. Auch viele andere methodologische und metatheoretische Begriffe (Gesetzesartigkeit, Bestätigung) scheinen keiner expliziten Definition allein mit den Mitteln der extensionalen Logik fähig, wie es der logische Empirismus anstrebte. Bunge (1987) entwickelt eine Konzeption der Kausalität, die den Gedanken der Verursachung bzw. des **Hervorbringens** nicht explizit definiert, sondern gemeinsam mit anderen Begriffen (Typen der Determination, Gesetzesartigkeit, zeitliche Abfolge, Wahrscheinlichkeit) im Rahmen einer (Meta-) Theorie einführt; der Kausalitätsbegriff wird auf diese Weise implizit näher bestimmt, so wie dies mit theoretischen Begriffen in den empirischen Wissenschaften geschieht.

Auch für den Fall statistischer Zusammenhänge findet der Kausalitätsbegriff Anwendung. Salmon (s. 4.2) bezeichnet die in seinem Erklärungsmodell vorkommenden Wahrscheinlichkeiten als „kausale Wahrscheinlichkeiten“. Suppes (1970) hat eine Konzeption entwickelt, die Kausalität auf Beziehungen zwischen Wahrscheinlichkeiten zurückführt. Ausgangspunkt ist dabei der Gedanke, daß für zwei aufeinanderfolgende Ereignisse A und B gelten kann:  $p(B/A) > p(B)$ , mit  $p(A) > 0$ . Dies ist noch nicht ausreichend für eine **probabilistische** Verursachung von B durch A, denn es könnte ein früheres Ereignis C geben, für das gilt:  $p(B/A \text{ und } C) = p(B/C)$ ; was eine Wirksamkeit von A in Frage stellen würde. Suppes führt einige Bedingungen an, unter denen nur eine **scheinbare** Ursache vorliegt und definiert anschließend durch Ausschluß dieser Bedingungen den Ursachenbegriff. - Ein damit verwandter Ansatz findet sich in Kap. 15, wo mit Hilfe probabilistischer bzw. korrelationsstatistischer Begriffe eine Definition der kausalen **Abhängigkeit** gegeben wird.

## **5. Eine Alternative zur Aussagenkonzeption: Der Non-Statement-View**

Zur Aussagenkonzeption wissenschaftlicher Theorien wurden eine Reihe von Alternativen entwickelt (vgl. Suppe, 1977). Die bekannteste davon ist der **Non-Statement-View**, inzwischen meist als **strukturalistische Theorienkonzeption** oder einfach **Strukturalismus** bezeichnet (Sneed, 1971; Stegmüller, 1973, 1979, 1986; Balzer, 1982). Herrmann (1976, s. auch Kap. 6) wendet den Strukturalismus im Rahmen seiner Konzeption psychologischer Forschungsprogramme an. Inzwischen wurden mehrere psychologische Theorien strukturalistisch rekonstruiert, unter anderem zu den Themen: Balancetheorie Heiders (Sukale, 1971), Theorie des operanten Verhaltens (Kraiker, 1977), Verhaltensinteraktionen in Zweierbeziehungen (Westmeyer, Eller, Winkelmann & Nell, 1982), allgemeine Theorie der Informationsverarbeitung (Ueckert, 1983), Theorie der kognitiven Dissonanz (Westermann, 1987), psychologische Nutzentheorie (Stephan, 1990). über weitere Arbeiten wird in Westmeyer (1989) berichtet. - Im folgenden können nur die Grundgedanken des Strukturalismus skizziert werden. Zusätzliche Ausführungen finden sich in den Kapiteln 6 und 10. Für ein eingehenderes Studium der strukturalistischen Vorgehensweise und ihrer Anwendung auf die Psychologie sei vor allem auf Westermann (1987) verwiesen. Eine kritische Analyse des Strukturalismus findet man in Gadenne (1984, 1986).

Eine Formulierung wie „ $T_e = M_e \times W_e \times A_e$ “ oder „Frustration führt zu Aggression“ hat, als Aussage verstanden, zwei Komponenten: Zum einen wird ein allgemeiner Zusammenhang beschrieben, den man je nach Inhalt als Ver-

haltungsmuster, mentalen Prozeß, Zusammenhang zwischen psychischen Eigenschaften usw. bezeichnen kann. Zum zweiten wird ausgesagt, daß das beschriebene Muster auf eine Klasse von Individuen zutrifft, daß jedes dieser Individuen die **Aussageform erfüllt**, durch die der Zusammenhang oder Prozeß dargestellt wird; erst dadurch wird aus der Aussageform eine **wahrheitsfähige Aussage**. Ein Individuum, das eine Aussageform erfüllt, heißt in der mathematischen Logik ein **Modell** dieser Aussageform. (Modelle in diesem Sinne sind zu unterscheiden von den mit Theorien verbundenen „idealen Modellen“ und weiterhin von „mathematischen Modellen“ im Sinne mathematisch formulierter Theorien in den Sozialwissenschaften. Letztere sind Aussagen bzw. Aussageformen, die „Modelle“ in der mathematischen Logik sind dagegen gerade umgekehrt die Entitäten, die Aussageformen erfüllen; vgl. zum Begriff des Modells auch Kap. 1, Abschnitt 2.1 dieses Bandes sowie Kap. 5). Wenn eine Theorie einen psychologischen Zusammenhang beschreibt, so gibt sie also implizit eine **Definition ihrer Modelle** und behauptet zugleich, daß alle Individuen (oder Gruppen) ihres Gegenstandsbereiches Modelle sind.

In der strukturalistischen Theorienkonzeption werden diese beiden Komponenten voneinander getrennt. Erstens werden durch **mengentheoretische Prädikate** Modelle definiert, wobei man sich der **informellen** (nicht voll formalisierten) Mengenlehre bedient (vgl. zur Axiomatisierung durch mengentheoretische Prädikate Suppes, 1960, 1967). Ein solches Prädikat hat die Form: Eine Entität  $x$  ist ein Modell der Theorie  $T$  genau dann, wenn . . . An dieser Stelle sind dann die Annahmen der jeweiligen Theorie anzugeben, einschließlich ihrer Anwendungsvoraussetzungen. Zweitens gibt es eine mit der Theorie verbundene **empirische Hypothese**, die aussagt, daß bestimmte empirische Fälle, etwa Individuen mit ihren Eigenschaften und Verhaltensmustern, als Modelle interpretierbar sind. Allein diese empirische Hypothese ist Gegenstand empirischer Forschung. Diese Trennung zwischen einer Theorie und einer zugehörigen empirischen Hypothese, die die Geltung der Theorie für bestimmte Fälle behauptet, ist nicht nur im Strukturalismus zu finden und ist auch nicht an die Axiomatisierung durch mengentheoretische Prädikate gebunden (vgl. etwa Toulmin, 1953).

Die empirische Hypothese ist in zweierlei Hinsicht eine schwächere Behauptung als eine theoretische Aussage im herkömmlichen Sinne. Letztere ist der Annahme äquivalent, daß **alle** Fälle eines zugrunde gelegten Gegenstandsbereiches Modelle sind. Die empirische Hypothese einer strukturalistisch rekonstruierten Theorie bezieht sich nur auf bestimmte Fälle des Gegenstandsbereiches, die **intendierten Anwendungen**. Diese bilden den „inhaltlichen“ Bestandteil einer Theorie: Wer eine empirische Theorie konstruiert und dazu gewisse abstrakte Formulierungen (eventuell Formeln) entwirft, faßt zugleich die Absicht, diese Formulierungen auf bestimmte empirische Fälle anzuwenden, im Rahmen der Leistungsmotivation etwa auf Aufgabenwahl, Ausdauer

und Anspruchsniveau. Von den intendierten Anwendungen wird aber nun nicht postuliert, daß sie Modelle sind (wie es der Aussagenkonzeption entsprechen würde), sondern daß sie „zu Modellen ergänzbar“ (entspricht etwa: als Modelle interpretierbar) sind. Die Aussage, daß ein empirischer Fall ein Modell **ist**, hält der Strukturalismus für nicht empirisch prüfbar.

Die strukturalistische Konzeption umfaßt ein spezielles Kriterium für Theoretizität bezüglich einer gegebenen Theorie T. Eine Variable ist T-theoretisch, wenn zur Bestimmung ihrer Werte die Annahme vorausgesetzt werden muß, daß ein Modell existiert. Das Kriterium läßt sich auch auf die Aussagenkonzeption übertragen. Dann gilt eine Variable als T-theoretisch, wenn zur Bestimmung ihrer Werte die Theorie T (jedenfalls ein Teil von ihr) vorausgesetzt werden muß. Es handelt sich um ein pragmatisches Kriterium. Daneben gibt es ein anderes, mehr formales Kriterium (vgl. Balzer, 1985), das jedoch für psychologische Theorien weniger geeignet sein dürfte (vgl. Westermann, 1987). Die Variablen „Erwartung“, „Anreiz“ und „Erfolgsmotiv“ sind nicht T-theoretisch, da zu ihrer Operationalisierung bzw. Messung die Axiome der Theorie Atkinsons nicht verwendet (vorausgesetzt) werden müssen. Dagegen sind die Variablen „Te“ und „Tm“ T-theoretisch, denn sie werden durch die Formeln A1 und A2 berechnet, die zu den zentralen Annahmen der Theorie gehören. Westermann zeigt, daß innerhalb von Festingers Theorie der kognitiven Dissonanz die Begriffe „dissonant“ und „konsonant“ (üblicherweise als „theoretisch“ eingestuft) nicht T-theoretisch sind, hingegen die Begriffe „Stärke kognitiver Dissonanz“ und „Reduktionsdruck“. Die Tatsache, daß Theorien T-theoretische Begriffe enthalten, ist der Grund dafür, daß nach strukturalistischer Auffassung Theorien im Sinne der Aussagenkonzeption nicht zirkelfrei prüfbar sind: Die Prüfung erfordert die Berechnung T-theoretischer Variablen, hierzu muß aber das, was geprüft werden soll (eben die Theorie T), verwendet werden.

Ein Gesetz wird im Strukturalismus als ein **Theorieelement** rekonstruiert, eine Theorie als ein Netz von **Theorieelementen**. Ein Theorieelement ist eine mathematische Struktur der folgenden Gestalt:  $\langle M, M_p, M_{pp}, C, I \rangle$ . Die ersten drei Zeichen stehen jeweils für eine bestimmte Menge von Modellen, C heißt Menge der „constraints“ (bei Westermann: „Eindeutigkeitsbedingungen“, unter anderem Skalenniveauanforderungen), I ist die Menge der intendierten Anwendungen. Eines dieser Theorieelemente, das sogenannte **Basiselement**, entspricht den Grundannahmen einer Theorie (Kernannahmen eines Forschungsprogramms), die anderen sind Rekonstruktionen spezieller abgeleiteter Aussagen. Ein Theorieelement ist selbst keine Aussage. Es ist überhaupt kein sprachliches Gebilde, also auch nicht zu verwechseln mit einem mengentheoretischen Prädikat. Allerdings wird ein Theorieelement jeweils mit Hilfe eines mengentheoretischen Prädikats definiert. Zwischen Theorieelementen können Relationen bestehen, unter denen die **Spezialisierung** (im strengen Sinne, vgl. Westermann, 1987) besonders wichtig ist. Ein Theorieelement  $\langle M', M_p', M_{pp}'$ ,

$C', I'$  heißt Spezialisierung eines Theorieelements  $\langle M, M_p, M_{pp}, C, I \rangle$ , wenn  $M', C'$  und  $I'$  jeweils Teilmengen von  $M, C$  und  $I$  sind, während  $M_p' = M_p$  und  $M_{pp}' = M_{pp}$ . Das erste Theorieelement entsteht aus dem zweiten dadurch, daß in das erste mengentheoretische Prädikat zusätzliche Bestimmungen aufgenommen werden. Um zu verstehen, was durch die Spezialisierungsrelation ausgedrückt werden kann, ist es hilfreich, die Aussagenkonzeption als Analogie heranzuziehen. Aus den Axiomen einer Theorie sind Theoreme ableitbar. Zum Teil sind diese Theoreme sehr allgemeiner Natur, zum Teil beziehen sie sich aber auf spezielle Anwendungssituationen, über die man mit Hilfe der Theorie etwas aussagen möchte. Wenn beispielsweise die Grundgesetze etwas über kognitive Dissonanz und ihre Reduktion im allgemeinen aussagen, können sich entsprechende abgeleitete Gesetze etwa auf den speziellen Fall der kognitiven Dissonanz nach Entscheidungen beziehen, auf Dissonanz nach forcierter Einwilligung, auf selektive Informationssuche nach kognitiver Dissonanz usw. Wie oben aufgezeigt wurde, erfordert die Ableitung solcher spezieller Gesetze in der Regel Zusatzannahmen, die bei der üblichen Formulierung einer Theorie nicht explizit erwähnt werden. Wenn man darüber hinaus „empirische“ Gesetze ableiten will, die mit Beobachtungssätzen konfrontierbar sind, werden zusätzlich noch Hilfsannahmen zur Operationalisierung benötigt. Wenn also von einer Theorie angenommen wird, daß sie auf eine bestimmte Klasse von Situationen (z.B. forcierte Einwilligung) erfolgreich anwendbar ist, so impliziert dies, daß für diese Situationen zum einen die Axiome, zum anderen die Zusatz- bzw. Hilfsannahmen gelten. Die Axiome bleiben für jede Anwendungssituation dieselben, während die Hilfsannahmen zum Teil situationsspezifisch sind. Nun wurde oben aufgezeigt, wie durch Axiome als Bestimmungsstücke eines mengentheoretischen Prädikats Modelle definiert werden können. Je mehr Bestimmungen das Prädikat enthält, desto enger ist die Klasse der Modelle. Im Strukturalismus werden die Grundgesetze einer Theorie durch das Basiselement rekonstruiert. Daher kann die Gesamtheit der Annahmen, die im Rahmen einer speziellen Anwendung gemacht werden, also die Axiome und die Zusatz- bzw. Hilfsannahmen, jeweils durch ein verschärftes mengentheoretisches Prädikat rekonstruiert werden, das dem mengentheoretischen Prädikat des Basiselements die erforderlichen zusätzlichen Bestimmungen hinzufügt. Das verschärfte mengentheoretische Prädikat definiert dann ein spezielleres Theorieelement  $\langle M', M_p', M_{pp}', C', I' \rangle$ . - Die verschiedenen Theorieelemente, die auf der Grundlage des Basiselements definierbar sind, bilden das erwähnte **Theoriennetz**. Das Netz entsteht durch die **Teilmengenbeziehungen** zwischen den Theorieelementen, ähnlich wie eine deduktive Struktur durch Ableitbarkeitsbeziehungen zwischen den „Elementen“ einer Theorie in herkömmlicher Darstellung, den Aussagen.

Wenn es sich im Verlauf empirischer Forschung zeigt, daß sich ein intendierter Anwendungsfall nicht als Modell interpretieren läßt (und dieser Mißerfolg



nicht auf Untersuchungs- oder Meßfehler zurückgeführt werden kann), so besteht die Möglichkeit, die bei dieser Anwendung involvierten speziellen Gesetze zu ändern oder aber den Anwendungsfall aus der Menge der intendierten Anwendungen auszuschließen. Die Grundgesetze gelten als gegenüber Falsifikation **immun**. Da der Strukturalist annimmt, daß Forscher die zentralen Annahmen ihrer Theorien in der Tat als nicht falsifizierbar betrachten, und daß diese Annahmen auch aus logischen Gründen nicht prüfbar sind, sieht er es als adäquat an, Theorieelemente als Strukturen zu rekonstruieren, die, anders als wahrheitsfähige Aussagen, „natürlicherweise“ immun sind. Die Immunität ist allerdings eine relative, vergleichbar derjenigen eines Paradigmas bei Kuhn oder eines Forschungsprogramms bei Lakatos: Wenn empirische Mißerfolge anhalten, kann es dazu kommen, daß auch die Grundannahmen aufgegeben werden, was einer wissenschaftlichen Revolution im Sinne Kuhns entspricht. Anders als Kuhn kennt der Strukturalismus allerdings einen „revolutionären wissenschaftlichen Fortschritt“, neben einigen Arten des Fortschritts in der „normalen“ Wissenschaft.

Daß der Strukturalismus stark durch Ideen Kuhns geprägt ist, zeigt sich nicht zuletzt an der These von der angeblichen **Immunität** von Theorien. Vor allem Stegmüller, dem der Strukturalismus seine Verbreitung im deutschsprachigen Raum verdankt, hat diesen Punkt immer wieder betont (1973, 1979, 1986) und erklärt, es wäre das „stärkste denkbare Argument gegen den Non-Statement-View“, wenn gezeigt werden könnte, daß Theorien als Aussagen empirisch prüfbar sind (1979, S. 15). Nun kann in der Tat gezeigt werden, daß eine Prüfbarkeit gegeben ist, wenn man nicht die isolierte Prüfung einer theoretischen Aussage fordert, sondern akzeptiert, daß mehrere Aussagen gleichzeitig einer Prüfung ausgesetzt werden können (Glymour, 1980; Gadenne, 1985, 1987; vgl. zur kontroversen Diskussion über diese Problematik: Balzer, 1987; Schurz, 1987a; Gadenne, 1988). - Dieser Aspekt darf allerdings in seiner Tragweite nicht überbewertet werden, da die strukturalistische Rekonstruktionsweise nicht mit ganz bestimmten methodologischen Regeln des Umgangs mit Theorien kombiniert werden muß, ebensowenig wie die Aussagenkonzeption etwa mit der Falsifikationstheorie Poppers gleichzusetzen ist. Westermann (1987) verbindet seine Version des Strukturalismus mit Elementen der Methodologie Poppers und hält den Strukturalismus im übrigen auch für neutral hinsichtlich Realismus und Instrumentalismus. Vergleiche zwischen beiden Theorienkonzeptionen dürfen daher nicht in den Fehler verfallen, die Kontroverse zwischen der Popperschen und der Kuhnschen Wissenschaftsauffassung zu reproduzieren.

Schurz (1987b) hält die Unterschiede zwischen Aussagenkonzeption und Strukturalismus für weit geringer, als gewöhnlich unterstellt wird und zeigt, wie man Aussagensysteme in mengentheoretische Strukturen überführen kann und umgekehrt. Insofern ist allerdings die von Kritikern gestellte Frage be-

rechtigt, welche Vorteile es haben soll, sich der Mühe der aufwendigen strukturalistischen Umdeutungsweise zu unterziehen (Feyerabend, 1977; Tuomela, 1978). Westermann (1987) sieht diese Vorzüge in der Möglichkeit einer „differenzierten Rekonstruktion“, die den Aufbau einer psychologischen Theorie durchschaubar macht. Ohne Zweifel tragen Rekonstruktionen wie diejenigen von Westmeyer et al. (1982), Westermann (1987) und Stephan (1990) zur Klärung des Aufbaus der betreffenden Theorie bis hin zu Fragen der Operationalisierung und Messung bei. Der Strukturalismus stellt ein Verfahren bereit, das zur Klärung einiger Fragen zwingt, die im Rahmen der in der Psychologie üblichen Darstellungsweise offen bleiben. Bisher ist allerdings nicht gezeigt worden, daß vergleichbare Klärungsleistungen nicht auch mit den Mitteln der Aussagenkonzeption möglich wären. Wer den Strukturalismus als Mittel in Betracht zieht, die Präzision psychologischer Darstellungen zu vergrößern, sei darauf hingewiesen, daß eine strukturalistische Rekonstruktion die Präzision einzelner theoretischer Annahmen nicht verändert. Eine mathematische Formel wie „ $T_e = M_e \times W_e \times A_e$ “ erscheint unverändert in einem mengentheoretischen Prädikat, dasselbe gilt aber auch für eine vage verbale Aussage; entsprechend unbestimmt ist dann die Menge der Modelle. Insofern ist es allerdings auch ein unbegründetes Vorurteil, daß der Strukturalismus nur dort anwendbar sei, wo mathematisierte Gesetze vorliegen.

## **Literatur**

- Albert, H. (1968). **Traktat über kritische Vernunft**. Tübingen: Mohr.
- Albert, H. (1987). **Kritik der reinen Erkenntnislehre**. Tübingen: Moor.
- Anderson, J. R. (1983). **The architecture of cognition**. Cambridge, Mass.: Harvard University Press.
- Armstrong, B. D. M. (1983). **What is a law of nature**. Cambridge: Cambridge University Press.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. **Psychological Review**, **64**, 359-372.
- Atkinson, J. W. (1964). **An introduction to motivation**. Princeton (N.J.): Van Nostrand.
- Atkinson, J. W. & Birch, D. (1970). **The dynamics of action**. New York: Wiley.
- Atkinson, J. W. & Litwin, G. H. (1960). Achievement motive and test anxiety conceived as motive to approach success and to avoid failure. **Journal of Abnormal and Social Psychology**, **60**, 52-63.
- Baars, B. J. (1988). **A cognitive theory of consciousness**. New York: Cambridge University Press.
- Balzer, W. (1982). **Empirische Theorien: Modelle - Strukturen - Beispiele**. Braunschweig: Vieweg.
- Balzer, W. (1985). **Theorie und Messung**. Berlin: Springer.

- Balzer, W. (1987). Sneeds Zirkel ist nicht wegzukriegen. Kritik an Volker Gadenne. **Conceptus**, 21, Nr. 52, 103-105.
- Barr, W.F. (1971). A syntactic and semantic analysis of idealizations in science. **Philosophy of Science**, 38, 258-272.
- Barr, W.F. (1974). A pragmatic analysis of idealizations in physics. **Philosophy of Science**, 41, 48-64.
- Bischof, N. (1981). Aristoteles, Galilei, Kurt Lewin - und die Folgen. In W. Michaelis (Hrsg.), **Bericht über den 32. Kongreß der Deutschen Gesellschaft für Psychologie in Zürich 1980** (Bd. 1, S. 17-39). Göttingen: Hogrefe.
- Block, N. (1980). Troubles with functionalism. In N. Block (Hrsg.), **Readings in Philosophy of Psychology**. Cambridge, Mass.: Harvard University Press.
- Bridgman, P. W. (1927). **The logic of modern physics**. New York: Macmillan.
- Bunge, M. (1967). **Scientific research, Bd. 1**. Berlin: Springer.
- Bunge, M. (1987). **Kausalität: Geschichte und Probleme**. Tübingen: Mohr.
- Carnap, R. (1939). **Foundation of logic and mathematics**. Chicago: University of Chicago Press.
- Carrier, M. & Mittelstraß, J. (1989). **Geist, Gehirn, Verhalten. Das Leib-Seele-Problem und die Philosophie der Psychologie**. Berlin: De Gruyter.
- Chalmers, A. F. (1986). **Wege der Wissenschaft**. Berlin: Springer.
- Churchland, P. (1981). Eliminative materialism and the propositional attitudes. **Journal of philosophy**, 78, 67-90.
- Churchland, P.M. (1984). **Matter and consciousness**. Cambridge: The MIT Press.
- Collingwood, R. G. (1940). **An essay on metaphysics**. Oxford: University Press.
- Coombs, C. H., Dawes, R. M. & Tversky, A. (1975). **Mathematische Psychologie**. Weinheim: Beltz.
- Cronbach, L.J. & Meehl, P. E. (1955). Construct validity in psychological tests. **Psychological Bulletin**, 52, 281-302.
- Cummins, R. (1983). **The nature of psychological explanation**. Cambridge, Mass.: MIT Press.
- Eimer, M. (1990). **Informationsverarbeitung und mentale Repräsentation**. Berlin: Springer.
- Einhorn, J. H. & Hogarth, R. M. (1986). Judging probable cause. **Psychological Bulletin**, 99, 3-19.
- Escalona, S. K. (1940). The effect of success and failure upon the level of aspiration and behavior in manie-depressiv psychoses. **University of Iowa Studies in Child Welfare**, 16, 199-302.
- Feather, N.T. (1961). The relationship of persistence at a task to expectation of success and achievement-related motives. **Journal of Abnormal and Social Psychology**, 63, 552-561.
- Feyerabend, P. (1976). **Wider den Methodenzwang**. Frankfurt am Main: Suhrkamp.
- Feyerabend, P. (1977). Changing patterns of recognition. **British Journal for the Philosophy of Science**, 28, 351-369.

- Fodor, J. (1968). **Psychological explanation**. New York: Random House.
- Fodor, J. (1975). **The language of thought**. New York: Crowell.
- Fodor, J.A. (1987). **Psychosemantics. The problem of meaning in the philosophy of mind**. Cambridge (Mass.): MIT Press.
- Gadene, V. (1984). **Theorie und Erfahrung in der psychologischen Forschung**. Tübingen: Mohr.
- Gadene, V. (1985). Theoretische Begriffe und die Prüfbarkeit von Theorien. **Zeitschrift für allgemeine Wissenschaftstheorie**, 16, 19-24.
- Gadene, V. (1986). Darstellung und Prüfbarkeit psychologischer Theorien aus der Sicht der Aussagenkonzeption und des Non-Statement-View. In M. Amelang (Hrsg.), **Bericht über den 31. Kongreß der Deutschen Gesellschaft für Psychologie in Heidelberg 1986**. Göttingen: Hogrefe.
- Gadene, V. (1987). Die These von der Zirkularität empirischer Theorien und der Non-Statement-View. **Conceptus**, 21, Nr. 52, 95-101.
- Gadene, V. (1988). Die methodologische Bedeutung des Sneed-Stegmüller-Zirkels. **Conceptus**, 22, Nr. 55, 123-125.
- Gadene, V. (1990). Unvollständige Erklärungen. In M. Sukale (Hrsg.), **Sprache, Theorie und Wirklichkeit**. Frankfurt am Main: Lang.
- Gadene, V. & Oswald, M. (1991). **Kognition und Bewußtsein**. Berlin: Springer.
- Glaserfeld, E. v. (1987). **Wissen, Sprache und Wirklichkeit**. Braunschweig: Vieweg.
- Glymour, C. (1980). **Theory and evidence**. Princeton, N.J.: University Press.
- Goldstein, I. & Papert, S. (1977). Artificial intelligence, language, and the study of knowledge. **Cognitive Science**, 1, 84-123.
- Graumann, C.F. (1965). Subjektiver Behaviorismus? **Archiv für die gesamte Psychologie**, 117, 240-251.
- Groeben, N. & Westmeyer, H. (1975). **Kriterien psychologischer Forschung**. München: Juventa.
- Haugeland, J. (1978). The nature and plausibility of cognitivism. **The Behavioral and Brain Sciences**, 2, 215-260.
- Heckhausen, H. (1980). **Motivation und Handeln**. Berlin: Springer.
- Hempel, C. G. (1958). Theoretician's dilemma. In H. Feigl, M. Scriven & G. Maxwell (Hrsg.), **Minnesota studies in the philosophy of science**, Bd. 2. Minneapolis: University of Minnesota Press.
- Hempel, C. G. (1977). **Aspekte wissenschaftlicher Erklärung**. Berlin: De Gruyter.
- Herrmann, T. (1973). **Persönlichkeitsmerkmale**. Stuttgart: Kohlhammer.
- Herrmann, T. (1976). **Die Psychologie und ihre Forschungsprogramme**. Göttingen: Hogrefe.
- Herrmann, T. (1979). **Psychologie als Problem**. Stuttgart: Klett.
- Herrmann, T. (1982). über begriffliche Schwächen kognitivistischer Kognitionstheorien: Begriffsinflation und Akteur-System-Kontamination. **Sprache und Kognition**, 1, 3-14.
- Herrmann, T. (1983). Nützliche Fiktionen. **Sprache und Kognition**, 2, 88-99.

- Hull, C. L. (1943). **Principles of behavior**. New York: Appleton-Century-Crofts.
- Hull, C.L. (1952). **A behavior System**. New Haven: Yale University Press.
- Jackendoff, R. (1987). **Consciousness and the computational mind**. Cambridge (Mass.): MIT Press.
- Kraiker, C. (1977). Behavioral analysis and the structural view of scientific theories. **European Journal of Behavioral Analysis and Modification**, 4, 203-213.
- Külpe, O. (1912, 1920, 1923). **Die Realisierung. Ein Beitrag zur Grundlegung der Realwissenschaften, 3 Bände**. Leipzig: Hirzel.
- Külpe, O. (1923). **Einleitung in die Philosophie**. Leipzig: Hirzel.
- Kuhn, T. S. (1962). **The structure of scientific revolutions**. Chicago: University of Chicago Press.
- Kuipers, T. A. F. (Hrsg.) (1987). **What is closer-to-the-truth?** Amsterdam: Rodopi.
- Lewin, K. (1936). **Principles of topological psychology**. New York: McGraw-Hill.
- Lewis, D. (1973). Causation. **Journal of Philosophy**, 70, 556-567.
- Lyons, W. (1986). **The disappearance of introspection**. Cambridge (Mass.): MIT Press.
- Mackie, J.L. (1974). **The cement of the Universe**. Oxford: Clarendon Press.
- Malewski, A. (1967). **Verbalten und Interaktion**. Tübingen: Mohr.
- Mandler, G. (1985). **Cognitive Psychology**. Hillsdale (N.J.): Lawrence Erlbaum.
- Maturana, H. R. (1985). **Die Organisation und Verkörperung von Wirklichkeit**. Braunschweig: Vieweg.
- McClelland, D.C. (1951). Measuring motivation in phantasy: The achievement motive. In H. Guetzkopf (Hrsg.), **Groups, leadership, and men**. Pittsburgh: Carnegie Press.
- McCorquodale, K. & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. **Psychological Review**, 55, 95-107.
- McCorquodale, K. & Meehl, P.E. (1954). Edward C. Tolman. In W.K. Estes et al. (Hrsg.), **Modern learning theory** (S.177-266). New York: Appleton-Century-Crofts.
- Miller, G. A., Galanter, E. & Pribram, K. H. (1960). **Plans and the structure of behavior**. New York: Holt, Rinehart and Winston.
- Moulton, R. W. (1965). Effects of success and failure on level of aspiration as related to achievement motives. **Journal of Personality and Social Psychology**, 1, 399-406.
- Murray, H. A. (1938). **Explorations in personality**. New York: Oxford University Press.
- Musgrave, A. (1979). Theorie, Erfahrung und wissenschaftlicher Fortschritt. In H. Albert & K. Stapf (Hrsg.), **Theorie und Erfahrung**. Stuttgart: Klett.
- Musgrave, A. (1981). Der Mythos vom Instrumentalismus in der Astronomie. In H.P. Duerr (Hrsg.), **Versuchungen: Aufsätze zur Philosophie Paul Feyerabends** (Bd. 2, S. 231-279). Frankfurt a.M.: Suhrkamp.
- Musgrave, A. (1982). Constructive empiricism versus scientific realism. **The Philosophical Quarterly**, 128, 262-271.
- Newell, A. & Simon, H. A. (1972). Human problem solving**. Englewood Cliffs (N.J.): Prentice-Hall.

- Norman, D.A. & Rumelhart, D. E. (1975). *Explorations in cognition*. San Francisco: Freeman.
- Nüse, R., Groeben, N., Freitag, B. & Schreier, M. (1991). *über die Erfindung/en des Radikalen Konstruktivismus*. Weinheim: Deutscher Studien Verlag.
- Oswald, M. (1980). *Jenseits der Methodologie des Behaviorismus*. Mannheim: Dissertationsdruck.
- Popper, K. (1972). *Objective knowledge*. Oxford: Clarendon Press.
- Popper, K. (1972). The aim of science. In K. Popper (Hrsg.), *Objective knowledge*. Oxford: Clarendon Press.
- Popper, K. (1983). *The Postskript to the logic of scientific discovery, Vol. 1: Realism and the aim of science*. Totowa, N.J.: Rowman and Littlefield.
- Putnam, H. (1960). Minds and machines. In S. Hook (Hrsg.), *Dimensions of mind*. New York: University Press.
- Pylyshyn, Z. (1984). *Computation and cognition*. Cambridge, Mass.: MIT Press.
- Rapaport, D. (1959). *Die Struktur der psychoanalytischen Theorie*. Stuttgart: Klett.
- Roth, G. (1986). Selbstorganisation - Selbsterhaltung - Selbstreferentialität: Prinzipien der Organisation der Lebewesen und ihre Folgen für die Beziehung zwischen Organismus und Umwelt. In A. Dress, H. Hendrichs & G. Küppers (Hrsg.), *Selbstorganisation (S. 149-180)*. München: Piper.
- Rorty, R. (1965). Mind-body identity, privacy, and categories. *Review of metaphysics*, **19**, 24-54.
- Salmon, W. C. (1971). *Statistical explanation and statistical relevance*. Pittsburgh: University of Pittsburgh Press.
- Salmon, W.C. (1984). *Scientific explanation and the causal structure of the world*. Princeton: University Press.
- Schlick, M. (1938). Die Kausalität in der gegenwärtigen Physik. In M. Schlick (Hrsg.), *Gesammelte Aufsätze (S. 41-82)*. Wien: Gerold.
- Schurz, G. (1987). Der neue Strukturalismus. *Conceptus*, **21**, Nr. 52, 113-127.
- Schurz, G. (1987). Der Sneed-Stegmüller-Zirkel. Bemerkungen zur Kontroverse Gadenne - Balzer. *Conceptus*, **21**, Nr. 52, 107-111.
- Searle, J. R. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences*, **3**, 417-457.
- Skinner, B. E. (1956). Critique of psychoanalytic concepts and theories. In P. G. Frank (Hrsg.), *The validity of scientific theories*. Boston: The Beacon Press.
- Sneed, J. D. (1971). *The logical structure of mathematical physics*. Dordrecht: Reidel.
- Stegmüller, W. (1973). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band 2,2. Halbband: Theorienstrukturen und Theoriendynamik*. Berlin: Springer.
- Stegmüller, W. (1979). *The structuralist view of theories*. Berlin: Springer.
- Stegmüller, W. (1983). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band 1, zweite, erweiterte Auflage: Erklärung - Begründung - Kausalität*. Berlin: Springer.

- Stegmüller, W. (1986). **Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band 2,3. Teilband: Die Entwicklung des neuen Strukturalismus seit 1973.** Berlin: Springer.
- Stephan, E. (1990). Zur **logischen Struktur psychologischer Theorien.** Berlin: Springer.
- Sukale, M. (1971). Zur Axiomatisierung der Balancetheorie: Eine wissenschaftstheoretische Fallstudie. **Zeitschrift für Sozialpsychologie, 2, 40-57.**
- Suppe, F. (1977). The search for philosophic understanding of scientific theories. In E. Suppe (Hrsg.), **The structure of scientific theories.** Urbana: University of Illinois Press.
- Suppes, P. (1960). **Axiomatic set theory.** Princeton: Van Nostrand.
- Suppes, P. (1967). What is a scientific theory? In S. Morgenbesser (Hrsg.), **Philosophy of science today** (S.55-67). New York: Basic Books.
- Suppes, P. (1970). **A probabilistic theory of causality.** Amsterdam: North-Holland.
- Tolman, E. C. (1932). **Purposive behavior in animals and men.** New York: Appleton-Century-Crofts.
- Toulmin, S. (1953). **Einführung in die Philosophie der Wissenschaft.** Göttingen: Vandenhoeck.
- Tuomela, R. (1978). On the structuralist approach to the dynamics of theories. **Synthese, 39, 211-231.**
- Ueckert, H. (1983). Computer-Simulation. In J. Bredenkamp & H. Feger (Hrsg.), **Enzyklopädie der Psychologie, Serie Forschungsmethoden der Psychologie, Band 5: Hypothesenprüfung.** Göttingen: Hogrefe.
- Van Fraassen, B. (1980). **The scientific image.** Oxford: Clarendon Press.
- Wann, T. W. (Hrsg.) (1964). **Behaviorism and Phenomenology.** Chicago: University of Chicago Press.
- Weiner, B. (1972). **Theories of motivation: From mechanism to cognition.** Chicago: Rand McNally.
- Weiner, B. (1980). **Human motivation.** New York: Holt, Rinehart and Winston.
- Wendel, H. J. (1990). **Moderner Relativismus.** Tübingen: Mohr.
- Westermann, R. (1987). **Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie.** Berlin: Springer.
- Westmeyer, H. (1973). **Kritik der psychologischen Unvernunft.** Stuttgart: Kohlhammer.
- Westmeyer, H. (1974). Statistische Analysen in der psychologischen Diagnostik. **Diagnostica, 22, 31-42.**
- Westmeyer, H. (Hrsg.) (1989). **Psychological theories from a structural point of view.** New York: Springer.
- Westmeyer, H., Eller, F., Winkelmann, K. & Nell, V. (1982). A theory of behavior interaction in dyads: A structuralist account. **Metamedicine, 3, 209-231.**
- Wundt, W. (1908). **Grundzüge der physiologischen Psychologie.** Leipzig: Engelmann.

## Heuristik der Theorienbildung

***Dietrich Dörner***

### ***1. Theorienbilden und Theorienprüfen als wissenschaftliche Tätigkeiten***

Die Tätigkeiten in den Wissenschaften lassen sich grob in zwei große Gruppen aufteilen, nämlich in die Tätigkeiten der Theorienprüfung und der Theorienbildung (s. auch Herrmann, Kapitel 6 dieses Bandes, der diese groben Kategorien erheblich ausdifferenziert). Die Verfahren der Theorieprüfung basieren auf logischen Prinzipien, sind daher formalisier- und algorithmisierbar und einer rationalen Diskussion leicht zugänglich. Aus diesem Grunde werden die Verfahren der Theorieprüfung in den Einzelwissenschaften und der Wissenschaftstheorie breit diskutiert, und oftmals wird die Methodik einer Wissenschaft sogar mit dem Instrumentarium zur Theorienprüfung identifiziert.

Mit der Theoriebildung verhält es sich anders. Theorien kommen, so meint Popper (1966, S. 7), aus dem Sumpf des irrationalen, amorphen Bereichs des menschlichen Geistes. Diesen Bereich nennt man oft „Intuition“ und verschleiert mit dieser Taufe eher, daß man über die damit gemeinten geistigen Prozesse nicht viel weiß. Kékulé (Chemieprofessor in Bonn in der zweiten Hälfte des 19. Jahrhunderts, einer der Mitbegründer der theoretischen organischen Chemie) soll die Struktur des Benzolmoleküls geträumt haben. Wie tief der Sumpf ist, aus dem Theorien stammen, kann man schon daraus ersehen, daß mir über die Entdeckung der Struktur des Benzolmoleküls durch Kékulé drei verschiedene Anekdoten bekannt sind, nämlich eine, in der sich die aus dem offenen Kamin in Bogen springenden Funken dem in die Flammen starrenden Kékulé zu Kreisen fügten, eine andere, in der dem Junggesellen Kékulé im Traum eine feurige Schlange erschien, und eine dritte, in der Kékulé auf der Fahrt durch Brüssel ein Kürschneremblem ins Auge fiel, auf dem sechs Äfflein einen Kreis bildeten, indem sie sich an den Händen faßten und dabei die Schwänze nach außen streckten! (Für keine dieser Anekdoten kenne ich die Quellen noch.) Mendelejeff fand die Struktur des Perioden-



Systems der Elemente beim Patiencelegen (s. Sergejew, 1970). Poincaré kam auf die Lösung eines mathematischen Problems, als er während einer Exkursion aus dem Omnibus stieg. Theorien stammen also aus den dunklen und chaotischen Bereichen des menschlichen Geistes, aus der Phantasie, aus dem Traum, dem plötzlichen Einfall, und daher scheint ihre Erzeugung weniger gut rational durchdringbar, formalisierbar und lehrbar zu sein als die Methoden der Theorienprüfung.

Es hat nicht an Versuchen gefehlt, den Sumpf zu drainieren und das Theorienbilden aus den Wissenschaften ganz zu eliminieren oder wenigstens in eine feste Form zu bringen. Ansätze zur Kanonisierung der Theorienbildung findet man schon bei Sir Francis Bacon, der vorschlug, statt zu spekulieren „im Buche der Natur zu lesen“, sich also auf die Beobachtung und - das wäre dann die Theorienbildung - die Generalisierung über die Einzelbeobachtungen zu beschränken. Ganz explizit wird dieses Verfahren im Positivismus von Comte und Mach vorgeschlagen.

Bei Popper (z.B. 1966, S.22) findet man die Tendenz, die Erkenntnistheorie, außer auf die „dogmatische“ Festsetzung der Basissätze (s. Popper 1966, S. 70) auf die „Erkenntnislogik“ der Überprüfung von Theorien zu beschränken und damit den Bereich der Theorienbildung ganz auszuklammern.

Auch wenn man Popper wohl zustimmen muß, wenn er sagt, daß die Tätigkeit des Aufstellens von Theorien „einer logischen Analyse weder fähig noch bedürftig“ sei (Popper, 1966, S. 6), sollte man ihm hinsichtlich der Einschränkung der Erkenntnistheorie auf die logische Überprüfung von Theorien nicht beipflichten. Die Theorienbildung ist in der Psychologie allgegenwärtig, daher sollte man sie nicht einfach im „Sumpf“ stecken lassen. Nicht nur der Forscher in der Psychologie, sondern auch der Praktiker ist ständig mit der Notwendigkeit konfrontiert, Theorien erfinden zu müssen. Welche Aspekte der Familienstruktur mögen wohl dafür verantwortlich sein, daß Frau X so depressiv ist? Oder liegt es gar nicht an der Familie)? Die eine Theorie, die alles erklärt, gibt es in der Psychologie nicht, und so muß man sich, ständig, Theorien neu erfindend, prüfend, revidierend, irgendwie durchwursteln. Wenn das aber so ist, darf man sich in der psychologischen Methodenlehre auf die Prüfmethode nicht beschränken.

Wie gewinnt man Theorien? In der Psychologie ist wohl der naiv-empiristische Gedanke, daß man mit Hilfe der Faktoren- oder Varianz- oder multiplen Regressionsanalyse theoretische Sätze „automatisch“ durch eine Art induktiver Logik aus den Daten gewinnen könne, noch weit verbreitet. Immer noch gibt es „faktorenanalytische Intelligenztheorien“. „Es ist also nicht so, daß wir unsere Erlebnisse sammeln, ordnen und so zur Wissenschaft aufsteigen; oder, wenn wir das mehr ‚formal‘ ausdrücken: daß wir, wenn wir Wissenschaft treiben wollen, zunächst Protokolle sammeln müssen“, meint Popper (1966, S.71) dazu.

Es gibt aber auch jenseits dieser Verfahren eine Reihe von Methoden, die man als Heuristiken für die Theoriengewinnung verwenden kann und üben und lehren sollte. Keineswegs als sichere Methoden auf dem Wege zu Erkenntnis, wohl aber als Hilfsmittel, um mit größerer Wahrscheinlichkeit zum Ziel zu kommen. „Logisch“ sind diese Verfahren bestimmt nicht; eine induktive Erkenntnislogik, die, nach festen Regeln angewendet, Theorien „automatisch“ aus den empirischen Daten gewinnt, kann es nicht geben. Dennoch lassen sich aber Verfahren zum Theorienbilden aufzählen, und man kann über ihre größere oder geringere Geeignetheit in einem bestimmten Fall auch „mit Gründen“, also rational, diskutieren.

Wir wollen uns im nachfolgenden Abschnitt zunächst einmal das Problem des Theorienkonstruktors vor Augen führen. Sodann wollen wir verschiedene Formen der Lösung dieses Problems untersuchen.

## **2. Die Welt der „schwarzen Kästen“**

Stellen Sie sich einen schwarzen Kasten vor. An der Frontseite sehen Sie - sagen wir - acht Lampen. Diese Lampen können in verschiedenen Farben leuchten und zeigen zu einem bestimmten Zeitpunkt eine bestimmte Farbkombination, sagen wir z.B. rot-blau-blau-rot-gelb-grün-blau-gelb. Eine kurze Inspektion des Kastens zeigt, daß diese Farben wechseln können. Manchmal tun sie dies anscheinend „spontan“. Aber man kann die Farbwechsel auch beeinflussen. Es befinden sich nämlich an dem Kasten einige Tasten, sagen wir 10. Betätigt man diese Tasten, so wirkt sich das auf die Lampen aus. Sie wechseln die Farben. Manche Tasten allerdings scheinen ohne Wirkung zu sein.

An sich könnte uns ja dieser schwarze Kasten reichlich gleichgültig sein, wir wollen aber einmal annehmen, daß wir in einer Welt leben, die nur aus schwarzen Kästen besteht. Dann müssen wir herausfinden, wie sie sich verhalten, damit wir mit ihnen kooperieren können. Wir müssen wissen, in welcher Weise der „Output“ (oder die „Reaktionen“) von dem „Tasten-Input“ (oder von den „Reizen“) abhängig sind.

Nun gibt es in unserer Kastenwelt verschiedenartige Kästen. Manche sind sehr groß und weit weg. Ihre Farbkombinationen sind ziemlich stabil. Tasten haben sie nicht, oder diese sind nicht erreichbar. Bei dem einen leuchten die ersten vier Lämpchen rot, die anderen blau. Ein anderer Kasten zeigt die Farbfolge rot-blau-rot-blau-rot-blau-grün-grün, ein dritter gelb-gelb-gelb-rot-gelb-blau-gelb-grün, usw. Manchmal verfällt einer der Kästen in eine Phase hektischer Aktivität und sendet eine chaotisch erscheinende Folge von farbigen Signalen. Dann beruhigt er sich wieder. Nennen wir diese Kästen „Astronomenkästen“.

Da man wissen will, wieso es in der Kastenwelt so geworden ist, wie es ist, müssen sich manche Leute auch mit Kästen befassen, die längst vermodert sind, und von denen nur noch Verhaltensprotokolle existieren. Man weiß aber nicht genau, ob diese Protokolle richtig sind. Meist sind sie auch lückenhaft. Man kennt auch die Tastenbetätigungen nicht genau, die mit dem jeweils protokollierten Farbwechsel verknüpft waren. Die Leute, die sich mit solchen längst vermoderten Kästen befassen, wollen wir „Historiker“ nennen.

Wieder andere Personen befassen sich mit Kästen, die prompt auf Tastenbetätigungen reagieren. Die Tasten sind gut zugänglich und sie haben eindeutige Wirkungen. Auch kann man die Kästen zersägen und die so entstandenen Teilkästen isoliert untersuchen, was die Erforschung der Gesetze, die Input und Output verknüpfen, erheblich erleichtert. Fügt man die Teilkästchen wieder zu einem Gesamtkasten zusammen, so zeigt es sich, daß man die Teilergebnisse verwenden kann, um das Verhalten des Gesamtkastens zu erklären und zu prognostizieren. Die einzelnen Teilkästchen scheinen auch recht homogen zu sein; kennt man einen wirklich genau, so auch alle anderen. Diese Forschungsobjekte wollen wir „Physikerkästen“ nennen.

Andere Forscher sehen sich mit Kästen konfrontiert, die ein verwirrenderes Verhalten zeigen. Manche Tasten-Lämpchen-Kombinationen dieser Kästen verhalten sich so ähnlich wie bei den Kästen der Physiker. Bestimmte Tastenbetätigungen haben bestimmte Veränderungen der Farbkombinationen zur Folge. Aber leider sind auch noch andere Tasten da, deren Betätigung entweder überhaupt nichts ändert, oder die manchmal so und manchmal ganz anders wirken. Taste 3 z.B. schaltet manchmal die Lampe 5 in den Zustand gelb. Manchmal bewirkt aber Taste 3 überhaupt nichts und manchmal tauscht sie anscheinend die Zustände der Lampen 4 und 8 aus. Taste 4 schaltet Lampe 6 nur, wenn Lampe 8 rot leuchtet.

Man hat auch den Eindruck, daß manche Tasten nicht direkt, sondern erst nach einer bestimmten Zeit, also mit Verzögerung wirken. Manche Lämpchen unseres schwarzen Kastens verändern sich anscheinend auch „von selbst“. Verwirrenderweise scheinen bestimmte Tasten auch gar nicht auf den Lampenzustand zu wirken, sondern die Wirkung anderer Tasten zu verändern. Nennen wir diese Kästen „Psychologenkästen“.

Was soll diese Kastenwelt? Sie soll demonstrieren, daß Theorienkonstrukteure vor ziemlich verschiedenen Aufgaben stehen, je nachdem, mit welcher Art von Kästen sie konfrontiert sind. Ehe man also über Theorienbildung redet, sollte man sich darüber klar werden, mit welcher Art von „schwarzem Kasten“ man sich auseinandersetzen hat. Davon ist die Art des Vorgehens abhängig. Wenn man Kästen zersägen kann, kann man anders mit ihnen umgehen, als wenn das nicht geht, weil zwischen den Teilkästen Interaktionen vorhanden sind.

Die Tatsache, daß man die „Art des Kastens“ beachten muß, ehe man die Forschungsstrategie festlegt, war bereits Aristoteles klar, wenn er sich fragt, ob man vernünftigerweise beim Studium psychologischer Sachverhalte „von oben“, also vom Ganzen ausgehend, oder „von unten“, also von den Teilen ausgehend, vorgehen sollte. („... Ferner aber, wenn nicht viele Seelen, sondern viele Seelenteile vorhanden sind, ob man zuerst die ganze Seele zu untersuchen hat oder ihre Teile. Schwer ist auch, bei diesen zu bestimmen, welche voneinander unterscheidbar sind, und ob man zuerst die Teile untersuchen muß oder ihre Leistungen; z.B. das Denken oder den denkenden Geist und das Wahrnehmen oder das Wahrnehmungsvermögen ...“. Aristoteles 1986, S. 6).

Bei unseren „Physikerkästen“ stellen sich diese Probleme nicht. Der Physiker (genauer gesagt: der Physiker des vergangenen Jahrhunderts) hat es relativ einfach. Die Tasten seines Kastens haben direkte Wirkungen, man braucht nicht anzunehmen, daß in dem Kasten irgendein verborgenes, eigenaktives Innenleben existiert. Offensichtlich gibt es im Inneren des Kastens nur einen Draht von der Taste direkt zur Lampe. Das Vorgehen des Physikers bei der Erforschung seiner Kästen kann ziemlich einfach sein. Er experimentiert mit seinem Kasten und merkt sich, welche Wirkungen jeweils die Inputs haben, und wenn er es sich nicht merken kann, da der Lämpchen und der Tasten zu viele sind, schreibt er es sich auf. Er kann dann aus seinen Protokollen ablesen, was es mit seinem Kasten auf sich hat. Input a produziert Output 1 - Input b produziert Output 2 - Input c produziert Output 3, usw.

Die Leute mit den Astronomenkästen haben es da schon schwerer. Sie sehen kaum Veränderungen, und sie können auch nicht auf Tasten drücken, trotzdem wollen sie gerne wissen, warum sich Kasten I im Zustand rot-gelb-blau-rot-gelb-blau-grün-grün befindet und Kasten II im Zustand gelb-gelb-gelb-rot-rot-rot-blau-blau. Wie kann man das herauskriegen? Nun, vielleicht sieht man, daß Kasten II doch einen glänzenderen Lack hat und keine abgestoßenen Ecken, also irgendwie „neuer“ aussieht als Kasten I. Also könnte man die Hypothese aufstellen, daß der jeweilige Output altersabhängig ist, und daß auch Kasten II sich einmal, wenn er entsprechend gealtert ist, in dem Zustand befinden wird, den Kasten I jetzt schon hat.

Ziemlich schwierig ist natürlich die Situation der Historiker. Sie müssen zunächst einmal versuchen, herauszubekommen, wie die Protokolle eigentlich „wirklich“ aussehen. Wie soll man da vorgehen? Vielleicht findet man zu dem einen Protokoll, welches man vervollständigen will, ein anderes, welches an der Stelle, wo Protokoll Nr. 1 Lücken hat, keine aufweist. Und wenn man nun annimmt, daß die beiden Protokolle von ähnlichen Kästen stammen, so kann man die Lücken des Protokolls Nr. 1 durch das, was man im Protokoll Nr. 2 vorfindet, auffüllen.

Der Theorienkonstrukteur ist mit einem „schwarzen Kasten“ konfrontiert, und es ist seine Aufgabe, herauszufinden, wie dessen innere Struktur aussieht. Eine allgemeine „Theorie der Theorienbildung“ müßte sich mit allen Arten von Kästen befassen und jeweils angeben, welche Methode der Forschung der jeweiligen „Natur“ des zu erforschenden Kastens am meisten gerecht wird. Eine solche allgemeine Theorie wäre selbstverständlich sehr wünschenswert und sehr reizvoll; wir wollen uns aber zunächst einmal auf die schwarzen Kästen der Psychologen beschränken. Was sollen Psychologen mit ihren verwirrenden Kästen machen?

### ***3. Wie kommt man in das Innere von schwarzen Kästen?***

Nach hundert Jahren Existenz einer eigenständigen akademischen Psychologie will es einem zwar nicht mehr so recht von den Lippen, daß die Psychologie eine „junge“ Wissenschaft ist. Dennoch kam sie später als viele andere, und die Versuchung lag nahe, sich der Denkmodelle älterer Wissenschaften zu bedienen. In unserem Kastengleichnis hieße das, daß die Psychologen in der Wissenschaft die Tendenz hatten (oder haben), ihre „Kästen“ so zu betrachten, als wären sie genau so beschaffen, wie die Kästen anderer Wissenschaften. Besonders die Physikerkästen haben auf die Psychologen immer einen so starken Reiz ausgeübt, daß Leahey (1987, S. 25) von einem „Physikneid“ der wissenschaftlich tätigen Psychologen als Berufssyndrom spricht. Die Analogisierung der eigenen „Kästen“ mit denen anderer Wissenschaften ist eine Methode, um in das Innere der Kästen zu gelangen. Es gibt aber noch andere.

Nachfolgend wollen wir über die Methoden sprechen, die man verwenden kann, um das Innenleben von schwarzen Kästen zu studieren. Zunächst soll uns die Form von Theorien interessieren (Abschnitt 3.1), dann wollen wir untersuchen, wie man das Hohlgefäß einer bestimmten Form mit Inhalten füllen kann.

#### **3.1 Die Form der Theorie**

Wenn man einen undurchsichtigen schwarzen Kasten hypothetisch mit einem Inhalt füllen soll, so ist es vernünftig, sich zunächst einmal zu überlegen, wie denn die Form oder Struktur dieses Inhalts beschaffen sein soll. Soll man nur einfache Verbindungen von den Tasten zu den Lampen annehmen oder kompliziertere Strukturen?

In der Psychologie gibt es hinsichtlich der Formen von Theorien eine reichhaltige Tradition. Die älteste Auffassung ist wohl die, daß sich im Inneren des schwarzen Kastens etwas „ganz anderes“ befindet. Die Tasten sind der Input-

Rand des Kastens; sie wirken im Inneren auf eine eigenständige, nichtmaterielle Entität, „Seele“ genannt; diese bedient sich einer Reihe von „Vermögen“, um die Inputs in die Outputs umzuwandeln. So etwa Thomas von Aquin, der diese Auffassung von Avicenna (alias Ibn Sina) übernahm (s. Leahey, 1987, S. 71ff.), der sie wiederum aus Aristoteles' Schrift über die Seele herausdestilliert hatte. Die Instanzentheorie des Aristoteles und das, was Thomas von Aquin daraus gemacht hat, ist im Zeitalter des Kognitivismus mit seiner Wiederbelebung von Instanzenlehren betrachtenswert (man lese in diesem Zusammenhang die Kritik Herrmanns, 1982), aber über einen immateriellen „Steuermann“ für diese Instanzen redet wohl heute - zumindest in der wissenschaftlichen Psychologie - kaum jemand mehr.

Eine der ersten Formhypothesen der wissenschaftlichen Psychologie ist der „Drahthypothese“ über die Verbindung von Taste und Lampe, die wir oben skizzierten, sehr ähnlich. Es war die Reflexbogen-, Reiz-Reaktions- oder - wie man sie auch genannt hat - die „Telefonzentralenhypothese“ des Behaviorismus. Von außen kommen in den Organismus „Anrufe“ hinein, diese werden auf den Ausgang umgeschaltet; die entsprechenden „Verdrahtungen“ liegen entweder fest oder aber werden durch Lernprozesse „umgestöpselt“. Vorbildstiftend war hier wohl die S-R-Theorie (Stimulus-Response) Thorndikes (s. Neel, 1969, S.116ff.). Ihm gelang es (seiner Meinung nach), einen Großteil „höherer“ Geistestätigkeiten auf Reiz-Reaktions-verknüpfungen zurückzuführen, und er stiftete damit eine Tradition, alle psychologischen Phänomene auf die Form von S-R-Schemata zurückzuführen. Das „Umstöpseln“ im Telefonamt der Seele, also die Veränderung von S-R-Schemata, ist natürlich besonders interessant, und daher legte der Behaviorismus einen großen Wert auf das Studium der Lernprozesse. Bei Guthrie reduzierte sich die gesamte Psychologie auf den einen Satz: „Eine Kombination von Reizen, die mit einer Bewegung einhergeht, pflegt beim erneuten Auftreten diese Bewegung nach sich zu ziehen“ (nach Bischof, 1981, S.21).

Eine reine Reiz-Reaktions-Psychologie ließ sich nicht durchführen und wurde bald abgelöst von einer anderen Strukturhypothese, die bis heute die Anfängervorlesungen in der Allgemeinen Psychologie beherrscht, nämlich vom S-O-R-Schema. Gemäß diesem Schema ist die Reaktion (also der „Output“) abhängig einmal von der gegebenen Reizsituation S und zum zweiten von „innerorganismischen“ Größen O. (Interessant ist hier, daß von innerorganismischen Größen geredet wird, also eine materielle Grundlage „innerer“ Prozesse angenommen wird.)

Eine Präzisierung dieses S-O-R-Schemas ist das Konzept des „Automaten“ der Automatentheorie (s. Aiserman et al., 1967, S.62ff.). Wir charakterisieren dieses Konzept in Abbildung 1. Gemäß diesem Konzept ist die „Reaktion“ r eines Systems abhängig von der Reizsituation zur gleichen Zeit und vom

inneren Zustand. Dabei ist der innere Zustand wiederum abhängig von der vorausgehenden Reizsituation und dem vorhergehenden inneren Zustand; seine Veränderbarkeit ist - im Gegensatz zum S-O-R-Schema - expliziert.

Formal ist ein Automat ein Quintupel  $[s, r, z, R, Z]$ . Dabei ist  $[s]$  die Menge der möglichen Reizeingänge,  $[r]$  die Menge der möglichen Reaktionen,  $[z]$  die Menge der möglichen inneren Zustände.

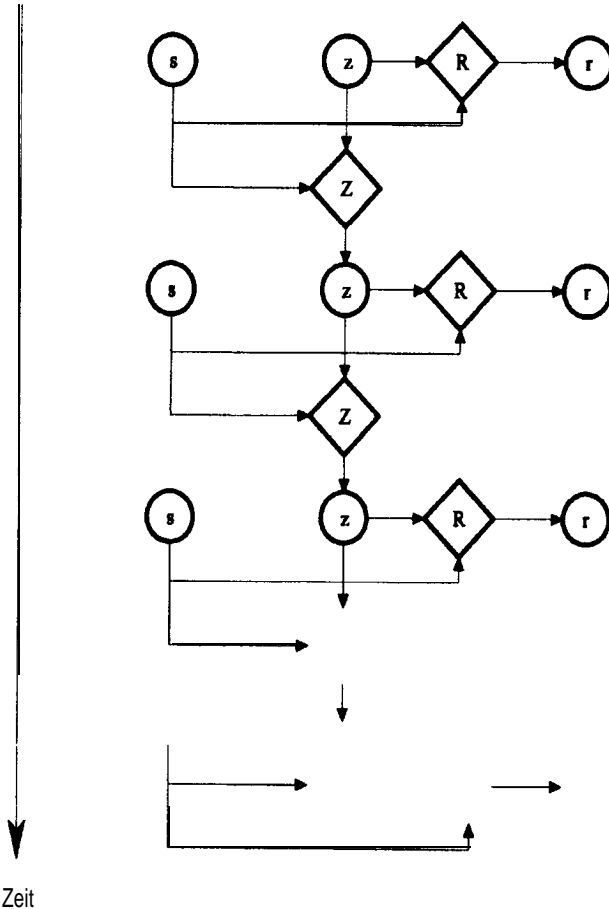


Abb. 1: Relationen und Zeitverhalten eines Automaten

$R$  ist die „Reaktionsfunktion“. Sie ist allgemein festgelegt als

$$r = R(s_t, z_t).$$

Die Reaktion zu einem bestimmten Zeitpunkt ist also von dem inneren Zustand und dem „Reiz“ zum gleichen Zeitpunkt abhängig.

Die „Zustandsfunktion“  $Z$  ist definiert als

$$z = Z (s_{t-1}, z_{t-1}).$$

Dies bedeutet also, daß der innere Zustand abhängig ist von der Reizsituation und dem inneren Zustand vorher. (Das was wir soeben festgelegt haben, ist ein sogenannter **Vergangenheits-Vergangenheits-Automat**. Es gibt auch andere Formen von Automaten, der Vergangenheits-Vergangenheits-Automat ist aber die allgemeinste Form, auf die sich die anderen zurückführen lassen.) Die automatentheoretische Form einer Theorie ist sehr allgemein verwendbar, aber natürlich auch außerordentlich grob. Zwischen verschiedenen Reizklassen, verschiedenen Arten von inneren Zuständen und verschiedenen Reaktionsformen wird nicht unterschieden.

Der all(zu-)gegenwärtige Gebrauch statistischer Methoden in der Psychologie hat ebenfalls zu bestimmten, allgemeinen Formen von Theorien geführt. Im Zusammenhang mit varianzanalytischen Verfahren und Verfahren der Faktorenanalyse und der multiplen Regression werden theoretische Sätze gern in die Form des „Allgemein Linearen Modells“ (ALM, s. z.B. Bortz, 1985, S. 578ff.) gebracht. Dieses lautet:

$$\begin{aligned} y = & x_1 b_1 + x_2 b_2 + x_3 b_3 + \dots + x_n b_n \\ & + x_1 x_2 b_{1,2} + x_1 x_3 b_{1,3} + \dots + x_1 x_n b_{1,n,1} \\ & + \dots \\ & + x_1 x_2 \dots x_n b_{1,2,\dots,n}. \end{aligned}$$

Das bedeutet: eine Variable ist von  $n$  anderen Variablen so abhängig, daß ihre Ausprägung sich aus der Summe der Ausprägungen der „Input-Variablen“, jeweils gewichtet mit einem spezifischen Faktor  $b_i$  ergibt; hinzu kommen die Zweier-, Dreier-, . . . - Interaktionen.

Gebraucht man Verfahren der multiplen Regression oder der Faktorenanalyse, so reduziert man meist dieses allgemeine Prinzip zu einer speziellen Form, nämlich zu:

$$y = x_1 b_1 + x_2 b_2 + x_3 b_3 + \dots + x_n b_n.$$

Die Interaktionen fallen weg (s. Bortz, 1985, S.584).

Der Gebrauch der genannten statistischen Verfahren zwingt die Daten „automatisch“ in dieses theoretische Gerüst. Daß dieses Schema keineswegs das einzig denkbare ist, ist den Verwendern solcher Verfahren sehr oft gar nicht klar. Das Schema enthält z.B. die bedenkliche Annahme der **Kompensation**; die niedrige Ausprägung der einen Prädiktorvariablen kann durch die hohe Ausprägung einer anderen kompensiert werden. Wenn man z. B. in faktorenanalytischen Intelligenztheorien die Annahme findet, daß die Leistung in einem bestimmten Intelligenztest von den Faktoren „Merkfähigkeit“, „Fähigkeit



zum Umgang mit sprachlichem Material“, usw. abhängt, so geht in dem kompensatorischen Modell des ALM die Unterscheidung „notwendiger“ und nur „hilfreicher“ Faktoren verloren. Faktisch wird sich ja eine bestimmte Intelligenzleistung meist nicht nur aus der Addition einzelner Fähigkeiten ergeben; fällt die Fähigkeit zum Behalten (von Zwischenergebnissen) aus, so ergibt sich keine Minderleistung, sondern überhaupt keine Leistung mehr. Zwischen Merkfähigkeit und den anderen Fähigkeiten, die zur Erbringung einer bestimmten Intelligenzleistung notwendig ist, existiert eine logische „und“-Beziehung: beides muß vorhanden sein. Arithmetisch wäre dieses „und“ ein „\*“ und kein „+“ (welches ein logisches „oder“ ist), wie in dem Allgemeinen Linearen Modell! Wir gehen im Abschnitt 3.3 („Positivismus“) noch einmal auf dieses Problem ein.

Jenseits umgreifender Formvorstellungen für Theorien in der Psychologie gibt es Schemata auch für Theoriedetails. Ein solches ist das TOTE-Schema. Dabei handelt es sich um eine psychologische Fassung des Regelkreisprinzips. TOTE ist ein Akronym für Test-Operate-Test-Exit und bedeutet: wenn festgestellt (Test) wird, daß ein Istzustand nicht dem Sollzustand entspricht, dann soll eine (auf die Erreichung des Sollzustandes gerichtete) Aktivität begonnen oder fortgesetzt (Operate) werden. Entspricht der Istzustand nun dem Sollzustand (Test), so wird die Aktivität beendet (Exit), sonst wird die Aktivität weiter fortgesetzt. Dieses Schema, welches von Miller, Galanter & Pribram (1961) in die Psychologie eingeführt wurde, brachte zaghaft in eine behavioristische Psychologie wieder so etwas wie Zweck und Ziel und Motivation zurück.

## 3.2 Die Introspektion und das „Verstehen“

Im vorangegangenen Abschnitt sind wir auf verschiedene Theorieformen eingegangen. Wir haben gewissermaßen die Form theoretischer „Gefäße“ betrachtet. In diesem und den folgenden Abschnitten wollen wir nun diskutieren, wie man die Formen mit Inhalten füllen kann.

Die in der Psychologie am meisten naheliegende Methode, einen schwarzen Kasten mit Inhalt zu füllen, ist die Introspektion. Man betrachtet sich selbst beim Denken, Fühlen und Wollen. Immerhin ist es ja ziemlich sinnfällig, daß man selbst dem schwarzen Kasten, den man studieren möchte, recht ähnlich ist. Also sollte man es doch schaffen, durch Selbstbetrachtung an allgemeine Sätze über das Innere des schwarzen Kastens zu kommen. Und wenn es nicht um einen selbst, sondern um einen anderen geht, so sollte eine analoge Nachkonstruktion des anderen Hypothesen über dessen Innenleben heranschaffen können. Wenn sich jemand in einer bestimmten Situation in einer bestimmten Weise verhält, so versuche ich mich in seine Situation zu versetzen und überlege mir, welche Motivationen und Gefühle und welche Art von Situations-

sicht ich in Abhängigkeit von diesen Motivationen und Gefühlen an der Stelle dieses Anderen hätte. Und wenn ich dann bei mir ein gleichartiges Verhalten ableiten kann, dann habe ich das Gefühl, ich hätte die andere Seele verstanden. Bei dieser Form des einführenden Verstehens mache ich mich zum Modell des anderen. Mein eigenes Seelenleben, meine Motive, Wünsche, Werthaltungen benutze ich, um die Werthaltungen, Denk- und Entscheidungsprozesse des anderen zu konfigurieren.

Gegen die Introspektion als Selbstanalyse oder auch in der Form der analogen Nachkonstruktion fremden Seelenlebens hat man vielerlei eingewendet, z. B. daß die Methode nicht „objektiv“ sei (s. z.B. Roth, 1984, S. 125 f.). Die noch so große subjektive Überzeugung von der Richtigkeit meiner introspektiven Erkenntnisse ersetzt keine objektive Prüfung.

Die Tatsache, daß die Ergebnisse der Introspektion nicht „objektiv“ sind, war wohl der Hauptgrund dafür, daß eine verächtlich als „Lehnstuhlwissenschaft“ bezeichnete introspektionistische Psychologie gegenüber der „empirischen“ Psychologie in die Hinterhand geriet. Dabei verkannte und verkennt man aber wohl, daß sich Introspektion und die Verwendung „objektiver“ (also bezüglich der Ergebnisse intersubjektiv überprüfbarer) empirischer Verfahren nicht zu widersprechen brauchen. Introspektion dient der Hypothesenfindung. Sie ist subjektive Empirie. Ich erfahre mich selbst, aber kein anderer kann meinen Selbsterfahrungen zustimmen oder sie ablehnen, da diese keinem außer mir zugänglich sind. Aber warum sollte ich meine Selbsterfahrungen nicht zur Hypothesengewinnung benutzen? Hat man die Hypothesen erst einmal, kann man sich ja um deren objektive Prüfung bemühen.

Ich kann z. B. das zukünftige Verhalten des anderen, den ich „verstehe“, voraussagen und damit prüfen, ob tatsächlich das sichtbare Verhalten folgt, was gemäß meinen Annahmen folgen müßte. Man sollte daher das introspektionistische Verfahren der Aufklärung der inneren Struktur des schwarzen Kastens nicht leichtfertig auf den Müllhaufen der „vorwissenschaftlichen“ Methoden werfen. Warum auch? Zumindest die innere Struktur meiner **eigenen** Seelenprozesse ist mir zum Teil zugänglich. Warum soll ich diese partielle Zugänglichkeit nicht verwenden, um die innere Struktur anderer „schwarzer Kästen“, die meinem „schwarzen Kasten“ doch ähnlich zu sein scheinen, hypothetisch zu erhellen? Bereite ich mir nicht durch den Verzicht auf diese Methode freiwillig Schwierigkeiten, die andere Wissenschaften (wie z. B. die Biologie) gern los wären (s. Traxel, 1968, S.49)? Als Methode der Hypothesengenerierung ist die Introspektion auf jeden Fall brauchbar. Kritisch wird es nur, wenn man in der Freude darüber, etwas „verstanden“ zu haben, vergißt, daß die Erfindung einer Hypothese kein Wahrheitsbeweis ist. Generierung und Prüfung von Theorien darf man nicht verwechseln. Die Tatsache, daß

dies oft doch geschah, hat wohl viel dazu beigetragen, daß die Methode der Introspektion in Mißkredit geriet.

Allerdings hat die Methode der Introspektion auch als Verfahren der Hypothesenerfindung ihre Grenzen, nämlich folgende:

1. Veränderung des Ablaufs. Das was ich bewußt betrachte, ist nicht das, was ohne Betrachtung ablaufen wurde.
2. Selbst meine eigenen Seelenprozesse sind mir nicht gänzlich zugänglich. Nicht erst seit Freud wissen wir von der Existenz unbewußter Prozesse.
3. Meine eigenen Seelenprozesse sind mir nur auf einem relativ niedrigen Grad der Auflösung zugänglich. Ich weiß, daß ich soeben den Kollegen L. auf dem Hof des Institutes gesehen habe. Ich weiß aber nicht, wie die Mustererkennungsprozesse im einzelnen abgelaufen sind, die mir diese Erkenntnis beschert haben.
4. Das Verhalten und die Seelenprozesse von Personen in Extremsituationen, z. B. in Situationen starken und andauernden Hungers, der Lebensbedrohung sind meiner „analogen Nachkonstruktion“ allenfalls zugänglich, wenn ich solche Situationen selbst schon einmal erlebt habe.
5. Die Seelenprozesse abnormer Personen sind mir nur in engen Grenzen zugänglich. Zwar meinte z. B. Goethe, es gäbe kein Verbrechen, als dessen Urheber er sich nicht vorstellen könnte, dennoch kann sich wohl kaum jemand in das Seelenleben eines Menschen, der in einem Zeitraum von 3 Monaten 12 kleine Mädchen in kleine Stücke zerschneidet, hineindenken. Auch die Erlebniswelt eines Schizophrenen bleibt uns wohl zum großen Teil verschlossen.

### 3.3 Der „Positivismus“

#### **3.3.1 Die „analytische Prozedur“ und Wechselwirkungen**

„Die Physik! Das ist eine vernünftige Wissenschaft!“, sagten und sagen sich viele Psychologen. „Genau wie sie müssen wir es auch machen!“ An vernünftige Theorien über schwarze Kästen kommt man, indem man sie zunächst einmal zersägt, um dann die einzelnen Segmente zu studieren. Da die Teilkästen homogen sind, muß man beim Zersägen keine große Sorgfalt walten lassen. Es ist also im Grunde gleichgültig, welchen der Kästen man untersucht. In der Physik sind Pendel, schiefe Ebene, die Bewegung des Mondes um die Erde im Grund alle verschiedene Realisationen eines Prinzips, und dieses Prinzip müssen wir in der Psychologie eben auch finden. (Zu dieser „Homogenitätsannahme“ in der Psychologie s. Bischof, 1981, S.21ff.). Nachdem man die Kästen zersägt hat, wählt man also einen von ihnen aus und erforscht sein Verhalten, indem man systematisch die verschiedenen Tasten und Tastenkom-

binationen durchprobiert. Man schreibt auf, welcher Effekt sich einstellt, und verallgemeinert dann. Wenn man bei Neumond und bei Vollmond, morgens und abends, im Frühling, Sommer, Herbst und Winter immer beobachten konnte, daß die Betätigung der Taste 1 die einzige Lampe unseres Kastensegmentes in den Zustand „rot“ versetzt, so soll man doch mit Fug und Recht in Zukunft behaupten können: „Immer wenn man Taste 1 betätigt wird, leuchtet die Lampe rot!“

Wissenschaften, die ihre schwarzen Kästen zersägen, nennt man „analytische Wissenschaften“ und das Verfahren der isolierten Variation der Inputs, um auf diese Weise die Gesetze der Input-Output-Beziehungen zu erfassen, nennt Bertalanffy (1968, S. 18f.) die „analytische Prozedur“.

Die Theorienbildung ist bei der „analytische Prozedur“ denkbar einfach. Theorienbildung bedeutet hier lediglich die Verallgemeinerung der Einzeleffekte über Raum und Zeit. Diese Art der Theorienbildung ist zuerst explizit im Rahmen des Positivismus (Comte, Mach, nach Leahey, 1987, S. 144ff.) aufgekomen. Besonders von Psychologen wird sie oft mit **der** naturwissenschaftlichen Methode gleichgesetzt. Und so setzt der empirisch arbeitende Psychologe die eine Hälfte seiner Versuchspersonen unter Stress, die andere Hälfte nicht, läßt die Versuchspersonen Dreisatzaufgaben lösen und verkündet hinterher: „Streß vermindert die Leistung beim Lösen von Mathematikaufgaben um 30 %!“

Das Hantieren mit den Kästen mag im einzelnen mühselig sein, da die Tasten schwer erreichbar sind oder die Lampen schlecht beobachtbar. Und man muß vielleicht viel Scharfsinn entfalten, um mit den Kästen in der angestrebten Weise experimentieren zu können. Für die Theorienbildung selbst aber braucht man nur wenig Scharfsinn. Dafür mag richtig sein, was Heidegger (nach Wandschneider, 1991, S.XVII) über den Unterschied von Philosophie und Naturwissenschaft meinte: „Die Naturwissenschaft denkt nicht!“ Das Verhalten kann von einer voreingestellten Meßapparatur registriert werden; es werden keine Hypothesen über Zusammenhänge erfunden. Die einzige theoretische Arbeit besteht in der Festlegung der Beobachtungskategorien (man entschließt sich zur Betrachtung der Tasten und Lampen und nicht zur Betrachtung der Lichtreflexe auf der Holzmaserung des Gehäuses) und in der Generalisierung über Raum und Zeit. Das ist nun zwar nicht nichts, aber auch nicht viel.

Es wird von Röntgen berichtet, daß er auf die Frage eines Reporters, was er denn so gedacht hätte, als er seine Knochenhand auf dem Projektionsschirm sah, indigniert geantwortet hätte: „ich dachte nicht, ich beobachtete!“ Einer solchen Aussage liegt die Doktrin von Sir Francis Bacon zugrunde, daß man die Naturgesetze im „Buche der Natur“ ablesen könne. Denken stört.

Nirgends sonst muß man auf die Möglichkeiten der Existenz sehr verschiedener „schwarzer Kästen“ so hinweisen, wie bei den Verfechtern dieser Art empirischer Forschung und Theorienbildung. Der Erfolg dieser Art von Forschung stellt sich nämlich nur dann ein, wenn die „Kästen“ eine bestimmte Struktur haben; sie müssen - kurz gesagt - interaktionsfrei sein (s. Bertalanffy, S. 19). Das Segment A eines Kastens muß sich isoliert genauso verhalten, wie als Teil eines größeren Kastens. Wenn dies nicht der Fall ist, wenn z.B. das Segment A sein Verhalten auch in Abhängigkeit von dem Zustand, in dem sich gerade das Segment B befindet, ändert, dann wird eine isolierte Betrachtung des Segmentes A zwar Ergebnisse bringen; diese sind aber im Hinblick auf seine „wahre“ Natur falsch.

Betrachten wir dies einmal etwas genauer. Nehmen wir einmal an, wir wurden uns mit dem Kasten der Abbildung 2 beschäftigen. Dieser Kasten ist so gestaltet, daß der Druck auf die Taste A dazu führt, daß die Lampe rot leuchtet, aber nur, wenn zugleich die Taste des Segments B gedrückt wird, sonst leuchtet die Lampe von A blau. Abbildung 2 zeigt die Struktur des Kastens als Blockdiagramm. Es ist die denkbar einfachste Form einer „Wechselwirkung“, d.h. der Wirkung einer Variablen auf eine zweite (hier der Taste A auf die Lampe), die in ihrer Form von einer dritten Variablen (der Taste B) beeinflusst wird. Die Wirkung der Taste A auf die Lampe wechselt also, je nach Zustand der Taste B.

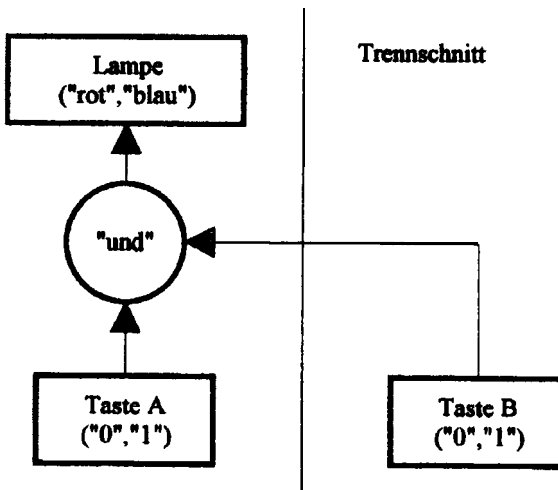


Abb. 2: Ein einfacher Fall einer Wechselwirkung

Zersägt man den Kasten und betrachtet A isoliert, dann kann verschiedenes der Fall sein. Wenn der mitzersägte Verbindungsdraht zwischen A und B z. B. nun dauernd so reagiert, als sei Taste B gedrückt, dann wird Taste A ein rotes Licht erzeugen, was sonst nur vorkäme, wenn die Taste B betätigt wird. Wenn

der zersägte Verbindungsdraht dauernd so reagieren würde, als sei Taste B nicht gedrückt, so wird die Taste A gar keinen Einfluß auf die Lampe haben; diese würde im Zustand „blau“ verharren. Es könnte auch sein, daß der Verbindungsdraht seine Durchtrennung übel nimmt und - „verrückt spielend“ - manchmal den einen, manchmal den anderen Zustand einnimmt. In diesem Fall hätte die Betätigung der Taste A manchmal den Effekt „rot“, manchmal keinen Effekt. (Die drei soeben geschilderten Annahmen über das Verhalten des durchtrennten Drahtes entsprechen den in der experimentellen Psychologie üblichen Methoden des Umgangs mit Variablen, die nicht systematisch variiert werden. Man hält sie konstant oder läßt sie zufällig variieren.)

Wichtig ist nun, daß jeder der isolierten Befunde, die man in diesen drei verschiedenen Fällen erhalten würde, nicht eine Teilerkenntnis über den Gesamtkasten zutage fördern würde, sondern drei falsche Erkenntnisse.

Der Satz:

A<sub>1</sub>) Immer wenn Taste A, dann „rot“!

ist genau so falsch wie

A<sub>2</sub>) Die Taste A hat keine Wirkung!

oder

A<sub>3</sub>) Immer wenn Taste A, dann „rot“ mit  $p = 0.5$ ,  
„blau“ (bzw. „keine Wirkung“) mit  $p = 0.5$ !

Es würde auch nichts helfen, wenn man später die Erkenntnisse über die gleichfalls isoliert untersuchte Wirkung der Taste B den Erkenntnissen über Taste A hinzufügte. Für die isoliert untersuchte Wirkung der Taste B auf die Lampe ergeben sich unter den oben geschilderten Annahmen drei analoge „Erkenntnismöglichkeiten“ B<sub>1</sub>, B<sub>2</sub> und B<sub>3</sub>. Die jeweiligen Erkenntnisse A<sub>i</sub> und B<sub>j</sub> muß man - gemäß der Annahme, daß die Tasten A und B isoliert wirken - entsprechend dem Additionstheorem der Wahrscheinlichkeiten durch das logische „oder“ zusammenfügen. Für die Addition einer „Erkenntnis“ B<sub>j</sub> zu einer „Erkenntnis“ A<sub>i</sub> gibt es 9 verschiedene Möglichkeiten. Sie sind im Hinblick auf die wahre Struktur des Kastens ausnahmslos falsch!

Es hilft nichts: wenn Wechselwirkungen vorliegen, darf man die Input-Variablen nicht isoliert untersuchen, wenn man Wert darauf legt, die Wahrheit über schwarze Kästen zu erfahren. Die experimentellen Verfahren der Konstanthaltung oder Zufallsvariation der nicht im Experiment betrachteten Variablen versagen hier.

Man kann den durch isolierende Betrachtung gewonnenen Ergebnissen auch nicht nachträglich Ergebnisse hinzufügen, um auf diese Weise durch die Einführung von „Korrekturfaktoren“ die „Teilerkenntnisse“ weiter zu präzisieren (s. Gadenne in diesem Band, Kapitel 6). Oder: Man kann es schon; es kommt

aber nicht viel dabei heraus, was die tatsächliche Struktur des Kastens erhellt. Wenn der jeweils durchtrennte Draht beispielweise „verrückt spielen wurde“, also in zufälliger Weise signalisieren wurde, daß die Taste A oder B jeweils gedrückt oder nicht gedrückt sei (= Zufallsvariation der nicht berücksichtigten Variablen), so erhielt man bei der isolierenden Erforschung unseres Kastens die beiden „Erkenntnisse“:

A<sub>3</sub>) „Wenn Taste A, dann „rot“ mit  $p = 0.5!$ “

B<sub>3</sub>) „Wenn Taste B, dann „rot“ mit  $p = 0.5!$ “

Daraus läßt sich ableiten:

„Wenn Taste A UND Taste B, dann „rot“ mit  $p = (0.5 + 0.5 - 0.5 * 0.5) = 0.75!$ “

Das aber ist falsch, da gemäß der Konstruktion des Kastens gilt:

„Wenn Taste A UND Taste B, dann „rot“ mit  $p = 1.0!$ “

Im Beispielsfall ist die Wechselwirkung denkbar einfach! Die vorgestellten Betrachtungen gelten aber für kompliziertere Formen von Wechselwirkungen in verstärktem Maße. Auf die Problematik der Verwendung der Verfahren der experimentellen Psychologie bei wechselwirkungsbehafteten Systemen habe ich an anderer Stelle hingewiesen (Dörner, 1988). Ich halte das Problem der für die Ermittlung der Struktur wechselwirkungsbehafteter Systeme angemessenen Forschungsstrategie für das methodische Problem der Psychologie. Die zeitgenössische Methodologie in der Psychologie widmet diesem Problem kaum Aufmerksamkeit; Gadenne (in diesem Band, Kapitel 6, insbesondere 6.3.3) streift dieses Problem kurz, wenn er über „Idealisierungen“ und „Vereinfachungen“ spricht, geht aber leider nicht genauer darauf ein und schiebt die entsprechenden Ausführungen von Bischof (1981) in den Keller der „vgl.“-Anmerkungen. Dagegen hält Herrmann (1989, S. 10) ausdrücklich eine nicht-experimentelle Strategie bei der Erforschung „vernetzter“ Systeme für angemessen. Eine solche Strategie muß in „langfristigem, geduldigem Beobachten und dem spielerischen Variieren von Bedingungen ohne die strengen Regeln des Experiments“ bestehen.

Quintessenz dieser Betrachtung ist: Die „analytische Prozedur“ bringt nur dann generalisierbare Resultate, wenn die Segmente des Kastens nicht durch Wechselwirkungen miteinander verbunden sind.

Wir haben in diesem Abschnitt die Merkmale wechselwirkungsbehafteter Systeme sehr stark betont. Diese Betonung scheint uns im Hinblick auf die zeitgenössische Methodenlehre wichtig. Man sollte aber natürlich nicht vergessen, daß es auch in der Psychologie nicht (oder schwach) wechselwirkungsbehaftete Systeme gibt, für die die analytische Prozedur durchaus eine angemessene Forschungsstrategie darstellen kann. Viele Vorgänge im elementaren

Bereich der psychischen Prozesse, z.B. elementare Wahrnehmungsprozesse, Figur-Wahrnehmung, Hell-Dunkel-Adaptation, auf der Outputseite die Produktion von Phonemen, standardisierte Handlungen, usw. lassen sich sehr wohl „analytisch“, d.h. unter weitgehender Vernachlässigung ihrer Einbettung untersuchen. Ob aber z.B. die Betrachtung von Denkabläufen ohne Berücksichtigung der emotionalen und motivationalen Einbettung sehr sinnvoll ist, erscheint mir fraglich. Bei der Untersuchung komplexer Denkabläufe findet man zahlreiche Phänomene, die sich ohne Bezug auf emotionale und motivationale Begleitprozesse nicht erklären lassen.

Die positivistische Art der Theoriengewinnung durch Korrelation von Input und Output und durch Generalisierung war in der wissenschaftshistorischen Situation um 1850 durchaus verständlich. Der Positivismus wendete sich gegen die überbordende Spekulation, die sich nicht an der Erfahrung orientierte, sondern „freischwebend“ Annahme auf Annahme häufte. Der radikale Empirizismus des Positivismus von Comte und Mach, der alles aus der Wissenschaft verbannen wollte, was nicht direkt beobachtbar ist, war der „roll back“ einer Wissenschaft, die sich allzu weit von der immer zu fordernden Kontrollierbarkeit theoretischer Konstruktionen entfernt hatte (s. Leahey, 1987, S. 144ff.). Pendel schlagen von einer Extremlage meist in die andere.

Wie alle Programme ist das positivistische Programm allerdings nie ganz durchgeführt worden, und zwar ganz einfach deshalb, weil es auch in der Physik überhaupt nicht auf alle Gegenstände paßte. (Man sollte unsere „Physikerkästen“, wie wir sie oben beschrieben haben, ja nicht mit tatsächlichen Forschungsgegenständen der Physiker verwechseln!) Von Physikern ist vielmehr zu hören: „Ich hoffe, daß niemand mehr der Meinung ist, daß Theorien durch zwingende logische Schlüsse aus Protokollbüchern abgeleitet werden, eine Ansicht, die in meinen Studententagen noch sehr in Mode war. Theorien kommen zustande durch ein vom empirischen Material inspiriertes Verstehen, welches am besten im Anschluß an Plato als zur Deckung kommen von inneren Bildern mit äußeren Objekten und ihrem Verhalten zu deuten ist“ (Pauli 1957).

Der Grund für solche nicht mit der positivistischen Methode der Theoriengewinnung im Einklang befindliche Stellungnahmen ist die Tatsache, daß besonders die Quantenphysik sowohl mit Wechselwirkungen als auch mit zeitlichen Abhängigkeiten konfrontiert wurde, die das Programm einer nur korrelativen Feststellung von Naturgesetzen zumindest in diesen Bereichen in Frage stellte.

Wie dem auch sei: Die positivistische Methode ist ein Verfahren zur Gewinnung von theoretischen Sätzen, welches bestimmten Gegenständen angemessen ist, anderen aber nicht. Ehe man sie anwendet, sollte man sich Gedanken über die Art des schwarzen Kastens machen, den man erforschen möchte.



Um Mißverständnisse zu vermeiden: Wir haben über die positivistische Methode als Methode zur Gewinnung theoretischer Sätze gesprochen. Wir haben hier nicht über die *Prüfung* theoretischer Sätze gesprochen. Auch wenn ein komplexes, wechselwirkungsbehaftetes System vorliegt, ist es oft möglich, ein einfaches Experiment durchzuführen, um die Richtigkeit von Aussagen über das System zu prüfen. Wir werden weiter unten, in einem anderen Zusammenhang (s. Abschnitt 3.5.2), dafür ein Beispiel zeigen. Wenn man exakte Annahmen über die Struktur eines Systems hat, lassen sich auch für komplexe, wechselwirkungsbehaftete Systeme einfache Prüfoxperimente ersinnen.

### 3.3.2 **Positivismus und Statistik**

Die Übernahme der positivistischen Forschungsstrategie der Physiker machte den Psychologen einige Schwierigkeiten: Eine davon ist, daß die psychologischen Kästen gewöhnlich nicht mit der Eindeutigkeit reagieren, wie das die physikalischen tun. Drückt man auf den Knopf 1, so leuchtet manchmal die dritte Lampe rot auf, manchmal aber auch grün, manchmal gelb, manchmal blau - **aber** doch meistens rot. Dasselbe gilt für andere Tasten und Tastenkombinationen. Es scheint, als wenn sich bei den „psychologischen Kästen“ zwischen Taste und Lampen gewissermaßen ein Wackelkontakt eingeschlichen hätte. Die „Wackelkontakttheorie“ ist in der Psychologie weit verbreitet; die gewöhnliche Reaktion darauf ist die Behandlung der Ergebnisse der empirischen Forschung mit Hilfe der Statistik, als einer spezifischen, für wackelkontaktbehaftete Systeme besonders geeigneten Form der Mathematik.

Wie geht man mit unsicherheitsbehafteten Daten um? Statt eines Gesetzes der Form

wenn  $x$ , so  $y$ ,

bekommt man durch Generalisierung über die Ergebnisse nun stochastische Gesetze der Form

wenn  $x$ , so  $y_1$  mit  $py_1$ ,  
 $y_2$  mit  $py_2$ ,  
 $y_3$  mit  $py_3$ .

Ist das die allgemeine Form eines psychologischen Gesetzes? Dagegen sträubt sich etwas! Soll man solche Wahrscheinlichkeitsgesetzmäßigkeit so interpretieren, als wäre die menschliche Seele ein Zufallsgebilde, welches mit zahlreichen Wackelkontakten versehen manchmal dieses, manchmal jenes Verhalten hervorbringt? Da die meisten Menschen wohl der Meinung sein werden, daß ihr Verhalten nur zu einem verschwindenden Teil vom Zufall bestimmt wird, liegt eine andere Antwort nahe: „In Wirklichkeit“ wirkt  $x$  auf  $y$  nicht über

einen Wackelkontakt. „In Wirklichkeit“ wirkt  $x$  auf  $y$  zusammen mit vielen anderen, im Augenblick nicht erfassbaren, verborgenen Faktoren (Kryptodeterminismus). Das können innere oder äußere Faktoren sein. Um bei unserer Kastenmetapher zu bleiben: Es mag sein, daß außer der Taste  $x$  noch andere Tasten zur gleichen Zeit gedrückt werden, was man aufgrund der Komplexität des Tastenfeldes nicht überschauen kann. Es mag aber auch sein, daß „innere“ Variable eine Rolle spielen, daß also die Reaktion  $y$  nicht allein durch die äußeren Umstände bestimmt wird, sondern außerdem noch durch verborgene Faktoren. Wenn man nun die Reaktion  $y$  genauer bestimmen will, so muß man versuchen, so viele der verborgenen Faktoren, wie nur immer möglich, zu erfassen und mit zu berücksichtigen.

Bleibt man bei der positivistischen Methode der Gewinnung theoretischer Sätze aus den Daten, so verwendet man nun das „Allgemeine Lineare Modell“ als Strukturierungsform für die Daten. Die Verwendung der Faktorenanalyse in der Psychologie ist das historisch am meisten bedeutsame Beispiel für die positivistische Form der Theorienbildung auf statistischer Grundlage (s. Bortz, 1985, S. 615ff., Herrmann, 1969, S. 98ff.). In der Faktorenanalyse verwendet man die empirisch gefundenen Korrelationen zwischen Variablen zur Gewinnung theoretischer Sätze und kam so z.B. zu den zahlreichen „faktorenanalytischen Intelligenztheorien“ (s. Hofstätter, 1960, S. 172ff.).

Gegen das Verfahren der Reduktion komplexer Zusammenhänge auf einzelne bzw. auf eine Addition einzelner Einflußgrößen wendet sich Bischof (1981) mit seiner Unterscheidung „aristotelischer“ und „galileischer“ Wissenschaften. Bischof vertritt die Auffassung, daß man allgemeine Gesetze in der Psychologie nicht durch eine fortschreitende Abstraktion von der Systemstruktur finden kann, also nicht durch Zerlegung des Gesamtsystems in einzelnen Kästchen. Bischof betont, daß in der Biologie und in der auf der Biologie basierenden Psychologie „Organisation“ das Kernprinzip ist und nicht „Ordnung“, wie in der Physik. Ordnung bedeutet bei Bischof Gleichförmigkeit, Symmetrie und Harmonie. Das System „Lebewesen“ ist nicht durch Symmetrien, Gleichförmigkeit und Harmonien gekennzeichnet, sondern durch eine teleonome, d.h. „zweckbestimmte“ Organisation, und diese ist in einem System miteinander vernetzter Variablen niedergelegt, die man nicht ohne die Folgen der Denaturierung des Gegenstandes in ihre Teile zerlegt. Der (bei Bischof, 1981, S. 32 zitierte) Versuch von Schneirla (1959), die Stärke einer aversiven Reaktion allein auf die Reizstärke zurückzuführen, ist als allgemeine Regel falsch. Ein leiser, aber naher Reiz wird u.U. eine viel stärkere Schreckwirkung entfalten, als das entfernte Donnerrollen. Unerwartetheit kommt als Determinante hinzu. Unerwartetheit, Nähe (indiziert durch das Überwiegen hoher Frequenzen im Klangbild des Reizes) und die Semantik des Reizes (zischt hier eine Schlange?) bilden ein System, dessen Komponenten man nicht isoliert betrachten darf.

Ganz davon abgesehen, daß man formale Bedenken gegen die Verwendbarkeit des „allgemeinen linearen Modells“ haben kann, stößt die empirische Erfassung von Zusammenhängen gemäß diesem Modell mit wachsender Zahl der untersuchten Variablen sehr bald auf Grenzen, die in der über alle Maßen wachsenden Anzahl von Daten liegt, die man dann braucht.

Es ist wohl unmöglich, das wechselwirkungsbehaftete Innenleben eines „schwarzen Kastens“ allein durch die Manipulation des Inputs, Beobachtung des Outputs und Generalisierung über die Ergebnisse empirisch zu erforschen. Was kann man sonst tun? Darauf wollen wir in den nächsten drei Abschnitten eingehen.

### 3.4 Die Heuristik der Zweckmäßigkeit

Eine wichtige Methode, um an theoretische Sätze zu kommen, ist die von Bischof (1981) so genannte „Heuristik der Zweckmäßigkeit“. Warum hören Leute bei dem Verzehr von Kartoffeln relativ bald nach der Deckung ihres Energiedefizits mit der Nahrungsaufnahme auf? Und warum ist beim Verzehr von Schokolade nicht das gleiche der Fall?

Wenn man diese Frage beantworten will, so kann man nach der biologischen Zweckmäßigkeit dieses Phänomens fragen. Man könnte z.B. folgendermaßen theoretisieren: Süßigkeiten traten in der natürlichen Umwelt des Menschen sehr selten auf. Zucker ist ein sehr guter Energielieferant. Also könnte die Natur den Menschen und andere Lebewesen so programmiert haben, daß man, wenn man diesen guten Energielieferanten nur irgend habhaft werden kann, ihn in möglichst großen Mengen zu sich nehmen sollte, um ihn in entsprechenden Fettpölsterchen zu speichern. Über Jahrhunderttausende war das ziemlich unproblematisch; die wenigen Bienennester und süßen Äpfel, die unsere in den Savannen umherstreifenden Vorfahren dann und wann einmal fanden, haben ihre „schlanke Linie“ nicht wesentlich gefährdet und ihr Herzinfarktrisiko nicht bedeutsam vergrößert.

Die Heuristik der Zweckmäßigkeit ist in der Psychologie und in der Biologie letztlich ein Art von Simulation der Evolution. Die Frage nach den Determinanten des Verhaltens wird beantwortet, indem man ein System konstruiert. Annahmen sind dann gerechtfertigt, wenn sie die Überlebenschancen des Systems vergrößern. Das ist das Kriterium! Warum gibt es Gefühle? Die stören doch nur! Der Wutanfall verdirbt meist mehr, als er nützt. Der Verliebte hat keine klare Realitätssicht mehr und geht unter. Ist das so? Sind dies nicht vielleicht nur die „Ausrutscher“ eines im Großen und Ganzen sehr gut funktionierenden Systems der Anpassung der Form der Informationsverarbeitung an die Umstände? Ist es nicht vernünftig, daß jemand, der mit 160 km/h auf

der Autobahn auf einen sich querstellenden Laster zurast, nur noch sehr grob wahrnimmt? Wo ist eine Lücke, durch die man entkommen kann? Wenn „Angst“ unter anderem eine Verringerung des Auflösungsgrades der Wahrnehmung ist, ist dann Angst nicht vernünftig? Wenn man solche Fragen stellt, verwendet man bereits die „Heuristik der Zweckmäßigkeit“.

Diese Heuristik besteht darin, daß man sich fragt, wie ein System beschaffen sein muß, damit es die auf die Befriedigung seiner Bedürfnisse gerichtete Aktivität möglichst effektiv reguliert. Das Kriterium für das Beibehalten oder Verwerfen einer Hypothese ist dabei nicht, ob diese empirisch belegt werden kann oder nicht. Das Kriterium ist, ob das System, wenn eine bestimmte Hypothese hinzugefügt oder weggelassen wird, seine Aufgaben besser erfüllt.

Dieses Kriterium ist gewissermaßen evolutionistisch. Eine Hypothese überlebt, wenn ihre Einführung oder Beibehaltung von Vorteil oder sogar notwendig für die Funktion des Systems ist. Hinter der Verwendung dieses Kriteriums steckt die Idee, daß die Evolution auch unsere Seelenausstattung nicht einfach willkürlich „erfunden“, sondern sich dabei „etwas gedacht“ hat. Im großen und ganzen (keineswegs immer!) sollten auch Seelenzustände und -prozesse einen Sinn für das bessere Funktionieren des Organismus haben.

### 3.5 Analogien

Die von Mary B. Hesse (1970) als „Königsweg der Fortentwicklung der Wissenschaft“ bezeichnete Methode der Erfindung von Annahmen über die innere Struktur von schwarzen Kästen ist der Analogieschluß. Was ist das?

Der Analogieschluß besteht darin, daß man an die Stelle der unbekannt inneren Struktur eines Realitätsausschnittes probenhalber eine bekannte Struktur einsetzt und auf diese Art und Weise neue Hypothesen über den Gegenstand entwickelt.

Wenn beispielsweise Rutherford eine Theorie über die innere Struktur des Atoms dadurch erfand, daß er das Atom mit einem Planetensystem analogisierte, so konstruierte er die unbekannt innere Struktur eines Atoms probenhalber entsprechend einer bekannteren aus dem Makrokosmos. Der Ausgangspunkt dieser Entwicklung war die Beobachtung, daß Strahlen durch eine Metallfolie, die nur wenige Atome dick war, manchmal ohne Ablenkung durch die Folie hindurchgingen, als ob diese überhaupt nicht vorhanden wäre. Manchmal wurden die Strahlen aber durch irgendetwas abgelenkt. Also mußten sich zwischen den Atomen „irgendwie“ Freiräume befinden. Wie soll man sich aber ein Medium vorstellen, welches „dichtgepackt“ aus Atomen besteht und manchmal Strahlung verschluckt oder reflektiert oder ablenkt, manchmal aber völlig unbehindert durchläßt?

Die Lösung von Rutherford und Bohr war die Annahme, daß Atome nicht kleine kompakte Masseklümpchen sind, sondern Gebilde, die selbst wieder aus kleineren Teilchen zusammengesetzt sind. Die Elektronen kreisen - wie die Planeten um die Sonne - um einen Atomkern. Die hilfreiche Analogie war hier das Sonnensystem.

Abbildung 3 zeigt das analoge Schließen bei der Erfindung neuer Hypothesen schematisch. Es existiert ein nur teilweise bekannter Sachverhalt, der aus Komponenten und Relationen besteht. Im oben genannten Beispiel sind das die Atome und die Strahlung und die Beziehungen zwischen ihnen. Man hat nun Grund zur Annahme, daß es noch andere Komponenten und Relationen gibt; die bislang ermittelte Struktur ist unbefriedigend. Um beim Beispiel zu bleiben: Irgendwie müssen zwischen Atomen und der Strahlung noch andere Beziehungen vorhanden sein als die eines Wasserstrahles, der auf einen Ziegelstein trifft. Und die Relationen zwischen den Atomen müssen auch „irgendwie“ anders sein, als die von Ziegelsteinen in einer Mauer zueinander. „Irgendwie“ haben die Ziegelsteine Löcher oder sind nicht dichtgepackt, so daß der auf die Ziegelmauer gerichtete Wasserstrahl oft völlig unabgelenkt die „Mauer“ passiert. Diesen Stand des Problems symbolisiert die Skizze A in Abbildung 3.

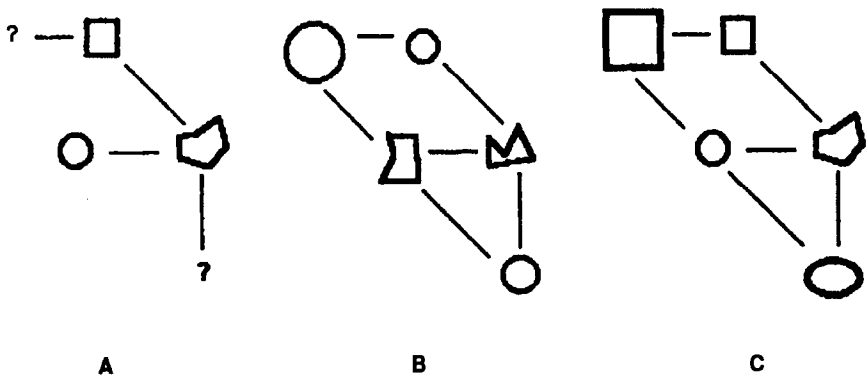


Abb. 3: Der Analogieschluß: C als Ergänzung von A aufgrund der Strukturähnlichkeiten von A und B

Hier symbolisieren die Figuren Objekte, die Striche Relationen, die Fragezeichen unbekannte Relationen zu unbekanntem Objekten.

Nun findet man einen anderen Sachverhalt, der dem ersten in mancher Beziehung strukturähnlich oder sogar strukturgleich ist. Aber er ist viel besser bekannt als der erste. Konkret: auf der Suche nach einem Ganzen, welches sehr viel freien Raum bietet, fällt der Blick auf ein Bild des Sonnensystems mit seinen Planeten. Da hat man es: ungeheuer viel freier Raum und dennoch

ein Ganzes! Atome als Mikrosonnensysteme? Die Abbildung 3B stellt die Struktur da, die A ähnlich ist (etwas Ganzes mit viel freiem Raum), aber doch besser bekannt. Und nun überträgt man die Strukturidee von B auf A und bekommt mit C ein vollständigeres Bild von A als neue Hypothese über die Beschaffenheit des Sachverhalts A.

Analogien gibt es zuhauf in den Wissenschaften. Das vielleicht prominenteste Beispiel in der Psychologie ist die Freudsche Psychoanalyse, die in ihrer Grundstruktur zahlreiche Übertragungen aus der Thermodynamik aufweist (s. Wyss, 1970, S. 30 f.). Die Seele ist eine Art von Dampfmaschine, mit einem brodelnden Triebkessel, der Energie erzeugt. Es gibt ein ganzes System von Ventilen, durch die der „Triebdampf“ in die richtigen Kanäle geleitet werden kann. Diese Ventile werden von einem Maschinisten namens „Ich“ betrieben, der die Ventile so zu steuern versucht, daß der Dampf in diejenigen Teile der Maschinerie gerät, die schließlich ein Ablassen des Dampfes ohne weitere Gefahren gestatten. Die Arbeit von „Ich“ wird dadurch nicht leichter, daß er auch noch Befehle vom Fabrikchef, dem „Überich“ entgegen nehmen muß. In dieser Triebmaschine kann alles mögliche schiefgehen: Der Triebdampf will in ein Rohr, welches ihm der Maschinist versperrt; vielleicht auf Weisung des Chefs. Wenn der Dampf nicht heraus kann, gibt es im Kessel Überdruck, und vielleicht entstehen Haarrisse im Kessel, durch die der Dampf zum Teil entweicht. Das ängstigt den Maschinisten, und er versucht, die Risse zuzuschmieren. Das aber führt nur dazu, daß sich Risse an anderen Stellen bilden... In der Freudianischen Psychoanalyse wird „Energie“ erzeugt und muß „abgefahren“ werden. Wenn dies nicht geht, z.B. aufgrund einer „Verdrängung“, dann muß die Energie irgendwie auf „illegalen“ Wegen raus. Theoretisch bedeutsam für das gesamte Freudianische Konzept ist die - wohl implizit geschehene - Übernahme des zweiten Hauptsatzes der Thermodynamik, nämlich des Satzes von der „Energieerhaltung“. Dieser Satz gilt für „richtige“ Energie. Ob er notwendigerweise auch für das gilt, was Freud metaphorisch „Energie“ nannte, kann man für Triebenergien, gerade auch für die psychoanalytisch so bedeutsame Sexualität, durchaus bezweifeln. Analogien bringen nicht nur weiter, sondern haben auch ihre Gefahren.

Woher kommen die Analogien? Der „Zeitgeist“ spielt hier wohl eine entscheidende Rolle; im 18. Jahrhundert, dem Zeitalter des Newtonschen Weltuhrwerks, spielte auch für die Psychologie die Vorstellung von der Seele als Uhrwerk eine große Rolle (s. LaMettrie, 1966). In der Entwicklungspsychologie taucht immer wieder die Metapher von der geologischen Schichtung auf, durch die phylo- und ontogenetisch frühere Strukturen durch spätere zugedeckt werden.

Auf die Telefonzentralenmetapher des Behaviorismus haben wir schon hingewiesen. Das Zeitalter des Kognitivismus wurde eingeleitet durch die sehr

explizite Übernahme der Computeranalogie. Plötzlich wurde aus dem Gedächtnis ein „Speicher“, aus dem Kurzzeitgedächtnis ein „Schieberegister“, das Bewußtsein wurde eine „CPU“ (Central Processing Unit).

Analogien können natürlich auch einfach aus der persönlichen Erfahrungswelt eines Individuums stammen. Die im Eingangsabschnitt erwähnten Beispiele von Mendelejeff und Kékulé gehören hierher. Gigerenzer (1988) demonstriert, in wie hohem Maße die in einer Wissenschaft zu speziellen Zwecken verwendeten Verfahren zu Quellen von Analogien werden können und damit unversehens eine Karriere vom Werkzeug zur Theorie machen. Gigerenzer zeigt an vielen Beispielen, wie die Verwendung statistischer Verfahren zur Hypothesenprüfung in der Psychologie Anlaß dazu gab, sich Denk-, Wahrnehmungs- und Entscheidungsprozesse als Prozesse der Inferenzstatistik vorzustellen. Die Seele wurde unversehens zu einem Gefüge von Signifikanztestsystemen, Wahrnehmen bedeutet das Ziehen einer Stichprobe, Erfahrungsgewinnung heißt Mittelwertbildung usw.

Die Verwendung von Analogien ist ein unschätzbare Werkzeug, um das Denken aus eingefahrenen Gleisen zu befreien und auf neue Ideen zu kommen. Auf der anderen Seite birgt sie Gefahren. Gewöhnlich sind nämlich die Analogien unvollständig oder sogar falsch. Kein Sachverhalt ist wohl einem anderen vollständig analog. Wenn man also einen Analogieschluß macht und Beziehungen von einem Bereich auf einen anderen überträgt, unterliegt man immer auch der Gefahr, Beziehungen fälschlicherweise zu übertragen. Wir haben dies oben schon am Beispiel des zweiten Hauptsatzes der Thermodynamik und der Freudianischen Psychoanalyse demonstriert. Solche falschen Übertragungen haben viel Unheil angerichtet und auch viel Forschungsenergie in die falsche Richtung gelenkt.

Die eigentlich wohl vernünftige Übernahme des Konzeptes der Informationsverarbeitung in die Psychologie war oft mit einer allzu weitgehenden Analogisierung des Computers mit psychischen Zuständen und Prozessen verbunden. Der Computerspeicher z.B. ist ein ganz anders aufgebautes System als das lebendig sich verändernde Gedächtnisnetzwerk beim Menschen. (Mit „Netzwerk“ sind wir schon wieder bei einer Analogie!) Man wird auf die falsche Fährte gelockt, wenn man das Gedächtnis als „Speicher“ betrachtet. Die Bezeichnung „Speicher“ für das Gedächtnissystem des Computers war auch eine Analogisierung! Man hat das Computer„gedächtnis“ nicht von ungefähr „Speicher“ und nicht „Gedächtnis“ genannt. Ein Speicher ist - wie auch ein Gedächtnis - ein System für den „Zeittransport“ von Dingen. Aber ein Speicher ist in landläufiger Bedeutung ein System für die Aufbewahrung von unverbundenen Materialien, z.B. von wahllos zusammengeschütteten Getreidekörnern oder Kohlebrocken. Im Computer bekommen die Speicherzellen ihre Verknüpfung erst durch das Programm; der Speicher für sich verbin-

det auch im Computer die Inhalte nicht miteinander, daher ist das Gedächtnis kein Speicher.

Die Arbeitsweise des Computers ist ganz anders als die Arbeitsweise des Gehirns. In den heutigen Computern findet die Informationsverarbeitung dadurch statt, daß Inhalte aus dem Speicher in die CPU geholt werden, hier verarbeitet und dann wieder in den Speicher zurücktransportiert werden. Die Annahme, daß menschliches Denken in ähnlicher Weise vor sich geht, bedeutet, daß man die „Computemetapher“ zu ernst nimmt. Vermutlich findet die Informationsverarbeitung im menschlichen Gehirn viel dezentralisierter statt als im Computer.

Es geschieht leicht, daß man Analogien allzu ernst nimmt. Um beim Gedächtnis zu bleiben: Wenn man entdeckt hat, daß es für bestimmte Zwecke ganz praktisch sein kann, die Hilfsmittel der Prädikatenlogik für die Beschreibung der Beziehungen von Gedächtnisinhalten zueinander zu verwenden, so kann es leicht geschehen, daß man die Prädikatenlogik reifiziert und plötzlich glaubt, daß die Gedächtnisstruktur „prädikatenlogisch“ sei. So kann dann eine „Tricode-Theorie“ des Gedächtnisses entstehen (s. Anderson, 1983, S. 45ff.; Wessels, 1984, S. 288ff.), in der die Sachverhalte und Situationen dieser Welt außer in Sprache und Bildern als Beziehungsstruktur in einem „konzeptuellen“, prädikatenlogischen System gespeichert sind.

## **3.6 Modelle und Simulation**

### **3.6.1 Modelle**

Jeder kennt Modelleisenbahnen oder Modellflugzeuge. Es sind Replikat von „Urbildern“, die bestimmte Merkmale der Urbilder meist in verkleinertem Maßstab aufweisen. Charakteristisch für Modelle ist es, daß sie nur bestimmte Aspekte der Urbilder replizieren, andere nicht. Im Inneren einer Modelllokomotive befinden sich gewöhnlich kein Dampfkessel und kein Heizer, sondern ein Elektromotor. Man kann zwar von den Längen-, Breiten- und Höhenmaßen des Modells unter Verwendung des Maßstabes die entsprechenden Maße des Urbildes errechnen, die Flächen- und Rauminhalte aber des Urbildes sind nicht einfach durch Multiplikation mit dem Maßstab zu gewinnen; man muß quadrieren bzw. die dritte Potenz wählen.

Die Modelle, von denen wir bislang gesprochen haben, sind Modelle von „statischen“ Gebilden. Das Modell einer ganzen Rangieranlage aber mit allen Weichen, Signalen und Verladeanlagen ermöglicht auch eine „dynamische“ Simulation des Urbildes „Güterbahnhof“. Man kann die Prozesse abbilden, die auch im Urbild ablaufen können. Eine solche „Simulation“ des Urbildes spielt



nun in der Psychologie und zunehmend in allen Sozialwissenschaften eine immer wichtigere Rolle als Heurismus bei der Theorienbildung, und deshalb wollen wir auf den Begriff des Modells und den Gebrauch von Modellen in der Psychologie zum Zwecke der Theorienbildung nun ausführlicher eingehen.

Die Simulation eines psychischen Prozesses wird gewöhnlich nicht durch ein mechanisches Gebilde geschehen; das geeignete Instrument ist heute der Computer, den man relativ einfach so einrichten kann, daß er jeden beliebigen, genau beschreibbaren Prozeß simulieren kann.

### Theorien, Modelle, empirische und theoretische Begriffe

Was ist ein Modell? Der Begriff wird in den Sozialwissenschaften mit verwirrend vielen verschiedenen Bedeutungen verwendet. Oft wird er als Synonym für „Theorie“ gebraucht. Boguslaw (1965) und Chapanis (1963) unterscheiden z.B. „Replikatmodelle“ (ein solches wäre eine Modellokomotive) von „symbolischen“ Modellen (ein solches wäre eine in einer formalen Sprache abgefaßte Theorie).

Warum verwendet man das Wort „Modell“ anstelle des Wortes „Theorie“? Braithwaite (1963) glaubt, drei Motive dafür feststellen zu können:

1. Eine Theorie ist so „klein“, so umfangarm, so speziell, daß der Ausdruck „Theorie“ dafür zu anspruchsvoll erscheint.
2. Halb- oder ganz formalisierte Theorien sind in den Sozialwissenschaften selten. Man bezeichnet solche Theorien als Modelle, um sie von unformalisierten Theorien abzugrenzen.
3. Der Ausdruck „Modell“ wird statt des Begriffs „Theorie“ verwendet, wenn es sich bei der Theorie im Grunde nur um die Vorform einer solchen handelt, also nur um ein Theorienschema. D.h. die Theorie enthält Leerstellen, die noch spezifiziert werden müssen. „Irgendwie ist die Intelligenzleistung von der Begabung, dem Sozialklima, den Anregungen im Elternhaus abhängig!“, das wäre eine solche „Hohlform“ für eine Theorie. Das „Denkmodell“ ist ein Vorläufer der eigentlichen Theorie.

Ich halte die Gleichsetzung von „Modell“ und „Theorie“ für unglücklich; sie verwischt die Unterscheidung von Sprache und dem, was mit Sprache bezeichnet wird. Wir wollen bei dem ursprünglichen und alltäglichen Gebrauch des Begriffs Modell bleiben. Ein Modell ist ein Replikat eines Urbildes und damit ein „Ding“. Eine Theorie ist dagegen ein Gefüge von Aussagen. Ein Modell ist also „jede Struktur, die wohl definiert, klar und deutlich und hinreichend isomorph zu einem anderen System ist“ (Ashby, 1966). Ähnlich formuliert Couffignal: „Ein Modell ist ein künstlicher Mechanismus, der gewisse

Analogien zu einem vorgegebenem Mechanismus aufweist und dessen Ziel es ist, andere Analogien aufzudecken“ (Couffignal nach Marx, 1963).

Eine Theorie ist immer ein Gefüge von Aussagen, ein sprachliches Gebilde, sie tut nichts, und sie verhält sich nicht, wenn nicht irgendwer da ist, der sie **versteht**, der also die Symbole in Vorstellungen von materiellen Dingen, Energieflüssen oder Informationsverarbeitungsprozessen umsetzen kann. Eine Theorie ist, wenn man so will, ein „Programm“ für die Herstellung des Objektes, über das theoretisiert wird. Ein Programm tut für sich allein genommen überhaupt nichts. Es bedarf eines Systems, welches das Programm „versteht“ und dann auch simulieren kann. Ein Modell dagegen ist auch da, wenn keiner es „versteht“. Es „verhält sich“ auch wenn keiner zusieht. Es hat eine Existenz für sich.

Was heißt es nun, daß ein Modell seinem Urbild „isomorph“ bzw. „hinreichend isomorph“ sein soll, wie Ashby (1966) es ausdrückt? Bei den oben dargestellten Beispielen für Modelle (Modelleisenbahnen, Puppen, usw.) ist die Modellrelation sinnfällig. Man „sieht“ die Ähnlichkeit. Man sieht auch, wie weit die Isomorphie geht. Sie umfaßt nicht das gesamte „Ding“. Eine Isomorphie liegt nur hinsichtlich bestimmter Teile von Urbild und Modell vor; sowohl Urbild als auch Modell haben Merkmale, die einander nicht entsprechen (s. hierzu im einzelnen Tack, 1969, S. 235ff.). Bei der Modellokomotive gibt es eine Isomorphie zum Urbild hinsichtlich der Relationen der Längenmaße zueinander; Höhe und Breite einer Modellokomotive verhalten sich so zueinander wie beim Urbild. Und - gewöhnlich - herrscht Isomorphie auch hinsichtlich der Farben. Eine Isomorphie der „Innereien“ ist nicht gegeben; auch das „sieht“ man.

Modelle müssen aber ihren Urbildern nicht in sinnfälliger Weise ähnlich sein. Zwei Sachverhalte können gänzlich verschieden aussehen und doch Modelle voneinander sein. Wie das? Ein Betrachtung von Braithwaite (1963) macht das klar.

Braithwaite unterscheidet eine „Modelltheorie“ von der Theorie. Man kommt von einer Theorie zu einer „Modelltheorie“, indem man die Terme der Theorie von ihren Bedeutungen trennt (Braithwaite spricht von „Disinterpretation“) und ihnen neue Bedeutungen zuweist („Reinterpretation“). Die Struktur von Theorie und Modelltheorie, nämlich das logische Gerüst der Ableitungsbeziehungen zwischen den Termen, also das, was man in der formalen Logik „Kalkül“ nennt, bleibt dabei unverändert.

Bei Braithwaite sind Theorie und Modelltheorie „interpretierte“ Kalküle, also logische Systeme, die formal gleich sind, deren Begriffe sich aber auf verschiedene Realitäten beziehen. Abbildung 4 zeigt anschaulich, wie man sich die Beziehung von Modell und Theorie nach Braithwaite vorstellen kann.



wird also nur dann „wahr“ ( $=1$ ), wenn beide Eingangsgrößen den Wert 1 haben, sonst „falsch“ ( $=0$ ). Die Variable  $b$  wird 1, wenn zwar die Glocke ertönt, aber kein Futter dargeboten wird (die Futtervariable wird durch die Negation  $N$  invertiert; aus 0 wird 1 und umgekehrt). Die Funktion  $S$  soll eine Summierung der Eingänge  $a$ ,  $b$  und  $c$  vornehmen, und zwar so, daß  $c := c+a-b$ . Ist also  $c$  zu Anfang  $= 0$  und werden fünf Mal hintereinander Glocke und Futter gemeinsam dargeboten, dann hat  $c$  danach den Wert 5.

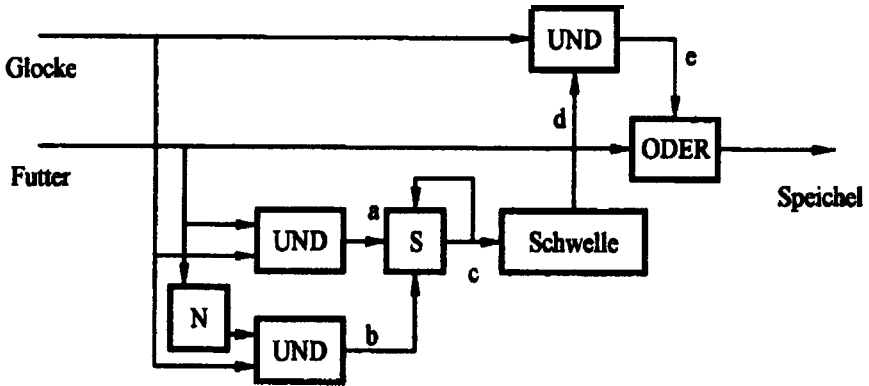


Abb. 5: Eine Theorie des bedingten Reflexes

Auch  $d$  soll eine zweiwertige (Boolesche) Variable sein. Sie ist von  $c$  gemäß der Funktion „Schwelle“ abhängig. Liegt  $c$  über einem bestimmten Schwellenwert  $\text{ALPHA}$  (ein fixer Parameter des Systems; er gehört zur Funktion  $\text{SCHWELLE}$ ), so wird  $d = 1$ , sonst wird  $d = 0$ . Wenn  $d = 1$  ist, so kann auch  $e = 1$  werden, wenn die Glocke ertönt. Dann wird Speichel fließen, auch wenn kein Futter dargeboten wird, denn die „Speichelvariable“ wird durch die „oder“-Funktion aktiviert, wenn entweder Glocke und  $d$  oder wenn Futter dargeboten wird, oder wenn beides geschieht. Unser System lernt also auf die Glocke allein mit Speichelfluß zu reagieren, wenn oft genug Futter und Glocke gemeinsam auftraten, und es verlernt dies auch wieder, wenn häufiger die Glocke ohne das Futter auftritt. (Als Theorie sollte man das Blockdiagramm nicht allzu ernstnehmen; wir haben die tatsächlichen Verhältnisse vereinfacht, damit sie leicht verstanden werden kann. Man müßte z.B.  $c$  - die Lernvariable - gegen eine obere und untere Grenze gehen lassen.)

In bezug auf das Urbild, nämlich den „echten“ Pavlov-Hund (s. Pavlov, 1972, S.73ff.), haben einige der Begriffe der Theorie eine klare Bedeutung; „Speichel“ bezieht sich auf eine Flüssigkeit, „Glocke“ auf ein bestimmtes Muster von Schalldruckschwankungen, „Futter“ auf Fleisch. Programmiert man nun einen Computer so, daß er den Pavlov-Hund simulieren kann, so wird man diese Begriffe uminterpretieren. „Futter“ wird dann z. B. die Eingabe  $F$  am Terminal bedeuten, „Glocke“ die Eingabe  $G$ , und statt zu speicheln wird der Computer

das Wort „Speichel“ auf den Bildschirm schreiben. Der Computer und der wirkliche Hund sind also einander sinnfällig überhaupt nicht ähnlich, dennoch kann man behaupten, daß der entsprechend programmierte Computer ein Modell des Hundes sei.

„Futter“, „Glocke“, „Speichel“ lassen sich disinterpretieren und reinterpretieren. Was aber ist mit den Variablen a bis e und den Funktionen? Sie sind ja auch Teile der Theorie! Sie sind zunächst uninterpretiert; in der Theorie haben sie die Funktion, ein Beziehungsnetz so zu konstituieren, daß das Ganze funktioniert. Mit solchen uninterpretierten Begriffen hat man Schwierigkeiten. Man kann nicht wahrnehmen, worauf sie sich beziehen, man kann nicht unmittelbar überprüfen, in welchem Zustand sich die Variablen befinden. Sie konstituieren zunächst nichts anderes als ein Beziehungsnetz. Im Gegensatz zu den „empirischen“ Begriffen „Speichel“, „Glocke“ und „Futter“ nennt man solche Begriffe auch „theoretische Begriffe“.

Nun sind auch theoretische Begriffe überprüfbar, nur nicht unmittelbar. Man kann sie mittelbar überprüfen, indem man nachsieht, ob das gesamte Gefüge der Terme der Theorie (die „operationale Komponente“ eines theoretischen Konstrukts, wie Hyland (1981) es nennt), nur dann die erwarteten Erklärungsleistungen erbringt, wenn die fraglichen Begriffe darin eingebaut sind. Ist dies der Fall und sind die entsprechenden Ableitungen ohne die theoretischen Begriffe nicht zu haben, so spricht einiges dafür, daß die theoretischen Begriffe nicht nur Chimären sind, sondern daß ihnen irgendwelche tatsächlichen Vorgänge oder Zustände entsprechen. Aber eben nur „irgendwelche“, die das entsprechende Verhalten produzieren. Nun kann zwar nicht jedes System alles, es lassen sich aber für eine operationale Komponente unendlich viele Äquivalente denken. Man könnte z.B. in der Theorie des bedingten Reflexes, die wir in Abbildung 5 geschildert haben, jede „UND“-Funktion durch eine „ODER“-Funktion ersetzen, die, wenn man Eingänge und Ausgänge negiert, sich genauso verhalten würde, wie eine „und“-Funktion.

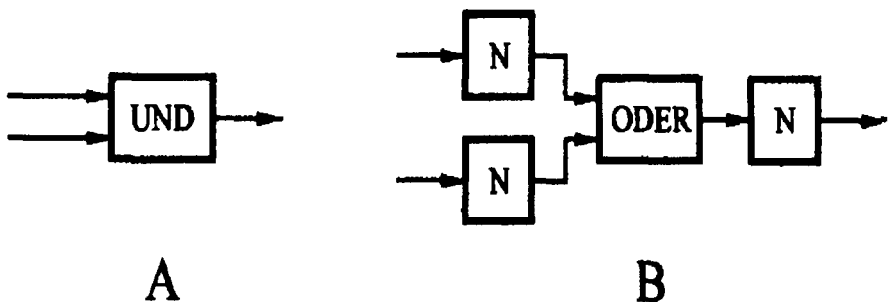


Abb. 6: Zwei verhaltens-, aber nicht strukturgleiche Systeme

Abbildung 6 zeigt zwei solche verhaltensgleiche, aber nicht strukturgleiche Systeme. Die Systeme der Abbildung 6 sind „Verhaltensmodelle“ voneinander, nicht aber „Strukturmodelle“. (Zur Unterscheidung von Verhaltens- und Strukturmodellen s. Klir & Valach, 1968, S. 45).

Man kann Strukturmodelle verschiedener Stufen unterscheiden. Wenn ein System ein Strukturmodell eines Urbildes ist, so heißt das, daß seine Teile **Verhaltensmodelle** der entsprechenden Teile des Urbildes sind. Gewöhnlich lassen sich die Teile noch weiter in Teile zerlegen. Wenn auch diese Teile Verhaltensmodelle der entsprechenden Teile des Urbildes sind, so könnten wir von einem Strukturmodell 2. Ordnung sprechen, usw.

Woher soll man nun wissen, ob sich „real“ in dem Pavlov-Hund ein Gefüge von „UND“-Funktionen oder ein Gefüge von „NICHT (NICHT... ODER NICHT...)“-Funktionen befindet? Wie soll man zwischen den unendlich vielen verhaltensäquivalenten Systemen unterscheiden? Durch reine Verhaltensbeobachtung geht das nicht.

Johnson-Laird (1988, S. 52) meint: „... ,since there is an infinity of different programs for carrying out any computable task, observations of human performance can never eliminate all but the correct theory. There will always be some alternative theories that are equally plausible. Theories are vastly undetermined by empirical data: they invariably go beyond what has been, and what can be observed“.

Durch Verhaltensbeobachtung allein kann man die verschiedenen Realisationen verhaltensäquivalenter Systeme nicht unterscheiden. Dennoch aber geht die Johnson-Lairdsche Folgerung, daß es aufgrund der Ununterscheidbarkeit durch Beobachtung immer „einige gleich plausible“ Theorien geben wurde, zu weit. Es gibt bestimmte Kriterien, die oft doch eine der alternativen „operationalen Komponenten“ am meisten plausibel erscheinen lassen. Und unter diesen Kriterien befinden sich sogar empirische.

Ein solches Kriterium ist das der **Einfachheit**. Kompliziertere Systeme sind im allgemeinen **störanfälliger** als einfache. Deshalb werden bei gleicher Leistung die einfachen Systeme die komplizierteren besiegen, wenn sie miteinander konkurrieren müssen. Und eine solche Konkurrenz gibt es z.B. in der biologischen (aber auch in der technischen) Evolution. Bei dem System B der Abbildung 6 kann einfach mehr kaputt gehen als bei dem System A (wenn man annimmt, daß ein UND-Gatter technisch genauso kompliziert ist, wie ein ODER-Gatter).

Wenn man also eine Theorie so einfach wie möglich baut, so hat man wohl eine größere Chance, das richtige zu treffen, als wenn man eine Theorie kompliziert baut.

Eine weitere Sicherung gegen inplausible Theorien liegt in der Einführung eines „Surplus“, wie Hyland (1981) es nennt. Was ist darunter zu verstehen? Ein Surplus ist eine Art tentativer Interpretation theoretischer Begriffe.

Man kann z.B. sagen, daß eine „UND“-Funktion ein Neuron mit einer hohen Erregungsschwelle darstellt, welches nur dann aktiv wird, wenn alle seine Eingänge aktiv sind. Eine „ODER“-Funktion ist dann ein Neuron mit einer niedrigen Schwelle, welches aktiv wird, wenn mindestens einer seiner Eingänge aktiv ist. Die Zählvariable  $c$  samt anhängender Funktion „SCHWELLE“ könnte man als synaptische Schwelle des Neurons interpretieren, welches die „UND“-Funktion mit  $e$  als Ausgang darstellt. Mit sinkender Schwelle wird dieses Neuron immer leichter durchgängig, bis schließlich der Glockenton „durchkommen“ und den Speichelfluß auslösen kann.

Eine solche Interpretation ist nur eine tentative; bei einem realen Hund wird man die neuronalen Prozesse in dem angenommenen Neuronennetz nicht tatsächlich beobachten können. Eine solche „theoretische“ Interpretation hat daher einen anderen Stellenwert als die Interpretation der empirischen Variablen, da sie nicht direkt prüfbar ist.

Hyland (1981) verlangt ein „Surplus“ für jede Einführung theoretischer Variablen in eine Theorie. Er unterscheidet drei verschiedene Arten von „Surplus“ für die Psychologie, nämlich „mechanistische“, „physiologische“ und „mentalistiche“. Unsere Interpretation der „Innereien“ der Theorie des bedingten Reflexes wäre ein „physiologisches Surplus“. Wenn wir die Summierungsvariable  $c$  z.B. als ein Wasserbecken interpretiert hätten, durch die über die Variable  $a$  ständig Wasser hineinläuft und durch die Variable  $b$  Wasser abgelassen werden kann, dann wäre eine solche analoge Interpretation ein mechanistisches Surplus. Ein mentalistisches Surplus läge vor, wenn man sich bei der Interpretation einer bestimmten Variable auf das eigene Erleben bezieht.

Betrachten wir ein Beispiel für die verschiedenen Formen von Surplus: Man könnte annehmen, daß die „Aggressionstendenz“ von Menschen oder Tieren der Aktivität eines bestimmten Gehirnareals, nämlich eines Aggressionszentrums im Hypothalamus (s. Ehrhardt, 1975, S. 67ff.), entspricht. In diesem Fall wurde man also sagen: „Das was ich eigentlich meine, wenn ich von Aggressionsneigung spreche, ist die Aktivität einer bestimmten Hirnstruktur. Das Ausmaß der Aktivität dieser Struktur entspricht dem Ausmaß der Tendenz, aggressiv zu reagieren.“

Ein mechanistisches Surplus liegt dann vor, wenn man einem theoretischen Begriff eine bestimmte physikalisch-mechanistische Interpretation unterlegt. Wenn man beispielsweise der Meinung ist, daß die „Triebspannung“ auf so etwas wie die Füllung eines Wasserbehälters zurückzuführen sei (je größer die Füllung, desto stärker die Tendenz zur Ausübung des triebentsprechenden

Verhaltens), so verwendet man ein mechanistisches Surplus. Wenn man dagegen z.B. von „Wut“ redet, so verwendet man ein *mentalistisches* Surplus. Man bezieht sich auf das, was man als „Wut“ tatsächlich bei sich erleben kann.

Die Beachtung der Existenz verschiedener Surplus ist sehr wichtig für die Theorienbildung. Hyland verlangt nicht nur ein Surplus für alle theoretischen Begriffe, sondern zusätzlich, daß die Vermischung verschiedener Surplustypen in ein- und derselben Theorie zu vermeiden sei. Warum das?

Der Grund für die beiden Vorschriften ist, daß die Vermischung verschiedener Surplus leicht zu einer „Realitätsverdoppelung“ führen kann. Wenn man z.B. weiß, daß Aggressivität mit der Aktivität eines hypothalamischen Zentrums XXX zusammenhängt, so liegt es nahe, folgenden „theoretischen“ Satz zu formulieren: „Die Aktivität des hypothalamischen XXX beeinflusst das Wutgefühl!“

Das klingt harmlos. Der Satz besagt, daß subjektiv das Gefühl der Wut oder des Ärgers um so größer ist, je größer die Aktivität des entsprechenden hypothalamischen Zentrums ist. Was soll dagegen einzuwenden sein?

Es ist dagegen einzuwenden, daß man auf diese Art und Weise leicht aus einer Sache zwei macht. Denn es wäre ja möglich, daß das Wutgefühl nichts anderes ist als die „Innenansicht“ der Aktivität des entsprechenden hypothalamischen Zentrums. Wenn aber dies der Fall wäre, dann würde man mit dem Satz „Die Aktivität des hypothalamischen Zentrums beeinflusst das Wutgefühl!“ eine Kausalrelation in die Theorie einführen, die überhaupt nicht existiert. Denn die Variablen, die man hier als Auslöser und Empfänger einer Kausalbeziehung nennt, wären ja dann gar nicht verschieden voneinander.

Dies macht deutlich, daß die Vorschrift von Hyland, keine theoretischen Variablen ohne Surplus zu benutzen, sinnvoll ist. Wenn man für theoretische Variablen kein Surplus einführen würde, dann könnte man nicht entscheiden, ob man nicht dabei ist, „eigentlich“ verschiedene Surplus zu vermischen. Die „Surplus - Reinheit“ ist also ein Kriterium für die Plausibilität einer operationalen Komponente.

Ein weiteres Kriterium, welches mitunter zur Stützung der Annahme herangezogen werden kann, daß eine operationale Komponente eine ganz bestimmte Struktur und nicht eine andere hat, ist die „Seiteneduktion“. Darunter wollen wir folgendes verstehen:

Man konstruiert eine Theorie zur Erklärung der Tatsache, daß: „wenn A, so B“ ( $A \rightarrow B$ ), daß also der Pavlov-Hund nach einer gewissen Anzahl von gemeinsam dargebotenen Schall- und Futterreizen schließlich auf den Schall so reagiert wie auf das Futter. Diese Theorie bildet ein System. Wenn man nun aus dem System außer dem zu erklärenden Satz  $A \rightarrow B$  noch einen anderen Satz  $A \rightarrow C$  ableiten kann, der sich aus einer bestimmten Annahme über die



innere Struktur ergibt, so stellt die empirische Prüfung, ob nun wirklich  $A \rightarrow C$ , eine Prüfung der Annahme über die innere Struktur dar.

Ich habe z.B. einmal eine Theorie über die Realisierung von Denkprozessen in Neuronennetzen aufgestellt (Dörner, 1974). Wir simulierten diese Theorie, und dabei ergab es sich, daß nicht nur Denkprozesse simuliert wurden, sondern - ganz unbeabsichtigterweise - auch EEG-Rhythmen. Wenn man das künstliche Neuronensystem, welches wir konzipiert hatten, in Ruhe ließ und ihm keine Aufgaben stellte, dann verfiel es in einen Schwingungszustand mit großen Amplituden und niedrigen Frequenzen. Wenn das Netz aber zu „denken“ begann, zeigte es unregelmäßigere Oszillationen mit geringerer Amplitude und höherer Frequenz (s. Dörner, 1974, S.236ff.). Das ist der Unterschied, den man zwischen Alpha- und Beta-Wellen im menschlichen EEG findet. Die Alpha-Wellen signalisieren beim Menschen einen „dösenden“ Ruhezustand, die Beta-Wellen hingegen geistige Aktivität. Die Tatsache, daß man aus unseren Strukturannahmen über denkende Neuronennetze EEG-Rhythmen ableiten konnte, ohne daß dies beabsichtigt war, machte die Annahmen plausibler (ohne sie zu beweisen). Man kann also auch über die verborgenen Bestandteile von Theorien „mit Gründen“ streiten.

### Simulation und das Turing-Kriterium

Warum konstruiert man in den Wissenschaften überhaupt Modelle? Das hat verschiedene Gründe, auf die wir im nächsten Abschnitt noch genauer eingehen werden. Ein prominenter Zweck des Modellbaus ist die Überprüfung einer Theorie. Wenn sich ein Modell genau so verhält wie ein Urbild, dann ist dies ein Indiz für die Richtigkeit der Theorie. Statt von „Modellverhalten“ spricht man gewöhnlich von „Simulation“. Ein Modell simuliert ein Urbild, wenn es sich genauso verhält wie dieses.

Anders Hanneman (1988, S. 84ff.), der als „Simulation“ jedes Realitätsobjekt betrachtet, welches dem Urbild zwar hinsichtlich der „Form“, nicht aber hinsichtlich der materiellen Substanz ähnlich ist. „A ‚simulation‘ is usually defined as a construct that has the appearance or form, but not the substance of some real Object.“ (Wieder eine Begriffsverwirrung; das was Hanneman hier „Simulation“ nennt, haben wir gerade „Modell“ genannt.)

Die Simulation verwendet man, um darüber zu entscheiden, ob das Modell tatsächlich eines ist. Die Computersimulation einer Theorie über einen psychischen Prozeß ist eine Methode der Kontrolle der Richtigkeit einer Theorie. Man programmiert einen Computer so, daß er die gleichen Informationsverarbeitungsprozesse durchführt, wie man sie im menschlichen Gehirn annimmt, und prüft dann, ob das Verhalten des Computers mit dem des Menschen

übereinstimmt. Dieses Kriterium ist als „Turing-Kriterium“ (nach dem englischen Mathematiker und Pionier der Computertechnik Allan Turing (z.B. 1956, 1959)) bekannt. Turing schlug es als Kriterium für die Frage vor, wann man von einer Maschine behaupten dürfe, sie weise Bewußtsein oder - allgemeiner - geistige Prozesse auf. Im Hinblick auf diese Frage wurde die Aussagekraft des Kriteriums oft diskutiert (s. z.B. Searle, 1984). Wir haben an dieser Stelle nicht den Raum, auf diese Diskussion einzugehen. Wir halten das Turing-Kriterium weiterhin für vernünftig für die Entscheidung der Frage, ob man eine bestimmte Theorie aufrechterhalten kann oder nicht.

Die Schwierigkeiten bei der Anwendung des Turing-Kriteriums liegen im Detail. Was soll man auf Gleichheit prüfen? Ist man berechtigt, die Programmstrukturen eines Schachcomputers mit menschlichen kognitiven Strukturen zu identifizieren, wenn man festgestellt hat, daß der Rechner so ähnlich Schach spielt wie ein Mensch?

Oder soll man zusätzlich verlangen, daß er sich auch über ein verlorenes Spiel ärgert, im Laufe der Zeit seine Spielstrategien aufgrund seiner Erfahrungen ändert oder uns seine Strategien erläutern kann?

Bei der Simulation psychischer Prozesse wird man wohl gut daran tun, ziemlich viel an „Gleichheit“ zu verlangen. Sonst gerät man in die Gefahr, die Art der Fortbewegung eines Autos mit der eines Hundes zu identifizieren, nur weil sich beide in ungefähr der gleichen Geschwindigkeit von A nach B bewegen. (Nebenbei: man kann heute aufgrund des Spielstils wohl noch ganz gut unterscheiden, ob man gegen einen Menschen oder einen Computer Schach spielt. Die heutigen Schachcomputer haben - das wird besonders im Endspiel sichtbar - Schwierigkeiten, „strategisch“ zu spielen, also Fernziele anzustreben.)

Noch eine Anmerkung zur „Gleichheit“ des Verhaltens von Modell und Urbild. Wenn wir sagen, daß das ComputermodeLL eines psychischen Prozesses „richtig“ sei, weil Computer und Mensch „gleiches“ Verhalten zeigen, so muß man sich an Abbildung 4 und die Beziehung zwischen Theorie und Modell erinnern. Ein Modell ist die Replizierung eines Urbildes in einem anderen Medium oder eine andere Realisierung der Theorie durch Wechsel der Interpretationen der empirischen (und theoretischen) Begriffe. Hat man ein Modell des bedingten Reflexes, welches auf einem Computer „läuft“, so wird anstelle von realem Fleisch, welches man dem Hund ins Maul schiebt, ein „F“ auf der Tastatur des Computers eingegeben. Dem Computer wird bei der Darbietung von „Futter“ auch kein Speichel aus den Disketten-Slots rinnen, sondern er wird schlicht auf dem Bildschirm das Wort „Speichel“ schreiben.

Die Interpretationen der Begriffe der Theorie, also die Bezüge der Begriffe zu der Welt der materiellen und energetischen Dinge, ändern sich, wenn man

vom Urbild zum Modell übergeht. Man kann aber das Verhalten des Modells durch Dis- und Reinterpretation der Begriffe in das Verhalten des Urbildes übersetzen und umgekehrt, und damit kann man die Frage beantworten, ob das Verhalten des Modells gleich dem Verhalten des Urbildes ist oder nicht. Keineswegs geht es bei der Gleichheit um die Identität des Verhaltens. Die Erscheinungsformen des Verhaltens können sehr verschieden sein.

### **3.6.2 Der Gebrauch von Modellen bei der Konstruktion von Theorien**

Modelle werden in den Wissenschaften häufig verwendet. Jeder kennt die aus bunten Plastikstücken zusammengesteckten Molekülmodelle, mit denen die Chemiker gerne operieren, um sich die räumliche Gestalt eines Moleküls zu verdeutlichen. Ingenieure und Architekten arbeiten mit Maschinen- und Gebäudemodellen.

Wozu braucht man Modelle und warum verwenden Chemiker und Architekten oft viel Zeit und Mühe auf die Modellerstellung? Warum bemüht man sich in der Psychologie zunehmend um die Konstruktion von Computermodellen? Warum reicht nicht die Theorie? Chapanis (1963) meint, daß die Konstruktion von Modellen für die Wissenschaft folgende Vorteile biete:

1. Modelle beschreiben ein theoretisches System und helfen uns, komplexe Systeme zu verstehen.
2. Modelle zeigen den Rahmen auf, innerhalb dessen neue Experimente unternommen werden können.
3. Modelle helfen uns, neue Beziehungen zu erkennen.
4. Modelle helfen bei der Voraussage, wenn Experimente unmöglich sind.
5. Modelle machen Spaß!

Was heißt das? Wir wollen diese Vorteile im Hinblick auf die Verwendung von Computersimulationen in der Psychologie etwas genauer diskutieren.

Die Vorteile 1, 2 und 3, die Chapanis nennt, kann man wohl folgendermaßen zusammenfassen: Modelle konkretisieren eine Theorie; dadurch verstehen wir sie besser, da wir ihre Implikationen besser überschauen. Wir können auch die Implikationen leichter bekommen; der Computer leitet sie uns ab. Dies bedeutet, daß wir innerhalb der Theorie neue Beziehungen entdecken können, und aufgrund dessen können wir uns neue Experimente ausdenken. Wir wollen dies an einem Beispiel verdeutlichen.

Abbildung 7 zeigt ein Variablengeflecht, welches wir zur Erklärung des Verhaltens von Versuchspersonen konstruiert haben, die die Aufgabe hatten, in einem Computersimulationsspiel den Einsatz einer Feuerlöschbrigade im schwedischen Urwald zu leiten (s. Dörner & Pfeifer, 1992). Die Versuchspersonen



Bei den nichtgestreßten Versuchspersonen gab es mehr „Schwerpunktfehler“, d.h. diese Versuchspersonen setzten ihre Einheiten häufig am falschen Ort ein, bekämpften ungefährliche Feuer und ließen gefährliche Feuer unbekämpft.

Das war das Muster der empirischen Befunde. Abbildung 7 zeigt die Zusammenführung dieser Befunde zu einer einheitlichen Theorie durch ein System von Hypothesen. Das Variablengeflecht der Abbildung 7 kam durch Versuche zustande, das Verhalten der Versuchspersonen zu simulieren, also ein Modell der Versuchspersonen zu bauen. Wir fragten uns zunächst, wie denn die einzelnen empirischen Ergebnisse zusammenhängen. Wieso machen die gestreßten Versuchspersonen mehr Dosierungsfehler, aber weniger Schwerpunktfehler? Wieso tendierten die nichtgestreßten Versuchspersonen mehr zu Massenbefehlen? Warum gaben die gestreßten Versuchspersonen mehr Befehle? Und warum zusätzlich relativ mehr Patrouillenbefehle?

Wenn man das Verhalten der Versuchspersonen simulieren will, muß man auf diese Fragen Antworten geben. Natürlich kann man sich diese Fragen auch stellen, wenn man kein Modell bauen möchte; das Vorhaben des Modellbaus aber erzwingt die exakte Beantwortung dieser Fragen, und dieser Zwang bestunde nicht, wenn man nur vorhätte, die empirischen Befunde „irgendwie“ zu interpretieren. In der Psychologie besteht natürlicherweise, nämlich weil man sich beim Theoretisieren oft auch selbst meint, eine hohe Neigung zum „impliziten Verstehen“ von Befunden. Man versteht „intuitiv“, daß Leute unter Streß mehr Dosierungsfehler machen und neigt dann nicht dazu, die genauen Ursachen dieses Zusammenhanges zu explizieren. Der Lärmstreß produziert Unaufmerksamkeit, und da ist es doch klar, daß die Versuchspersonen ungenauer dosieren! In Wirklichkeit ist es wohl komplizierter (s.u.), aber das wurde erst durch den Simulationsversuch klar. Mit dem „intuitiven Verständnis“ wurde ein Computer nicht arbeiten; man muß ihm schon genau sagen, was er tun soll.

Abbildung 7 zeigt, was bei dem Versuch herauskam, die oben gestellten Fragen zu beantworten. Kern der Theorie ist die Annahme, daß die gestreßten Versuchspersonen sich durch die ständig verändernde Geräuschkulisse gestört fühlten, da die unvoraussagbaren und z.T. sehr lauten oder schrillen Töne ihren problemgerichteten Denkablauf unterbrachen. Dies führte dazu, daß sie das Gesamtproblem der Befehlsgebung für ihre Brandbrigade in kleine Teilprobleme zerlegten. Sie gaben ihre Befehle hauptsächlich für einzelne Einheiten und setzten außerdem den Auflösungsgrad bei der Betrachtung der Situation und bei der Analyse der Bedingungen und der Folgen für die Befehlsgebung herab.

Dies führte dazu, daß sie die Situationen nur ziemlich grob analysierten, was einerseits zu Fehlern in der Dosierung führte, andererseits aber auch dazu, daß die - wie durch eine Sonnenbrille betrachtete - Situation ihnen ihre Konturen deutlicher zeigte. Dies verminderte die Anzahl der Schwerpunktfehler

bei den gestreßten Versuchspersonen. Sie setzten ihre Einheiten mehr entsprechend der Gefährlichkeit der Feuer ein.

Die Tatsache, daß der problemgerichtete Denkablauf bei den gestreßten Versuchspersonen ständig unterbrochen wurde, führte aber außerdem dazu, daß sie sich den Anforderungen des Gesamtproblems weniger gewachsen fühlten. Dies hatte eine „reaktive Anspannungssteigerung“ zur Folge; die damit verbundene größere Aktivierung mündete in einer größeren Anzahl von Aktionen überhaupt. Das absinkende Kompetenzzempfinden führte weiterhin dazu, daß die unter Lärmstreß stehenden Versuchspersonen ein Bestreben zur Kompetenzdemonstration entwickelten. Dies schlug sich in einer größeren Anzahl von (ziemlich nutzlosen) Patrouillenbefehlen nieder; die dadurch ständig in der Gegend herumfahrenden Einheiten gaben den Versuchspersonen das Gefühl: „Da tut sich was, und ich habe das bewirkt!“. Außerdem führte diese Tendenz zu einer größeren Anzahl von Befehlen überhaupt und zu mehr Fehlern der Überdosierung.

Dieses System von Annahmen fügt die empirischen Ergebnisse durch theoretische Ergänzungen zu einem Gesamtsystem zusammen, welches vollständig ist und das Verhalten der Versuchspersonen der beiden Gruppen erklärt. Die Modellkonstruktion führt aber nicht nur dazu, daß sich das „intuitive Verständnis“ in ein explizites System von Kausalbeziehungen verwandelte; zusätzlich erzwang sie Präzision. Variablen und Funktionen mußten genau bestimmt werden. Die Plus- und Minuszeichen in der Abbildung 7, die Relationen vom Typ „je mehr desto mehr!“, bzw. „je weniger desto weniger!“ (+), bzw. „je mehr desto weniger!“ und „je weniger desto mehr!“ (-) symbolisieren, sind für die Computersimulation natürlich gänzlich ungeeignet. Wie hängt das Ausmaß der erwarteten Störungen nun genau mit dem Auflösungsgrad zusammen? Hängt der Auflösungsgrad logarithmisch vom Ausmaß der erwarteten Störungen ab? Oder ist der Zusammenhang besser durch eine Potenzfunktion beschreibbar? Oder ist eine lineare Funktion angemessen? Man muß sich hier schon festlegen. Und was heißt eigentlich überhaupt genau „Auflösungsgrad“? Bezieht sich der Auflösungsgrad nur auf die Wahrnehmung und bedeutet ein niedriger Auflösungsgrad einfach eine grobe (aber konturenverstärkende) „Sonnenbrillenbetrachtung“? Oder bezieht sich der Auflösungsgrad auch auf die Elaboration der Konsequenzen einer bestimmten Entscheidung, die bei niedrigem Auflösungsgrad eben nicht sehr weit getrieben wird? Solche Betrachtungen zerlegen die globale Variable „Auflösungsgrad“ in zwei verschiedenen Komponenten.

Die Modellkonstruktion erzwingt also Vollständigkeit, Explizitheit und Präzision der theoretischen Annahmen. Die Explizierung der theoretischen Annahmen führt oft dazu, daß man Implikationen der Theorie sieht, die zunächst unsichtbar waren. So ergab es sich bei der Konstruktion des Annahmengebäudes der Abbildung 7, daß man aus der Theorie ableiten konnte, daß die

Reaktionszeiten der gestreßten Versuchspersonen kürzer sein müßten als die der nichtgestreßten Versuchspersonen. Dies ist ableitbar aus der Annahme, daß die gestreßten Versuchspersonen das Gesamtproblem in höherem Maße in Teilprobleme zerlegen als die nichtgestreßten Versuchspersonen. Außerdem ist es ableitbar aus der Annahme, daß die gestreßten Versuchspersonen mit einem niedrigeren Auflösungsgrad arbeiten als die nichtgestreßten Versuchspersonen.

Wir hatten aber die Reaktionszeiten weder bei der Konzeption der Theorie erwogen, noch hatten wir sie ausgewertet. In den Computerprotokollen des Verhaltens der Versuchspersonen aber waren sie registriert und so konnten wir die Hypothese nachträglich prüfen. Die Hypothese stimmte.

Nebenbei bemerkt: Dies ist ein Beispiel dafür, daß man mitunter ein Gefüge von theoretischen Variablen, welches man konstruierte, um den Zusammenhang  $A \rightarrow B$  („wenn Lärmstress, dann mehr Patrouillenbefehle, mehr Überdosierung, ...!“) zu erklären, einem empirischen Test unterwerfen kann, da man plötzlich feststellt, daß auch  $A \rightarrow C$  („wenn Lärmstress, dann kürzere Reaktionszeiten!“) aus der Theorie abgeleitet werden kann.

Eine Folge der durch die Modellkonstruktion erzwungenen größeren Explizitheit und Vollständigkeit des theoretischen Systems ist auch, daß Widersprüche im theoretischen System leichter entdeckt werden können. Aus unseren theoretischen Annahmen war abzuleiten, daß die gestreßten Versuchspersonen **mehr** Dosierungsfehler machen sollten als die nichtgestreßten Versuchspersonen. Dies ergibt sich aus dem geringeren Auflösungsgrad. Denken und Wahrnehmen mit einem niedrigen Auflösungsgrad führt dazu, daß die Versuchspersonen in einer Gefahrensituation weniger gut entscheiden können, wieviele Feuerwehreinheiten zur Brandbekämpfung wohl notwendig sind. Und das bedeutet Unter- oder Überdosierung.

Auf der anderen Seite wurde die Hypothese aufgestellt, daß die gestreßten Versuchspersonen gerade wegen ihres geringeren Auflösungsgrades einen besseren Überblick über die Gesamtsituation haben wurden. Daraus kann man ableiten, daß die gestreßten Versuchspersonen **weniger** Dosierungsfehler machen wurden. Denn ein besserer Überblick über die Gesamtsituation sollte zu einer besseren Einschätzung der jeweils notwendigen Anzahl von Feuerwehreinheiten führen. Aus der einen Annahme ergab sich also die Folgerung: „Mehr Dosierungsfehler!“, aus der anderen die Folgerung: „Weniger Dosierungsfehler!“ Das ist zunächst einmal ein Widerspruch. Die Explizitheit des theoretischen Systems zeigt diesen Widerspruch auf und ermöglicht seine Beseitigung. Diese kann in verschiedener Weise erfolgen. Z.B. könnte man annehmen, daß beide Tendenzen vorhanden sind, sich aber wechselseitig aufheben. Man könnte auch entweder die eine oder auch die andere Hypothese fallenlassen. Eine genaue Betrachtung der Dosierungsfehler zeigt, daß die gestreßten Versuchspersonen tatsächlich mehr Dosierungsfehler machen als die

nichtgestreßten; sie neigen aber nur zu mehr Überdosierungsfehlern („Nicht kleckern! Klotzen!“); nicht zu mehr Unterdosierungen. Der Grund für die größere Anzahl von Dosierungsfehlern bei den gestreßten Versuchspersonen liegt also nicht in der oberflächlicheren Analyse. Zugleich vermeiden die gestreßten Versuchspersonen Dosierungsfehler nicht wegen der besseren „Übersicht“. Die größere Anzahl von Überdosierungsfehlern liegt bei den gestreßten Versuchspersonen vielmehr an ihrer Tendenz zur „Kompetenzdemonstration“. So verschwindet der Widerspruch.

Formalisiert man das theoretische System der Abbildung 7 (die Abbildung enthält nur die Grobform der Theorie), so erhält man ein recht umfangreiches System von Annahmen, welches auch Rückkopplungen enthält. Z.B. wirken sich die bemerkten Fehler auf das Kompetenzzempfinden und dieses wieder auf die Fehlertendenzen aus. Das System ist ohne Computersimulation überhaupt nicht hantierbar. Man ist ohne sie gänzlich außerstande, das Geflecht von 20 oder mehr Variablen und ihre zum Teil sich in der Zeit verändernden Beziehungen zueinander zu überschauen. Zwar sind die in die Theorie eingehenden Annahmen im einzelnen primitiv, das Gesamtgefüge aber, welches man aus solchen einfachen Annahmen erstellt, verhält sich unter Umständen keineswegs in der Weise, wie man glaubt, daß es sich verhalten müsse. Dies ist die Bedeutung der Behauptung von Chapanis, daß man eine Theorie oft erst durch Modellbildung versteht, da man das, was in ihr steckt, ohne Computersimulation nicht herausfinden kann.

Psychologische Phänomene sind oft nicht auf eine oder wenige Ursachen reduzierbar. Weiterhin verändern sich oft die Bedingungen für das Verhalten oder die internen Prozesse während des Verhaltens und während des Ablaufs der Prozesse. Es ist nicht einfach, Theorien, die sich auf solche komplexe und plastische Sachverhalte beziehen, zu prüfen. Ohne Computersimulation kommt man bei der Prüfung von Aussagen über solche Gebilde kaum aus. Man hat es dann ja nicht mit einfachen Kausalketten zu tun, sondern mit Prozessen, die durch ein Geflecht verschiedenartiger Beziehungen zustandekommen. Solchen Kausalgeflechten wird man mit den „Alltagsmitteln“ der logischen Deduktion „von Hand“ oder der Auflösung von Gleichungssystemen nicht mehr gerecht. „In many cases the available mathematical and logical tools are simply insufficient to give determined answers to questions about the consequences of changes in variables and the overall behavioural tendency of the theory under various conditions“, meint Hanneman (1988, S. 84) in bezug auf solche komplizierte theoretische Gebilde.

Aber nicht nur als Deduktionshilfen sind Computermodelle nützlich. Der Gebrauch von Modellen macht eine Theorie **anschaulich**. Das Modell produziert Verhalten, genau wie eine „richtige“ Versuchsperson; auf diese Weise wird die Theorie plastisch, und man sieht unmittelbar, ob sie überhaupt ein



aussichtsreicher Kandidat für die Erklärung des Verhaltens ist. Die Computersimulation macht oft komplexe *und* exakte Theorien erst möglich! Ein Modell ermöglicht es, mit der Theorie „zu spielen“, mal diesen, mal jenen Parameter zu verändern um so zu sehen, welches Gewicht die verschiedenen Annahmen haben.

A propos „Spiel“: Chapanis meint: „Modelle machen Spaß“. In der Tat stellt die Modellierung eines komplizierten psychischen Prozesses die Unmittelbarkeit der Beobachtung wieder her. Man kann mit der Theorie gewissermaßen so umgehen wie mit einer Versuchsperson. Oder besser gesagt: wie mit vielen verschiedenen Versuchspersonen. Läßt man die Theorie (genauer gesagt: das Modell) Feuerwehrhauptmann spielen, so stellen sich ganz schnell Assoziationen zu tatsächlichen Versuchspersonen ein. Es ist reizvoll, zu beobachten, welche „Persönlichkeitstypen“ man durch die Variation bestimmter Parameter erzeugen kann, und entsprechende „Spielereien“ mit der Theorie machen in der Tat Spaß, halten dadurch die Motivation der Beschäftigung mit dem System aufrecht und führen so zu neuen Einsichten.

Im Grunde ist es merkwürdig, wie langsam die „Modellmethode“ in der Psychologie vordringt. Seit Ende der 50er Jahre gibt es Computermodelle für psychische Prozesse; in den 60ern hatte die Tendenz, Computermodelle zu konstruieren, bereits einen bestimmten Höhepunkt erreicht. Dennoch ist diese Methode keineswegs gebräuchlich. Wahrscheinlich sind die technischen Schwierigkeiten, der Erwerb des erforderlichen Knowhows zum Umgang mit dem Computer und mit entsprechenden Programmiersprachen trotz der Tatsache, daß man heutzutage ein Universitätsrechenzentrum der 60er Jahre in der Aktentasche mit sich herumtragen kann und trotz der Verfügbarkeit sehr leicht zu verwendender Programmiersprachen, die zum Teil ganz explizit auf die Simulation von Prozessen in den Sozialwissenschaften ausgerichtet sind (z.B. die Sprache DYNAMO von Forrester und Mitarbeitern, s. Richardson & Pugh, 1981) noch immer schwer überwindbar.

Dabei ist die Computersimulation für die Psychologie und auch viele Sozialwissenschaften das Mittel der Wahl zur Theorienkonstruktion und -prüfung. Gewöhnlich hat man es in der Psychologie nicht mit einfachen Prozessen zu tun, die sich zwischen jeweils zwei Variablen abspielen.

Die Modellkonstruktion hilft aber nicht nur; sie hat auch ihre Gefahren. Chapanis (1963) nennt folgende:

1. Modelle verführen zu vorschneller Generalisierung.
2. Die Beziehungen zwischen den Modellvariablen können andere sein, als die zwischen den Prototypvariablen.
3. Modelle werden oftmals nicht validiert.

4. Modelle können nützliche Forschungsenergie in unproduktiver Aktivität vergeuden.

Man sollte die in diesen vier Behauptungen von Chapanis steckenden Gefahren der Modellkonstruktion ernst nehmen. In der Tat kann die Freude über die auf den ersten Blick gelungene Modellierung eines komplizierten psychischen Ablaufes so groß sein, daß man das endlich lauffähig gewordene Produkt seiner Bemühungen eben wegen dieser Bemühungen auch schon für gültig hält. Die Geschichte der Computersimulation in der Psychologie ist voll von solchen Übergeneralisierungen. Das wohl bislang immer noch berühmteste Beispiel für die Simulation menschlichen Denkens, nämlich der GPS (General Problem Solver, Newell & Simon, 1972) wurde niemals dadurch validiert, daß man überprüfte, ob das System tatsächlich die Denkabläufe einer größeren Anzahl von Versuchspersonen prognostizieren kann. Übereinstimmungen zwischen dem Verhalten des GPS und dem Verhalten von „echten“ Versuchspersonen wurde nur streckenweise gefunden, und die Validierung des GPS geschah gewissermaßen durch Anekdoten. So sieht es mit sehr vielen Computersimulationen aus.

Chapanis' Anmerkungen 2 und 3 hängen mit der Anmerkung 1 eng zusammen. Wenn man ein „lauffähiges“ Modell schon allein deswegen, weil es tut, was der Konstrukteur will, für auch gültig hält, wird man nicht dazu neigen, große Anstrengungen zu machen, das Modell zu validieren. Und damit schleicht sich die Gefahr ein, daß die Gültigkeit des Modells nur von recht geringer Reichweite ist.

Der Umgang mit Computermodellen komplizierter (psychischer) Prozesse ist sehr spannend. Der Computer ist für denjenigen, der mit ihm umgehen kann, ein dynamisches Spielzeug von hohem Unterhaltungswert. Und daher kann es sich leicht ereignen, daß der Modellkonstrukteur Aufwand und Nutzen nicht mehr balanciert und viel Forschungsenergie in fruchtlosen Spielereien verloren geht.

#### **4. *Schlußbemerkungen***

Wie im Eingangskapitel betont, ist das Thema „Theorienbildung“ für die Psychologie wohl wichtiger als für andere Wissenschaften. Der Ingenieur und wohl auch der Arzt können bei der Lösung ihrer praktischen Probleme in hohem Maße auf „vorgefertigte“ Verfahrensweisen zurückgreifen. Von Psychologen aber wird fast immer verlangt, daß sie sich über die Hintergründe der Phänomene, mit denen sie sich befassen müssen, Gedanken machen. Eine Depression kann diese oder jene Gründe haben; dasselbe gilt für die Drogensucht oder für Alkoholismus. Schulschwierigkeiten können am Klassenklima,

am Elternhaus, an der individuellen Begabung oder an einer Kombination aller dieser Faktoren zusammen liegen. Die multiple und variable Bedingtheit psychischer Phänomene stellt den Psychologen sehr oft vor die Notwendigkeit, neue Hypothesen erfinden zu müssen.

Im Hinblick auf die wieder und wieder beklagte Unverbundenheit der Ergebnisse der psychologischen Forschung, im Hinblick auf den Fragmentarismus der theoretischen Ansätze ist nicht nur in der angewandten sondern auch in der wissenschaftlichen Psychologie das Theorienbilden sehr notwendig. Das reine Sammeln von Daten, fälschlicherweise oft mit wissenschaftlicher Tätigkeit an sich identifiziert, führt nicht weit; wir brauchen übergreifende theoretische Ansätze. Und wir brauchen solche Ansätze nicht nur, sondern haben heute für viele Bereiche der Psychologie auch bessere Mittel für die Produktion von Theorien als früher. Der menschliche Geist ist überfordert, wenn er die Interaktionen von 20 oder mehr Variablen zugleich betrachten soll; hier bieten sich der Computer und die Modellbildung und Simulation als „theoretische Werkbank“ an. Aus diesem Grunde haben wir in diesem Kapitel auch viel Wert auf das Thema „Computermodele“ gelegt. Wenn man es etwas überspitzt ausdrücken möchte, könnte man sagen, daß für viele Bereiche der Psychologie ein Theoretisieren, welches diesen Namen verdient, erst seit der Existenz des Computers möglich ist.

Auf der anderen Seite wird sich das Theorienbilden wohl nie so exakt und formal betreiben lassen wie das Theorienprüfen. Gigerenzer (1991, S.1) hat wohl recht, wenn er meint, daß derjenige Teil des Ozeans der wissenschaftlichen Tätigkeiten, der das Prüfen von Theorien betrifft, in heller Sonne liegt, wohingegen der Teil, der die Theorienbildung betrifft, von dichten Nebelbänken bedeckt ist. Die Nebelbänke werden wohl nie ganz verschwinden. Beim Theorienbilden schafft man etwas Neues, und man weiß vorher nicht, wie das Ergebnis aussehen wird. Daher kann das Bilden von Theorien nicht nach einem festen Schema vor sich gehen, obwohl man das in der Psychologie (s. Faktorenanalyse) manchmal geglaubt hat. Aber einige Heuristiken, auf die man sich besinnen kann, wenn es um das Bilden von Theorien geht, kann man doch nennen. Und die Kenntnis dieser Verfahren mag hilfreich sein. Der Theoriekonstrukteur in der Psychologie, der sich der Methode der Computersimulation bedienen will oder muß, muß sich allerdings darauf einstellen, daß er in Zukunft einiges über die Theorien der Informationsverarbeitung und die dahinter stehenden Bereiche der formalen Logik und der Mathematik wissen muß.

## **Literatur**

Aiserman, M.A., Gussew, L. A., Rosonoer, L. I. Smirnowa, I. M. & Tal, A. A. (1967): **Logik - Automaten - Algorithmen**. München: Oldenbourg.

- Anderson, J.R. (1983): **The Architecture of Cognition**. Cambridge, Mass.: Harvard University Press.
- Aristoteles (1986): **über die Seele** (übersetzt von Willy Theiler). Berlin: Akademieverlag.
- Ashby, W. R. (1966): Mathematical Models and Computer Analysis of the Functions of the Central Nervous System. **American Review of Physiology**, **28**, 89-106.
- Bertalanffy, L. von (1968): General **System Theory**. New York: Braziller.
- Bischof, N. (1968): Kybernetik in Biologie und Psychologie. In S. Moser & S. J. Schmidt (Hrsg.), **Information und Kommunikation** (S. 68-96). München: Oldenbourg.
- Bischof, N. (1981): Aristoteles, Galilei, Lewin - und die Folgen. In W. Michaelis (Hrsg.), **Bericht über den 32. Kongreß der Deutschen Gesellschaft für Psychologie, Zürich** 1980 (S. 17-39). Göttingen: Hogrefe.
- Bischof, N. (1985): **Das Rätsel Ödipus**. München: Piper.
- Boguslaw, R. (1965): The New Utopies: **A Study of System Design and Social Change**. Englewood Cliffs, N.J.: Prentice Hall.
- Bortz, J. (1985): **Lehrbuch der Statistik**. Berlin: Springer.
- Braithwaite, R. B. (1963): Models in the Empirical Sciences. In A. Nagel, P. Suppes & A. Tarski (Eds.), **Logic, Methodology and the Philosophy of Science** (pp. 224-231). Stanford: Stanford University Press.
- Chapanis, A. (1963): Men, Machines, and Models. In M. Marx (Ed.), **Theories in Contemporary Psychology** (pp. 232-250). New York: Macmillan.
- Dörner, D. (1974): **Die kognitive Organisation beim Problemlösen**. Bern: Huber.
- Dörner, D. (1988): Die kleinen grünen Schildkröten und die Methoden der experimentellen Psychologie. **Sprache und Kognition**, **8**, 86-97.
- Dörner, D. & Pfeifer, E. (1992): Strategisches Denken, Streß und Intelligenz. **Sprache und Kognition**, **11**, 75-90.
- Ehrhardt, K. J. (1975): Neuropsychologie **motivierten Verhaltens**. Stuttgart: Enke.
- Gigerenzer, G. (1988): Woher kommen Theorien über kognitive Prozesse? **Psychologische Rundschau**, **39**, 91-100.
- Gigerenzer, G. (1991): From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology. **Psychological Review**, **98**, 1-13.
- Hanneman, R. A. (1988): **Computer Assisted Theory Building: Modeling Dynamit Social Systems**. Newbury Park: Sage.
- Herrmann, T. (1969): **Lehrbuch der empirischen Persönlichkeitsforschung**. Göttingen: Hogrefe.
- Herrmann, T. (1982): über begriffliche Schwächen kognitivistischer Kognitionstheorien: Begriffsinflation und Akteur-System-Kontamination. **Sprache und Kognition**, **1**, 3-14.
- Herrmann, T. (1990): Die Experimentiermethodik in der Defensive? **Sprache und Kognition**, **9**, 1-11.
- Hesse, M. B. (1970): **Models and Analogies in Science**. Notre Dame, Indiana: University Press.

- Hofstätter, P. (1960): **Psychologie**. Frankfurt/Main: Fischer.
- Hyland, M. (1981): **Introduction to Theoretical Psychology**. London: Macmillan.
- Johnson-Laird, P. N. (1988): **The Computer und the Mind: An Introduction to Cognitive Science**. Cambridge, Mass.: Harvard University Press.
- Klir, J. & Valach, M. (1967): **Cybernetic Modelling**. London: Iliffe Books.
- Leahey, T.H. (1987): **A History of Psychology: Main Currents in Psychological Thought**. New York: Prentice-Hall.
- Marx, M. (1963): **Theories in Contemporary Psychology**. New York: MacMillan.
- Miller, G.A., Galanter, E. & Pribram, K.H. (1960): **Plans und the Structure of Behavior**. New York: Holt, Rinehart & Winston.
- Neel, A. F. (1974): **Handbuch der psychologischen Theorien**. München: Kindler.
- Newell, A. & Simon, H. A. (1972): **Human Problem Solving**. Englewood Cliffs, N.J.: Prentice Hall.
- Pauli, W. (1957): Phänomen und physikalische Realität. **Dialectica, 11, 36-48**.
- Pavlov, I.P. (1972): **Die bedingten Reflexe**. München: Kindler.
- Popper, K. (1966): **Logik der Forschung**. Tübingen: Mohr (Paul Siebeck).
- Richardson, G.P. & Pugh, A.L. (1981): **Introduction to System Dynamics Modeling with DYNAMO**. Cambridge, Mass.: MIT Press.
- Roth, E. (Hrsg.) (1984): **Sozialwissenschaftliche Methoden**. München: Oldenbourg.
- Schneirla, T. C. (1959): An Evolutionary and Developmental Theory of Biphasic Processes Underlying Approach and Withdrawal. In M.R. Jones (Ed.), **Nebraska Symposium on Motivation** (vol. 7). Lincoln: University of Nebraska Press.
- Searle, J. (1984): **Minds, Brains and Science**. Harvard: Harvard University Press.
- Sergejew, J. (1970): Psychologische Hintergründe großer Entdeckungen. **Bild der Wissenschaft, 7, 456-553**.
- Tack, W. H. (1969): Mathematische Modelle in der Sozialpsychologie. In C. F. Graumann (Hrsg.), **Handbuch der Psychologie, Band 7: Sozialpsychologie, 1. Halbband: Theorien und Methoden** (S.232-218. Göttingen: Hogrefe.
- Traxel, W. (1968): **über Gegenstand und Methode der Psychologie**. Bern: Huber.
- Turing, A. M. (1956): Can a Machine Think? In: **The World of Mathematics** (p.2099). New York: Simon & Suster.
- Turing, A.M. (1959): Computing Machinery and Intelligence. **Mind, 59, 236-433**.
- Wandschneider, D. (1991): Naturphilosophisch querdenkend. Vorwort zur deutschsprachigen Ausgabe von R. Penrose: **Computerdenken - Des Kaisers neue Kleider oder die Debatte um Künstliche Intelligenz, Bewußtsein und die Gesetze der Physik** (S. XVII-XX). Heidelberg: Spektrum-Verlag.
- Wegener, H. & Dörner, D. (1973): Simulation als Forschungstechnik: Bericht über ein Symposium. In G. Reinert, **Bericht über den 27. Kongreß der Deutschen Gesellschaft für Psychologie, Kiel 1970** (S. 69-78). Göttingen: Hogrefe.
- Wessels, M.G. (1984): **Kognitive Psychologie**. New York: Harper & Row.
- Wyss, D. (1970): **Die tiefenpsychologischen Schulen von den Anfängen bis zur Gegenwart**. Göttingen: Vandenhoeck und Ruprecht.

## 9. Kapitel

# Theoriebewertung

*Volker Gadenne*

Grundlage der Bewertung von erfahrungswissenschaftlichen Theorien sind eine Reihe von sehr allgemeinen Zielen. Besonders häufig werden in den Wissenschaften und in der Wissenschaftstheorie folgende Ziele genannt: Wahrheit bzw. empirische Adäquatheit, logische Konsistenz (Widerspruchsfreiheit), semantische Einheitlichkeit, Einfachheit, Informationsgehalt, Erklärungskraft, empirische Prüfbarkeit, Tiefe. Es handelt sich bei diesen Zielen allerdings noch nicht um exakte Kriterien, sondern um regulative Ideen, die allein nur eine grobe Orientierung ermöglichen. Sie müssen durch Kriterien ergänzt werden, aus denen sich Handlungsanweisungen für konkrete wissenschaftliche Problemsituationen ergeben. Man kann sich diesen Zusammenhang in Form einer Ziel-Mittel-Hierarchie vorstellen, an deren Spitze die allgemeinsten Ziele stehen, während die darunter angeordneten Kriterien jeweils Mittel darstellen. So kann z.B. empirische Bestätigung als Kriterium für die Wahrheit einer Theorie herangezogen werden. Um empirische Bestätigung zu erreichen, müssen kritische empirische Tests durchgeführt werden. Dieses Zwischenziel erfordert wiederum Bedingungen wie Validität der Messung und Kontrolle von Störfaktoren, die sich abermals in noch konkretere Maßnahmen aufschlüsseln ließen. Alle Knotenpunkte einer solchen Ziel-Mittel-Hierarchie können als methodologische Regeln (oder Prinzipien, Empfehlungen) formuliert werden: „Konstruiere Theorien mit hohem Informationsgehalt“; „Akzeptiere nur intern valide Untersuchungsergebnisse“ usw. Die allgemeineren Prinzipien dieser Art, die in allen Erfahrungswissenschaften anzutreffen sind, pflegt man eher als methodologisch zu bezeichnen, die spezielleren eher als methodisch. Letztere stehen in unmittelbarem Zusammenhang zu den speziellen Untersuchungsverfahren der einzelnen Wissenschaften und sind oft gegenstandsspezifisch.

An dieser Stelle empfiehlt es sich, etwas über die Natur methodologischer Prinzipien (Regeln, Empfehlungen) zu sagen (vgl. hierzu auch Kap.1). Nach einer verbreiteten Auffassung sind solche Prinzipien Normen wissenschaftlichen Vorgehens. Woher aber könnte die Methodologie die Berechtigung nehmen, Normen zu setzen? Überzeugender ist es, das Hauptanliegen der Methodo-

logie folgendermaßen zu deuten (Albert, 1987): Sie rekonstruiert die Vorgehensweisen der Wissenschaften und analysiert, inwieweit diese Vorgehensweisen geeignet sind, die von den Wissenschaften selbst gesetzten Ziele zu erreichen. Darüber hinaus können auch die Ziele selbst diskutiert werden, etwa unter dem Gesichtspunkt ihrer Erreichbarkeit oder Vereinbarkeit miteinander. Hauptsächlich geht es jedoch um die Adäquatheit der verwendeten Strategien bis hin zu konkreten Methoden, wobei die vom Gegenstand abhängigen Problemsituationen innerhalb der Wissenschaften zu berücksichtigen sind. Hier berührt sich die methodologische Analyse mit der Methodendiskussion in den Einzelwissenschaften, von der sie ohnehin nicht scharf zu trennen ist. - Von einer so verstandenen Methodologie können kritische Impulse für die Theorienbildung und empirische Forschung ausgehen. Möglicherweise meinen viele Autoren dies, wenn sie sagen, daß die Methodologie eine normative Komponente habe. Zutreffender ist es jedoch, dieses Anliegen als ein technologisches zu bezeichnen, denn methodologische Prinzipien lassen sich offenbar in die Form zweckrationaler Aussagen bringen: Dieses Vorgehen ist geeignet (geeigneter als jenes), um die zugrunde gelegten Ziele zu erreichen. Es ist üblich, spezielle methodische Verfahren, wie einen Versuchsplan oder psychologischen Test, nach ihrem instrumentellen Wert zu beurteilen. In entsprechender Weise kann man z.B. das Falsifikationsprinzip danach bewerten, wie sehr es dazu beiträgt, Theorien im Hinblick auf empirische Adäquatheit zu verbessern; die strukturalistische Rekonstruktionsweise von Theorien kann als Mittel zu dem Zweck angesehen und beurteilt werden, Theorien präzise zu formulieren und ihre Struktur offenzulegen usw. Zur Beurteilung methodologischer Prinzipien können auch Ergebnisse der Einzelwissenschaften selbst herangezogen werden (z.B. der Wahrnehmungspsychologie zu den Möglichkeiten und Irrtümern der Beobachtung). - Bei diesem Unternehmen muß nun allerdings die Methodologie an den aktuellen Erkenntnisstand und die verfügbaren Methoden der einzelnen Wissenschaften anknüpfen, und insofern ist sie auch auf eine zutreffende Deskription bzw. Rekonstruktion dieser Wissenschaften, ihrer Probleme, Ziele und Vorgehensweisen angewiesen. Die Aussagen, die im folgenden zur Bewertung von Theorien gemacht werden, sind zum Teil deskriptiver und zum Teil zweckrationaler Art. (Es wäre allgemein zu empfehlen, den Begriff der „Metatheorie“ für die deskriptive Komponente und den der „Methodologie“ für die zweckrationale zu verwenden.)

Ein zweiter Punkt betrifft die Art und Weise, in der methodologische Prinzipien verwendet werden. Für alle im folgenden behandelten Kriterien (vielleicht mit Ausnahme der logischen Konsistenz) gilt, daß Wissenschaftler sie in bestimmten Fällen absichtlich umgangen und dennoch bedeutsame Leistungen vollbracht haben: Theorien werden manchmal eingeschränkt, ihr Gehalt wird also vermindert, um sie mit den Fakten besser in Einklang zu bringen. Manchmal werden Theorien aufrechterhalten, denen eindeutige Fakten widersprechen. Manchmal werden theoretische Entwicklungen vorgenommen, die

sich keineswegs durch Einfachheit auszeichnen. Dies alles ist unumstritten und hat einige Wissenschaftstheoretiker zu der Auffassung gebracht, daß in der Wissenschaft keine methodologischen Regeln existieren und daß sie auch nur hinderlich wären (Feyerabend, 1976). Dabei darf jedoch nicht übersehen werden, daß die oben genannten Ziele und entsprechenden Kriterien in allen empirischen Wissenschaften als regulative Ideen anerkannt und berücksichtigt werden. Dies gilt selbst für die umstrittene Einfachheit und die kaum klar definierbare semantische Konsistenz. Die Erklärungskraft einer Theorie, also eine Eigenschaft, die allgemein positiv bewertet wird, könnte unbegrenzt gesteigert werden, indem man der Theorie beliebige weitere Aussagen als Axiome hinzufügt. Anhand der Tatsache, daß ein entsprechendes Gebilde allen Wissenschaftlern völlig „unattraktiv“ erscheinen würde, läßt sich erkennen, daß sie sich sehr wohl an den Ideen der Einfachheit und semantischen Konsistenz orientieren: Obwohl nicht genau festgelegt werden kann, wie einfach und semantisch konsistent eine Theorie sein muß, gilt ein Vorgehen, das gegen diese regulativen Ideen in krasser Weise verstößt, schlicht als indiskutabel. Weiterhin wäre es absurd, zu behaupten, daß empirische Ergebnisse überhaupt keinen Einfluß darauf hätten, wie Theorien bewertet werden und wie weiterhin mit ihnen verfahren wird. Nur läßt sich dieser Einfluß möglicherweise nicht in Form strikter Regeln etwa der Induktion oder Falsifikation darstellen, die ausnahmslos eingehalten werden. Die Art des Umgangs mit methodologischen Kriterien ist eher der Verwendung von heuristischen Prinzipien der Problemlösung vergleichbar, die etwa empfehlen, die Beschaffenheit des Problems genau zu analysieren, Teilziele zu setzen, analoge Aufgaben zu studieren usw. (vgl. zur Heuristik Kap.8 dieses Bandes). Solche Empfehlungen gelten als nützlich, obwohl sie nicht eindeutig festlegen, wie zu verfahren ist und eine Problemlösung nicht garantieren. In ähnlicher Weise kann man methodologische Regeln auffassen: Sie lenken die Phantasie des Forschers in eine bestimmte Richtung, ohne den Lösungsweg bis ins einzelne zu bestimmen (Albert, 1987). Sie lassen also Raum für zusätzliche, von der Situation abhängige Gründe, die gegebenenfalls ein Abweichen von einer Regel als sinnvoll erscheinen lassen. Nichtsdestoweniger kann es für den Fortschritt der Wissenschaften entscheidend sein, daß diese Regeln als Richtlinien zur Orientierung grob eingehalten werden. Im folgenden werden methodologische Prinzipien in diesem Sinne verstanden.

## *1. Logische und semantische Bewertungskriterien*

### *1.1 Logische Konsistenz*

*Logische Konsistenz* oder *Widerspruchsfreiheit* ist sicherlich unter allen Bedingungen, die an Theorien gestellt wurden, die eindeutigste und am wenigsten umstrittene. Dies gilt jedenfalls für den Bereich der empirischen Wissenschaft-



ten und die auf sie bezogene Wissenschaftstheorie. (Von Versuchen, Widersprüche im Rahmen eines dialektischen Denkens zu rechtfertigen, sei hier abgesehen.) Nicht nur jede Art der empirischen Prüfung einer Theorie, sondern darüber hinaus jede rationale Argumentation setzen das Prinzip voraus, Widersprüche zwischen Aussagen zu vermeiden. Eine Menge von Aussagen ist logisch inkonsistent, wenn aus ihr sowohl eine Aussage A als auch deren Negation Non-A logisch ableitbar ist; ist dies nicht der Fall, gilt die Aussagenmenge als logisch konsistent. Die Forderung nach Widerspruchsfreiheit steht im Dienste der Wahrheit einer Theorie (verstanden als Aussagenmenge). Zwar ist logische Konsistenz keine Garantie für die Wahrheit einer Theorie. Umgekehrt ist aber logische Inkonsistenz eine Garantie dafür, daß die Theorie falsch ist: Entweder A, oder aber Non-A muß falsch sein. Daher enthält eine Theorie, in der sowohl A als auch Non-A vorkommen, auf jeden Fall eine falsche Aussage und ist somit als Konjunktion von Aussagen falsch. Selbstverständlich wird niemand absichtlich ein Axiom und anschließend die Negation dieses Axioms formulieren. Es kann allerdings sein, daß es die Axiome zusammen erlauben, im Verlauf vieler Ableitungsschritte einerseits zu A, andererseits zu Non-A zu gelangen. Bei wenigen theoretischen Variablen und Axiomen von relativ einfacher Form ist diese Gefahr nicht allzu groß, etwa, wenn eine Theorie aus den zwei Axiomen besteht: Wenn A dann B; Wenn B dann C. Man kann ohne formale Analyse „überschauen“, daß diese theoretischen Annahmen vereinbar sind. Anders ist es, wenn eine Theorie aus vielen Aussagen besteht, die z.B. komplexe kognitive Prozesse und ihre Wechselwirkungen beschreiben. Weiterhin darf nicht übersehen werden, daß die Forderung nach Widerspruchsfreiheit auch die Hilfsannahmen miteinbeziehen sollte, jedenfalls diejenigen, die im Rahmen einer bestimmten Anwendung der Theorie vorausgesetzt, meist aber nicht explizit formuliert werden. Es ist möglich, daß solche Hilfsannahmen, die oft anderen Theorien entstammen, teilweise miteinander unvereinbar sind bzw. sich zusammen mit den Axiomen als unvereinbar erweisen.

Bisweilen wird betont, daß aus einer widerspruchsvollen Theorie jede Aussage logisch folgt, die Theorie somit auch jedem empirischen Ergebnis Rechnung zu tragen vermag. In der Tat folgt nach den Regeln der deduktiven Logik aus A zusammen mit Non-A jede beliebige Aussage. Freilich wird niemand diese Möglichkeit ernsthaft dazu benutzen, Beliebiges vorherzusagen oder eine Theorie auf diese Weise zu bestätigen. Das eigentliche Problem besteht darin, daß im Falle eines Widerspruchs einerseits eine Garantie dafür gegeben ist, daß mindestens ein Axiom bzw. eine Hilfsannahme falsch ist, und daß andererseits in diesem Fall auch garantiert ist, daß (unendlich viele) falsche empirische Vorhersagen aus der Theorie ableitbar sind. Daher ist Widerspruchsfreiheit auch bei instrumentalistischer Auffassung von Bedeutung. Zwar geht es hier nicht um die Wahrheit der Theorie, doch sollen jedenfalls die empirischen Folgerungen der Theorie wahr sein.

Im allgemeinen gibt es keinen Algorithmus, mit dessen Hilfe die Widerspruchsfreiheit einer erfahrungswissenschaftlichen Theorie nachgewiesen werden könnte. Ein solcher Algorithmus existiert nur dann, wenn die Theorie in der Prädikatenlogik mit einstelligem Prädikaten formulierbar ist. Die meisten Theorien enthalten aber quantitative oder zumindest relationale Begriffe. Die einzige Möglichkeit einer Prüfung auf Widerspruchsfreiheit besteht dann in dem Versuch, zwei miteinander inkonsistente Aussagen abzuleiten. Wenn die von einer Theorie beschriebenen Prozesse auch als Computerprogramm formuliert wurden, so bietet die Simulation dieser Prozesse auf dem Computer eine Hilfe, Widersprüchen auf die Spur zu kommen (vgl. dazu Kap.5).

Soweit wurde die Bedingung der Widerspruchsfreiheit auf jeweils eine Theorie bezogen. Zusätzlich zu dieser *internen* Konsistenz kann *externe* Konsistenz gefordert werden: Die verschiedenen Theorien einer Wissenschaft bzw. das gesamte wissenschaftliche Wissen sollten konsistent sein (Bunge, 1967). Dies ist insofern begründet, als zu Erklärungs- und Vorhersagezwecken häufig mehrere Theorien gleichzeitig angewendet werden müssen. In diesem Fall gelten die bereits angestellten Überlegungen. Allerdings ergeben sich für die interne und externe Konsistenz jeweils verschiedene methodologische Regeln, wenn man weitere Gesichtspunkte hinzuzieht. Im Falle der internen Inkonsistenz kann empfohlen werden, die betreffende Theorie aufzugeben oder so zu verändern, daß der Widerspruch beseitigt ist. In bezug auf die externe Konsistenz ist die Situation komplizierter. Bunge (1967) empfiehlt, Konsistenz zu verlangen und die jeweils besser gesicherten Theorien zum Test für die anderen zu machen. Nach seiner Auffassung sollte beispielsweise eine psychologische Theorie ohne Rücksicht auf ihren empirischen Erfolg aufgegeben werden, wenn sie mit einer biologischen oder gar physikalischen in Widerspruch steht. Diese Auffassung tritt allerdings in Konflikt mit der Erkenntnis, daß Wissenschaft keineswegs in einer reinen Kumulation von Wissen besteht, sondern (auch) revolutionären Charakter hat. Um die Probleme und Irrtümer innerhalb des herkömmlichen Wissens zu finden, ist es unbedingt erforderlich, dieses Wissen durch neue Theorien herauszufordern, die mit den vorhandenen Theorien und manchmal auch Fakten unvereinbar sind, in einem gewissen Sinne also *kontrainduktiv* vorzugehen (Feyerabend, 1976). Die Empfehlung, konkurrierende Theorien zu entwickeln, ist gleichzeitig eine Empfehlung, Widersprüche zu produzieren. Allerdings werden diese Widersprüche nur zu dem Zweck produziert, die irrtümlichen Bestandteile des Gesamtwissens zu entdecken, um sie zu eliminieren. Diese Eliminierung ist wiederum am Kriterium der logischen Konsistenz orientiert.

Diese Überlegungen lassen es durchaus auch als sinnvoll erscheinen, Theorienbildung in Einklang mit bestimmten anderen Theorien zu betreiben, also z.B. bei der psychologischen Theorienbildung die Erkenntnisse der Physiologie zu berücksichtigen. In diesem Falle folgt man der heuristischen Strategie,

vorläufig dasjenige Wissen zum Maßstab zu machen, das als besser gesichert erscheint. Wegen der grundsätzlichen Fehlbarkeit allen Wissens darf diese Strategie eben nur nicht zu einer strikten, ausnahmslos verbindlichen Norm gemacht werden.

In bezug auf die externe Konsistenz ergeben sich aus realistischer und instrumentalistischer Sicht unterschiedliche Bewertungen. Für den Instrumentalisten ist nur die Konsistenz derjenigen Annahmen notwendig, die gemeinsam zum Zweck der logischen Ableitung weiterer Annahmen verwendet werden sollen. Aus realistischer Perspektive muß dagegen die Konsistenz aller Annahmen angestrebt werden, die sich auf denselben Ausschnitt der Wirklichkeit beziehen. Wenn beispielsweise eine Theorie mentale Strukturen oder Prozesse annimmt, für die es nach vorhandenem Wissen keine physiologische Grundlage gibt, so ist dies nur dann ein Problem, wenn man von der Voraussetzung ausgeht, daß sich beide Theorien auf die (eine) Wirklichkeit beziehen.

## 1.2 Semantische Einheitlichkeit

Eine wissenschaftliche Theorie sollte semantisch konsistent sein, ihre Begriffe sollten eine gewisse *Einheitlichkeit* aufweisen (Bunge, 1967). Die Aussage „Magnetfelder sind lieblich“ wird nicht in erster Linie als falsch, sondern als semantisch unsinnig empfunden. Herrmann (1982, 1987) hat die Idee der semantischen Konsistenz im Bereich der kognitiven Psychologie angewendet. Er verweist auf Probleme, die sich ergeben, wenn man Begriffe z.B. einer systemtheoretischen und einer handlungstheoretischen Perspektive vermengt. Es kann dann zu Aussagen kommen, denen zufolge z.B. „ein kognitives Schema etwas sieht“, „der zentrale Prozessor meint, sein Ziel nicht erreichen zu können“ oder „das Bewußtsein die kognitiven Prozeduren steuert“. Wenn man solche Aussagen anders als metaphorisch deutet, so scheinen sie semantisch nicht korrekt zu sein.

Nun gibt es allerdings kein präzises Kriterium für semantische Konsistenz. Ob gegebene Aussagen als konsistent empfunden werden, hängt in hohem Maße von den zur jeweiligen Zeit akzeptierten Theorien ab. Churchland (1984, S. 30) erläutert die Problematik anhand der Identitätstheorie zum Leib-Seele-Problem, aus der sich folgern läßt, daß ein Gedanke (identisch mit einem Gehirnvorgang) einen Ort haben muß und ein Gehirnvorgang einen Inhalt haben kann - nach Meinung einiger Kritiker eine semantisch absurde Konsequenz. Churchland verteidigt die Identitätstheorie mit dem Hinweis, daß die früheren Kritiker der Kopernikanischen Theorie folgendermaßen hätten argumentieren können: Es ist semantisch unsinnig, zu behaupten, daß sich die Erde bewege. Denn sich *zu bewegen*, *bedeutet* nichts anderes, als seine Position relativ zur Erde zu verändern. - Diese Argumentation ist so lange über-

zeugend, als man eine bestimmte Theorie über das Universum voraussetzt. Sobald aber diese Theorie aufgegeben wird, ist es notwendig, einen anderen Begriff der Bewegung zu bilden bzw. die semantischen Regeln für den Ausdruck Bewegung zu ändern (vgl. dazu Feyerabend, 1976). Vergleichbares gilt für Leib-Seele-Theorien. Allgemein gesprochen: Gewisse Theorien über die Welt, die allgemein akzeptiert und mit der Zeit zu Selbstverständlichkeiten werden, fließen in die Sprache als semantische Regeln ein (vgl. dazu Wendel, 1987, Kap.4). Aber manchmal erweist es sich als notwendig, diese Theorien durch andere zu ersetzen und sich damit auch an neue semantische Regeln zu gewöhnen.

Eine als semantisch inkonsistent empfundene Aussage ergibt zusammen mit den theoretischen Aussagen, die der Inkonsistenz zugrunde liegen, eine *logische Inkonsistenz*. Zum Beispiel steht die Aussage, daß „dieses Magnetfeld lieblich“ sei, in einem logischen Widerspruch zu der von den meisten (Ausnahme: die Panpsychisten) akzeptierten Annahme, daß Magnetfelder keine mentalen Eigenschaften haben. Bei einigen der von Herrmann kritisierten Beispielen läßt sich ebenfalls eine bestimmte logische Inkonsistenz entdecken (Gadenne & Oswald, 1991): Die kognitive Psychologie, die Personen als informationsverarbeitende Systeme auffaßt, postuliert gesetzmäßige Beziehungen zwischen den internen Zuständen eines kognitiven Systems. Viele Handlungstheorien behaupten jedoch explizit, daß Handlungen und Handlungsgründe nicht unter Gesetze fallen.

Will man eine semantische Inkonsistenz beseitigen, so muß entweder die problematische Aussage oder die ihr zugrunde liegende Theorie verworfen werden. Meist wird man dazu neigen, eher eine neue Aussage aufzugeben als eine grundlegende Theorie, die vielfältig im Gesamtwissen verankert ist. Dies gilt natürlich insbesondere dann, wenn die neue Aussage sich nur als Resultat einer nachlässigen Formulierung oder einer metaphorischen Ausdrucksweise herausstellt, so daß das Problem allein durch eine andere Darstellungsweise gelöst werden kann. Doch mag auch der Fall vorkommen, daß eine neue Idee explizit in Konkurrenz zu herkömmlichen Theorien entwickelt wurde, und in diesem Fall sollte ein entsprechendes Vorgehen nicht durch das methodologische Prinzip untersagt werden, jede neue Aussage müßte auf der Grundlage des herkömmlichen Wissens semantisch konsistent sein. Die Forderung nach semantischer Konsistenz kann nur einem heuristischen Zweck dienen. Sie dient zum einen dazu, zahllose unsinnige Aussagen, die keinerlei Chance auf Bestätigung hätten, von vornherein aus der Betrachtung auszuschalten. Wenn dann bei den ernsthaft vorgeschlagenen Aussagen dennoch der Eindruck einer Inkonsistenz entsteht, so ist dies ein Hinweis auf ein Problem. Wie aufgezeigt wurde, gibt es in diesem Fall verschiedene Lösungswege, von denen keiner als der einzig richtige vorgeschrieben werden kann.

### 1.3 Informationsgehalt, Prüfbarkeit, Tiefe

Der *Informationsgehalt* (kurz *Gehalt*) einer Theorie T gibt den Reichtum an Information an, den T vermittelt. Gehalt kann syntaktisch definiert werden als Menge aller Sätze, die aus T ableitbar sind, semantisch als Klasse aller Fälle, in denen T nicht erfüllt ist, die also von T ausgeschlossen werden (Carnap, 1934, 1942). Popper (1966) definierte den *empirischen Gehalt* von T als Klasse der Basissätze (Beobachtungssätze), die von T ausgeschlossen werden; die mit T vereinbaren Basissätze bilden den *Spielraum* von T. Bei Vergleichen des Gehalts verschiedener Theorien führen die Definitionen in den wesentlichen Punkten zum selben Ergebnis.  $T_2$  hat einen höheren Gehalt als  $T_1$ , wenn  $T_1$  aus  $T_2$  logisch ableitbar ist,  $T_2$  aus  $T_1$  jedoch nicht. Eine Theorie, die eine andere zu erklären vermag, hat mehr Gehalt als diese. Eine Theorie hat also mehr Gehalt als die speziellen Hypothesen, die sie erklärt. Daraus wird deutlich, daß die Forderung nach Theorien, die möglichst viele Ergebnisse eines Fachgebietes zu systematisieren vermögen und in diesem Sinne eine hohe *Erklärungskraft* besitzen, als eine Forderung nach einem hohen Gehalt interpretiert werden kann.

In einer gewissen Hinsicht ist der Gehalt einer Theorie nicht genau bestimmt: Die interessanten Folgerungen einer Theorie ergeben sich oft nicht aus der Theorie allein, sondern erst aus der Theorie zusammen mit bestimmten Hilfsannahmen. Beispielsweise folgen die Aussagen über Risikowahl, Ausdauer und Anspruchsniveau, die in Kap. 7 zitiert wurden, nicht aus der Theorie der Leistungsmotivation allein. Dies spricht dafür, zur Definition des Gehalts einer Theorie die Hilfsannahmen mit hinzuzuziehen. Die zahlreichen Hilfsannahmen, die im Zusammenhang mit verschiedenen Anwendungen einer Theorie gemacht werden, sind jedoch manchmal schwer zu überblicken, und es werden ständig neue gebildet. Entsprechend ist auch der Gehalt unbestimmt bzw. in ständiger Veränderung.

Wenn man davon ausgeht, daß Theorien aus Gesetzhypothesen in Wenn-Dann-Form bestehen (vgl. Kap. 7), läßt sich folgendes feststellen: Der Gehalt einer Theorie steigt, wenn man ihr Gesetzhypothesen (konjunktiv) hinzufügt, und er nimmt ab, wenn man Gesetzhypothesen entfernt. Der Gehalt einer einzelnen Gesetzhypothese steigt mit der *Allgemeinheit* ihrer Wenn-Komponente und der *Präzision* (oder *Bestimmtheit*) ihrer Dann-Komponente. Mit der Allgemeinheit der Wenn-Komponente ist hier zum einen gemeint, wie umfassend die Klasse der Individuen und Situationen ist, auf die sich eine Aussage bezieht. Die Dann-Komponente ist manchmal abermals eine Aussage der Form „Wenn A, dann B“. In diesem Fall gilt, daß der Gehalt der gesamten Gesetzesaussage steigt, wenn A eine möglichst weit gefaßte Klasse von Bedingungen (eventuell Ursachen) beschreibt und B ein möglichst eng bestimm-

tes Ereignis ist. Die Wenn-Komponente „ $A_1$  oder  $A_2$ “ ist allgemeiner als  $A$ , allein; „ $A_1$  und  $A_2$ “ ist dagegen noch weniger allgemein als  $A_1$ . „ $B_1$  und  $B_2$ “ ist präziser als  $B_1$  allein; „ $B_1$  oder  $B_2$ “ ist weniger präzise als  $B_1$ .

Betrachten wir zwei Beispiele: In der Sozialpsychologie wurde im Zusammenhang mit dem Begriff der „Reaktanz“ folgende Hypothese formuliert und geprüft: Wenn ( $A_1$ ) man die Freiheit einer Person, zwischen verschiedenen Handlungsmöglichkeiten zu wählen, einschränkt, dann ( $B_1$ ) wird sie die eliminierten (oder bedrohten) Handlungsmöglichkeiten in ihrer Attraktivität aufwerten und versuchen, ( $B_2$ ) die ursprüngliche Freiheit wiederherzustellen (Brehm, 1966; Grabitz-Gniech & Grabitz, 1973). Zunächst sei darauf hingewiesen, daß der Gehalt dieser Hypothese geringer wäre, wenn nur  $B_1$  oder nur  $B_2$  behauptet würde; und er wäre noch geringer mit der Dann-Komponente „ $B_1$  oder  $B_2$ “. Tatsächlich scheinen die Reaktanz-Forscher anzunehmen, daß  $B_2$  nicht immer auftritt. Beachten wir jedoch vor allem die Wenn-Komponente. Aufgrund einiger Experimente fand man, daß  $A_1$  nicht immer ausreicht, um  $B_1$  bzw.  $B_2$  eintreten zu lassen (vgl. Irle, 1975). Zusätzlich muß gelten: ( $A_2$ ) Die Person  $q$ , die die Freiheitseinschränkung der Person  $p$  vornimmt, darf nicht zur Bezugsgruppe von  $p$  gehören; ( $A_3$ )  $q$  darf keine hohe Attraktivität für  $p$  besitzen. Die ursprüngliche Reaktanz-Hypothese mußte eingeschränkt werden, indem die neue Wenn-Komponente „ $A_1$  und  $A_2$  und  $A_3$ “ gebildet wurde. Dadurch wurde der Gehalt der Hypothese vermindert.

Ein zweites Beispiel: In einigen Untersuchungen zur Leistungsmotivation mit weiblichen Vpn konnten die von Atkinsons Theorie postulierten Zusammenhänge (vgl. Kap.7) nicht nachgewiesen werden. Es besteht die Möglichkeit, die Theorie einzuschränken und sie nur noch auf Männer zu beziehen. Dies würde ihren Gehalt einschränken. Man beachte hierzu, daß es (für die Theorie) völlig unproblematisch wäre, *wenn* Frauen lediglich extreme Werte im Leistungsmotiv hätten; dies müßte die Gültigkeit der Theorie nicht tangieren. Wenn jedoch etwa angenommen wird, daß Frauen eine „Furcht vor Erfolg“ haben, so wird damit die ursprüngliche Theorie zunächst nicht mehr auf Frauen bezogen, also in ihrem Gehalt vermindert.

Damit sind gleichzeitig einige methodologische Regeln angegeben worden, nach denen man verfahren kann, um den Gehalt von  $T$  zu erhöhen: den Anwendungsbereich von  $T$  erweitern,  $T$  präziser machen, zu  $T$  neue Aussagen hinzufügen; besonders die dritte Strategie darf allerdings nicht nach Belieben gebraucht werden, da sie dem Ziel einer möglichst einfachen Theorie mit wenigen Grundannahmen zuwiderlaufen würde.

Es gibt kein numerisches Maß für den Gehalt, das es erlauben würde, Theorien über verschiedene Gegenstandsbereiche zu vergleichen, etwa Atkinsons Theorie mit derjenigen Festingers. Voraussetzung für einen Vergleich sind logische Beziehungen. Am eindeutigsten ist die Situation dann, wenn  $T_1$  aus  $T_2$  ableit-

bar ist. Dies ist z.B. dann der Fall, wenn  $T_2$  durch eine einfache Veränderung von  $T_1$  entstanden ist, wie in den gegebenen Beispielen, oder wenn  $T_2$  eine umfassendere Theorie ist, die  $T_1$  zu erklären vermag. Kruglanski (1980) erklärt verschiedene spezielle Gesetzmäßigkeiten der Attribution durch eine allgemeine Theorie der Informationsverarbeitung; seine Theorie ist daher gehaltvoller als jede dieser Gesetzhypothesen. Bei Erklärungen wird allerdings die erklärte Theorie oder Hypothese  $T_1$  oft nicht wirklich abgeleitet, sondern zugleich *korrigiert*, d.h. es wird etwas abgeleitet, das  $T_1$  annäherungsweise entspricht: Aus  $T_2$  folgt, daß  $T_1$  einen Grenzfall von  $T_2$  darstellt und unter bestimmten Bedingungen zutrifft (s. auch Kap.7, 4.1). Malewski (1967), der mit methodologischen Argumenten für eine Integration einzelner sozialwissenschaftlicher Befunde durch umfassende Theorien eingetreten ist, hat selbst solche Integrationen vorgenommen, indem er mit seiner Version der Verhaltenstheorie verschiedene spezielle Gesetzmäßigkeiten aus der Sozialpsychologie und Soziologie erklärte. Dabei werden diese Gesetzmäßigkeiten zugleich korrigiert, indem gezeigt wird, daß sie nur unter solchen sozialen Randbedingungen zutreffen, unter denen bestimmte Zustände als Belohnungen bzw. Bestrafungen wirken. Schmid (1993) zeigt wiederum, daß die Theorie des subjektiv erwarteten Nutzens die Verhaltenstheorie zu erklären und zu korrigieren vermag.

Oft ist es nicht möglich, zwischen konkurrierenden Theorien eine solche Korrespondenz aufzuzeigen, da die Theorien ontologisch verschieden sind: Obwohl sie sich auf dieselben Phänomene beziehen, etwa Gedächtnisphänomene, handelt  $T_1$  z.B. von mehreren Gedächtnisspeichern und der entsprechenden Informationsübertragung,  $T_2$  aber von verschiedenen Ebenen oder Formaten der Informationsverarbeitung. In diesem Fall ist ein Vergleich dann möglich, wenn  $T_2$  alle empirischen Tatsachen erklärt, die  $T_1$  erklären kann, und darüber hinaus eine Tatsache erklärt, die  $T_1$  nicht erklären kann;  $T_2$  hat gegenüber  $T_1$  einen *Gehaltsüberschuß*. Popper (1972) formulierte diesen Gedanken auch folgendenmaßen:  $T_2$  beantwortet jede empirische Frage mindestens so genau wie  $T_1$ , aber nicht umgekehrt. - Sollte allerdings auch  $T_1$  etwas erklären können, was  $T_2$  nicht erklärt, so sind die Theorien *inkommensurabel*.

Diese Idee des Gehaltsvergleichs ist vielfach kritisiert worden (Laudan, 1977; Glymour, 1980). Es ist natürlich nicht immer eindeutig auszumachen, was die relevanten, „zu erklärenden Tatsachen“ oder „zu beantwortenden Fragen“ sind, die einem Vergleich zugrunde liegen sollen. Wenn  $T_1$  und  $T_2$  eine verschiedene Ontologie haben, was bei konkurrierenden Theorien meist der Fall ist, so ist es immer möglich, aus  $T_1$  eine Aussage  $S$  abzuleiten, die aus  $T_2$  nicht folgt, z.B. über Gedächtnisspeicher, die in  $T_2$  nicht mehr vorkommen. Man könnte daraufhin einwenden, daß  $S$  keine empirische Tatsache beschreibe und deshalb für den Erfolg von  $T_1$  nicht maßgeblich sein könne. Es wäre also zu entscheiden, ob  $S$  als empirisches Faktum anerkannt wird oder nicht. Kuhn

(1962) und Feyerabend (1976) vertreten die Auffassung, daß sich mit der Änderung umfassender Theorien (Paradigmen, Weltbilder) auch die Überzeugungen darüber ändern, was die empirischen Tatsachen sind. Kuhn behauptet beispielsweise, daß die Vertreter der Phlogiston-Theorie regelrecht Phlogiston „sahen“. Andersson (1988) zeigt allerdings, daß die diesbezüglichen Beispiele von Kuhn und Feyerabend durchweg fragwürdig sind und daß die Befürworter verschiedener „Weltbild-Theorien“ sich sehr wohl darüber einig waren, was die empirischen Gegebenheiten (z.B. Resultate chemischer Versuche) waren, zu deren Erklärung sie verschiedene theoretische Konstruktionen (wie Phlogiston) bevorzugten.

Trotz der Schwierigkeiten, mit denen ein Gehaltsvergleich behaftet ist, erscheint es immerhin möglich, eine heuristische Empfehlung zu geben (die die „Unschärfen“ der Analyse zugesteht):  $T_2$  stellt (unter ansonsten gleichen Bedingungen) einen *Fortschritt* gegenüber  $T_1$  dar, wenn  $T_2$  einen *Gehaltsüberschuß* hat. Ein Gehaltsüberschuß in diesem Sinne ist etwa dasselbe, was in den empirischen Wissenschaften als größere *Erklärungskraft* bezeichnet und geschätzt wird.

Soll in der psychologischen Theorienbildung ein möglichst hoher Gehalt angestrebt werden? Im Prinzip ist eine Theorie um so besser, je höher ihr Informationsgehalt ist. Jedoch führt der Versuch, eine Theorie zu entwickeln, die „alles erklären“ soll, mit größter Wahrscheinlichkeit zu einem Produkt, das sich als empirisch völlig inadäquat erweist. Auch kann es sein, daß das Streben nach übergroßer Allgemeinheit es notwendig macht, die Gesetzhypothesen so unpräzise zu belassen, daß tatsächlich gar kein hoher Gehalt erzielt wird. Die Empfehlung muß also lauten, in einer dem Wissenstand angemessenen Weise zu versuchen, einen Gehaltsüberschuß, d.h. eine größere Erklärungskraft, anzustreben. Was in dieser Hinsicht angemessen ist, hängt vom jeweiligen Gegenstandsbereich, dem Wissensstand und nicht zuletzt der Kompetenz des Forschers ab. Der von wissenschaftstheoretischer Seite gelegentlich erhobene Vorwurf, die Sozialwissenschaften seien dem Empirismus verhaftet, weil sie keine echten, „großen“ Theorien konstruieren würden, entspringt größtenteils einer Unkenntnis des Gegenstandes und der methodischen Probleme.

In der „Logik der Forschung“ hat Popper die methodologische Forderung erhoben, Theorien nicht dadurch vor einer Falsifikation zu bewahren, daß man ihren Gehalt vermindert. Sicherlich ist eine fortwährende Einschränkung einer Theorie nicht erwünscht und wird auch in keiner empirischen Wissenschaft angestrebt. Allerdings wäre es auch nicht sinnvoll, eine strikte Befolgung dieser Regel zu verlangen. Man muß immer im Auge behalten, daß nicht alle Bewertungskriterien optimiert werden können. Eine etwas eingeschränkte, jedoch empirisch adäquate Theorie ist letztlich besser als eine falsche. So war



z. B. Berkowitz' (1962) Einschränkung der Frustrations-Aggressions-Hypothese durchaus überzeugend, obwohl sie den Gehalt der ursprünglichen, aber unhaltbaren Hypothesen verringerte. Auch in dieser Hinsicht sollte also dafür plädiert werden, die Bewertungskriterien flexibel zu handhaben.

Popper (1966) setzte den empirischen Gehalt einer Theorie mit ihrem Prüfbarkeitsgrad und diesen wiederum mit dem Falsifizierbarkeitsgrad gleich. Der empirische Gehalt wird dadurch gleichzeitig zu einem Abgrenzungskriterium, das die empirische Wissenschaft von den Formalwissenschaften einerseits und der Metaphysik andererseits trennt: Eine empirisch-wissenschaftliche Theorie hat im Unterschied zu einer formalwissenschaftlichen und metaphysischen einen empirischen Gehalt, sie kann an der Erfahrung scheitern. Die Metaphysik sollte dadurch nicht, wie durch die logischen Positivisten, als sinnlos abgewertet, jedoch von der empirischen Wissenschaft abgegrenzt werden. Dabei wurde die mögliche heuristische Fruchtbarkeit metaphysischer Auffassungen für die Wissenschaft nicht gelehnt. Die Betonung der Abgrenzung war allerdings doch durch den Geist des Positivismus beeinflusst. Die Bedeutung einer solchen Abgrenzung ist inzwischen vielfach bezweifelt worden (vgl. Bartley, 1968). Später hat Popper hervorgehoben, daß metaphysische Auffassungen, wenn auch nicht empirisch falsifiziert, so doch rational diskutiert werden können, z.B. durch logische Analyse und durch Konfrontation mit bewährten wissenschaftlichen Theorien, die ihnen widersprechen können. Bunge (1967) vertritt die Meinung, daß jede Wissenschaft auf metaphysischen Voraussetzungen beruht, zu denen etwa der Realismus oder die Annahme von Naturgesetzen gehören. Und selbst der konsequenteste Empirismus akzeptiert immerhin die Voraussetzung, daß Erfahrung für die Erkenntnis grundlegend ist; auch dies ist eine metaphysische Hypothese, die weder logisch begründet noch im selben Sinne empirisch geprüft werden kann wie eine empirische Hypothese.

Agassi (1975) hat sogar zu zeigen versucht, daß die Wissenschaft im wesentlichen immer versucht hat, die von der Metaphysik vorgegebenen Probleme zu lösen: Auch wenn vielleicht der einzelne Wissenschaftler sich ganz der Lösung eines fachspezifischen Detailproblems widmet, hängen die größeren Fragen, in deren Rahmen er arbeitet, doch mit Weltbild-Problemen zusammen, etwa mit der Struktur des Universums oder der Natur der Materie. In der Psychologie könnte eine solche metaphysische Frage unter anderem die nach der „Natur des Geistes“ oder auch der Intelligenz sein: Können diejenigen menschlichen Fähigkeiten, die einmal dazu Anlaß dazu gaben, den menschlichen Geist zu postulieren und ihn dem Bereich der Natur bzw. der Naturgesetze gegenüberzustellen, eventuell naturgesetzlich (nomologisch, manche sagen „mechanistisch“) erklärt werden?

Es empfiehlt sich, verschiedene Arten von Konzeptionen zu unterscheiden, die alle nicht im engeren Sinne empirisch-wissenschaftlich sind. *Erstens* gibt es

die schon angesprochenen erkenntnistheoretischen und methodologischen Voraussetzungen wissenschaftlicher Tätigkeit. Die *zweite* Kategorie bilden Konzeptionen, die ebenso wie wissenschaftliche Theorien mit dem Ziel der Erklärung der Wirklichkeit entworfen werden, die jedoch aus technischen Gründen kaum oder noch gar nicht empirisch prüfbar sind. Hierzu gehörte einmal die Theorie der Atome von Demokrit und Leukipp, die später zu einer physikalischen Theorie wurde. Zwischen solchen Konzeptionen und wissenschaftlichen Theorien besteht keine sehr scharfe Abgrenzung. Gute wissenschaftliche Theorien haben allein, d.h. ohne Hilfshypothesen, in einem gewissen Sinne auch keinen empirischen Gehalt, da sie nur zusammen mit letzteren Folgerungen über (relativ gut) beobachtbare Sachverhalte zulassen. Daher sind neue wissenschaftliche Entwürfe manchmal schwer prüfbar, da ihre Prüfung neue Hilfshypothesen, Methoden und Instrumente erfordert. Viele Eigenschaften von psychologischen Theorien stehen einer möglichst eindeutigen Prüfbarkeit entgegen: die Verwendung theoretischer Begriffe, die Unvollständigkeit, die statistische (statt deterministische) Form. Unmittelbare und sofort erkennbare empirische Prüfbarkeit ist nicht unbedingt ein guter Indikator für die Fruchtbarkeit eines theoretischen Ansatzes und auch nicht für den Informationsgehalt, den die betreffende Theorie nach einiger Zeit der Entwicklung haben wird. Daraus ergibt sich, daß es weder eindeutig möglich noch notwendig ist, Wissenschaft von metaphysischen Konzeptionen dieser zweiten Art abzugrenzen. Wichtiger ist es vielmehr, die methodologische Empfehlung zu akzeptieren, Theorien so zu entwickeln und mit Operationalisierungen auszustatten, daß sie möglichst gut prüfbar werden.

Eine *dritte* Art der Metaphysik ist den Zielen und Methoden der Wissenschaft jedoch entgegengesetzt. Gemeint sind alle Konzeptionen, die so angelegt oder mit solchen Verteidigungsstrategien (*Immunsierungsstrategien*) verbunden sind, daß eine rationale Kritik sie nicht zu treffen vermag. In erster Linie ist hier an Glaubenssysteme, unter anderem religiöser Art, zu denken, deren Anhänger im allgemeinen eine Prüfung nach logischen oder empirischen Gesichtspunkten weder für möglich noch für erstrebenswert halten. Im übrigen können aber die verschiedensten Denkansätze mit Immunsierungsstrategien einhergehen, im einfachsten Fall mit dem Argument, die behaupteten Phänomene würden sich entziehen, wenn man sie systematisch zu erforschen versuche. (Goethe äußerte zu Newtons Farbexperimenten, die seiner eigenen Farbenlehre widersprachen: „Die Natur verstummt auf der Folter.“) - Entsprechend immunisierte Konzeptionen sind von empirisch-wissenschaftlichen nicht unbedingt durch ihren Inhalt oder ihre logische Form, jedoch durch die *Methode des Umgangs* mit den jeweiligen Inhalten abgegrenzt.

Als letzte Eigenschaft ist in diesem Zusammenhang noch die *Tiefe* einer Theorie zu nennen (vgl. Bunge, 1967; Popper, 1972; vgl. auch Kap.5, Abschnitt 4.1.2 dieses Bandes). Mit dem Begriff der Tiefe soll ausgedrückt werden, daß

eine Theorie dazu in der Lage ist, einen Zusammenhang zwischen Variablen zu erklären, indem sie die *Mechanismen* oder *Prozesse* aufzeigt, die diesem Zusammenhang **zugrunde** liegen. Es sei beispielsweise G ein Gesetz, das einen Zusammenhang zwischen einer situativen Bedingung und einem Verhalten beschreibt. T sei eine Theorie, die etwas über motivationale oder kognitive Prozesse aussagt und G zu erklären vermag. Und T' sei schließlich eine Theorie, die T erklärt, indem die physiologischen Strukturen und Prozesse spezifiziert werden, die den von T postulierten psychischen Vorgängen zugrunde liegen. In diesem Fall hat T eine größere Tiefe als G und T' wiederum eine größere Tiefe als T. Ein besonderes Anzeichen größerer Tiefe ist nach Poppers Auffassung eine Erklärung, bei der die ältere Theorie durch die neue korrigiert wird. - Die Eigenschaft einer Theorie, um die es hier geht, steht in Beziehung zu der in der Psychologie häufig gebrauchten Metapher der *Black Box*, mit der ein Organismus verglichen werden kann. Das Ziel, vertiefte Erklärungen zu erreichen, bedeutet, das Innere der Black Box ergründen zu wollen, über deren beobachtbares Verhalten schon einiges bekannt ist. Eine erklärende Theorie postuliert etwa Gedächtnisspeicher und damit verbundene Vorgänge der Informationsverarbeitung. Auch ein Gedächtnisspeicher kann aber wieder als Black Box angesehen werden, und man kann im Rahmen einer noch tiefergehenden Analyse z.B. nach den physiologischen Vorgängen suchen, die dem zugrunde liegen, was zuvor als „Speicherung“ oder „Übertragung“ von Information beschrieben wurde. Selbstverständlich gibt es auch die Möglichkeit, die vertiefte Erklärung wiederum auf psychologischer Ebene zu suchen.

Eine tiefere Theorie hat einen höheren Informationsgehalt bzw. höhere Erklärungskraft. Die Umkehrung gilt aber nicht. Man kann den Gehalt vergrößern, indem man die Allgemeinheit oder Präzision seiner Aussage erhöht, auch wenn dabei die Analyseebene dieselbe bleibt. Größere Tiefe verlangt erstens, daß T durch T' erklärt wird, und zweitens, daß T' die zugrunde liegenden Mechanismen nennt. Der zweite Punkt ist im ersten nicht schon enthalten, denn nach dem allgemeinen Erklärungsbegriff ist eine Erklärung auch durch bloße *Subsumption* unter ein (gehaltvolleres) Gesetz möglich, ohne daß sich die erklärenden Aussagen auf eine tiefere Analyseebene beziehen müssen. Manche Wissenschaftler verwenden einen Erklärungsbegriff, der den zweiten Punkt einschließt: Im Unterschied zur bloßen *Beschreibung* eines Zusammenhanges wird manchmal erst dann von Erklärung gesprochen, wenn auf einer vertieften Analyseebene aufgeklärt werden kann, welche Strukturen und Prozesse den beschriebenen Zusammenhang erzeugen bzw. aufrechterhalten.

Tiefe ist ein Teil der Zielsetzung der nomologischen Wissenschaften. Sie präsentiert einen zentralen Aspekt des Erkenntnisinteresses der Grundlagenforschung: zu ergründen, was sich hinter den beobachtbaren Phänomenen verbirgt und sie hervorbringt. Auch hier gilt, was oben in bezug auf den Informationsgehalt gesagt wurde: Versuche der vertieften Erklärung müssen an das

gegebene Wissen anknüpfen. Wenn man sich zu frühzeitig den elementarsten Prozessen zuwendet (in der Psychologie etwa: Nervenzellen, Aktionspotentiale), kann es sein, daß es nicht gelingt, eine Theorie auszuarbeiten, die zu erklären vermag, was zur Erklärung ansteht (z. B. Entscheiden, Problemlösen). Freilich können auch metaphysische Spekulationen nützlich sein, insbesondere wenn sie von der Zielvorstellung geleitet sind, sie zu empirischen Theorien zu entwickeln. Doch sollten (z. B. physiologisch-psychologische) *Erklärbarkeitsbehauptungen* nicht mit tatsächlichen Erklärungen verwechselt werden.

## 1.4 Einfachheit

Das Ziel der Theorienbildung enthält die Idee der Einfachheit: Es sollen möglichst viele Befunde durch *möglichst wenige, einfache Annahmen* erklärt werden. Bei gleicher Erklärungskraft ist die einfachere Theorie vorzuziehen. Das vielzitierte *Sparsamkeitsprinzip* (Occams razor) zielt in dieselbe Richtung: Nur so viele theoretischen Entitäten und Annahmen sollen eingeführt werden, wie zur Erklärung der Phänomene notwendig sind (vgl. zum Aspekt der Einfachheit auch Kap. 1 dieses Bandes, Abschnitte 2.2.2 und 3.1). - Obwohl diese Forderung in den Wissenschaften einen hohen Grad an Akzeptanz genießt, ist sie außerordentlich schwer präzisierbar. Zunächst handelt es sich nicht um eine eindimensionale Größe, vielmehr müssen wenigstens 3 Aspekte unterschieden werden, die unabhängig voneinander variieren können (vgl. Kutschera, 1972). Gefordert werden: möglichst wenige Grundannahmen einer Theorie, möglichst wenige theoretische Begriffe, möglichst einfache Zusammenhangsannahmen. Alle drei Forderungen sind nicht in völlig eindeutige Kriterien umsetzbar. Aussagen wie Begriffe können zusammengesetzt sein. So kann man z.B. die beiden logisch unabhängigen Aussagen „Frustration führt zu Aggression“ und „Aggression setzt immer Frustration voraus“ zusammenziehen zu „Aggression genau dann wenn Frustration“. Die dritte Aussage sagt dasselbe wie die beiden ersten zusammen; hat man es nun bei der dritten Aussage mit einer oder mit zwei Aussagen zu tun? - In bezug auf den dritten Punkt gibt es verschiedene Vorschläge, unter welchen Bedingungen eine funktionale Beziehung als einfacher gelten soll, etwa: wenn die entsprechende Kurve im Durchschnitt weniger gekrümmt ist; wenn die Kurve weniger von einer Geraden abweicht; wenn der Polynomgrad niedriger ist. Popper (1966) hat zur Einfachheitsbeurteilung vorgeschlagen, davon auszugehen, wie viele elementare Basissätze (Beobachtungssätze) notwendig sind, um eine theoretische Aussage widerlegen zu können. So braucht man z.B. weniger Beobachtungen, um eine Aussage zu widerlegen, die als funktionale Beziehung eine Gerade postuliert, als für eine Aussage, die eine Parabel annimmt. Zur Widerlegung der Aussage „Wenn  $A_1$  und  $A_2$ , dann B“ braucht man mindestens drei elementare Beobachtungssätze, die jeweils aussagen, ob  $A_1$ ,  $A_2$  und B der Fall

ist oder nicht. Zur Widerlegung von „Wenn A, dann B“ genügen dagegen zwei Beobachtungen. Folglich ist die zweite Aussage die einfachere. Popper kommt zu dem Ergebnis, daß Einfachheit auf den Grad der Prüfbarkeit (Falsifizierbarkeit) zurückführbar ist, Dies gilt allerdings nicht allgemein, sondern nur für die logische Struktur einzelner theoretischer Aussagen. In bezug auf eine ganze Theorie läßt sich der Gehalt und damit die Prüfbarkeit nämlich erhöhen, indem man neue Aussagen (konjunktiv) hinzufügt, während der Einfachheitsaspekt „Zahl der Axiome“ dann sinkt.

In der Psychologie ist Einfachheit oft als Grundlage der Bewertung herangezogen worden (vgl. Battig, 1962). Am bekanntesten ist wohl die Regel C. Lloyd Morgans, zur Erklärung eines Verhaltens nicht auf höhere psychische Funktionen Bezug zu nehmen, wenn auch eine Erklärung durch elementarere Funktionen möglich ist; dies ist eine spezielle Anwendung der Forderung nach Einfachheit, sofern man davon ausgeht, daß man die Annahme der betreffenden elementaren Funktionen für die weniger hochentwickelten Lebewesen bereits akzeptiert hat und sich die Annahme der „höheren“ Funktionen eventuell ersparen kann. - In der Forschung zur Theorie Hulls wurde mehrfach der Versuch gemacht, ohne Einbuße der Erklärungskraft auf das Triebkonstrukt D zu verzichten, also mit weniger theoretischen Begriffen auszukommen. Miller interpretierte einen Trieb als speziellen Stimulus und versuchte, die allgemeine Aktivierungsfunktion von D als Stimulusgeneralisierung bezüglich dieses Stimulus zu erklären. (Die Begriffe „Stimulus“ und „Generalisierung“ kommen in der Theorie bereits vor.) - Birch, Burnstein & Clark postulierten die antizipatorische Zielreaktion als den Wirkmechanismus von D, wodurch D als eigenständiges Konstrukt überflüssig wird (vgl. Atkinson, 1964, Kap.7). - Das folgende Beispiel bezieht sich auf die Notwendigkeit bzw. Überflüssigkeit theoretischer Annahmen: Nisbett und Ross (1980, S.12) sehen es als Verstoß gegen das Prinzip der Einfachheit an, einen direkten Einfluß psychodynamischer Faktoren auf die Kognition (z.B. Wunschdenken) zu postulieren, da nachweislich ohnehin so viele rein kognitiv bedingte Irrtümer begangen werden. Als Erklärungsgrundlage genüge die Annahme, daß Personen in bestimmten Situationen aus Irrtum zu inadäquaten Strategien greifen. Bestritten wird weder das Vorhandensein motivationaler Faktoren noch deren Einfluß auf das Verhalten, sondern nur ein direkter Einfluß auf die Kognition, eine nach Meinung der Autoren überflüssige Annahme. - Auf ein weiteres Beispiel sei lediglich hingewiesen: Anderson (1976) gibt als Grund für die Entwicklung des ACT-Modells unter anderem an, daß es einfacher sei als das frühere HAM-Modell. - Das Einfachheitsprinzip ist auch der Hauptgrund dafür, daß in der heutigen Psychologie fast durchweg die Annahme abgelehnt wird, es gäbe eine Psyche im Sinne einer vom physischen Organismus unabhängig existierenden Substanz (Substanzdualismus). Um den beobachtbaren Tatsachen Rechnung zu tragen, ist die

Annahme ausreichend, daß das Zentralnervensystem der substanzielle Träger psychischer Zustände und Ereignisse ist; der Substanzdualismus behauptet also eine Klasse von (psychischen) Entitäten, die zu Erklärungszwecken (und übrigens auch zur Beschreibung der Bewußtseinszustände) nicht benötigt werden.

In der neueren Wissenschaftstheorie ist Einfachheit als Kriterium stark in Zweifel gezogen worden. Bunge (1963) spricht vom „Mythos der Einfachheit“ und argumentiert anhand der Wissenschaftsgeschichte, daß dieses Kriterium eine sehr geringe Rolle gespielt habe. Diese Kritik zeigt allerdings nur, daß Einfachheit nicht das einzige Kriterium ist und daß im Zweifelsfalle der Erklärungskraft und empirischen Adäquatheit der Vorzug gegeben wird. Man darf dennoch nicht übersehen, daß immer Erklärungskraft und empirische Adäquatheit bei gleichzeitiger Beschränkung hinsichtlich der Komplexität der theoretischen Konstruktionen angestrebt wird. Anders läßt sich die allgemein übliche Art der Theorienkonstruktion nicht deuten. Erklärungsleistung allein ließe sich nämlich durch Vermehrung der Grundannahmen und durch komplexe Annahmen beliebig steigern. Einfachheit ist kein exaktes Kriterium, und sie ist sicherlich auch nicht das gewichtigste Kriterium zur Beurteilung von Theorien. Es handelt sich eher um ein grobes heuristisches Prinzip. Als solches wird Einfachheit aber nachweislich anerkannt, sei es explizit oder implizit.

Wie läßt sich begründen, daß man sich um einfache theoretische Konstruktionen bemüht? Berühmte Wissenschaftler, darunter Kepler, Newton und Einstein, waren der Überzeugung, daß die Natur einfach sei. Da wir aber schwerlich a priori beurteilen können, ob dieser Glaube zutrifft, kann Einfachheit als Kriterium der Forschung auf diese Weise kaum begründet werden. Überzeugender ist ein methodologisches Argument auf der Grundlage der oben dargelegten Beziehung zwischen Einfachheit und *Prüfbarkeit*: Aussagen, die einfachere Beziehungen zwischen Variablen annehmen, sind mit weniger Aufwand prüfbar; also ist es (unter sonst gleichen Bedingungen) vernünftig, zunächst solche einfacheren Aussagen zu erproben. - Auf diese Weise läßt sich zumindest ein bestimmter Aspekt der Einfachheit begründen. Zusätzlich dürften dem Einfachheitsideal auch pragmatische Ziele zugrunde liegen: Unter sonst gleichen Bedingungen versuchen wir zuerst diejenigen theoretischen Lösungen, von denen wir es vorziehen würden, daß sie sich als erfolgreich erweisen, nämlich solche, die man (bei der begrenzten Kapazität des menschlichen Arbeitsgedächtnisses) ohne allzu großen Aufwand erlernen und anwenden kann. (Wer würde es nicht vorziehen, daß Fragen des physikalischen Weltbildes noch mit Hilfe der euklidischen Geometrie adäquat behandelt werden könnten?) Sollte es sich doch als notwendig erweisen, muß man mit komplizierteren Konstruktionen Vorlieb nehmen.

## 2. Empirische Bewertungskriterien

Während die Beurteilung einer Theorie hinsichtlich der bisher angesprochenen Ziele unabhängig von der Erfahrung erfolgen, also noch vor der Planung empirischer Untersuchungen vorgenommen werden kann, sind *Wahrheit* und *empirische Adäquatheit* mit der empirischen Bewertung verknüpft. Zwar erlaubt auch schon die Prüfung der logischen Konsistenz manchmal das Urteil, daß eine Theorie nicht wahr sein kann, doch muß jederzeit damit gerechnet werden, daß eine konsistente Theorie die Wirklichkeit unzutreffend darstellt. Wenn in den empirischen Wissenschaften davon gesprochen wird, daß Theorien anhand der Erfahrung geprüft und gegebenenfalls *bestätigt* oder *entkräftet* werden, so wird offenbar vorausgesetzt, daß die Ergebnisse empirischer Untersuchungen dazu geeignet sind, Theorien im Hinblick auf ihre Wahrheit und Falschheit zu bewerten. Allen Erfahrungswissenschaften liegt die erkenntnistheoretische Annahme zugrunde, daß unsere Beobachtungen trotz ihrer grundsätzlichen Fehlbarkeit im großen und ganzen einen größeren Anspruch auf Glaubwürdigkeit erheben können als unsere noch nicht geprüften theoretischen Spekulationen (Musgrave, 1991). Nach welchen *Kriterien* können nun Theorien anhand der Beobachtung bewertet werden? Vermag ein empirischer Befund die Wahrheit einer Theorie nachzuweisen (*Verifikation*) oder sie wenigstens zu bestätigen? Kann ein Befund die Falschheit einer Theorie nachweisen (*Falsifikation*) oder sie wenigstens entkräften? Wie können empirische Ergebnisse dazu verwendet werden, Theorien im Hinblick auf Wahrheit zu verbessern, und welche empirischen Informationen sind hierfür besonders relevant?

Eine Behandlung solcher Fragen muß an die wissenschaftstheoretische Debatte über Fragen der *Induktion* (s. auch Kap. 10), der *Falsifikation* und des *Erkenntnisfortschritts* anknüpfen. In der Geschichte der Erkenntnis- und Wissenschaftstheorie stand lange Zeit das Problem der Verifikation von Theorien im Zentrum der Aufmerksamkeit, wobei vor allem die Induktion als adäquate Methode der Verifikation angesehen wurde. Heute gibt es eine weitgehende Übereinstimmung darüber, daß Theorien durch Induktion nicht verifiziert werden können. Wenn von endlich vielen Beobachtungssätzen induktiv auf eine Theorie geschlossen wird, so ist es immer möglich, daß die Beobachtungssätze wahr sind und die Theorie dennoch falsch ist. Versuche, für Theorien eine *induktive Wahrscheinlichkeit* als *Bestätigungsgrad* zu definieren, haben sich ebenfalls nicht als erfolgreich erwiesen. Alle hierfür entwickelten formalen Kalküle führen zum Teil zu unakzeptablen Bewertungen und sind nur unter hochgradig idealisierten Bedingungen anwendbar, die mit den existierenden Wissenschaften wenig gemeinsam haben. Die Kritik am Induktivismus, vor allem durch Hume, hat viele Wissenschaftler an der Rationalität der Erkenntnis zweifeln lassen und zu verschiedenen Formen des Skeptizismus und

Relativismus geführt. Der von Karl Popper begründete *kritische Rationalismus* versucht, die Möglichkeit *rationaler Erkenntnisgewinnung* bei gleichzeitiger *prinzipieller Ungewißheit (Fallibilismus)* aufzuzeigen: Nachdem eine Theorie mit möglichst hohem empirischem Gehalt entworfen wurde, sollte mit allen Mitteln versucht werden, sie zu falsifizieren. Hält eine Theorie einem derartigen *strengen* Prüfversuch stand, so gilt sie als vorläufig *bewährt*, andernfalls als *falsifiziert*. Falsifizierende Befunde weisen den Weg zu neuen Theorien, die mit den Tatsachen besser übereinstimmen. Unter den nicht falsifizierten Theorien ist jeweils die am strengsten geprüfte und daher am meisten bewährte vorzuziehen. - Die wissenschaftstheoretische Diskussion in den letzten dreißig Jahren bestand zu einem wesentlichen Teil in der Auseinandersetzung mit diesem Falsifikationismus. Die Positionen etwa von Kuhn, Feyerabend und Lakatos sind entscheidend durch eine bestimmte Auffassung zu den Gedanken der *kritischen Prüfung* und der Rationalität bestimmt, und dasselbe gilt für die modernen Versionen des *Instrumentalismus* sowie für den *Non-Statement-View (Strukturalismus)*. Heute liegt auch der Falsifikationismus selbst in vielen Varianten vor. Es ist hier nicht möglich, alle diese Positionen einzeln zu referieren. Die Diskussion muß sich auf die eingangs erwähnten zentralen Fragen zur Bewertung von Theorien konzentrieren. Diese sind in etwa folgender Form kontrovers diskutiert worden: Sind Theorien falsifizierbar, sind sie überhaupt prüfbar? Wenn ja, in welchem Sinne? Besteht empirische Wissenschaft (adäquaterweise) aus kritischen Prüfversuchen oder eher aus einer Suche nach Bestätigungen? Können empirische Resultate eine Theorie bestätigen? Ist Bestätigung ein Kriterium für Wahrheit oder ein Anhaltspunkt für einen Fortschritt in Richtung auf Wahrheit? - Diese Fragen sollen nun mit Bezug auf psychologische Problemstellungen behandelt werden. Dabei können wir vorerst offen lassen, ob Wahrheit der gesamten Theorie oder nur Wahrheit ihrer beobachtbaren Konsequenzen (empirische Adäquatheit) gemeint ist.

## 2.1 Die Möglichkeit empirischer Prüfung und Kritik

Das Ergebnis einer empirischen Untersuchung kann mit den theoretischen Erwartungen *übereinstimmen* oder von ihnen *abweichen*. Ist es sinnvoll, einen abweichenden Befund als Grund für eine Falsifikation der Theorie zu betrachten? Popper hat darauf hingewiesen, daß ein Beobachtungsbefund eine allgemeine Aussage zwar nicht verifizieren, jedoch falsifizieren kann. Wenn die Aussage „Hier befindet sich ein weißer Rabe“ wahr ist, so muß die allgemeine Aussage „Alle Raben sind schwarz“ falsch sein. Eine Schwierigkeit entsteht allerdings dadurch, daß empirisch-wissenschaftliche Beobachtungen im allgemeinen mehr Probleme aufwerfen als die Beobachtung der Farbe eines Raben. Dies wird deutlich, wenn man singuläre Sätze betrachtet, die notwen-



dig wären, um ohne *Zusatzannahmen* eine (nicht-triviale) wissenschaftliche Theorie zu falsifizieren. Wenn wir das Beispiel der Leistungsmotivation aus Kap. 7 zugrunde legen, müßte ein singulärer falsifizierender Satz  $E_t$  die theoretischen Begriffe der Theorie enthalten und z.B. lauten: „p hat ein starkes Leistungsmotiv, ein schwach ausgeprägtes Mißerfolgsmotiv, hatte aber in der Situation  $s$  die *Tendenz*, eine als leicht wahrgenommene Aufgabe zu wählen“. Eine zuverlässige Entscheidung über das Zutreffen von  $E_t$  wäre äußerst schwierig, so daß man als psychologische Beobachtungsaussagen eher Befunde über besser beobachtbare Eigenschaften wählt, wie etwa  $E_i$ : „p hat im verwendeten Test bei 15 Fragen die Antwort ja angekreuzt“. Das Problem der Falsifikation verschiebt sich dadurch allerdings nur: Aus dem Problem der schwierigen Beurteilung von  $E_t$  wird das Problem der Wahrheit der Hilfsannahmen, die man voraussetzen muß, damit zwischen  $E_o$  und der Theorie eine logische Beziehung entsteht. Dies wurde bereits von Duhem (1908) klar aufgezeigt: Beobachtungsaussagen widersprechen (im Falle interessanter, nicht-trivialer Theorien) nicht einer isolierten theoretischen Aussage, sondern nur einer Menge von theoretischen Aussagen und Hilfsannahmen. Zur Verdeutlichung der Vielfalt solcher Hilfsannahmen betrachten wir die einzelnen Schritte bei der Prüfung einer psychologischen Theorie (vgl. Kap. 14).

a) Meist kann eine Grundannahme einer Theorie (z.B.  $T_e = M_e \times W_e \times A_e$ ) nicht isoliert geprüft werden; vielmehr wird aus mehreren Grundannahmen und einigen Hilfsannahmen (vgl. die „Zusatzannahmen“ in Kap.7) eine bestimmte zu prüfende Hypothese  $H$  abgeleitet (z.B. „Erfolgsmotivierte wählen bevorzugt mittelschwere Aufgaben“). Zur Ableitung von  $H$  muß auch eine *Ceteris-paribus*-Annahme akzeptiert werden (vgl. zur Problematik der *Ceteris-paribus*-Annahmen auch Kap. 1, Abschnitt 3.2.1 sowie Kap. 15): Die Theorie ist *unvollständig* in bezug auf bestimmte Variablen der zu prüfenden Hypothese, so daß angenommen und (durch Wahl eines geeigneten Untersuchungsplans) begründet werden muß, daß in der Untersuchungssituation störende Faktoren (andere Motive, nicht-psychologische Störfaktoren) keine beträchtliche Rolle spielen (vgl. Kap. 7). Auch viele „selbstverständliche“ Hilfsannahmen werden vorausgesetzt, etwa, daß die  $V_{pn}$  in „normalem“ Zustand sind, daß sie die Instruktion verstanden haben, zur Mitarbeit bereit sind usw.

b) Die theoretischen Variablen innerhalb von  $H$  werden operationalisiert, indem (weitere) Hilfsannahmen akzeptiert werden, die diese Begriffe (z. B. „Leistungsmotiv“) zu besser beobachtbaren Variablen in Beziehung setzen (z.B. Punktzahl in einem Test). Diese Hilfsannahmen sind zum Teil für den betreffenden Kontext spezifisch (z. B., daß Phantasien einer Person auf die Ausprägung ihres Leistungsmotivs schließen lassen), zum Teil allgemeinerer Natur (z. B. Skalenniveau, meßtheoretische Annahmen).

c) Auch nach der Operationalisierung wird H meist nicht unmittelbar mit einer einzelnen Beobachtungsaussage der Art  $E_b$  konfrontiert, sondern mit einem Befund, der sich aus einer *statistischen Verarbeitung* von „Rohdaten“ wie  $E_b$  ergibt. Wegen der zu erwartenden systematischen und zufälligen Beobachtungsfehler ist eine statistische Analyse meist auch dann erforderlich, wenn H selbst keine statistische Hypothese ist (vgl. Kap. 14). Aufgrund der Rohdaten wird mit Hilfe eines Testverfahrens eine statistische Hypothese akzeptiert (oder verworfen), die mit H in Einklang steht oder ihr widerspricht. Was in der Wissenschaftstheorie gewöhnlich einfach als „Akzeptieren eines Beobachtungssatzes“ bezeichnet wird, ist in der Praxis der Sozialwissenschaften in der Regel ein statistischer Test, der abermals gewisse Hilfsannahmen (z. B. Verteilungsvoraussetzungen) erfordert.

Wenn nun ein von der theoretischen Vorhersage abweichender Befund vorliegt, so fordert die Logik, die entstandene Inkonsistenz zu beseitigen. Hierzu gibt es eine Vielzahl von Möglichkeiten: Auch wenn man ausschließt, daß der Befund allein auf Zufallsfehlern bei der Beobachtung oder Randomisierung beruht, so ergibt sich daraus nur, daß *wenigstens eine* der zur Vorhersage verwendeten theoretischen Annahmen oder Hilfsannahmen verworfen werden muß. Nun kann für Hilfsannahmen der angesprochenen Art kein Gewißheitsanspruch erhoben werden. Die Hilfshypothesen, in denen die Grundbegriffe der Theorie vorkommen, können nicht unabhängig von weiteren Hilfsannahmen oder Grundannahmen der Theorie getestet werden. Selbst ein garantiert fehlerfreier empirischer Befund wurde also nicht zur Folge haben, daß eine ganz bestimmte theoretische Annahme als falsifiziert betrachtet werden muß.

Die Vertreter des Strukturalismus haben ein weiteres Argument gegen die Falsifizierbarkeit von Theorien vorgebracht: Um eine Theorie zu prüfen, müsse man T-theoretische Variablen messen, und hierzu müsse die Theorie selbst verwendet werden (vgl. Kap.7). Die Prüfung der Theorie erfordere also zugleich ihre *Voraussetzung*, was einen (unendlichen Regreß oder) *erkenntnistheoretischen Zirkel* zur Folge habe. - Diese Problemsicht entsteht allerdings nur dann, wenn man Sätze der Art  $E_t$  als Prüfinstanzen heranzieht (was nicht dem üblichen Vorgehen entspricht); in der Tat kann man nur mit Hilfe der Grundannahmen der Theorie die Größe „Tendenz zur Wahl einer Aufgabe“ berechnen. Dagegen kann man Beobachtungsaussagen wie  $E_b$  unabhängig von der Theorie überprüfen, so daß sich die angebliche Zirkularität auf die bereits geschilderte Duhemsche Problemsituation reduziert.

Popper hat diese Problematik berücksichtigt und betont, daß ein allgemein anerkannter, abweichender Beobachtungsbefund jederzeit neu überprüft und verworfen werden kann; auch hat er immer darauf hingewiesen, daß Theorien *nicht* als falsch *erweisbar* und in diesem Sinne *nicht falsifizierbar* sind. In dem Werk „Logik der Forschung“ finden sich im Zusammenhang mit der Falsifi-

kation folgende methodologischen Empfehlungen: 1) Zur Verteidigung einer Theorie sind nur solche Änderungen von Hilfshypothesen zugelassen, die den empirischen Gehalt des Gesamtsystems von Grundannahmen und Hilfshypothesen nicht vermindern, sondern erhöhen; nicht zugelassen sind z.B. solche Verteidigungen, die den abweichenden Befund ohne überprüfbare Begründung auf mangelnde Kompetenz oder auf eine Täuschungsabsicht des Untersuchers zurückführen. 2) Eine Falsifikation einer Theorie, die auf einem reproduzierbaren, intersubjektiv nachprüfbaren Befund beruht, sollte „im allgemeinen“ als „endgültig“ betrachtet werden (S.214). - Später hat Popper allerdings den heuristischen Charakter der Falsifikation betont und eingeräumt, daß die Falsifizierbarkeit „nicht als ein sehr strenges Kriterium gelten kann“ (1979, S.53). Jedoch hat er daran festgehalten, daß Falsifikationen trotz ihrer Revidierbarkeit „in der Geschichte der Wissenschaft eine Hauptrolle“ spielen.

Falsifikationen einer Theorie aufgrund eines empirischen Befundes werden unter anderem Kepler und Rutherford zugeschrieben; und Einstein legte sich zumindest darauf fest, seine allgemeine Relativitätstheorie als widerlegt anzusehen, wenn eine bestimmte Prognose (Rotverschiebung) sich nicht bewahrheiten sollte. In der Psychologie gibt es Falsifikationen in dem Sinne, daß eine einzelne Hypothese oder Theorie manchmal durch abweichende Befunde als *erheblich entkräftet* und fortan als kaum haltbar angesehen wird. So wurde beispielsweise die zitierte Frustrations-Aggressions-Hypothese bald als unhaltbar angesehen, weil es eindeutige Befunde gab, wonach einerseits Frustration nicht immer zu Aggression führt und andererseits Aggression nicht immer durch Frustration bewirkt sein muß (Berkowitz, 1962). - Das Gedächtnismodell von Atkinson und Shiffrin (1968) enthält die theoretische Annahme, daß Information im Kurzzeitspeicher ausschließlich akustisch kodiert ist. Diese Annahme (die von Bedeutung für die Unterscheidung der verschiedenen Gedächtnisspeicher ist) wurde durch Experimente stark entkräftet, die eine semantische und visuelle Kodierung im Kurzzeitspeicher demonstrierten (Lachman et al., 1979). - Es gibt viele Fälle dieser Art, in denen überzeugende empirische Kritik dazu geführt hat, daß eine Theorie als in der vorliegenden Form nicht länger haltbar angesehen wurde, worauf es dann zu Einschränkungen, Modifikationen, oder aber neuen Ansätzen kam.

Kuhn (1962) hat allerdings anhand von Beispielen aus der Geschichte der Physik darauf hingewiesen, daß *umfassende Theorien* wegen einzelner abweichender Befunde („Anomalien“) zunächst nicht aufgegeben wurden. Auch im psychologischen Bereich lassen sich zahlreiche Fälle finden, in denen psychologische Theorien durch Veränderung von Hilfsannahmen verteidigt wurden; einige seien erwähnt: In vielen Versuchen gelang es Pawlow nicht, sein Gesetz der klassischen Konditionierung zu bestätigen. Als Ursache hierfür gab er eine Reihe von störenden Bedingungen an: Das Versuchstier war zu müde

oder zu gesättigt; es waren ablenkende Reize vorhanden; der bedingte Reiz war selbst Auslöser eines biologisch wichtigeren Reflexes, was verhinderte, daß er mit dem zu konditionierenden Reflex assoziiert wurde (Pawlow 1953a, S. 187f.; 1953b, S. 21ff.) (Vgl. zur Kritik an Pawlows Argumentationstrategien Wettersten, 1974.) - Underwood und Ekstrand (1966) prognostizierten mit Hilfe der Interferenztheorie des Vergessens, daß in ihrem Experiment die von den Vpn früher gelernten Assoziationen zunächst einer Extinktion unterliegen und danach zu einer proaktiven Hemmung der im Experiment gelernten Liste führen würden. Als dies nicht nachgewiesen werden konnte, argumentierten sie, daß die früher erworbenen Assoziationen wohl durch langes und verteiltes Lernen zu sehr gefestigt waren, um durch das Lernen der neuen Liste beeinträchtigt werden zu können.- Atkinson und Litwin (1960) fanden, daß Personen mit überwiegender Furcht vor Mißerfolg Aufgaben von mittlerer Schwierigkeit nicht vermeiden, was sie gemäß der Theorie tun mußten. Atkinson erklärte dies später dadurch, daß neben den leistungsbezogenen Motiven noch andere Motive (z.B. Anschlußmotiv) wirksam waren, die auch Mißerfolgsmotivierte zu mittelschweren Aufgaben tendieren lassen.

Wie sind Poppers methodologische Empfehlungen 1 und 2 unter dem Gesichtspunkt der Zweckrationalität zu beurteilen? Die erste scheint eine sehr starke Anforderung zu sein, wenn man sie als strenge Regel interpretiert. Als heuristische Empfehlung erscheint sie aber überzeugend. Es kann zwar manchmal sinnvoll sein, eine Theorie durch Einschränkung ihrer Allgemeinheit besser mit den Daten in Einklang zu bringen. Fortwährende Einschränkungen stehen jedoch mit dem Ziel des empirischen Gehalts in Konflikt und sind in keiner empirischen Wissenschaft üblich. Wenn man dies voraussetzt, so kann man Empfehlung 1 durch die abgeschwächte Forderung ersetzen, daß die Annahmen zur Verteidigung einer Theorie *empirisch* prüfbare *Konsequenzen haben* müssen. Die bloße Behauptung einer unbekanntenen Störung des Untersuchungsablaufs hat keine prüfbaren Konsequenzen. Wenn vermutete störende Bedingungen dagegen konkret benannt werden, so kann man versuchen, deren Einfluß in weiteren Untersuchungen zu kontrollieren. Dabei muß nicht in jedem Fall eine Erhöhung des empirischen Gehalts des Gesamtsystems erreicht werden.

Die zweite Regel ist der Hauptgegenstand der Kontroversen um den Falsifikationismus gewesen. Oft haben Kritiker nur diesen Punkt beachtet und überdies noch vereinfacht, so daß die Kritik einen „naiven“ Falsifikationismus zum Gegenstand hatte, den niemand je vertreten hat. Nichtsdestoweniger läßt sich kaum begründen, insbesondere nicht durch Poppers sonstige Ausführungen, daß eine Falsifikation endgültig sein kann. Aufgrund der dargestellten Problemsituation ist das Risiko einer irrtümlichen Falsifikation nicht unbedeutend. (Nach Poppers späteren Äußerungen ist es „fast ebenso gewagt, eine Widerlegung (vorläufig) zu akzeptieren, wie eine Hypothese.“) Bedeutende

wissenschaftliche Theorien, etwa die Kopernikanische und die Newtonsche, waren von Anfang an schwerwiegenden empirischen Einwänden ausgesetzt, von denen sich jedoch nach längerer Zeit zeigte, daß sie auf falschen Hilfs-hypothesen beruhten. Zusammen mit neuen theoretischen Ansätzen werden immer auch neue Hilfsannahmen gebildet, unter denen einige vermutlich un-zutreffend sind. Es ist also durchaus begründet, im Falle abweichender Be-funde zunächst zu versuchen, Operationalisierungen zu ändern und Störfak-toren zu identifizieren, bevor man an der Adäquatheit der Grundannahmen zu zweifeln beginnt. Gemäß Lakatos' „Methodologie wissenschaftlicher For-schungsprogramme“ (1970) werden die Grundannahmen einer Theorie (sie entsprechen den „Kernannahmen“ des Programms) so lange beizubehalten, bis das Programm zu „degenerieren“ beginnt. Dies ist dann der Fall, wenn die Veränderungen und Einschränkungen der Hilfshypothesen, die man zur Verteidigung der Kernannahmen einführen muß, den empirischen Gehalt des Gesamtsystems fortwährend vermindern. Auch diese Methodologie vermag allerdings keine eindeutige Bedingung zu nennen, unter der eine Theorie bzw. ein Forschungsprogramm endgültig als falsifiziert betrachtet und aufgegeben werden sollte. Es kann niemals völlig ausgeschlossen werden, daß eine weitere Entwicklung des Forschungsprogramms doch noch zu spektakulären neuen Erfolgen geführt hätte.

Das Fehlen eines eindeutigen Kriteriums für das endgültige Verwerfen einer Theorie ist nun nicht mit dem Vorschlag gleichzusetzen, nach Belieben zu verfahren. Bislang wurden die heuristischen Empfehlungen akzeptiert (weitere werden hinzukommen), Theorien mit der Erfahrung zu konfrontieren, auf-tretende logische Widersprüche im Gesamtsystem von Grundannahmen, Hilfsannahmen und Daten zu beseitigen und dabei nur Änderungen zuzulas-sen, die neue prüfbare Konsequenzen haben. Wenn eine Abweichung von der Vorhersage eintritt, ist es der Phantasie des Wissenschaftlers überlassen, an einer bestimmten Stelle dieses Gesamtsystems einen Fehler zu vermuten, eine Änderung vorzunehmen und zu prüfen, ob sich mit den veränderten Gesamt-annahmen zutreffendere Vorhersagen erzielen lassen. Ebenso wie es hierbei kein eindeutiges Kriterium dafür gibt, eine Grundannahme zu revidieren, läßt es sich auch nicht zwingend verlangen, eine relativ frühzeitige Verwerfung der Grundannahmen zu verbieten, wie dies Lakatos' „negative Heuristik“ vor-sieht. (Musgrave (1976) zeigt anhand des Newtonschen Forschungspro-gramms, daß entgegen Lakatos' Metatheorie der ‚harte Kern‘ des Programms keineswegs unangetastet gelassen wurde.) In dieser Hinsicht sind verschiedene Forschungspläne denkbar, darunter auch solche, die bei gut fundierten Hilfs-annahmen frühzeitig eine Modifikation der Grundannahmen versuchen. Es kann sogar vorteilhaft sein, wenn Wissenschaftler in dieser Hinsicht verschie-dene Vorgehensweisen wählen und deren Erfolg vergleichen. Im Falle eines neuen theoretischen Ansatzes, der auch viele neue Hilfsannahmen umfaßt,

kann allerdings empfohlen werden, zunächst eher Änderungen der Hilfsannahmen zu versuchen.

Diese Art von Forschung erlaubt es, abweichende Befunde als Orientierungspunkte für neue Untersuchungen und verbesserte Theorien zu verwenden, ohne daß es zu einem ganz bestimmten Zeitpunkt klar sein müßte, daß eine bestimmte Menge von Grundannahmen unwiderruflich überholt ist. Ein wissenschaftlicher Fortschritt verlangt keine derartigen Urteile.

Eine Theorie ist *prüfbar* in dem Sinne, daß fortgesetzte Mißerfolge beim Versuch, mit veränderten Hilfshypothesen zu besseren Vorhersagen zu kommen, immer mehr die Option nahelegen, Grundannahmen zu ändern. Dies dürfte auch die in der Forschungspraxis überwiegend vertretene Auffassung sein. Die Tatsache, daß Wissenschaftler Theorien trotz abweichender Fakten aufrecht erhalten, zwingt keineswegs zu der Interpretation, daß sie diese Theorien als etwas nicht Prüfbares ansehen (wie dies der Strukturalismus vorschlägt). Ihr Vorgehen ist auch durch die Annahme interpretierbar, daß sie die Prüfung als einen schwierigen und langwierigen Prozeß auffassen, dessen Ausgang nicht durch einen einzelnen Wissenschaftler zu einem bestimmten Zeitpunkt endgültig beurteilt werden kann, sondern sich auf längere Sicht durch die gemeinsame Diskussion dadurch ergibt, daß die kontroversen Meinungen konvergieren.

Unabhängig von der Frage, ob es ein eindeutiges Falsifikationskriterium gibt, besteht heute eine Übereinstimmung darüber, daß eine erfolgreiche Theorie nicht aufgegeben werden sollte, so lange keine bessere zur Verfügung steht, die zusätzlich den abweichenden Befunden Rechnung trägt. Auch eine falsche Theorie kann von beträchtlichem Nutzen sein, insbesondere dann, wenn die Bedingungen identifizierbar sind, unter denen sie zu falschen Vorhersagen führt. Auch ist es oft möglich, abweichende Vorhersagen mit Hilfe von zusätzlichen Theorien über Störfaktoren zu korrigieren (vgl. Barr, 1974). Bekanntlich wird in der Technik erfolgreich mit der klassischen Mechanik gearbeitet, obwohl Übereinstimmung darüber besteht, daß die allgemeine Relativitätstheorie die Gesamtheit der Fakten besser zu erklären vermag (jedoch praktisch aufwendiger wäre). Ebenso ist es denkbar, zu bestimmten Zwecken etwa Skinners Verhaltenstheorie zu verwenden, auch wenn man der Überzeugung ist, daß sie mit bestimmten Fakten in Widerspruch steht.

## 2.2 Strategien der empirischen Prüfung

Da Datenerhebungen in der Regel aufwendig und kostspielig sind, ist es wichtig, wie man *informative* Daten finden kann, die zur Lösung der gegebenen Probleme, z. B. zur Bewertung von Theorien, möglichst viel beitragen. Nach

der Auffassung des kritischen Rationalismus sollte man versuchen, eine Theorie mit allen Mitteln *zu falsifizieren* und hierzu nach abweichenden Fakten suchen (wobei nicht festgelegt werden muß, daß im Falle abweichender Befunde die Theorie sofort aufzugeben ist). Diese Regel ist häufig als „realitätsfern“ empfunden worden. Nach einer verbreiteten Auffassung, die durch Kuhns Darstellung der „Normalwissenschaft“ bestärkt wurde, versuchen Wissenschaftler ihre Hypothesen und Theorien eher zu bestätigen, jedenfalls nicht zu falsifizieren. Nach den Ergebnissen psychologischer Forschung gibt es eine *Bestätigungstendenz* in bezug auf eigene akzeptierte Urteile, die allerdings nicht unter allen Bedingungen auftritt, wie häufig behauptet wurde, und die keineswegs den Charakter eines „Naturgesetzes“ hat (vgl. Gadenne & Oswald, 1986). Ohne Zweifel kann man davon ausgehen, daß die Schöpfer von Theorien ihre Produkte im allgemeinen lieber bestätigt als widerlegt sehen. Dies könnte es freilich um so dringlicher erscheinen lassen, ein Prinzip der *kritischen Prüfung* zur wissenschaftlichen Norm zu machen. Da Wissenschaft ein gemeinschaftliches und kommunikatives Unternehmen ist, ist es im Prinzip nicht erforderlich, daß Forscher ihre eigenen Hypothesen zu widerlegen versuchen; andere werden dies um so lieber tun, sofern diese Tätigkeit Anerkennung als wissenschaftliche Leistung verspricht.

Zur Diskussion der Frage, ob und unter welchen Bedingungen der Falsifikationsversuch eine zweckrationale Strategie darstellt, soll nun zunächst genauer bestimmt werden, worin ein solcher Versuch besteht. Wie kann man nach abweichenden bzw. übereinstimmenden Fakten suchen, ohne diese Fakten bereits zu kennen? Personen haben häufig Vermutungen darüber, ob eine Informationsquelle eher die eine oder die andere Art von Fakten liefern wird, sie können also eine entsprechende Informationssuche betreiben. (Damit befaßt sich die Forschung zur „selektiven Informationssuche“; vgl. Frey, 1981). Auch in der Wissenschaft kommt es häufig vor, daß die verschiedenen empirischen Folgerungen einer Theorie im Lichte des schon vorhandenen, akzeptierten Wissens (Hintergrundwissen, kurz HW) unterschiedlich plausibel erscheinen. Wenn aus der Theorie T (sowie den Anfangsbedingungen und Hilfs-hypothesen A) der empirische Sachverhalt E ableitbar ist, so sind in dieser Hinsicht verschiedene Untersuchungssituationen denkbar: 1) Das HW (zu dem auch A gehört) enthält Annahmen (z.B. eine konkurrierende Theorie T'), aus denen E ebenfalls ableitbar ist. 2) Das HW enthält keine Annahmen, die etwas über E folgern lassen. 3) Das HW enthält Annahmen (z.B. eine konkurrierende Theorie T'), aus denen sich Non-E ableiten läßt. - Situation 1 stellt keine echte Prüfung von T dar, da E aufgrund von plausibel gehaltenen anderen Annahmen auch dann zu erwarten ist, wenn T falsch sein sollte; im Extremfall ist E ein bereits allgemein anerkannter Sachverhalt, der möglicherweise sogar Ausgangspunkt bei der Konstruktion von T war. Fall 2 ist eine echte Prüfung, da hier immerhin die Möglichkeit besteht, daß ein von der

theoretischen Erwartung abweichender Sachverhalt eintritt. In Fall 3 schließlich ist E eine *neuartige*, riskante Vorhersage, die dem herkömmlichen Wissen widerspricht. Die *Strenge der empirischen Prüfung* von T nimmt von 1 nach 3 zu. - Da man es in der Psychologie mit *unvollständigen* Theorien zu tun hat (vgl. Kap. 7), tritt zu den erwähnten Untersuchungssituationen noch eine weitere hinzu, durch die die Strenge der Prüfung auf ein Maximum gebracht wird: 4) Man kann in die Untersuchung einen Wirkungsfaktor einführen, der gemäß den Annahmen des HW dem von der Theorie angenommenen psychologischen Vorgang entgegenwirkt, es diesem Vorgang und damit der Theorie also besonders „schwer macht“, bestätigt zu werden. Wenn E unter diesen Umständen dennoch eintritt, so spricht dies in besonderem Maße dafür, daß dieses Ergebnis kein bloßes Zufallsprodukt oder die Folge eines Untersuchungsfehlers ist.

Die Strenge einer Prüfung kann als die Wahrscheinlichkeit interpretiert werden, mit der in der gegebenen Untersuchungssituation abweichende Befunde auftreten, vorausgesetzt, T ist falsch. Da es im allgemeinen jedoch keine Möglichkeit gibt, eine solche Wahrscheinlichkeit zu berechnen, ist es wichtiger, die Unterschiede in der Strenge anhand der dargestellten, qualitativ verschiedenen Untersuchungssituationen aufzuzeigen, die in der Forschungspraxis auch identifiziert werden können. Aus dem Prinzip der kritischen Prüfung ergibt sich für empirische Untersuchungen die Empfehlung, sie als möglichst strenge Prüfungen anzulegen.

Mit Hilfe dieser Konzeption läßt sich nun deutlich machen, daß in der psychologischen Forschung bereits Regeln verankert sind, die auf strenge empirische Tests abzielen: Die Regel, Untersuchungssituationen der Art 1 zu vermeiden, weil hier das empirische Ergebnis auch durch andere Theorien erklärt werden kann, wird in der Methodenlehre der Psychologie als die Forderung formuliert, *Störvariablen zu kontrollieren*, um eine eventuelle Variation der AV auf die Variation der W zurückführen zu können (vgl. dazu Gadenne, 1976). Die Elimination und Konstanthaltung von Störfaktoren, die Randomisierung im Experiment, quasiexperimentelle Versuchspläne, Techniken der statistischen Kausalanalyse dienen alle dem Zweck, alternative Erklärungen der Variation der AV auszuschalten, um die empirischen Ergebnisse eindeutiger zur Bewertung der zu prüfenden Hypothese bzw. Theorie verwenden zu können (s. hierzu Kap. 12). Wenn ein Untersucher nichts weiter erreichen wollte als empirische Ergebnisse, die mit der Theorie übereinstimmen, so könnte er es sich sehr leicht machen, indem er einige Störfaktoren derart variieren ließe, daß dadurch die vorhergesagte Veränderung der AV „künstlich“ erzeugt wird. Eine derartige Untersuchung wurde allgemein als methodisch unakzeptabel beurteilt werden. Daran zeigt sich, daß die häufig geäußerte Meinung, in der Forschung wurde man letztlich nur nach Bestätigungen seiner Theorien suchen, recht unreflektiert ist. Kontrolle von Störfaktoren dient der Vermeidung



von irreführenden Bestätigungen. In der psychologischen Methodik gilt die Regel: Soviel Kontrolle, wie es der zu untersuchende Gegenstand, die Operationalisierung sowie die praktischen Gesichtspunkte (Durchführbarkeit, Ethik) zulassen.

Gehört es auch zu den impliziten Regeln der psychologischen Methodik, Untersuchungssituationen der Art 3 bzw. 4 solchen der Art 2 vorzuziehen? Fall 3 ist als *Entscheidungsexperiment* bekannt: Sollte E eintreten, so spricht dies für T und gegen T'; tritt E nicht ein, so ist es umgekehrt (wodurch natürlich, der Bezeichnung „Entscheidungsexperiment“ zum Trotz, nicht definitiv entschieden werden kann, daß eine der Theorien wahr ist). Aus methodologischer Sicht sind solche Untersuchungen deshalb wertvoll, weil sie empirische Information über zwei Theorien liefern. E bedeutet einen Erfolg der neuen Theorie und eine Korrektur des herkömmlichen Wissens; Non-E verweist auf einen Irrtum innerhalb des neuen Ansatzes und trägt zu seiner Verbesserung bei. Wenn man Theorien gegeneinander testet, so erhält man auf jeden Fall einen Befund *gegen* eine der beiden Theorien, und gegenteilige Befunde sind aus falsifikationstheoretischer Sicht der Motor des Erkenntnisfortschritts. Bekannte Entscheidungsexperimente wurden z. B. von Spence (1937), Tolman et al. (1946) und Bern (1967) durchgeführt. - Malewski (1967, S. 89ff.) hat die Theorie der kognitiven Dissonanz unter den besonders strengen Bedingungen der Art 4 getestet. Er stellte hierzu eine Situation her, in der die Vpn bestimmte Kosten in Kauf nehmen mußten, um kognitive Dissonanz zu reduzieren. Es zeigte sich, daß Personen bereit sind, ihre Selbsteinschätzung in negativer Richtung zu verändern, wenn sie auf diese Weise kognitive Dissonanz reduzieren können.

Obwohl es nun aber allgemein für wichtig gehalten wird, Theorien gegeneinander zu testen, und obwohl viele entsprechende Untersuchungen einen besonderen Grad an Bekanntheit erlangt haben, gehört dieses Verfahren doch nicht zu den unbedingten Geboten empirischer Forschung, vergleichbar der Kontrolle alternativer Erklärungen. In dieser Hinsicht geht das methodologische Prinzip der kritischen Prüfung über die schon bestehenden Normen der Praxis hinaus: Nach diesem Prinzip müßte zusätzlich zur selbstverständlichen Kontrolle von Störfaktoren viel mehr versucht werden, Theorien gegeneinander zu testen, um auf diese Weise die Zahl falsifizierender Befunde zu vermehren und die Beseitigung von Irrtümern zu forcieren. Ware es empfehlenswert, immer und in jeder Hinsicht so zu verfahren?

Aus Gründen, die mit der *Unvollständigkeit* von Theorien zusammenhängen, muß diese Frage verneint werden. Wenn zur Prüfung von T eine bestimmte Operationalisierung durchgeführt und eine bestimmte Beobachtungsvariable zur Vorhersage gewählt wird, so gilt immer, daß diese Variable durch sehr viele Faktoren mitdeterminiert wird, über die T nichts aussagt. T bezieht sich

auf einen der Wahrnehmung „verborgenen“ Vorgang V, dessen Existenz und Beschaffenheit anhand von gut beobachtbaren Sachverhalten E überprüft werden soll. Je besser beobachtbar E aber ist, desto länger ist die Kausalkette zwischen V und E, und desto mehr zusätzliche Einflüsse schieben sich dazwischen und erschweren die Interpretation. Die Gefahr, daß T zutreffend ist, daß V aber im „Rauschen“ der Störfaktoren nicht entdeckt wird, ist von Vertretern der Falsifikationstheorie meist zu wenig beachtet worden. Es gehört zur Kunstfertigkeit eines empirischen Forschers, eine Untersuchungssituation zu finden, in der V möglichst störungsfrei als vorhanden oder nicht vorhanden nachgewiesen werden kann. Neben der Wahl einer geeigneten Operationalisierung, die unter Umständen mehrfach abgeändert werden muß (ohne daß dies der Theorie angelastet wird), gibt es hierzu folgende Maßnahmen: 1) Die Intensivierung von V; man spricht manchmal von einer „durchschlagskräftigen“ AV. 2) Die Wahl einer für V *empfindlichen* Beobachtungsvariablen. 3) Die Kontrolle von Störfaktoren, die V entgegenwirken könnten; die meisten Kontrolltechniken, die dazu dienen, eine irrtümliche Bestätigung von T zu vermeiden (Konstanthaltung, Randomisierung), leisten gleichzeitig auch eine Kontrolle der gegenwirkenden Faktoren, die eine irrtümliche Falsifikation von T zur Folge haben könnten. Eine irrtümliche Falsifikation muß ebenso vermieden werden wie eine irrtümliche Bestätigung (Gadenne, 1976). Westermann (1987) fordert, eine Hypothese nicht nur streng, sondern auch „wohlwollend“ zu prüfen. Eine Prüfung von T kann als um so wohlwollender definiert werden, je kleiner die Wahrscheinlichkeit ist, mit der ein abweichendes Ergebnis auftritt, obwohl T wahr ist.

Die zuletzt dargestellte Problematik hat besonderes Gewicht, wenn es um einen neuen, besonders unvollständigen theoretischen Ansatz geht, mit noch wenig präzisierten Ideen und wenig erprobten Operationalisierungen. Die Ausgangsidee einer solchen Theorie könnte etwa besagen, daß der Wahrnehmungsvorgang durch Bedürfnisse, Werte und Persönlichkeitseigenschaften mitbestimmt wird, oder daß metakognitive Vorgänge einen Einfluß auf kognitive Leistungen haben. Wenn in diesem Stadium erste empirische Untersuchungen geplant werden, so ist es nicht zu empfehlen, bereits allzu strenge Tests gegen andere Theorien vorzunehmen oder gar gegenwirkende Faktoren einzuführen. Es wäre auf diese Weise leicht, Situationen zu finden, in denen die Bedürfnisse bzw. Metakognitionen *keinen* nachweisbaren Einfluß haben, doch wären solche Daten zunächst wenig informativ. Es geht vielmehr darum, Anwendungssituationen zu finden, in denen die von der Theorie postulierten Zusammenhänge eine Chance haben, „sichtbar“ zu werden. Diese Problematik existiert keineswegs nur in den „weniger präzisen“ Wissenschaften: Nachdem Faraday die zutreffende Idee hatte, daß ein sich bewegender Magnet Elektrizität erzeugt, erhielt er zunächst nur „falsifizierende“ Ergebnisse, und es dauerte fast 10 Jahre, bis er eine Versuchsanordnung fand, die empfindlich genug

war, den von ihm vermuteten Kausalzusammenhang anzuzeigen und seine Hypothese zu bestätigen (Williams, 1965). - Ein neu entworfener theoretischer Ansatz muß entwickelt werden, Vereinfachungen müssen beseitigt, Idealisierungen zum Teil zurückgenommen werden (Lakatos' „positive Heuristik“), und es wäre nicht sinnvoll, das Wissen um diese Vereinfachungen sofort zur „Falsifikation“ der Theorie einzusetzen. Pähler (1986, S. 136) fordert, daß Theorien nicht nur streng, sondern auch *fair* geprüft werden sollten, weil „Theorien *entwickelt* werden müssen und in der Regel genausowenig auf Anhieb die bisher beste Theorie herausfordern können wie ein Boxer den amtierenden Weltmeister. Beide müssen sich qualifizieren, d.h. zunächst bescheidenere Erfolge nachweisen.“ - Ist eine Theorie ausgearbeitet, und hat sie sich bereits in vielen Untersuchungen bewährt, so sind Prüfungen von zunehmender Strenge angebracht, die dazu dienen, die fehlerhaften Bestandteile der Theorie zu ermitteln. Solche Tests führen bei positivem Ausgang zu besonderen Erfolgen der Theorie, und bei negativem Ausgang helfen sie, Irrtümer zu ermitteln und weisen damit den Weg zu verbesserten Theorien.

Diese Überlegungen erweisen die Ideen der Falsifikation und der kritischen Prüfung keineswegs als inadäquat, verlangen aber, daß sie differenziert angewendet und durch zusätzliche Gesichtspunkte ergänzt werden. Strenge im Sinne der Kontrolle von Störfaktoren und der Ausschaltung von Alternativ-erklärungen ist unbedingt erforderlich; Strenge im Sinne der Untersuchungssituationen 3 und 4 sind eher im fortgeschrittenen Stadium der Theorienbildung und -prüfung angemessen. Das Prinzip, eine Theorie „mit allen Mitteln zu falsifizieren“, erfaßt nicht alle Aspekte, die im Zusammenhang mit der Suche nach informativen empirischen Ergebnissen wichtig sind.

### 2.3 Empirische Bestätigung und Wahrheit

Wenn Induktion es nicht vermag, Theorien zu verifizieren oder wahrscheinlich zu machen, *was* für ein *positives* Urteil ist dann im Hinblick auf Wahrheit möglich? Anders gefragt, welche unter den noch nicht falsifizierten Theorien soll man als beste Erklärung betrachten und zu praktischen Zwecken anwenden? Auf der Grundlage der in diesem Kapitel entwickelten Auffassung läßt sich folgende Antwort auf diese Fragen geben: Es ist rational, die am *besten bestätigte* (bewährte) Theorie vorzuziehen. (Im kritischen Rationalismus wird gewöhnlich der Begriff der „Bewährung“ verwendet, um eine Verwechslung mit „induktiver Bestätigung“ auszuschließen; im folgenden soll jedoch der üblichere Ausdruck „Bestätigung“ gebraucht werden, der inzwischen kaum mehr automatisch mit Induktion in Verbindung gebracht wird.) Am besten bestätigt ist die am *strengsten geprüfte und nicht falsifizierte* Theorie. Das

Ausmaß der Bestätigung hängt weniger von der Zahl der bestandenen Tests, als von deren Strenge und Verschiedenheit ab.

Als Minimalforderung für eine strenge Prüfung wurde oben die Tatsache angegeben, daß die empirische Evidenz E aus T zusammen mit A, nicht jedoch aus dem Hintergrundwissen ohne T ableitbar sein darf. Diese rein logische Bestimmung läßt jedoch in forschungspragmatischer Hinsicht verschiedene Auslegungen zu, die zu unterschiedlichen Konzeptionen der Strenge und Bestätigung führen: 1) E muß eine Vorhersage einer *neuen* Tatsache sein (und eintreten); E darf weder eine schon bekannte Tatsache noch aus akzeptierten Annahmen ableitbar sein. 2) E darf eine schon bekannte Tatsache sein, doch darf E nur aus T und A und nicht aus anderen Theorien (zusammen mit zutreffenden Anfangsbedingungen) ableitbar sein. 3) E darf nicht zweimal beansprucht werden, also nicht zugleich als Ausgangspunkt zur Konstruktion von T und als Bestätigung für T.

Die erste Auffassung, die von Popper und Watkins vertreten wird (zuvor von Whewell gegen Mill bzw. gegen den Induktivismus Bacons und Newtons), will definitiv ausschließen, daß etwas zur Bestätigung von T angeführt wird, was eventuell den *Anlaß zur Konstruktion* von T gab. Ware T konstruiert worden, um E zu erklären, so ist E kein echter Test, da Non-E ausgeschlossen ist, T also nicht an Non-E scheitern kann. Bestätigung kann aber nur aus einem gescheiterten Falsifikationsversuch resultieren. Es kommt darauf an, daß ein ernsthafter Versuch der Widerlegung von T unternommen wurde. E muß daher eine bislang von niemandem gekannte Tatsache sein. Diese Auffassung betrachtet Bestätigung also nicht als eine bestimmte Art von deduktiver oder induktiver Beziehung, sondern bezieht pragmatische Elemente mit ein: Es ist die *Entdeckung* einer neuen Tatsache, die T bestätigt.

Die zweite Konzeption (Lakatos, 1968; Musgrave, 1974) findet das Kriterium der Neuartigkeit zu streng und nicht entscheidend. Eine Theorie  $T_2$  kann gegenüber  $T_1$  einen empirischen Fortschritt darstellen, auch ohne zu einer Entdeckung geführt zu haben. Es ist denkbar, daß eine bekannte Tatsache E nur durch T erklärt werden kann und durch keine bisherige Theorie; in diesem Falle sollte man E als Bestätigung für T bewerten. Nicht der Entdeckungscharakter in bezug auf E, sondern der *empirisch belegte Gehaltsüberschuß*, die größere Erklärungskraft macht die Bestätigung von T aus. - Die Unterschiede seien anhand des Hauptbeispiels aus Kap. 7 erläutert. Als Atkinson seine Theorie T der Leistungsmotivation konstruierte, waren ihm aus eigenen Untersuchungen sowie aus Untersuchungen McClellands folgende Befunde bekannt: Die Leistung einer Person nimmt mit dem Anreiz (Gewinn) und mit der Gewinnerwartung bei einer Tätigkeit zunächst zu, bei sehr hoher Gewinnerwartung sinkt sie jedoch wieder.- Kinder mit hoher Leistungsmotivation wählen Aufgaben von mittlerer Schwierigkeit häufiger als Kinder mit

niedriger Leistungsmotivation. - Atkinson hat sich beim Entwurf seiner Theorie eindeutig von diesen Ergebnissen E leiten lassen. Als seine Theorie vorlag, gab es allerdings keine andere Theorie, die E hätte erklären können. Nach der ersten Bestätigungsauffassung, die *Neuartigkeit* verlangt, kann E nicht als Bestätigung von T akzeptiert werden. E ist allerdings eine Bestätigung im Sinne der zweiten Auffassung, nach der es genügt, daß E ausschließlich durch T erklärbar ist. Als Bestätigung gemäß der ersten Auffassung kann andererseits die von Feather (1961) entdeckte Tatsache gewertet werden, daß bei relativ leichten Aufgaben Erfolgsmotivierte und bei relativ schwierigen Aufgaben Mißerfolgsmotivierte ausdauernder sind.

Wie ist es nun, wenn E zwar zum Zeitpunkt der Konstruktion von T eine anerkannte Tatsache war, T jedoch ohne Berücksichtigung von E konstruiert wurde? Kein Theoretiker ist sich jederzeit aller anerkannten Fakten bewußt. Man könnte also nachträglich entdecken, daß T an einer schon lange bekannten Tatsache scheitert. Nach der dritten Auffassung kann auch eine bekannte Tatsache E als Bestätigung gewertet werden, wenn sie nicht bereits Ausgangspunkt zur Entwicklung von T war. Ein Problem hierbei ist allerdings, wie es objektiv festgestellt werden kann, ob E zur Konstruktion von T tatsächlich nicht verwendet wurde, nicht einmal unbewußt. (Die „positive Heuristik“ Lakatos' bietet eine Möglichkeit, das subjektive Element zu beseitigen, doch funktioniert diese Lösung vermutlich nur bei hochentwickelten, mathematisch formulierten Theorien; vgl. Musgrave, 1980).

Anstatt eine der drei Konzeptionen zu wählen, ist es auch möglich, sie zu integrieren und zu sagen, daß T durch die folgenden drei Arten von Befunden bestätigt wird: 1) bekannte, allein durch T erklärbare Tatsachen, zu deren Erklärung T entwickelt wurde; 2) bekannte Tatsachen, von denen man nachträglich feststellt, daß sie nur durch T erklärbar sind; 3) neue Tatsachen, die mit Hilfe von T vorhergesagt (entdeckt) wurden. Fall 2 kann man als eine „Entdeckung“ auffassen, deren Echtheit nicht so gut belegt ist wie im Fall 3. Ob man in 2 und 3 eine größere Bestätigung als in 1 sieht, ist letztlich nicht von allzu großer Tragweite, da alle Konzeptionen darin übereinstimmen, daß bei der Theorienbildung sowohl *höhere Erklärungskraft* als auch *Vorhersagen neuer Tatsachen* angestrebt werden sollten.

Gemäß diesen Regeln läßt sich die Bestätigung einer Theorie T entweder durch eine neue zutreffende Vorhersage erhöhen oder durch die Erklärung eines schon bekannten Sachverhalts, der bisher nicht mit T in Verbindung gebracht wurde. Ein Vergleich von verschiedenen Theorien hinsichtlich ihrer Bestätigung ist dann möglich, wenn  $T_2$  alles erklärt, was  $T_1$  erklären kann und darüber hinaus eine Tatsache E erklärt, die  $T_1$  nicht erklären kann. Ein Beispiel: Folgendes empirische Ergebnis E widerspricht der Theorie Atkinsons (T): Mißerfolg bei einer Aufgabe bewirkte bei den Vpn mit hoher resultie-

render Leistungsmotivation eine Zunahme der Motivation bei nachfolgenden Aufgaben; bei den Vpn mit niedriger resultierender Leistungsmotivation bewirkte Mißerfolg dagegen eine Abnahme der Motivation. E wird erklärbar, wenn man  $T_1$  ändert und ihr die Annahme einer „Trägheitstendenz“ der beiden angeregten Motive hinzufügt. Die dadurch entstandene Theorie  $T_2$  hat einen höheren Gehalt als  $T_1$ . Weiner (1965) testete in einem Experiment  $T_1$  und  $T_2$  gegeneinander, wobei es zu einer Bestätigung von  $T_2$  kam.  $T_2$  stellt gegenüber T, einen *Erkenntnisfortschritt* dar, der sowohl *theoretischer* (Gehalt) als auch *empirischer* Art (Bestätigung) ist. (Eine Erhöhung der Bestätigung *einer* Theorie allein ist allerdings auch eine Art des Erkenntnisfortschritts.) - Der Vergleich zwischen Theorien führt zu keinem eindeutigen Ergebnis, wenn auch  $T_1$  eine Tatsache erklärt, die  $T_2$  nicht erklären kann. In den Wissenschaften wird man freilich auch diese Prinzipien heuristisch verwenden und eventuell eine Überlegenheit von  $T_2$  auch dann zugestehen, wenn  $T_2$  zwar nicht in jeder Hinsicht, jedoch in den meisten Punkten besser abschneidet. Nicht durchführbar ist auf der Grundlage dieser methodologischen Regeln ein Vergleich von Theorien über verschiedene Gegenstandsbereiche, z. B. Leistungsmotivation und Einstellungsänderung. Ein solcher Vergleich wäre allerdings auch kaum von methodologischer Bedeutung.

Ein grundlegendes Problem aller Bestätigungskonzeptionen betrifft den Zusammenhang zwischen *Bestätigung* und *Wahrheit*. Was sagt Bestätigung, wie immer man sie definiert, über Wahrheit aus? Popper hat einerseits hervorgehoben, daß Bestätigung (Bewährung) nicht mehr sei als ein zusammenfassender Bericht über die Tests, ihre Art und Strenge, die T bestanden hat. Bestätigung sage nichts über zukünftige Erfolge von T aus und gebe keinen Grund dazu, T zu glauben, T für wahr zu halten bzw. mehr für wahr zu halten als eine andere Theorie. Andererseits sagt er aber auch, daß es rational sei, die besser bewährte Theorie zur Erklärung und praktischen Anwendung vorzuziehen; und daß wir Grund zu der Annahme hätten, daß die besser bewährte Theorie *der Wahrheit näher* sei (1982).

Diese Aussagen sind kaum miteinander in Einklang zu bringen: Entweder man verzichtet konsequent auf die Annahme, daß Erfahrung eine Theorie in irgendeinem Sinne zu bestätigen vermag. Oder man akzeptiert Bestätigung als einen Anhaltspunkt für Wahrheit, bekennt sich damit aber unvermeidlich zu Annahmen, die im weitesten Sinne des Wortes induktivistisch sind (Salmon, 1968). Lakatos (1974) ist der Auffassung, daß ein „Hauch“ Induktivismus unvermeidlich sei: Zwar läßt sich die Konstruktion von Theorien nicht durch formal-induktive Regeln beschreiben, und es gibt für Theorien auch keinen überzeugenden Kalkül einer induktiven Wahrscheinlichkeit; unverzichtbar ist jedoch die Annahme, daß Erfahrung dazu beizutragen vermag, einer Theorie irgendeine Art von Bestätigung im Hinblick auf Wahrheit (oder empirische Adäquatheit) zu verleihen bzw. Theorien in dieser Hinsicht einander vorzu-

ziehen. Ob man dies „Induktion“ nennt oder nicht, es handelt sich jedenfalls um eine *erkenntnistheoretische Annahme*, die nicht auf deduktiv-logische Regeln zurückgeführt werden kann.

Musgrave (1991) spricht sich dafür aus, eine entsprechende methodologische Regel zu akzeptieren (sie sei „R“ genannt): Es ist vernünftig, aus einer Reihe konkurrierender Hypothesen diejenige als wahr zu akzeptieren bzw. zu glauben, die der Kritik am besten standgehalten hat (etwa der Kritik durch Prüfung der Erklärungskraft und des Zutreffens neuer Vorhersagen) und die in diesem Sinne am besten bestätigt ist.

Es scheint kaum eine vernünftige Alternative zu R zu geben. Diese Regel ist allerdings so grundlegend, daß man sie vermutlich nicht begründen oder auch nur rational diskutieren kann, ohne sie selbst oder ein ähnliches Prinzip bereits verwenden zu müssen. Denn R gibt ja selbst an, was es heißt, eine Hypothese oder Regel unter (kritisch-) rationalen Gesichtspunkten vorzuziehen. Um eine deduktiv-logische Regel handelt es sich nicht. Eine Argumentation für R muß also Bezug nehmen auf methodologische Überlegungen, wissenschaftshistorische Betrachtungen und eventuell auf erfahrungswissenschaftliche Befunde. Wie auch immer man aber argumentiert, letztlich kann man nichts anderes tun, als R der Kritik Verschiedenster Art zu unterziehen, um dann gegebenenfalls zu dem Ergebnis zu kommen, daß R *der Kritik besser standhält als* konkurrierende Prinzipien (z. B. solche eines probabilistischen Induktivismus). Damit verwendet man aber R, um zu rechtfertigen, warum man R akzeptieren sollte. Die Frage, ob diese Art der Rechtfertigung einen Sinn macht oder unzulässig ist, kann hier nicht eingehender analysiert werden (vgl. dazu Musgrave, 1991).

Einige Wissenschaftstheoretiker verfolgen stattdessen diejenige Version des Fallibilismus weiter, die sich radikal *gegen* jeder Art von Begründungsversuch richtet. Agassi (1975) verzichtet konsequenterweise auf die Konzeption der Bestätigung. Er hält es für überflüssig, unter den nicht-falsifizierten Theorien Unterschiede in der Bestätigung zu machen. Glaube und Bestätigung seien zwar notwendig im Kontext der Technologie und des praktischen Handelns, nicht jedoch in der Wissenschaft. - Eine Schwäche dieser Auffassung besteht darin, daß handlungsrelevantes Wissen zum Teil aus wissenschaftlichen Theorien besteht bzw. sich aus ihnen herleitet.

Watkins (1984) vermeidet ebenfalls jeglichen Induktivismus und nimmt dafür eine Umdefinition des Zieles der Wissenschaft in Kauf: Das höchste erreichbare Ziel ist *mögliche Wahrheit* (possible truth); was die Wissenschaft als Ergebnis liefern kann, sind Theorien, die *noch nicht falsifiziert* sind, d.h. die in sich und mit der empirischen Evidenz konsistent sind. Theorien, die dieser Bedingung (die immerhin strenge Tests involviert) genügen, sind nichtsdesto-

weniger ohne jegliche positive Stützung („floating in the ocean of uncertainty“).

Es stellt sich die Frage, ob mit diesen Beschränkungen des Erkenntnisanspruchs in positiver Hinsicht viel gewonnen wird, wenn andererseits an der *Möglichkeit der Falsifikation* nach wie vor festgehalten werden soll. Eine Theorie wegen widersprechender Beobachtungsaussagen und Hilfsannahmen zu verwerfen, scheint nur dann einen Sinn zu machen, wenn man wenigstens diese Beobachtungsaussagen und Hilfsannahmen (trotz ihrer prinzipiellen Fehlbarkeit) für in irgendeinem Sinne *positiv gestützt* hält.

Einige Wissenschaftstheoretiker bestreiten dies jedoch. Bartley (1987) vertritt einen „pankritischen“ Rationalismus, der die in der Tradition für grundlegend gehaltene Frage „Wie kann man diese Aussage begründen?“ schlichtweg für rational nicht beantwortbar erklärt und durch die Frage ersetzt: „Wie kann man diese Aussage am besten kritisieren?“ - wobei *nicht* auf dem Umweg über die Kritik doch wieder so etwas wie partielle Begründung, Bestätigung oder ähnliches gewonnen werden soll. Rationalität ist mit Offenheit für Kritik und kritischem Vorgehen gleichzusetzen.

Auch Miller (1987) wendet sich gegen jeglichen Versuch, (vollständige oder unvollständige) Gründe für irgendetwas anzugeben. Er hält es im übrigen für einen Irrtum, zu meinen, daß im Zusammenhang mit einer Falsifikation etwas begründet werden müsse: Die empirische Evidenz Non-E, die T widerspricht, wird *ohne positive Begründung* gegen T angeführt. Jedem, der dies wünscht, steht es frei, eine weitere Annahme anzuführen, die wiederum gegen Non-E spricht usw. Nach dieser Konzeption von Rationalität ist es vorgesehen, Aussagen als wahr oder falsch zu bewerten, ohne daß dies begründet werden muß.

Es ist allerdings schwer einzusehen, worin sich diese Konzeption noch vom Skeptizismus unterscheidet. Es ist auch kaum zu verstehen, was es im Rahmen dieser Auffassung überhaupt noch bedeuten kann, daß Wahrheit das *Ziel* der Wissenschaft ist (woran Miller festhält). Im Zusammenhang mit einer Rationalitätskonzeption ein Ziel zu haben heißt, bestimmte Vorgehensweisen für *zielführend* (oder zumindest geeigneter als andere) zu halten; dies läuft aber unvermeidlich darauf hinaus, etwas zu *glauben*, etwas für besser *begründet* zu halten als etwas anderes. Wer also das Ziel rationaler Erkenntnis weiterhin verfolgen will, scheint nicht umhin zu kommen, wie Musgrave ein methodologisches Prinzip zu akzeptieren, das *gute Gründe* (wenn auch nicht sichere) für möglich hält.

Würden sich die zuletzt angeführten Probleme verringern, wenn Theorien als *nicht wahrheitsfähig* aufgefaßt und nur nach ihrer *empirischen Adäquatheit* bewertet werden? Qualitativ verändert sich die Problematik dadurch nicht, denn auch das Ziel der empirischen Adäquatheit wirft das Induktionsproblem



auf: Der Instrumentalist ist natürlich vorwiegend daran interessiert, zukünftige Ereignisse vorhersagen zu können. Und auch eine instrumentalistisch interpretierte Theorie bezieht sich auf unbegrenzt viele Fälle.

Legt man die *strukturalistische* Theorienkonzeption zugrunde, so erübrigt sich das Problem, die gesamte Theorie empirisch belegen oder entkräften zu müssen. Die Problematik der empirischen Bewertung stellt sich hier nur in bezug auf die einzelnen Hypothesen, die jeweils für eine Anwendungssituation behaupten, daß die Gegebenheiten der Situation als *Modelle der Theorie* interpretierbar sind. Eine solche Hypothese könnte etwa behaupten, daß eine Anzahl von Meßwerten für das Leistungsmotiv sowie einige Messungen der Ausdauer von Personen bei bestimmten Aufgaben, die in einer Untersuchungssituation gegeben wurden, mit den Annahmen der Theorie Atkinsons in Einklang stehen. Würde nun aber der Strukturalist lediglich solche Hypothesen prüfen, so würde die strukturalistische Forschung nicht mehr liefern können als ein Wissen über *vergangene Einzelfälle*. Jegliche Möglichkeit der Vorhersage wäre damit aufgehoben. Auch könnte der Tatsache nicht Rechnung getragen werden, daß in der empirischen Forschung tatsächlich *allgemeine Hypothesen* geprüft werden und daß *Kontrolltechniken* verwendet werden (s. Kap. 12). Letztere dienen der Kausalinterpretation, die immer über den Einzelfall hinausgeht. Wer nur singuläre Sätze prüfen wollte, könnte sich eine Anwendung dieser Techniken ersparen, denn die bloße Frage, ob die Daten bestimmte theoretische Annahmen erfüllen, kann völlig unabhängig von der eventuellen Möglichkeit alternativer Erklärungen beantwortet werden. Der Strukturalismus wird also seine Anwendungshypothesen sinnvollerweise als *allgemeine Hypothesen niedriger Allgemeinheitsstufe* konzipieren müssen, beispielsweise bezogen auf jede Risikowahl, alle Anspruchsniveausetzungen, alle Ausdauerleistungen usw. Will man derartige allgemeine Hypothesen strikt vermeiden, so bliebe nur die Möglichkeit eines erneuten Versuches, eine Art induktive Logik zu entwickeln, die einen Schluß von untersuchten Einzelfällen auf noch unbekanntes gestattet (vgl. Kap. 10).

## Literatur

- Agassi, J. (1975). *Science in flux*. Dordrecht: Reidel.
- Albert, H. (1987). *Kritik der reinen Erkenntnislehre*. Tübingen: Moor.
- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, N.J. : Erlbaum.
- Andersson, G. (1988). *Kritik und Wissenschaftsgeschichte*. Tübingen: Mohr.
- Atkinson, J. W. (1964). *An introduction to motivation*. Princeton (N.J.): Van Nostrand.
- Atkinson, J.W. & Litwin, G.H. (1960). Achievement motive and test anxiety conceived as motive to approach success and to avoid failure. *Journal of Abnormal and Social Psychology*, 60, 52-63.

- Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. W. Spence (Hrsg.), **Advances in the psychology of learning and motivation research and theory, Bd. 2**. New York: Academic Press.
- Barr, W.F. (1974). A pragmatic analysis of idealizations in physics. **Philosophy of Science, 41, 48-64**.
- Bartley III, W. W. (1968). Theories of demarcation between science and metaphysics. In I. Lakatos & A. Musgrave (Hrsg.), **Problems in the philosophy of science (S. 40-119)**. Amsterdam: North-Holland.
- Bartley III, W.W. (1987). **Flucht ins Engagement**. Tübingen: Mohr.
- Battig, W. F. (1962). Parsimony in psychology. **Psychological reports, 11, 555-572**.
- Bern, D.J. (1967). Self-perception: An alternative interpretation of a cognitive dissonance phenomena. **Psychological Review, 74, 183-200**.
- Berkowitz, L. (1962). **Aggression: A socialpsychological analysis**. New York: McGraw-Hill.
- Brehm, J. W. (1966). **A theory of psychological reactance**. New York: Academic Press.
- Bunge, M. (1963). **The myth of simplicity**. Englewood Cliffs, N.J.: Prentice-Hall.
- Bunge, M. (1967). **Scientific research, Bd. 1**. Berlin: Springer.
- Carnap, R. (1934). **Logische Syntax der Sprache**. Wien: Springer.
- Carnap, R. (1942). **Introduction to semantics**. Cambridge, Mass.: Harvard University Press.
- Churchland, P.M. (1984). **Matter und consciousness**. Cambridge: The MIT Press.
- Duhem, P. (1908). **Ziel und Struktur der physikalischen Theorien**. Leipzig: Barth. Nachdruck Hamburg: Meiner, 1978.
- Feather, N.T. (1961). The relationship of persistence at a task to expectation of success and achievement-related motives. **Journal of Abnormal and Social Psychology, 63, 552-561**.
- Feyerabend, P. (1976). **Wider den Methodenzwang**. Frankfurt am Main: Suhrkamp.
- Frey, D. (1981). **Informationssuche und Informationsbewertung bei Entscheidungen**. Bern: Huber.
- Gadenne, V. (1976). **Die Gültigkeit psychologischer Untersuchungen**. Stuttgart: Kohlhammer.
- Gadenne, V. & Oswald, M. (1986). Entstehung und Veränderung von Bestätigungstendenzen beim Testen von Hypothesen. **Zeitschrift für experimentelle und angewandte Psychologie, 33, 360-374**.
- Gadenne, V. & Oswald, M. (1991). **Kognition und Bewußtsein**. Berlin: Springer.
- Glymour, C. (1980). **Theory and evidence**. Princeton, N.J.: University Press.
- Grabitz-Gniech, G. & Grabitz, H. J. (1973). Psychologische Reaktanz: Theoretisches Konzept und experimentelle Untersuchungen. **Zeitschrift für Sozialpsychologie, 4, 19-35**.
- Irle, M. (1975). **Lehrbuch der Sozialpsychologie**. Göttingen: Hogrefe.

- Kruglanski, A. W. (1980). Lay epistemo-logic - process and contents: Another look at attribution theory. *Psychological Review*, **87**, 70-87.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Kutschers, F. v. (1972). *Wissenschaftstheorie*, Bd. 2. München: Fink.
- Lachman, R., Lachman, J.L. & Butterfield, E.C. (1979). *Cognitive psychology and information processing*. Hillsdale (N.J.): Lawrence Erlbaum.
- Lakatos, I. (1968). Changes in the problem of inductive logic. In I. Lakatos (Hrsg.), *The problem of inductive logic* (S. 315-417). Amsterdam: North Holland.
- Lakatos, I. (1974). Popper on demarcation and induction. In P.A. Schilpp (Hrsg.), *The philosophy of Karl Popper, Bd. 1*. La Salle: Open Court.
- Laudan, L. (1977). *Progress and its problems*. London: Routledge & Kegan Paul.
- Malewski, A. (1967). *Verhalten und Interaktion*. Tübingen: Mohr.
- Miller, D. (1987). A critique of good reasons. In J. Agassi & I.C. Jarvie (Hrsg.), *Rationality: The critical view* (S. 323-358). Dordrecht: Nijhoff.
- Musgrave, A. (1974). Logical versus historical theories of confirmation. *British Journal for the Philosophy of Science*, **25**, 1-23.
- Musgrave, A. (1976). Method or madness? In R.S. Cohen, P. Feyerabend & M. W. Wartofsky (Hrsg.), *Essays in memory of Imre Lakatos* (S.457-491). Dordrecht: Reidel.
- Musgrave, A. (1980). Stützung durch Daten, Falsifikation, Heuristik und Anarchismus. In G. Radnitzky & G. Andersson (Hrsg.), *Fortschritt und Rationalität der Wissenschaft* (S. 199-220). Tübingen: Mohr.
- Musgrave, A. (1991). What is critical rationalism? In A. Bohnen & A. Musgrave (Hrsg.), *Wege der Vernunft* (S. 17-30). Tübingen: Mohr.
- Nisbett, R.E. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgement*. Englewood Cliffs (N.J.): Prentice-Hall.
- Pähler, K. (1986). *Qualitätsmerkmale wissenschaftlicher Theorien*. Tübingen: Mohr.
- Pawlow, J. P. (1953a). *Sämtliche Werke, Bd. II/1: Zwanzigjährige Erfahrungen mit dem objektiven Studium der höheren Nerventätigkeit (des Verhaltens) der Tiere*. Berlin: Akademie-Verlag.
- Pawlow, J. P. (1953b). *Sämtliche Werke, Bd. IV: Vorlesungen über die Arbeit der Großhirnhemisphären*. Berlin: Akademie-Verlag.
- Popper, K. (1966). *Logik der Forschung*, 2. Aufl. Tübingen: Mohr (Originalausgabe Wien: Springer, 1935).
- Popper, K. (1972). *Objective knowledge*. Oxford: Clarendon Press.
- Popper, K. (1972). The aim of Science. In K. Popper (Hrsg.), *Objective knowledge*. Oxford: Clarendon Press.
- Popper, K. (1979). *Ausgangspunkte*. Hamburg: Hoffmann und Campe.
- Salmon, W.C. (1968). The justification of inductive rules of inference. In I. Lakatos (Hrsg.), *The problem of inductive logic* (S.24-43). Amsterdam: North Holland.

- Schmid, M. (1993). Verhaltenstheorie versus Nutzentheorie: Zur Systematik einer theoretischen Kontroverse. *Zeitschrift für allgemeine Wissenschaftstheorie* (im Druck).
- Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, **44**, 430-444.
- Tolman, E. C., Ritchie, B. F. & Kalish, D. (1946). Studies in spatial learning, I. Orientation and the short-cut. *Journal of experimental Psychology*, **36**, 13-24.
- Underwood, B. J. & Ekstrand, B. R. (1966). An analysis of some shortcomings in the interference theory of forgetting. *Psychological Review*, **73**, 540-549.
- Watkins, J. (1984). *Science and scepticism*. Princeton, N.J.: University Press.
- Weiner, B. (1965). The effects of unsatisfied achievement motivation on persistence and subsequent Performance. *Journal of Personality*, **33**, 428-442.
- Wendel, H. J. (1987). *Benennung, Sinn, Notwendigkeit*. Frankfurt: Athenäum.
- Westermann, R. (1987). *Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie*. Berlin: Springer.
- Wettersten, J.R. (1974). Methods in psychology; a critical case study of Pavlov. *Philosophy of the Social Sciences*, **4**, 17-34.
- Williams, L.P. (1965). Michael *Faraday*. London: Chapman and Hall.

## 10. Kapitel

# Induktion

### ***Rainer Westermann und Peter Gerjets***

Wissenschaftliche Erkenntnisse, so wie sie in Hypothesen, Theorien oder Modellen formuliert werden können, zeichnen sich nach traditioneller Ansicht gegenüber dem Alltagswissen dadurch aus, daß sie rational entstanden, begründet oder überprüft sind. Als optimale Form der Rationalität gilt dabei die deduktive Logik: Sie erlaubt, Aussagen so zu begründen, daß sie wahr sein müssen, wenn bestimmte andere Aussagen wahr sind, und sie erlaubt im Prinzip die Ableitung von Prüfinstanzen, die den zwingenden Nachweis der Falschheit bestimmter Aussagen ermöglichen. Dieses Rationalitätsparadigma ist allerdings insofern beschränkt, als weder die deduktiven Schlußregeln noch die Wahrheitswerte von basalen Aussagen über Sinneserfahrungen deduktiv begründet werden können. Vor allem aber können nicht alle wesentlichen Teile wissenschaftlichen Denkens oder Argumentierens mittels deduktiver Logik adäquat rekonstruiert werden. Ein Grund dafür liegt darin, daß wissenschaftliche Erkenntnisse einerseits auf dem Fundament empirischer Beobachtungen gegründet sein sollen, andererseits aber die Phänomene unserer Welt in allgemeiner Weise beschreiben und durch tieferliegende Ursachen erklären sollen. Damit müssen wissenschaftliche Erkenntnisse über das hinausgehen, was wir aufgrund von Sinneserfahrungen und deduktiv-logischen Ableitungen wissen können.

So finden sich in der Wissenschaft regelmäßig Argumentationen, die eindeutig nicht deduktiv-logisch sind. Dazu gehören vor allem Formulierungen, in denen beschränkte Befunde über Zusammenhänge zwischen empirischen Variablen genutzt werden, um weitergehende Aussagen abzuleiten oder zu stützen. Diese Argumente können in drei Gruppen eingeteilt werden:

- Extrapolationen (auf einzelne nicht untersuchte Situationen oder Personen),
- Generalisierungen (über alle interessierenden Situationen oder Personen),
- Abduktionen (auf nicht direkt beobachtbare theoretische Variablen und Zusammenhänge).

Diese verschiedenen Arten von Inferenzen sind typische Beispiele für induktive Argumente. Ihren Ergebnissen kommt zwar nicht die Sicherheit deduktiver Ableitungen zu, trotzdem erscheinen sie uns aber in der Regel nicht gänzlich unbegründet und willkürlich. Worauf diese Intuition beruht und ob sie berechtigt ist, soll im Verlauf dieser Arbeit näher untersucht werden.

Induktive Argumente werfen vor allem zwei Fragen auf:

1. Läßt sich eine in irgendeinem Sinn „vernünftige“ Begründung oder Rechtfertigung für die „Ableitung“ einer allgemeineren Aussage aus Aussagen über begrenzte Beobachtungen angeben?
2. Wird eine derartige weitergehende Aussage durch die zugrundeliegenden Beobachtungsaussagen in bestimmtem Maße „bestätigt“ oder „gestützt“?

Diese Fragen nach der Rechtfertigung induktiver Verallgemeinerungen und nach dem Grad der Bestätigung von Vorhersagen, Hypothesen oder Theorien durch empirische Daten umreißen das sog. Induktionsproblem, das eines der ältesten und am intensivsten diskutierten Probleme der Erkenntnistheorie und Wissenschaftsphilosophie ist. Wir werden die mit induktiven Argumenten verbundenen Probleme im ersten Abschnitt dieser Arbeit näher charakterisieren und in den folgenden Abschnitten verschiedene Ansätze zur Behandlung des Induktionsproblems darstellen, die uns für die Methodologie der psychologischen Forschung wichtig erscheinen.

Rudolf Carnap und Karl Popper sind die Urheber der beiden bekanntesten und häufig kontrovers diskutierten neueren wissenschaftsphilosophischen Ansätze zu diesem Induktionsproblem.

Carnap (1950) verfolgte zunächst das Ziel, auf der Basis der Wahrscheinlichkeitstheorie eine „induktive Logik“ zu entwickeln, so daß Maße für die Bestätigung von Hypothesen durch Aussagen über Beobachtungen aufgrund rein syntaktischer Kriterien abgeleitet werden können (siehe dazu Abschnitt 2). Die „Logik der Forschung“ von Popper (1934, 1984) hingegen ist von dem Bestreben durchzogen, sich eindeutig von diesen probabilistischen Lösungsversuchen des Induktionsproblems abzugrenzen. Nach seiner Überzeugung soll die Wissenschaft sich jeder Art von induktiven Argumenten und Bestätigungen enthalten und ausschließlich deduktiv vorgehen (siehe Abschnitt 3).

Während die logisch-syntaktisch orientierten Autoren nach formalen und kontextunabhängigen Regeln zur Identifizierung korrekter Induktion suchen, wird in den pragmatischen Ansätzen besonders der Verwendungskontext induktiver Argumente beachtet und ihre Nützlichkeit für konkrete Entscheidungen und Handlungen betont (Abschnitt 4). Eng verbunden mit den pragmatischen Überlegungen zur Induktion sind psychologische Beschreibungen und Simulationen tatsächlichen induktiven Denkens, die auf der Basis kognitiver Theorien der Informationsverarbeitung vorgenommen wurden (Abschnitt 5).

Im Abschnitt 6 schließlich wollen wir aus der Sicht der strukturalistischen Wissenschaftsphilosophie (Balzer, Moulines & Sneed, 1987; Westermann, 1987a) die induktiven Anteile in psychologischen Forschungsprozessen genauer lokalisieren und zusammenfassend gute Gründe für die Akzeptierung bestimmter induktiver Argumente formulieren.

## 1. *Charakteristika und Probleme induktiver Argumente*

Wie können induktive Argumente von deduktiven abgegrenzt werden? Charakteristisch für ein korrektes deduktives Argument ist, daß es nicht gehaltserweiternd ist: Der Gehalt der Konklusion, d.h. die Klasse aller Aussagen, die deduktiv aus der Konklusion ableitbar sind, ist im Gehalt der Prämissen enthalten (siehe Tarski, 1977). Da in diesem Falle die Wahrheit der Konklusion notwendigerweise aus der Wahrheit der Prämissen folgt, sind deduktive Schlüsse stets wahrheitskonsentierend. Das wesentliche Charakteristikum induktiver Argumente besteht hingegen darin, daß sie ampliative Inferenzen sind, d.h. daß bei ihnen der Gehalt der Konklusion über den Gehalt der Prämissen hinausgeht (siehe Essler, 1973, S. 10; Stegmüller, 1971, S. 16).

Gehaltserweiternd sind allerdings nicht nur induktive Argumente, sondern auch Argumente, die man besser als inkorrekte deduktive Schlüsse bezeichnen würde. Dazu gehört beispielsweise der „unzulässige Umkehrschluß“ (von  $A \Rightarrow B$  auf  $B \Rightarrow A$ ). Deshalb ist eine befriedigende Abgrenzung von induktiven und deduktiven Schlüssen vermutlich nur pragmatisch möglich (Essler, 1973; Machina, 1985; Salmon, 1979): Ein Argument ist induktiv, wenn der Vertreter dieses Argumentes behauptet, daß

1. die Wahrheit der Konklusion nicht notwendigerweise aus der Wahrheit der Prämissen folgt, weil der Konklusionsgehalt über den Prämissengehalt hinausgeht, daß aber
2. die Prämissen dennoch die Konklusion in irgendeiner Weise unterstützen, ihr Gewicht verleihen oder sie wahrscheinlicher machen.

Man sollte daher möglichst vermeiden, von induktiven „Schlüssen“ zu sprechen, um keine zu enge Parallelität mit den deduktiven Schlüssen und ihrer inferentiellen Sicherheit zu suggerieren.

### 1.1 Enumeration und andere Arten von Induktion

Obwohl die Klasse der induktiven Argumente damit prinzipiell unbegrenzt ist, meint man mit diesem Begriff häufig vor allem die verschiedenen Arten der sog. enumerativen Induktion (Kutschera, 1972, S. 189-195; Pollock, 1984). Typische Beispiele sind:

- $F(a_1), F(a_2), \dots, F(a_n) \approx F(a_{n+1})$ ,  
d.h. der (noch nicht näher spezifizierten und hier mit  $\approx$ ) bezeichnete) Übergang von der Aussage, daß die Objekte  $a_1$  bis  $a_n$ , die Eigenschaft  $F$  haben („Prämissen“), auf die Aussage, daß das nächste Objekt diese Eigenschaft hat („Konklusion“),
- $F(a_1) \ \& \ G(a_1), \dots, F(a_n) \ \& \ G(a_n) \approx \forall_i: F(a_i) \Rightarrow G(a_i)$ ,  
d.h. der Übergang vom gemeinsamen Auftreten von  $F$  und  $G$  in  $n$  Fällen auf eine für alle Objekte ( $\forall_i$ ) gültige implikative Beziehung ( $\Rightarrow$ ) zwischen beiden Merkmalen.

Diesen induktiven Argumenten liegt die Annahme zugrunde, daß ein Ereignis, das  $n$  mal aufgetreten ist, auch die folgenden Male auftreten wird. Sie können als Spezialfälle eines allgemeinen Induktionsprinzips (Reichenbach, 1938, S. 340-341) betrachtet werden, nach dem die bisherige relative Häufigkeit eines Ereignisses als Schätzung für seine Wahrscheinlichkeit genommen werden soll.

Wie Cohen (1989) ausführlich darstellt, haben aber bereits F. Bacon und J.S. Mill betont, daß wissenschaftliche Induktion keineswegs nur einfach enumerativ, sondern auch eliminativ ist. Das Ausmaß, in dem eine allgemeine implikative Aussage durch Beobachtungen gestützt ist, hängt danach nicht so sehr von der Anzahl der bestätigenden Instanzen ab, sondern viel mehr von der Vielfalt an Umständen und Faktoren, die berücksichtigt worden sind, um alternative Erklärungen auszuschließen. Wenn wir Mills Grundmethoden der eliminativen Induktion verwenden, können wir auf eine gesetzmäßige (ursächliche) Beziehung zwischen  $F$  und  $G$  schließen, falls

1.  $F$  und  $G$  regelmäßig miteinander auftreten, unabhängig von der Variation anderer möglicher Einflußfaktoren (Übereinstimmungsmethode) und falls
2. Situationen, in denen  $G$  bzw. nicht- $G$  eintritt, sich nur dadurch unterscheiden, daß im ersten Fall auch  $F$  vorgelegen hat, im zweiten Fall aber nicht (Differenzmethode).

Für die Wissenschaft charakteristisch sind außerdem die induktiven Argumente, die sich nicht in Prämisse und Konklusion auf die gleichen beobachtbaren Variablen beziehen, sondern von Aussagen über beobachtbare Ereignisse zu bestmöglichen Hypothesen oder Theorien über deren unbeobachtbare Ursachen führen (Abduktionen; vgl. Jason, 1985; Laudan, 1981). Bedeutsam sind auch induktive Argumente, die sich auf Analogien stützen, beispielsweise bei der Übertragung theoretischer Vorstellungen auf andere, ähnliche Kontexte. Diese Form analogen Rasonierens spielt eine wichtige Rolle bei der Einschätzung der Gültigkeit von Hintergrundannahmen in hypothesenprüfenden empirischen Untersuchungen und bei Theorienanwendungen in praktischen Kontexten.



## 1.2 Starkes und schwaches Induktionsproblem

Es ist zweckmäßig, mindestens zwei Spielarten des Induktionsproblems zu unterscheiden (Kemeny, 1963; Popper, 1965; Thagard & Nisbett, 1982; Will, 1982). Das starke Induktionsproblem besteht darin, eine globale, logische oder formale Rechtfertigung der Induktion zu finden, die zeigt, daß induktive Argumente generell die Wahrheit ihrer Konklusionen etablieren (wobei die Konklusion auch einen Wahrscheinlichkeitsausdruck enthalten kann). Seit David Humes pointierter Induktionskritik im „Treatise on Human Nature“ aus dem Jahre 1739 ist klar, daß Induktion in diesem starken Sinne nicht zu begründen ist (Popper, 1984; Salmon, 1979):

- Eine deduktive Rechtfertigung der Induktion ist nicht möglich, denn könnten wir deduktiv zeigen, daß aus den Prämissen die Konklusion folgt, wäre das Argument nicht mehr gehaltserweiternd, also nicht mehr induktiv.
- Der Versuch einer induktiven Rechtfertigung der Induktion (etwa aufgrund der Erfolge bisheriger Induktionen) führt in einen Zirkel oder unendlichen Regreß, weil das zu rechtfertigende Induktionsprinzip schon vorausgesetzt werden muß.
- Eine apriorische Rechtfertigung der Induktion (etwa durch die Einführung eines Uniformitätspostulates) ist nicht möglich, da es trotz der Versuche Kants als unmöglich gilt, synthetische Sätze a priori zu begründen.

Deshalb hat das starke Induktionsproblem nur eine triviale negative Lösung: „Es gibt keine wahrheitskonservierenden Erweiterungsschlüsse“ (Stegmüller, 1973, S. 76-78). Die Überzeugung, daß wir neben unserem Beobachtungswissen, das sich auf die Gegenwart und Vergangenheit bezieht, auch genauso sicheres Wissen über die Zukunft oder die Wahrheit allgemeiner Aussagen besitzen können, ist also logisch nicht zu begründen. Damit ist aber nur die Selbstverständlichkeit gezeigt, daß Induktion keine Deduktion ist (Kronthaler, 1984, S. 280; van Wright, 1957, S. 180): Aussagen über etwas (noch) nicht Bekanntes können per definitionem nicht mit Sicherheit aus Aussagen über Bekanntes abgeleitet werden, weil sie andernfalls keine Aussagen über Unbekanntes mehr wären.

Das schwache Induktionsproblem betrifft die Frage, ob wir trotz der negativen Beantwortung des starken Induktionsproblems zurecht der Überzeugung sind, daß es möglich und vernünftig ist, aufgrund vergangener Beobachtungen Aussagen über Nichtbeobachtetes zu treffen, auch wenn wir uns der Wahrheit dieser Aussagen nicht gewiß sein können. Dementsprechend gibt es in der Philosophie eine ganze Reihe von Autoren, die sich weiter intensiv mit dem (schwachen) Induktionsproblem auseinandersetzen und untersuchen, ob es Regeln und Prinzipien gibt, die unser tatsächliches induktives Vorgehen be-

schreiben und als rational begründen können (siehe z.B. Cohen, 1980, 1982, 1989; Cohen & Hesse, 1980; Kyburg, 1964; Kyburg & Nagel, 1963; Stegmüller, 1973). In den folgenden Abschnitten 2, 4 und 5 werden wir die wichtigsten dieser Induktionskonzeptionen näher betrachten.

Innerhalb des schwachen Induktionsproblems können wir zwei Aspekte unterscheiden, die Stegmüller (1973, S. 82-83) als das praktische und das theoretische Nachfolgeproblem des Humeschen Induktionsproblems bezeichnet:

- Welche Normen für rationale Entscheidungen unter Risiko lassen sich rechtfertigen?
- Wie ist der Begriff der Bestätigung, Stützung oder Bewährung einer Hypothese adäquat zu definieren?

Die Einführung eines adäquaten Bestätigungsbegriffs scheint auf den ersten Blick sehr einfach zu sein: Der Bestätigungsgrad einer allgemeinen Aussage der Form „ $\forall_i: F(a_i) \Rightarrow G(a_i)$ “ wird durch positive Instanzen „F & G“ erhöht und durch negative Instanzen „F &  $\sim G$ “ verringert. Dieser plausible Ansatz führt jedoch zu verschiedenen unakzeptablen Konsequenzen, von denen das Raben- und das Grue-Paradox die bekanntesten sind. Wir wollen hier die intensive Diskussion dieser Probleme in der philosophischen Literatur nicht nachzeichnen, da alle rein logischen Argumentationen nicht zu akzeptablen Lösungen geführt haben (siehe Barker & Achinstein, 1960; Cohen, 1980, 1989; Y. Cohen, 1979; Cox, 1986; Gärdenfors, 1990; Hesse, 1974; Hoppe, 1975; Kutschera, 1972; Rosenkrantz, 1982; Shirley, 1981; Stemmer, 1975). Umgehen kann man diese Paradoxien der Bestätigung durch eine pragmatische Betrachtungsweise (siehe Abschnitt 4).

## ***2. Probabilistische Explikationssysteme der Induktion***

Seit Entwicklung des Wahrscheinlichkeitsbegriffs durch B. Pascal ist es üblich, induktive Argumente probabilistisch zu interpretieren (Cohen, 1989). Zwar folgt bei ihnen die Konklusion nicht mit logischer Sicherheit aus der Prämisse, man kann aber annehmen, daß die Prämisse A die Konklusion B in irgendeinem Sinn wahrscheinlicher macht oder ihre Wahrscheinlichkeit  $r$  über eine a priori Wahrscheinlichkeit  $r'$  anhebt:

$$P[B|A] = r \text{ (mit } r \approx 1 \text{ oder } r > r')$$

Interpretiert wird P dabei üblicherweise nicht als objektive, sondern als subjektive oder personelle Wahrscheinlichkeit, die den Grad des rationalen Glaubens an das Eintreten des Ereignisses B ausdrückt (Kutschera, 1972; Stegmüller, 1973). Expliziert werden kann die subjektive Wahrscheinlichkeit für das Eintreten eines Ereignisses B dadurch, daß man untersucht, mit welchem Ein-

satz auf das Eintreten von B gewettet wird. Als rational kann man die subjektiven Überzeugungen der Person nur dann bezeichnen, wenn ein faires oder kohärentes Wettsystem vorliegt, d.h. wenn das Wettverhalten der Person nicht notwendigerweise zu Verlusten führt (Carnap, 1971, S. 105-116; Essler, 1973, S. 17; Kemeny, 1963, S. 720). Damit diese Bedingung erfüllt ist, muß - neben anderen Voraussetzungen - die subjektive Wahrscheinlichkeitsfunktion der Person die Axiome des mathematischen Wahrscheinlichkeitskalküls erfüllen (Kutschera, 1972, S. 70-88).

## 2.1 Carnaps induktive Logik

Von subjektiven Wahrscheinlichkeiten ausgehend hat sich Carnap (1950, 1952, 1963, 1971) um die Begründung eines Begriffs der induktiven Bestätigung von Hypothesen und Theorien bemüht und dazu den Begriff der induktiven oder logischen Wahrscheinlichkeit entwickelt (siehe im einzelnen Cohen, 1977, 1989; Cohen & Hesse, 1980; Kemeny, 1963; Kutschera, 1972; Stegmüller, 1973).

Nach Auffassung des frühen logischen Empirismus ist die Wissenschaft durch ein induktives Vorgehen gekennzeichnet, das darin besteht „aus ein- oder mehrmaliger Beobachtung eines gewissen Bedingungsverhältnisses auf seine allgemeine Gültigkeit zu schließen“ (Carnap, 1926, S.8). Trotz aller sonstigen Gegensätze stimmt Carnap mit Popper jedoch darin überein, daß es keine induktive Methodologie im Sinne eines algorithmischen Verfahrens geben kann, das quasi mechanisch von Beobachtungssätzen zu wahren theoretischen Aussagen führt (Carnap, 1946; Hempel, 1965, S.6; Stegmüller, 1978, S.468-470). Seine „induktive Logik“ bezieht sich daher nicht auf den Entdeckungs-, sondern auf den Rechtfertigungszusammenhang im Sinne von Reichenbach (1938).

Die von Carnap verfolgte Grundidee besteht darin, durch die Einführung zusätzlicher Rationalitätsbedingungen die verschiedenen möglichen subjektiven Wahrscheinlichkeitszuordnungen eindeutig auf eine bestimmte ausgezeichnete logische Wahrscheinlichkeit  $c$  zu reduzieren, die als partielle logische Implikation und damit als eine rein analytische Relation zwischen Hypothesen  $H$  und Sätzen über empirische Evidenzen  $E$  interpretiert werden soll (Kutschera, 1972, S. 123; Stegmüller, 1971, S. 53-56; Suppe, 1977, S. 626). So wie im deduktiven Schluß die Prämissen die Konklusion implizieren, weil die Konklusion im Gehalt der Prämissen enthalten ist, so sollen im induktiven Argument die Prämissen die Konklusion partiell implizieren, weil sich der Gehalt der Prämissen und der Konklusion teilweise überschneiden.

Carnaps Projekt, mit Hilfe logischer Wahrscheinlichkeiten den Begriff der induktiven Bestätigung zu explizieren, kann hier nicht im Detail nachgezeichnet werden. Nach allgemein akzeptierter Ansicht ist es aber mit erheblichen Schwierigkeiten verbunden (Cohen, 1977, 1980, 1989; Essler, 1980; Kronthaler, 1984; Kummer, 1982; Kutschera, 1972; Lakatos, 1968; Levi, 1979; Rivadulla, 1987; Salmon, 1982; Stegmüller, 1971, 1973):

1. Eine Bestimmung der Werte von  $c$  ist abhängig von der jeweils verwendeten Sprache, die nicht ohne willkürliche Entscheidungen festgelegt werden kann. Dabei können nur Sprachen verwendet werden, die im Vergleich zur tatsächlichen Wissenschaftssprache sehr primitiv sind. So wählt Carnap, um ein Analogon für die Ereignismenge (Sigma-Algebra, vgl. Steyer, Kapitel 15 in diesem Band) zu finden, auf dem die Wahrscheinlichkeiten zu definieren sind, zunächst eine Sprache  $L$ , die sich auf einen endlichen Gegenstandsbereich mit einer begrenzten Anzahl von Prädikaten bezieht und nur die Prädikatenlogik 1. Stufe enthält.
2. Die  $c$ -Funktion soll eine analytische Funktion sein, die auf dieser Sprache  $L$  definiert ist. Diese Analytizität der  $c$ -Funktion läßt vollkommen offen, wie eine Aussage über die Vergangenheit, nämlich  $E$ , zusammen mit einer analytischen Funktion  $c(H|E) = r$  zu Wissen über Zukünftiges, nämlich über  $H$ , führen kann (Stegmüller, 1971, S.59-60).
3. Die Werte von  $c$  sind abhängig von den A-priori-Wahrscheinlichkeiten für die Hypothesen. Da diese nicht empirisch zu gewinnen sind, müssen Annahmen wie z.B. das Indifferenzprinzip akzeptiert werden: „Wenn keine Gründe dafür bekannt sind, eher mit dem Eintreten eines von mehreren möglichen Ereignissen zu rechnen, dann sind die Ereignisse als gleichwahrscheinlich anzusehen“ (Kutschera, 1972, S. 127). Derartige Annahmen haben Carnap große theoretische Schwierigkeiten und den Vorwurf eines willkürlichen Apriorismus eingebracht.
4. Carnaps Ansatz ist praktisch nur im Fall von mehrfach replizierbaren Zufallsexperimenten anwendbar, bei denen die relativen Häufigkeiten der verschiedenen möglichen (Elementar-) Ereignisse als empirische Evidenzen zur Beurteilung der Bestätigung von Hypothesen über den Ausgang weiterer Wiederholungen dienen.
5. Um die Klasse der zulässigen  $c$ -Funktionen einzuschränken, müssen über die üblichen Eigenschaften (subjektiver) Wahrscheinlichkeiten hinaus zusätzliche Rationalitätsforderungen definiert werden, die nicht weiter gerechtfertigt werden können.
6. Auch unter diesen sehr einschränkenden Bedingungen läßt sich die Klasse der zulässigen  $c$ -Funktionen bestenfalls bis auf einen Parameter  $\lambda$  bestimmen (das Kontinuum der induktiven Methoden). Dieser kann als Vorsichtigkeitsindex interpretiert werden: Er drückt Vermutungen über die Merkmalsheterogenität bzw. Uniformität der Objekte aus, über die

generalisiert werden soll, und gibt an, inwieweit man sich eher auf die A-priori-Wahrscheinlichkeiten von Ereignissen oder mehr auf ihre beobachteten relativen Häufigkeiten verlassen will. Die spätere Version des Carnap-Ansatzes kann auch zu c-Funktionen führen, die nicht zur Lambda-Familie gehören, sondern zusätzlich von einem Parameter  $\epsilon$  zur Berücksichtigung von Analogieeinflüssen abhängen.

7. Der Bestätigungsgrad  $c(H|E)$  wird auch durch Beobachtungsaussagen erhöht, die nur Replikationen eines bereits in E enthaltenen positiven Befunds darstellen. Er formalisiert also im wesentlichen die einfache enumerative Induktion, wobei unberücksichtigt bleibt, daß Beobachtungen stärkere bestätigende Wirkung haben, wenn sie unter verschiedenartigen Umständen erfolgen und dadurch alternative Ursachenfaktoren ausschließen.
8. Alle c-Funktionen weisen allgemeinen Gesetzen und Hypothesen eine Wahrscheinlichkeit von Null zu. Dieser Sachverhalt wurde als sehr gravierendes Problem empfunden, weil er zu implizieren scheint, daß die Definition eines probabilistischen Bestätigungsbegriffes für Gesetze und Hypothesen prinzipiell nicht möglich ist (siehe unten).

Insgesamt sind Carnaps Bemühungen um eine eindeutige c-Funktion ohne Erfolg geblieben, und die induktive Hypothesenbeurteilung ist in seinem System immer noch subjektiv und in gewissem Grade willkürlich. Deshalb beschäftigt sich der von Stegmüller (1973) rekonstruierte Carnap-II-Ansatz (Carnap, 1971), bei dem die linguistische zugunsten einer modell- und maßtheoretischen Herangehensweise aufgegeben wurde, gar nicht mehr mit dem theoretischen Induktionsproblem der Bestätigung von Hypothesen und Theorien, sondern ist eine normative Theorie des induktiven Rasonierens für praktische Entscheidungen unter Unsicherheit.  $C(H|E)$  kann dabei weiterhin als „der Grad, in dem eine rationale Person an H bei gegebenem Datum E glaubt“ (Stegmüller, 1973, S. 103) interpretiert werden. Als Hypothese H können hier jetzt allerdings keine allquantifizierten Aussagen mehr auftreten, sondern nur noch Aussagen über singuläre Ereignisse, z.B. Vorhersagen (Stegmüller, 1973, S. 534). Daß keine eindeutige Bestimmung der C-Funktion möglich ist, erscheint unter der entscheidungstheoretischen Perspektive nicht unbedingt als Nachteil, denn die primäre Aufgabe einer rationalen Theorie ist die Errichtung negativer „Normativitätswälle“, und sie besteht nicht darin, den Freiheitsspielraum rationaler Personen auf Null zu reduzieren (Stegmüller, 1971, S.69).

Selbst wenn sich Werte für c oder eine andere induktive Wahrscheinlichkeits- oder Bestätigungsfunktion eindeutig genug bestimmen ließen, wäre damit noch nicht die Frage beantwortet, ob die betreffende Hypothese aufgrund der vorliegenden empirischen Evidenz akzeptiert oder verworfen werden soll (Hempel, 1981). Unglücklicherweise führen einfache Akzeptanzregeln, die mit festen Kriteriumswerten arbeiten, dazu, daß logisch inkonsistente Hypothesen

akzeptiert werden müssen (sog. Lotterierparadox, Kyburg, 1961; Cohen, 1980, S. 356). Deshalb sollten bei der Frage der Akzeptierung von Hypothesen die (epistemischen) Kosten und Nutzen der richtigen und falschen Entscheidungen systematisch mitberücksichtigt werden (Hempel, 1981; Levi, 1967). Dieser Vorschlag weist den Weg in Richtung auf eine pragmatische Behandlung des Induktionsproblems (siehe Abschnitt 4).

## 2.2 Einzelfallbestätigung

Ein bereits angesprochenes und gravierendes Problem der probabilistischen Explikationen von Induktion liegt darin, daß sie nicht sinnvoll auf generelle Hypothesen anwendbar sind, obwohl ein primäres Interesse an der Bestätigung von allgemeinen Gesetzaussagen besteht (Kutschera, 1972, S. 211-212; Lakatos, 1968; Popper, 1984): Alle adäquaten  $c$ -Funktionen führen zu Werten von Null für alle Gesetzhypothesen und Theorien, die die Form unbegrenzter Allsätze haben. Damit wären diese Aussagen alle gleich unwahrscheinlich und nicht bestätigungsfähig. Insbesondere wäre es nicht möglich, den Bestätigungs- oder Wahrscheinlichkeitsgrad  $c(H|E) = p(H|E)$  mit Hilfe des Bayes-Theorems aus der Likelihood  $p(E|H)$  der Daten und der A-priori-Wahrscheinlichkeit  $p(H)$  der Hypothese zu berechnen. Auf diese Schwierigkeiten läßt sich auf zweifache Weise reagieren.

(1) Man versucht, Carnaps Theorie so zu verbessern, daß sich eine von Null verschiedene Wahrscheinlichkeit für allgemeine Aussagen ergibt (Hintikka, 1966; Hintikka & Niiniluoto, 1976; Kuipers, 1978). Das sog.  $K$ -dimensionale System von Hintikka und Niiniluoto bezieht sich auf Hypothesen und Stichprobenevidenzen über eine potentiell unendliche Objektmenge, die in  $K$  durch prädikatenlogische Ausdrücke beschreibbare Klassen eingeteilt werden kann. Es führt zu einem Wahrscheinlichkeitsmaß  $p(H|E)$ , das allerdings nicht nur von einem, sondern von  $K$  Parametern abhängt. Als Spezialfälle enthält es sowohl das ursprüngliche zweiparametrische Alpha-Lambda-System von Hintikka (1966) wie das Lambda-Kontinuum von Carnap (1952). Es ist nicht auf rein enumerative Induktion beschränkt, sondern kann auch die Anzahl der unterschiedlichen Klassen von Beobachtungsevidenzen berücksichtigen und ist durch Einführung von Distanz- und Ähnlichkeitsmaßen zwischen den Prädikaten sogar auf analoge Induktionen anwendbar (Niiniluoto, 1980, 1981; zur Kritik siehe Kuipers, 1984; Spohn, 1981).

Diesen Vorteilen stehen aber einige gravierende Nachteile entgegen, die es weiterhin fraglich erscheinen lassen, ob man auf diesem Weg zu einer adäquaten Rekonstruktion induktivistischer Inferenzen in der Wissenschaft kommt. Insbesondere können auch die Systeme von Hintikka und Niiniluoto nur auf glücksspielähnliche Situationen angewendet werden, die in einer vorher fest-

gelegten relativ primitiven Sprache beschrieben werden, und sie können die unterschiedliche Bedeutung von bloßen Replikationen und von neuen Evidenzen unter veränderten Bedingungen für die Hypothesenbestätigung nur sehr vereinfacht widerspiegeln (Cohen, 1980, S. 361-362, 1989, S. 120-130; siehe dagegen Niiniluoto, 1980, S.226). Zudem sind die Wahrscheinlichkeitsbewertungen in diesen Systemen durch die Einführung weiterer unbestimmter Parameter noch willkürlicher als bei Carnap.

(2) Man akzeptiert, daß allgemeinen Gesetzen tatsächlich vernünftigerweise nur die Wahrscheinlichkeit Null zugeschrieben werden kann: Auf die Gültigkeit unbegrenzter Allsätze kann nicht sinnvollerweise gewettet werden, da der positive Ausgang der Wette niemals festgestellt werden kann (Stegmüller, 1971, S. 70). Wenn der Wert eines Gesetzes damit nicht in seiner hohen allgemeinen Geltungswahrscheinlichkeit liegen kann, müssen allerdings auch die Maße für die induktive Bestätigung spezieller definiert oder interpretiert werden.

Carnap (1962, S. 572-574) hat dazu die sog. Einzelfallbestätigungen („instance confirmations“) eingeführt. Nehmen wir an, eine Hypothese „ $\forall_i: F(a_i) \Rightarrow G(a_i)$ “ habe sich für eine große Zahl  $n$  von Fällen insofern bestätigt, als immer dann, wenn in einem Fall  $F(i)$  vorlag, auch  $G(i)$  beobachtet wurde. Dieser Sachverhalt werde in einem Beobachtungssatz  $E$  beschrieben. Die Einzelfallbestätigung ist dann gegeben durch

$$c_i(H|E) = p [ G(y) | F(y) \& E ],$$

wobei  $y$  ein neues, noch nicht in  $E$  vorkommendes Objekt ist (Kutschera, 1972, S.212-213; Lenzen, 1974, S.97).  $c_i$  soll also im Gegensatz zur einfachen („Bestätigungs-“)Funktion  $c(H|E)$  nicht den rationalen Glauben an die generelle Gültigkeit von  $H$  bei gegebener Erfahrung  $E$  ausdrücken, sondern nur den rationalen Glauben daran, daß bei gegebener Erfahrung  $E$  auch der nächste untersuchte Einzelfall eine Bestätigung der Hypothese erbringt.

Damit drückt die Instanzen- bzw. Einzelfallbestätigung einer Hypothese oder Theorie so etwas wie ihre Verlässlichkeit aus: Der Naturwissenschaftler mag nicht bereit sein, auf die unbeschränkte Allgemeingültigkeit einer seiner Theorien zu wetten, sie hat für ihn also einen Bestätigungsgrad von  $c = 0$ ; der Ingenieur kann trotzdem ziemlich sicher sein, daß genau diese Theorie auch bei der nächsten Anwendung sehr genaue Prognosen erlaubt, er wird ihr also einen hohen Verlässlichkeitsgrad  $c_i$  zuschreiben (Lakatos, 1982, S. 136-137).

Ein ernstes Problem des Carnapschen Konzeptes der Instanzenbestätigung besteht darin, daß die Bestimmung der Verlässlichkeit einer Hypothese unabhängig davon geschieht, wie weit die Variationsbreite der getesteten Instanzen ist. Eine Hypothese über Säugetiere kann z.B. nur durch die Testung von Vorhersagen über Katzen beliebig an Instanzenbestätigung gewinnen (Kut-

schera, 1972, S.440). Der von Carnap beschrittene Ausweg aus diesem Problem besteht in der Forderung, bei der Anwendung des Konzeptes der Instanzenbestätigung stets das gesamte gegenwärtig verfügbare Erfahrungswissen zu berücksichtigen. Dieses Prinzip der totalen Evidenz kann für die übliche tatsächliche Wissenschaftssituation aber wohl nur ein weit entferntes und unerreichbares Idealziel sein.

Das Konzept der Einzelfallbestätigung ist auch aus anderen Gründen kritisiert worden. So sieht Lakatos (1982) im Übergang von der Bestätigung allgemeiner Aussagen zur Bestätigung partikulärer Aussagen ein wesentliches Indiz für die degenerative Problemverschiebung innerhalb des Carnapschen Forschungsprogramms (zur Kritik siehe Adler, 1980). Stegmüller (1971, S. 56) sieht den Begriff der Einzelfallbestätigung vor allem wegen Poppers Einwand, daß ein längst widerlegtes Gesetz eine hohe Einzelfallbestätigung besitzen kann, als Irrweg an. Diese Divergenz ist aber lediglich eine Folge der unterschiedlichen Bedeutungen von Poppers und Carnaps Bestätigungsbegriffen und der unterschiedlich epistemischen Ansprüche, die mit ihnen verbunden sind: Gesetze wie die Newtonschen, die im Popperschen Sinn als falsifiziert gelten können und damit unter der theoretischen Zielvorstellung der Wahrheitsfindung unbrauchbar geworden sind, können dennoch zu recht eine hohe Einzelfallbestätigung besitzen, weil sie in einem weiten Bereich von Anwendungen korrekte Vorhersagen erlauben.

Trotz der angedeuteten Probleme weist Carnaps Konzept der Einzelfallbestätigung unseres Erachtens in eine Richtung, die für die Behandlung des Induktionsproblems in der Wissenschaft sehr fruchtbar sein kann. Wir werden im Abschnitt 6 sehen, daß nach den Ergebnissen strukturalistischer Analysen wesentliche Teile der mit Theorien verbundenen empirischen Behauptungen tatsächlich als induktive und einzelfallbezogene Erweiterungen des Wissens interpretiert werden können. Wenn wir die Berechtigung derartiger Vermutungen untersuchen, werden wir allerdings insofern von Carnap abweichen, als wir den Grad, in dem eine empirische Vermutung durch das bisherige Wissen gestützt wird, nicht als logisch-analytische Beziehungen zwischen Sätzen verstehen, sondern als subjektive Erwartungen von Wissenschaftlern über die erfolgreiche Anwendbarkeit ihrer Theorien, die auf Informationsverarbeitungsprozessen beruhen, die sich nicht in einer rein syntaktischen Weise beschreiben lassen.

### ***3. Deduktivistische Umgebung der Induktion***

Wie Salmon (1979) betont, kann Wissen, das allein auf enumerativer Induktion beruht, häufig bestenfalls vorwissenschaftlich genannt werden. Daß morgen die Sonne aufgeht, müssen wir beispielsweise nicht damit begründen, daß sie



bisher jeden Tag aufging, sondern wir können dies deduktiv aus wissenschaftlichen Gesetzen und Theorien ableiten, die mehr sind als einfache Generalisierungen aus Beobachtungen und die sogar in anderen Begriffen formuliert sind als die Beobachtungsaussagen.

Da es keine theoriefreien Beobachtungen gibt und da insbesondere jede Wahrnehmung einer Regelmäßigkeit oder Wiederholung bereits gewisse Abstrahierungen, Kategorisierungen und Erwartungen voraussetzt, kann es nach Auffassung von Popper (1965, 1984) überhaupt keine induktive Erfahrungsbildung geben, sondern nur ein deduktives Lernen durch Vermutungen und Widerlegungen (vgl. Campbell, 1974; zur Kritik Lane & Lane, 1984). Popper (1965, S. 64-65) hält es deshalb für obsolet, daß man nach Kriterien für zuverlässige und erfolgreiche induktive Argumente sucht, auch wenn man dabei zugesteht, daß induktives Denken nicht zu sicheren Schlußfolgerungen, sondern nur zu wahrscheinlich richtigen allgemeinen Aussagen führt.

### 3.1 Poppers deduktivistische Methodologie

Popper (1965, S. 54-55; 1974, S. 1023) sieht nur eine Möglichkeit, das Streben der Wissenschaft nach allgemeinen Gesetzen, das Grundprinzip des Empirismus und Humes Erkenntnisse über Induktion in Einklang zu bringen. Statt nach Verifikationen oder (als Ersatz) induktiven Bestätigungen zu streben, sollte man nach Popper anerkennen, daß

- alles theoretische Wissen in Vermutungen besteht,
- von empirischen Beobachtungen nicht auf die Wahrheit, sondern lediglich auf die Falschheit von Theorien geschlossen werden kann, und daß
- eine Vermutung dann besser ist als eine andere, wenn sie sich in strengen Tests besser bewährt hat.

Dementsprechend besteht eine der Basisannahmen von Poppers kritisch-rationalistischer Wissenschaftsmethodologie darin, daß nur die strenge Prüfung und versuchte Falsifikation von Theorien und Hypothesen zu Erkenntnisgewinn führen kann, und seine Forschungslogik beschäftigt sich mit den Festsetzungen und Beschlüssen, die als methodologische Regeln diese Tätigkeit leiten können (siehe Gadenne, in diesem Band, Kapitel 7 und 9).

Wenn die aus der Theorie abgeleiteten Vorhersagen auf die empirischen Beobachtungen nicht zutreffen, ist nach Popper eine Veränderung der Theorie zu fordern, durch die die Übereinstimmung zwischen Theorie und Empirie erhöht wird. Ansonsten bleibt die Genese von Theorien in Poppers Wissenschaftsphilosophie weitgehend ausgeklammert, d.h. es wird nicht untersucht, welche weiteren Faktoren die Auswahl, Ableitung, Formulierung und Modi-

fikation der später zu überprüfenden Hypothesen leiten und wie ihre Anwendungs- und Geltungsbereiche festgelegt, erweitert oder eingeschränkt werden.

Der im deduktivistischen Sinne verstandene Grad der Bewährung (corroboration,  $C_d(T|E)$ ) einer Hypothese oder Theorie T hängt nach Popper (1984, S. 211-213) von ihrer Übereinstimmung mit anerkannten Beobachtungssätzen E ab, vor allem aber von der Strenge der Prüfungen, denen die Hypothese unterworfen worden ist. Voraussetzung für einen positiven Bewährungsgrad ist also, daß zumindest ein Teil der theoriekonformen Beobachtungssätze als Resultate ernstzunehmender Widerlegungsversuche anerkannt sind. Damit vermeidet Popper unerwünschte Eigenschaften und Konsequenzen, die mit den vor allem auf Hempel (1943, 1945) zurückgehenden deduktiven Bestätigungsbegriffen verbunden sind (siehe Goodman, 1973; Lenzen, 1974).

### 3.2 Induktive Anteile in der deduktivistischen Methodologie

Wenn Popper die Induktivisten kritisiert, konstruiert er als gegnerische Position einen sehr starken Induktivismus, wie er seit langer Zeit nicht mehr vertreten wird. Seine Kritik richtet sich vor allem gegen drei Thesen, die einen solchen starken Induktivismus kennzeichnen:

1. Es gibt ein algorithmisches Verfahren zur Ableitung allgemeiner Aussagen aus Einzelaussagen.
2. Allgemeine Aussagen können durch Einzelaussagen zumindest teilweise verifiziert werden.
3. Diese partielle Verifikation oder induktive Bestätigung läßt sich als Wahrscheinlichkeit ausdrücken.

Trotz dieser Kritik basiert Poppers eigene Methodologie jedoch keineswegs allein auf deduktiven Schlüssen, vielmehr können in ihr auch gewichtige induktive Anteile identifiziert werden (Cohen, 1989; Hesse, 1974, S. 95; Hübner, 1980, S.275-276; Jones & Perry, 1982; Levinson 1974; Mackie, 1979; Niiniluoto & Tuomela, 1973, S. 203; Putnam, 1974, S. 222-223; Salmon, 1968, 1978; Will, 1982). Sie finden sich vor allem in der Suche nach allgemeingültigen Gesetzen und in den Konventionen, die zur Prüfung von Theorien notwendig sind.

(1) Wenn man wissenschaftliche Hypothesen, Gesetze oder Theorien, die laut Popper unbeschränkte Allsätze sind, zum Gegenstand strenger Prüfungen und deduktiver Bewährungen macht, trifft man implizit auch Annahmen über die Uniformität der Natur, weil man von der Voraussetzung ausgeht, daß die Natur so beschaffen ist, daß sie sich als prinzipiell gesetzmäßiger Ablauf beschreiben läßt. Beispielsweise kann durch Beobachtung eines blauen Schwans natürlich die Hypothese falsifiziert werden, daß alle Schwäne weiß sind. über

die in der Regel viel interessantere Hypothese, daß alle später zu beobachtenden Schwäne weiß sind, ist damit aber ohne Uniformitätsannahme noch gar nichts ausgesagt. Die Annahme, daß eine falsifizierte Hypothese auch unter künftigen vergleichbaren Umständen nicht gilt, ist eine induktive Annahme, aber ohne diese Annahme über die Uniformität der Welt gibt es keinen Grund, sich bei wissenschaftlichen Vorhersagen nicht weiterhin auf bereits falsifizierte Hypothesen zu verlassen.

Nicht nur die Verwendung des Gesetzesbegriffes, sondern auch die des Begriffes der strengen Prüfung hat solche Uniformitätsvoraussetzungen. Denn eine Hypothese wird dann streng geprüft, wenn die von ihr in einer konkreten Situation gemachten Vorhersagen unter der Berücksichtigung anderer und bisher gut bewährter Gesetze sehr unwahrscheinlich sind. Man muß also annehmen, daß die Erwartungen aufgrund des bisherigen Hintergrundwissens berechtigt sind. Berechtigte Erwartungen kann es aber nur in einer uniformen Welt geben, die Extrapolationen erlaubt.

Ebenso ist der Bewährungsgrad, der nach Popper nur ein zusammenfassender Bericht über vergangene (strenge) Prüfungen ist, ziemlich wertlos, wenn er nicht zu Annahmen über die Gültigkeit der Hypothese in noch nicht beobachtbaren Fällen führt. Derartige Annahmen sind ohne Zweifel induktiv, ohne sie gibt es aber keinen Grund, Vorhersagen und praktische Entscheidungen auf der Theorie basieren zu lassen, die den höchsten Bewährungsgrad hat.

(2) Auch bei der konkreten Prüfung einer Theorie muß Popper induktive Voraussetzungen machen. Zwar ist die Falsifikation einer Theorie  $T$  aufgrund einer empirischen Beobachtung  $B$  bei grober Betrachtung eine deduktiv-logische Argumentation, die Prämissen, die in diese Argumentation eingehen, sind aber häufig Konklusionen induktiver Argumente. Eine Aussage  $\sim B$ , die das Eintreten des Ereignisses  $B$  „verbietet“, kann aus einer wissenschaftlichen Theorie  $T$  in aller Regel nur abgeleitet werden, wenn bestimmte Zusatzannahmen  $Z$  zu Hilfe genommen werden. Sie betreffen die Operationalisierung der vorkommenden theoretischen Begriffe, das Vorliegen der notwendigen Anfangsbedingungen für die Anwendung der Theorie, die Ceteris-paribus-Bedingungen, die Abwesenheit von Störfaktoren usw. (vgl. Hager & Westermann, 1983). Das Vertrauen darin, daß diese Annahmen im gegebenen Fall erfüllt sind, ist zumindest dann, wenn die Annahmen nicht eigens überprüft werden, ein Ergebnis induktiver Erwartungsbildung. Die Vermutung, daß eine Theorie durch bestimmte Beobachtungen falsifiziert wurde, ist nur so berechtigt wie die Vermutung, daß die verwendeten Hintergrundannahmen zutreffend waren. Wird diese Vermutung über zutreffende Hintergrundannahmen nicht als gut begründet betrachtet, so kann bei einer erwartungswidrigen Beobachtung  $B$  deshalb nur geschlossen werden, daß  $T$  oder  $Z$  falsch ist. Dies illustriert die sog. Duhem-Quine-These, daß Hypothesen nicht isoliert, son-

dern nur im Zusammenhang mit einem umfassenderen Korpus des Wissens geprüft werden können (Balestra, 1979; zur Kritik siehe Grünbaum, 1963). Damit die erwartungswidrige Beobachtung B eindeutig der Theorie zur Last gelegt werden kann, empfiehlt Popper, die Zusatzannahmen per Konvention nicht in Zweifel zu ziehen. Die Hypothesen der Menge Z können und sollen zwar empirisch gut bewährt sein, ihre empirische Überprüfung kann aber in der Regel nur in anderen Untersuchungen erfolgen. Nimmt man an, daß sie auch im vorliegenden Fall gelten, so ist dies das Ergebnis von gehaltserweiterndem induktivem Rasonieren.

Poppers Position wird damit in sich widersprüchlich. Einerseits schreibt er der Wissenschaft durchaus die Aufgabe zu, die theoretischen Grundlagen für Prognosen zu schaffen, der Glaube an derartige Prognosen besitzt für ihn aber keinerlei Begründung (Popper, 1973, S. 384). Andererseits ist nach Popper der Glaube an Prognosen über das Zutreffen von Hintergrundannahmen bei Theorienprüfungen notwendig, um die Falsifikation zu einer begründeten Entscheidung zu machen.

Aus der Abhängigkeit der Falsifikation von durchaus willkürlichen Konventionen folgern Jones und Perry (1982), daß Popper auch induktive Argumente methodisch zulassen müßte. Wenn er nämlich induktive Argumente aufgrund ihrer Inkonklusivität a priori ablehnt, müßte er aus den gleichen Gründen auch die Falsifikation ablehnen; und wenn er Falsifikation mit methodologischen Überlegungen rechtfertigt, d.h. mit guten, pragmatischen oder konventionalistischen, aber nicht logisch zwingenden Gründen, muß er gute Gründe auch zur Begründung induktiven Vorgehens zulassen.

### 3.3 Deduktive und induktive Bestätigung

Insbesondere Popper (1981) und Popper und Miller (1983, 1987) haben zu zeigen versucht, daß ein probabilistischer induktiver Bestätigungsbegriff nicht sinnvoll ist. Ihre Argumente sind zwar formal korrekt, beruhen aber auf einer inadäquaten Explikation des empirischen Gehaltes, den eine Hypothese über die Beobachtungsevidenz hinaus hat (Gaifman, 1985; Mura, 1990). Poppers Argumentationen gegen einen induktiven und für einen rein deduktiven Bestätigungsbegriff sind vor allem deshalb nicht plausibel, weil er die induktiven Bestätigungsbegriffe generell an Kriterien mißt, die diese gar nicht erfüllen können, die vielmehr nur innerhalb seines eigenen Ansatzes notwendige Adäquatheitsbedingungen darstellen. So stellt es Popper (1965, S. 283) beispielsweise als Nachteil heraus, daß die Carnapsche Instanzenbestätigung für eine Theorie, die in jedem n-ten Versuch falsifiziert wurde, sich nicht dem Wert 0, sondern dem Wert  $1-1/n$  nähert. Dabei überträgt er natürlich Adäquatheitsbedingungen, die für seinen deduktiven Bewährungsbegriff sinnvoll sind,

nämlich daß eine falsifizierte Theorie einen Bewährungsgrad von 0 erhalten sollte, unhinterfragt auf den Bestätigungsbegriff Carnaps. Dies weist darauf hin, daß Popper davon ausgeht, daß er und Carnap bei ihren Bemühungen um Präzisierung des Bestätigungsgrades einer Hypothese annähernd das gleiche Explikandum haben, sich also auf den gleichen „Gegenstand“ beziehen (siehe Popper, 1984, S.345).

Besser begründet ist u. E. die Auffassung, daß Carnaps und Poppers Bestätigungs- bzw. Bewährungskonzepte sich auf ganz verschiedene Problembereiche beziehen. Poppers scheinbar rein logische Konzeption bezieht sich auf den pragmatischen Kontext einer Person, die ausschließlich theoretisch an deduktiven (und damit sicheren) Konsequenzen ihrer Überzeugungen interessiert ist, die alle nicht deduktiv ableitbaren Aussagen ignorieren kann und die Überlegungen darüber anstellt, wie gut sich eine Theorie bei strenger Prüfung bewährt hat und wie sie näher an die Wahrheit zu bringen ist.

Carnaps Ziel hingegen ist (zumindest in seiner späteren Theorieversion) eine normative Theorie für rationale Entscheidungen unter Risiko (Stegmüller, 1973, S. 78-90, 531-540). Er bezieht sich damit auf eine Person, die eine Theorie als Basis für epistemische oder praktische Handlungen benutzt, obwohl sie weiß, daß ihre Überlegungen nicht die Sicherheit einer deduktiven Argumentation haben,

Deshalb sind die verschiedenen Bestätigungsbegriffe „nicht verschiedene Explikationsversuche eines eindeutig bestimmbar vorwissenschaftlichen Begriffs der Bestätigung, sondern sie stellen mögliche Präzisierungen von verschiedenen plausiblen Ansichten über Bestätigung dar“ (Lenzen, 1974, S. 94): Für jede dieser zugrundeliegenden Ansichten gibt es spezielle und plausible Adäquatheitsbedingungen, die zwar vom jeweiligen Explikat erfüllt werden, gegen die andere Explikate aber zumindest teilweise verstoßen.

Beim Vergleich beider Klassen von Bestätigungsbegriffen kommt Lenzen (1974) zu dem Schluß, daß „die Theorie deduktiver Bestätigung zwar im Bereich deterministischer Hypothesen ein adäquates Modell wissenschaftlicher Bestätigung darstellt, daß sie aber dennoch um induktive Überlegungen ergänzt werden muß, wenn sie eine umfassende und der wissenschaftlichen Praxis gerecht werdende Rekonstruktion dieses Begriffes liefern will“ (a.a.O., S.208). Notwendig sind solche ergänzenden induktiven Überlegungen insbesondere bei probabilistischen Hypothesen und bei deterministischen Hypothesen, die sich nicht in allen Prüfungen bewährt haben. In diesen beiden wissenschaftlichen Standardsituationen kann trotz fehlender deduktiver Bewährung die induktive Bestätigung im Sinne der Instanzenbewährung relativ hoch sein. Für alle Theorien, deren empirische Konsequenzen teilweise nicht bestätigt werden konnten, kann der deduktive Bewährungsgrad im Sinne Poppers konsequenterweise nur gleich Null gesetzt werden. Wenn keine voll be-

währten Theorien zur Verfügung stehen, wenn es aber trotzdem Veranlassung gibt, aus einer Menge von derartigen teilweise bewährten und (noch) nicht falsifizierten Theorien die beste Prognose- oder Handlungsgrundlage auszuwählen, ist es durchaus vernünftig, diese Auswahl nach anderen Kriterien vorzunehmen (Kutschera, 1972, S.460-466; Salmon, 1979). Zu denken ist dabei insbesondere an die induktive Instanzenbestätigung im Sinne von Carnap.

Wenn man die Poppersche Methodologie von induktiven Anteilen freihalten will (Miller, 1980), muß man sich auf ein sehr enges Bild von Funktion und Wert wissenschaftlicher Theorien zurückziehen. Insbesondere muß man alle Ereignisvorhersagen und Akzeptanzentscheidungen als gleichermaßen unbegründet und unbegründbar bezeichnen und sie deshalb aus dem Aufgabenbereich der Wissenschaft ausschließen. Dies läuft aber eindeutig dem tatsächlichen und gewünschten Gebrauch wissenschaftlicher Aussagen zuwider. Hypothesen und Theorien, die (noch) nicht falsifiziert und durch bessere ersetzt sind, sind nicht nur Objekte eines intellektuellen Spiels, die es wert sind, weiteren strengen Prüfungen ausgesetzt zu werden. Wissenschaftler und Anwender von Wissenschaft nehmen vielmehr in der Regel an, daß ihre Theorien auch in künftigen und hinreichend ähnlichen Fällen gelten werden, und sie bauen ihre Überzeugungen und Handlungen auf derartigen (natürlich veränderbaren) Annahmen auf. Alle Autoren, die das (schwache) Induktionsproblem behandeln, haben das Ziel, diese Haltung als vernünftig zu begründen. Eine ausschließlich deduktive Methodologie löst dieses Induktionsproblem nicht, sondern ignoriert es einfach durch Ausgrenzung. Man kann die Poppersche Methodologie so interpretieren, daß sie aufzuzeigen versucht, wieviel (bzw. wie wenig) sich ohne Bezugnahmen auf Induktion erreichen läßt. Uns scheint es jedoch nicht befriedigend, sich mit dem so Erreichbaren zu begnügen.

#### ***4. Pragmatische Rechtfertigungen der Induktion***

Als pragmatisch wollen wir die Ansätze zur Behandlung des Induktionsproblems bezeichnen, die nicht nach globalen, rein logisch-syntaktischen und kontextunabhängigen Regeln für korrekte Induktion streben, sondern die den jeweiligen Inhalt und Verwendungszusammenhang von induktiven Argumenten beachten und ihre Nützlichkeit im Vergleich mit anderen Vorgehensweisen betonen (Salmon, 1979).

Pragmatisch orientierte Induktionsansätze haben typischerweise keinen hohen Anspruch an die Begründbarkeit von Induktion. So wird betont, daß induktive und deduktive Logik im gleichen Sinne nicht zu begründen sind, weil sie bei der Begründung ihrer Regeln sich selbst bereits voraussetzen (Amico, 1986; Goodman, 1973, S. 62). Wenn wir trotzdem allgemein davon ausgehen,

daß zwar die induktive, aber nicht die deduktive Logik einer Rechtfertigung bedürfe, dann deshalb, weil das deduktive System in sich konsistent ist und eindeutige Festlegungen umfaßt, was eine korrekte deduktive Ableitung ist, während starke Zweifel bestehen, wie und ob induktive Regeln wohlformuliert werden können (Hesse, 1974, S. 92-93; Jones, 1982).

#### 4.1 Reichenbachs pragmatische Argumentation

Die vermutlich bekannteste pragmatische Rechtfertigung der Induktion geht auf Reichenbach (1935, 1938) zurück (vgl. de Olivera, 1985; Salmon, 1968, 1979, 1982). Nach seiner bereits erwähnten Induktionsregel (siehe S. 431) wird, vereinfacht gesagt, die bisherige relative Häufigkeit eines Ereignisses als beste Schätzung für seine Wahrscheinlichkeit angenommen. Reichenbach akzeptiert Humes Argument, daß es unmöglich ist, logisch zu begründen, daß irgendein induktives Argument jemals eine wahre Konklusion haben wird. Dennoch glaubt er, daß die Induktionsmethode zu rechtfertigen ist: Auch wenn ihr Erfolg als Methode der Vorhersage nicht im Vorhinein begründet werden kann, so kann man doch zeigen, daß sie jeder anderen Methode der Vorhersage zumindest gleichwertig ist (Salmon, 1979, S.52). „Die Induktionsregel ist die günstigste Setzung, weil sie die einzige Setzung ist, von der wir wissen: wenn es überhaupt möglich ist, Zukunftsaussagen zu machen, so werden wir sie durch diese Setzung finden... Freilich ist der naive Wunsch nicht erfüllbar, daß sich der Induktionsregel ein Garantieschein für sicheren Erfolg beifügen läßt ...“ (Reichenbach, 1935, S. 418).

Reichenbachs Begründung geht von zwei möglichen Fällen aus:

- 4 Die Natur könnte in einer Weise uniform sein, die es erlaubt, erfolgreiche induktive Vorhersagen zu machen. In diesem Fall wird die induktive Methode funktionieren, denn das Wesentliche an ihr besteht darin anzunehmen, daß die Zukunft der Vergangenheit ähnelt. Eine beliebige andere Methode X könnte genauso gut funktionieren, muß es aber nicht.
- b) Die Natur könnte nicht in der angegebenen Weise uniform sein. Dann wird die induktive Methode nicht funktionieren. Es kann aber auch keine alternative Methode X funktionieren, denn wenn X konsistent funktionieren würde, also zutreffende Vorhersagen machen würde, dann würde darin eine wichtige Uniformität bestehen, die durch induktive Methoden genutzt werden könnte.

Reichenbach greift zurecht nicht den seit Kant geläufigen Vorschlag auf, induktive Argumente durch Einfügen eines Uniformitätsprinzips in deduktive Schlüsse umzuwandeln. Das Problem der Begründung von Induktion wurde sich damit nur auf die Begründung des Uniformitätsprinzips verschieben, und dies scheint weder induktiv noch deduktiv möglich zu sein (Edidin, 1984;

Kronthaler, 1984; Salmon, 1979; Stegmüller 1971, S. 17-18). Selbst wenn das Prinzip als a priori wahr angesehen werden könnte, wurden damit konkrete induktive Argumente nur dann gerechtfertigt sein, wenn es eine absolute Uniformität ausdrücken würde. Dies ist aber höchst unplausibel, denn die Zukunft ist vielleicht in gewisser Hinsicht der Vergangenheit ähnlich, aber nicht in allen Aspekten, und die Schwierigkeit bestünde darin, herauszufinden, in welcher Hinsicht die Natur uniform wäre (Salmon, 1966, S. 42). Reichenbachs Vorgehen kann eher so beschrieben werden, daß er eine Uniformität annimmt, weil dies die einzige Möglichkeit ist, die uns dem Ziel der Ableitung begründeter Zukunftserwartungen näher bringen kann. Dabei bleibt aber das Problem bestehen, welche der verschiedenen möglichen Uniformitätssetzungen zu bevorzugen ist.

## 4.2 Anpassung von Prinzipien und Praxis

Pragmatische Überlegungen zur Rechtfertigung der Induktion stammen auch von Strawson (1952) und anderen Vertretern der Philosophie der normalen Sprache, die im Anschluß an Wittgenstein und im Unterschied zur analytischen Philosophie Begriffsklärung nicht durch die logische Präzisierungen erreichen will, sondern durch die Analyse der Gebrauchsregeln von Begriffen in der Alltagssprache (vgl. Schlesinger, 1981, S. 199-200). Zusammenfassend gelangt Strawson zu der Auffassung, daß die Ableitung von Erwartungen aufgrund induktiver Argumente rational ist, wenn sie den per Konvention akzeptierten Regeln des normalen induktiven Schließens gehorchen: Leitet man aus der Beobachtung einer großen Anzahl schwarzer Raben ab, daß alle Raben schwarz sind, ist das ein rationaler Schluß, für den eine weitere Rechtfertigung durch grundsätzlichere Prinzipien weder notwendig noch möglich ist. Selbst wenn der nächste beobachtete Rabe nicht schwarz sein sollte, kann es immer noch rational gewesen sein, unter den gegebenen Umständen erwartet zu haben, daß er schwarz sein würde. Wie Salmon (1978) kritisch anmerkt, ist das Induktionsproblem damit aber keineswegs endgültig umgangen. Zum einen könnte man mit dieser Argumentation auch verschiedene „pathologische“ Vorgehensweisen als rational rechtfertigen, zum anderen ist Induktion zwar tatsächlich kaum durch tieferliegende Prinzipien zu rechtfertigen (sog. validation), wohl aber dadurch, daß sie der Erreichung bestimmter Ziele dient (sog. vindication).

Deshalb ist es wohl angemessener, entsprechend den Vorschlägen von Goodman (1973) allgemeine induktive Regeln und spezielle induktive Inferenzen gemeinsam zu rechtfertigen, indem man sie sukzessiv in eine möglichst große Übereinstimmung (Kohärenz) bringt. Dazu ist zu zeigen, daß diese induktiven Argumente mit generellen Regeln des Argumentierens übereinstimmen und



daß diese Regeln mit tatsächlich verwendeten und allgemein akzeptierten Argumenten übereinstimmen.

Dabei sollten wir uns nach Goodman primär der Frage zuwenden, wie normalerweise akzeptierte induktive Argumente konstruiert sind, und versuchen, die Differenz zwischen akzeptierten und nicht akzeptierten Vorhersagen zu definieren. Eine generelle Regel soll dann aufgegeben werden, wenn sie eine Inferenz erlaubt, die wir nicht akzeptieren wollen, und eine Inferenz wird zurückgewiesen, wenn sie eine Regel verletzt, die wir nicht aufgeben wollen.

Diese beständige wechselseitige Anpassung von Prinzipien und Praxis, bei der Beschreibung und Begründung Hand in Hand gehen, ist eine grundlegende Methode für pragmatische Lösungen des (schwachen) Induktionsproblems. Sie ist zwar rekonstruktiv ausgerichtet, erlaubt aber auch normative Aussagen darüber, wie Personen induktiv argumentieren sollten, zumindest dann, wenn sie Erklärungen dafür anbieten kann, warum tatsächlich stattfindende Überlegungen nicht zu optimalen Ergebnissen führen (Thagard, 1988, 113-115; Thagard & Nisbett, 1982, S.392).

Insgesamt betrachtet man aus pragmatischer Sicht induktive Argumente nicht als (quasi)logische Begründungen von Wissen, sondern als Entscheidungen unter Unsicherheit und als (möglicherweise) regelgeleitete Ableitungen von (revidierbaren) Erwartungen und Hypothesen.

Auch von der Induktion kann man nicht verlangen, unerreichbares Wissen zu erreichen, z.B. darüber, ob eine bestimmte Vorhersage korrekt sein wird oder nicht. Die Frage, ob ein Ereignis E eintreten wird oder nicht, ist eindeutig faktischer Natur und kann niemals allein durch (logische) Begriffsklärungen und -beziehungen beantwortet werden. Induktives Denken kann nicht auf die Erlangung von Zukunftswissen abzielen, sondern nur auf eine rationale Organisation unserer Zukunftserwartungen (Kutschera, 1972, S. 215).

### 4.3 Einbeziehung von Hintergrundwissen

Im Unterschied zu deduktiven Schlüssen kann man nach pragmatischer Auffassung die Adäquatheit eines induktiven Schlusses nicht allein aufgrund syntaktischer Kriterien beurteilen, wie sie das Programm einer formalen induktiven Logik erbringen sollte (Thagard & Nisbett, 1982). Sie hängt vielmehr ganz wesentlich vom Inhalt der Prämissen und Konklusionen und vom Verwendungszusammenhang ab. Konkret wird die Berechtigung einer induktiven Wissenserweiterung nicht nur von der gesamten verfügbaren empirischen Evidenz bestimmt, sondern auch von der vorliegenden Situation, der untersuchten Fragestellung, dem aktivierten Hintergrundwissen, den Zielen, die die Person aktuell verfolgt, von dem erwarteten Wissenszuwachs und seinem mög-

lichen Nutzen, von den drohenden Fehlannahmen und ihren schädlichen Konsequenzen usf. (Holland, Holyoak, Nisbett & Thagard, 1986, S.5; Levi, 1979, S.340, 1982, S. 186-189).

Für die Induktion besonders bedeutsame Aspekte des Hintergrundwissens sind Informationen über die Merkmalsvariabilität (Thagard & Nisbett, 1982): Inwieweit verschiedene Beobachtungen eine induktive Verallgemeinerung rechtfertigen oder (umgekehrt betrachtet) inwieweit eine allgemeine Aussage durch einzelne Beobachtungen gestützt werden kann, hängt zwar stets auch von der Anzahl der positiven Instanzen ab. Hauptsächlich hängt der Bestätigungsgrad aber vom Hintergrundwissen darüber ab, wie stark sich die Arten von Objekten, über die eine allgemeine Aussage getroffen werden soll, hinsichtlich der betrachteten Art von Merkmalen zu unterscheiden pflegen.

Induktiven Argumenten liegt also nicht eine generelle Annahme über die globale Uniformität der Natur zugrunde, sondern sie basieren auf differenzierten Variabilitäts- und Konsistenzverfahren. Dadurch ist auch zu erklären, weshalb alle Versuche der Philosophie, Induktion durch allgemeine, quasi-logische Argumente zu rechtfertigen, letztendlich ohne durchschlagenden Erfolg geblieben sind: Ob ein induktives Argument als adäquat oder rational bezeichnet werden kann, hängt nicht primär von generellen, logischen und kontextfreien Kriterien ab, sondern vom konkreten Gegenstandsbereich und unserem spezifischen Wissen.

Die Größe dieser Variabilität hängt davon ab, auf welche Referenzklasse die betrachteten Objekte bezogen werden. Je weiter sie definiert wird, desto größer wird die Merkmalsvariabilität und desto schwächer die Rechtfertigung der Induktion. Auf der anderen Seite erlauben sehr spezifische Bezugsmengen zwar eine problemlosere Induktion, diese beziehen sich aber eben auch nur auf einen relativ kleinen Objektbereich. Wie allgemein oder speziell eine Referenzklasse zu wählen ist, kann nicht generell festgelegt werden, sondern hängt von den jeweils verfolgten Zielen, insbesondere vom intendierten Verallgemeinerungsbereich ab.

Betrachtet man die Frage der Adäquatheit induktiver Argumente pragmatisch, d.h. berücksichtigt man dabei insbesondere Ausgangssituation und Ziele, verlieren die von Philosophen sehr intensiv diskutierten Paradoxien der Bestätigung (siehe Seite 433) ihre Relevanz. Sie zeigen nur, daß rein syntaktische Betrachtungen des Induktionsproblems unzureichend sind, weil es ihnen aufgrund der Beschränkung auf logische Äquivalenzen nicht gelingen kann, zwischen unterschiedlich verallgemeinerungsfähigen und bestätigungsrelevanten Begriffen und Situationen zu unterscheiden (Cohen, 1989; Y. Cohen, 1979; Gärdenfors, 1990; Holland et al., 1986, S. 7, 234-235; Stemmer, 1975; Thagard & Nisbett, 1982; Will, 1982).

Grundsätzlich können nur solche Objekt- und Ereignisarten als Referenzklassen gewählt werden, deren intensionale oder extensionale Abgrenzung uns ausreichend klar ist und über die wir bereits hinreichend Wissen erworben haben. Spätestens an dieser Stelle wird deutlich, daß pragmatische Induktionskonzeptionen wie die von Thagard und Nisbett explizit auf kognitionspsychologischen Vorstellungen aufbauen. Insbesondere beruhen sie auf der menschlichen Fähigkeit, Objekte in Kategorien einzuteilen, ihnen Eigenschaften und gesetzmäßige Verhaltensdispositionen zuzuschreiben, diese Konzepte in semantischen Netzwerken zu hierarchisieren usw. Wir werden auf elaborierte psychologische Induktionskonzeptionen im Abschnitt 5 näher zu sprechen kommen.

Wichtig für die Berechtigung induktiver Erweiterungen sind aber nicht nur Informationen über bisherige Merkmalskonstanzen, -variabilitäten und -zusammenhänge, die relativ direkt empirisch beobachtbar sind, sondern vor allem auch Kenntnisse oder Annahmen über die Gleichheit oder Analogie der zugrundeliegenden Faktoren und Prozesse. Wie Hesse (1974) betont, ist die Güte der Vorhersagen, die aus einer Theorie abgeleitet werden, d.h. die Wahrscheinlichkeit ihrer Bestätigung, vom Ausmaß der Analogie zwischen den bereits vorliegenden theoriekonformen Beobachtungen und den noch nicht untersuchten Anwendungsfällen abhängig. Beispielsweise konnte Newton so verschiedene Phänomene wie Planetenbahnen und Gezeitenwechsel aus seiner Theorie vorhersagen, weil er zurecht davon überzeugt war, daß die beteiligte Materie sich in allen Anwendungen im wesentlichen analog zu massiven Körpern verhält (Hesse, 1980, S. 215). Newtons Theorienentwicklung begann zwar mit enumerativer Induktion, konnte aber nur durch eine ständige Kombination von Mustererkennung, analogem Denken und induktiver Inferenz erfolgreich fortschreiten (Gaifmann, 1985). Der damit angesprochene Faktor der Ähnlichkeit von bisherigen und zukünftigen Anwendungsfällen spielt bei der strukturalistischen Analyse wissenschaftlicher Prozesse eine entscheidende Rolle (siehe Abschnitt 6).

## ***5. Psychologische Analysen der Induktion***

In diesem Abschnitt wollen wir unsere primär analytisch-philosophischen Überlegungen zur Induktion durch einige Blicke auf empirisch ausgerichtete Induktionskonzeptionen erweitern. Der Induktionsbegriff wird in Psychologie und Kognitionswissenschaft allerdings wesentlich breiter verstanden als nach unserer Charakterisierung zu Anfang dieser Arbeit. Er bezeichnet nicht-deduktive Inferenzen aller Art, beispielsweise Konzeptlernen, Kategorienbildung, Regelidentifikation, Kovariationsentdeckung, Konditionierung, Bildung mentaler Modelle und analoges Problemlösen. Allerdings beziehen sich die einschlägigen

Untersuchungsparadigmata auf recht einfache Gegenstandsbereiche und sind zur Analyse wissenschaftlicher Prozesse nur geeignet, wenn sie auf komplexere Arten von Regeln, Hypothesen und Rückmeldungen ausgedehnt werden (Klahr & Dunbar, 1988).

Schon Wertheimer (1959) hat aus seiner gestaltpsychologischen Sicht wissenschaftliche Problemstellungen als Suchprozesse durch eine Menge von Wissenszuständen beschrieben. Später hat insbesondere Simon (1966, 1973; Newell, Shaw & Simon, 1958) dafür plädiert, sich bei der Analyse wissenschaftlicher Prozesse auch dem Entdeckungszusammenhang zuzuwenden und die Suche nach wissenschaftlichen Regel- oder Gesetzmäßigkeiten als spezielle Problemlösungsprozesse zu betrachten und mit Hilfe von Computern zu simulieren.

Speziell für die Beschreibung und Simulation wissenschaftlicher Problemlösungsprozesse sind eine Reihe von Konzeptionen und Computerprogrammen vorgestellt worden. So simuliert (um nur zwei paradigmatische Beispiele **ZU** nennen) BACON (Langley, Simon, Bradshaw & Zytkow, 1987) erfolgreich die Ableitung von Konzepten, Hypothesen und Gesetzen aus empirischen Daten, während KEKADA (Kulkarni & Simon, 1988) auch die Planung und Durchführung von Experimenten, die Ableitung von Bewährungsmaßen für die überprüften Hypothesen und die Modifikation von Theorien beschreibt.

Eine interessante Perspektive dürfte durch die Simulation induktiven Denkens mit Hilfe von konnektionistischen oder PDP-Modellen (McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986) entstehen, die im Unterschied zu den vorliegenden Produktionssystemen mit subsymbolischen Einheiten und paralleler Verarbeitung operieren. Insbesondere kann dadurch ohne Schwierigkeiten modelliert werden, wie implizite Regeln allein aus dem Umgang mit Beispielen entstehen. Es scheint aber noch unklar zu sein, wie explizite Regelabstraktionen modelliert werden können (Strube, 1990, S. 139).

## 5.1 Eine allgemeine Induktionskonzeption

Holland, Holyoak, Nisbett und Thagard (1986; vgl. Holyoak & Nisbett, 1988) haben den bisher wohl ambitioniertesten theoretischen Rahmen zur Erklärung und Beschreibung von Induktion entwickelt. Er umfaßt verschiedene bekannte Konzeptionen und Ergebnisse zur Psychologie der Informationsverarbeitung (siehe Opwis & Spada, Kapitel 5 in diesem Band). Konkretisierungen findet diese Theorie in den beiden Programmsystemen CS (classifier system) und PI (process **of** induction, vgl. Thagard, 1988), mit denen unterschiedliche Arten induktiven Verhaltens kognitiver Systeme simuliert werden. Sie reichen von der klassischen Konditionierung bei Tieren (Holyoak, Koh & Nisbett,

1989), über Begriffsbildungen und soziale Beurteilungen bis zum theorieentwickelnden Wissenschaftler. Allerdings konnten bisher noch nicht alle betrachteten Phänomene adäquat modelliert werden, und viele postulierte theoretische Prinzipien sind noch nicht ausreichend empirisch geprüft worden (Holland et al, 1986, S. 351). Eine Konkretisierung dieser Rahmenkonzeption, die bisher nur auf relativ allgemeinem Niveau formuliert ist, für alle von den Autoren ins Auge gefaßten Anwendungen, wird sicherlich noch erhebliche Anstrengungen kosten, verspricht jedoch andererseits interessante Einsichten in induktive Denkprozesse. Die für uns wichtigsten Grundideen dieser Induktionskonzeption sollen im folgenden kurz nachgezeichnet werden.

Die Hauptaufgabe des kognitiven Systems kann darin gesehen werden, durch adaptive Anpassung Unsicherheit über die Umwelt zu reduzieren. Dies erfolgt nach Auffassung der Autoren dadurch, daß kognitive Systeme quasi-homomorphe mentale Modelle der Umwelt konstruieren und diese aufgrund von Umweltruckmeldungen induktiv verändern. Wesentliche Bausteine dieser mentalen Modelle sind Produktionen („Wenn . . . dann ...“-Regeln. Anderson, 1983; Newell & Simon, 1972). Ihre Aktionskomponenten können Veränderungen sowohl der Umwelt wie der individuellen Wissensbasis herbeiführen, und ihre Hauptfunktion besteht darin, Erwartungen, Vorhersagen und Hypothesen zu erzeugen.

Induktion ist aus der Sicht der Autoren eine problembezogene Erweiterung und Veränderung der Wissensstruktur durch neue Hypothesen und Regeln. Sie wird in Abhängigkeit von der aktuellen individuellen Zielhierarchie durch problematische innere oder äußere Ereignisse ausgelöst und umfaßt vor allem zwei schwierige, miteinander verknüpfte Arten von Aufgaben:

- die Bewertung, Verbesserung und Auswahl bestehender Regeln und
- die Erzeugung von neuen Regeln.

Um die Bewertung des vergangenen Erfolgs von Regeln beschreiben zu können, wird in Produktionssystemen für jede Regel ein Stärkewert angenommen, der sich nach erfolgreichen Anwendungen erhöht und nach falschen Vorhersagen vermindern kann.

Die Generierung neuer und erfolgsversprechender Regeln wäre viel zu ineffizient, wenn sie zufällig erfolgen würde, und scheint deshalb durch eine Reihe von Heuristiken gesteuert zu sein (Holyoak & Nisbett, 1988). Insbesondere bestimmen auslösende Bedingungen, Urteilsschemata und die kombinierbaren vorhandenen Regelemente die Ergebnisse. Damit werden im Gegensatz zu Poppers Auffassung selektive Mechanismen zur Auswahl brauchbarer Hypothesen nicht erst bei deren empirischer Prüfung wirksam, sondern bereits in der Phase ihrer Entstehung. Entdeckungs- und Rechtfertigungszusammenhang sind aus dieser Perspektive wesentlich enger verknüpft als üblicherweise in der Wissenschaftstheorie angenommen wird (Thagard, 1988, S. 51).

Die in den Simulationssystemen hauptsächlich verwendeten Mechanismen zur Regelgenerierung sind verschiedene Arten von Generalisierungen und Spezialisierungen von Bedingungs-Aktions-Regeln, beispielsweise **Generalisierungen** durch Verallgemeinerung von einigen auf alle Fälle. Diese Generalisierungen werden nach den Ergebnissen verschiedener empirischer Untersuchungen um so eher akzeptiert

- je mehr positive Instanzen und je weniger Gegenbeispiele aufgefunden worden sind und
- je geringer die Variabilität der betrachteten Objekte hinsichtlich des betrachteten Merkmals eingeschätzt wird (Holyoak & Nisbett, 1988).

Zum Beispiel hielten die Versuchspersonen von Nisbett, Krantz, Jepson und Kunda (1983) allgemeine Aussagen über chemische Eigenschaften eines unbekanntes Elements für berechtigter als allgemeine Aussagen über Persönlichkeitseigenschaften eines unbekanntes Volksstamms.

Die Produktionsregeln innerhalb eines mentalen Systems können in Standardhierarchien (**default hierarchies**) geordnet sein, die durch die über- und Unterordnungsbeziehungen zwischen den beteiligten Konzepten bestimmt werden (Holland et al., 1986, S. 18-25). Sie führen zu den normalen mit Kategorien verbundenen Erwartungen (z. B. „wenn etwas ein Vogel ist, kann es fliegen“), können aber durch spezifischere Regeln überlagert werden (z.B. „wenn etwas ein Pinguin ist, kann es nicht fliegen“). Diese Regelhierarchien repräsentieren sowohl nützliche Generalisierungen als auch (auf den untergeordneten Ebenen) wesentliche Ausnahmen. Sie stellen damit die Information dar, die eine Person über die **Uniformität** und die **Variabilität** der betrachteten Objekte hinsichtlich der betrachteten Merkmale hat. Je weniger Ausnahmen es nach bisherigem Wissen zu einer Regel gibt, desto weniger gleichlautende Beispiele genügen, um induktiv auf die Eigenschaften aller Objekte einer Klasse zu schließen.

Auch **wissenschaftliche Theorien** können nach Holland et al. (1986) als mentale Modelle aufgefaßt werden. Sie beziehen sich auf Unbeobachtbares und erklären verschiedene Einzelgesetze und beobachtbare Regelmäßigkeiten aus einem breiten Bereich. Sie sind Teil des kognitiven Apparats des Wissenschaftlers und werden so lange verwendet, bis es bessere Werkzeuge zur Lösung der interessierenden Probleme gibt. Wissenschaftliche Gesetze werden als Regeln innerhalb dieser (Produktions-)Systeme aufgefaßt, die nur unter sonst gleichen Bedingungen (**ceteris paribus**) gelten und für die es auch Ausnahmen gibt.

Die Entwicklung von neuen Hypothesen und Gesetzen mit theoretischen Begriffen kann nicht mit einfachen induktiven Verallgemeinerungsmechanismen erklärt werden, sondern nur mit Hilfe der für Problemlösungsprozesse typischen Kombinationen von aufsteigenden und absteigenden, **daten- und konzeptgeleiteten** Prozessen. Dabei sind oft gründliche Rekonzeptualisierungen

des Problemraums notwendig, und insbesondere zur Simulation von „revolutionären“ Umwälzungen im Sinne von Kuhn (1981) gilt es, theoretische Vorstellungen darüber zu entwickeln und umzusetzen, wie ein altes konzeptuelles Netzwerk durch ein neues ersetzt werden kann (Thagard, 1990). Eine bedeutende Leitfunktion kommt bei diesen Entwicklungsprozessen Analogien zu anderen Wissensgebieten zu, wie beispielsweise die „Wellentheorie“ des Lichts (Analogie zwischen Wasser und Licht) und Darwins „natürliche Zuchtwahl“ (Analogie zwischen Evolution und Tierzucht) (siehe Clement, 1988; Dreistadt, 1968; Gaifman, 1985; Hesse, 1966).

## 5.2 Induktionspsychologie und -philosophie

Die angesprochenen kognitiven Induktionskonzeptionen und -simulationen können sowohl Antworten auf deskriptive wie normative Fragestellungen geben:

1. Durch welche Regeln und Heuristiken können die menschlichen Fähigkeiten zur Erfindung und Bewertung neuer Hypothesen, Erklärungen und Konzepte beschrieben und erklärt werden?
2. Welche (induktiven) Regeln und Heuristiken erzeugen ein maximal effektives und adaptives kognitives System, d.h. ein System, das die brauchbarsten Hypothesen erzeugt, die umfassendsten Erklärungen gibt oder die besten Vorhersagen macht?

Kognitionspsychologische Untersuchungen zeigen, daß die **Effizienz** kognitiver Systeme bei der Antizipation zukünftiger Beobachtungen durch heuristische Informationsverarbeitungsregeln, die über deduktives Denken hinausgehen, deutlich gesteigert wird. Um das System nicht zu überlasten und ineffizient zu machen, muß in den Simulationsprogrammen die Hypothesengenerierung aber stark **restringiert** werden. Dies erfolgt erstens durch Beschränkung auf Hypothesen, die den vorherrschenden Zielen des Systems entsprechen und die mit den jeweiligen Hintergrundinformationen und Erklärungsschemata verträglich sind, und zweitens durch Bevorzugung von Generalisierungen, analogen Übertragungen oder Modifikationen bestehender Hypothesen. Auch kann ein kognitives System stets nur einen kleinen Teil aller möglichen deduktiven Inferenzen ausführen. Zwar erbringt deduktive Informationsverarbeitung Resultate, die mit größerer Sicherheit wahr sind als die Resultate induktiver Informationsverarbeitung, der größte Teil der deduktiven Resultate ist für die Erreichung der Ziele des Systems aber völlig irrelevant (Holland et al., 1986, S.4).

Die **Ableitung normativer Konsequenzen** und Empfehlungen für die Wissenschaft aus derartigen Effektivitätsstudien ist sicher nicht unproblematisch. Mit der gebührenden Vorsicht können aber Informationsverarbeitungsstrategien, die sich in diesen Untersuchungen bewährt haben, auch für die Zukunft als

begründete und vielleicht bestmögliche Wege zur Generierung erfolgreicher Hypothesen und Vorhersagen empfohlen werden. Sie setzen zwar eine prinzipielle Uniformität der Umwelt voraus, diese Annahme liegt aber allen methodologischen Regeln und Empfehlungen zugrunde. Darüber hinaus können Informationsverarbeitungsstrategien, die sich in entsprechenden Simulationen als adaptiv gegenüber veränderten Bedingungen erwiesen haben, auch bei partiellen Verletzungen der Uniformitätsbedingung empfohlen werden.

Kognitionswissenschaftliche Induktionskonzeptionen können die **philosophischen Analysen** der Induktion damit in verschiedener Weise ergänzen und bereichern:

1. Durch nomologische Modelle und Regelsystem kann offensichtlich auch der **Entdeckungsprozeß** allgemeiner Gesetze beschrieben werden, also gerade die Art von Induktion, die z.B. Carnap und Popper aus ihren Überlegungen ausgeschlossen haben, weil sie ihnen nicht regelgeleitet erschien.
2. Dabei zeigt sich, daß normative Kriterien und kognitive Heuristiken bereits bei der Generierung von Konzepten und verallgemeinernden Hypothesen eine Rolle spielen, nicht erst in daran anschließenden Evaluationsprozessen.
3. Die deduktive und induktive Informationsverarbeitung muß durch die aktuelle **Zielstruktur** des Systems gesteuert werden. Die von Popper nahegelegte Vorstellung einer völlig freien Produktion von Hypothesen ist deshalb kognitionstheoretisch fragwürdig.
4. Die in der Philosophie diskutierten generellen Annahmen über die Uniformität der Welt können durch spezielles **Hintergrundwissen** über regelmäßige Merkmalskonstanzen, -variabilitäten und -zusammenhänge ersetzt werden. Die gegenüber der deduktiven Logik reicheren Repräsentationsformalismen der Kognitionspsychologie könnten sich dabei als adäquatere Möglichkeit der Beschreibung und Präzisierung induktiver Annahmen und Prozesse erweisen.
5. Die induktiven Argumenten zugrundeliegenden Analogie- und **Ähnlichkeitsurteile** können durch Modelle der kognitiven Repräsentation präziser und vor allem psychologisch angemessener dargestellt werden als in den vorliegenden philosophischen Konzeptionen.
6. Mit Hilfe empirisch-psychologischer Untersuchungen kann man die Regeln ermitteln, die tatsächlichen und akzeptierten induktiven Argumenten zugrundeliegen, und typische Fehler wie etwa die Tendenz zur Bestätigung eigener Hypothesen empirisch analysieren. Damit sind die Voraussetzungen geschaffen, um das von Goodman (1973, S.66) beschriebene Ziel eines **reflektiven Äquilibriums** zwischen allgemeinen Induktionsregeln und tatsächlich akzeptierten induktiven Argumenten zu verfolgen und darüber hinaus eine maximale Kohärenz mit dem verfügbaren Hintergrundwissen und den Zielen des Systems anzustreben (Thagard, 1988).



Allerdings müssen auch die Beschränkungen der vorliegenden psychologischen Analysen induktiven Denkens beachtet werden. Erstens kann kaum geprüft werden, ob mit den Computersimulationen die Informationsverarbeitungsmechanismen bei wissenschaftlichen Induktionsprozessen adäquat beschrieben werden, da die zum Vergleich notwendigen detaillierten Protokolle wissenschaftlicher Problemlösungsprozesse im allgemeinen fehlen. Deshalb ist es sinnvoll, bei den Simulationen zumindest die Ergebnisse der **quasi-empirisch** ausgerichteten Wissenschaftsphilosophie und insbesondere der **strukturalistischen** Rekonstruktionen tatsächlicher Wissenschaftsprozesse (siehe Abschnitt 6) mit einzubeziehen. Zweitens sind Komponenten der Simulationsmodelle oft übersimplifiziert. Vor allem werden die Probleme bei der genauen Definition des Grades der Bewährung oder Bestätigung von Hypothesen ausgeklammert oder nur unvollständig behandelt.

Drittens schließlich operieren viele kognitionspsychologische Simulationen in einer **künstlichen Welt**, deren Fakten und Regularitäten vom Konstrukteur der Simulation relativ willkürlich festgelegt werden können. Es muß deshalb üblicherweise offenbleiben, inwieweit die Erfolge der Simulation beim Erkennen und Vorhersagen dieser Fakten und Regularitäten auf andere mögliche Welten, insbesondere auf die reale Welt, übertragen werden können.

Um Hypothesen generieren zu können, die möglichst zutreffende Voraussagen erlauben, enthalten die meisten Simulationssysteme **Uniformitätsvermutungen**. So geht man gemeinhin davon aus, daß in ähnlichen Situationen sich am ehesten ähnliche Hypothesen bewähren werden, daß oft und gut validierte Hypothesen sich auch weiterhin bewähren werden usw. Hinterfragt oder problematisiert werden derartige Annahmen üblicherweise nicht weiter. Vielmehr werden die Uniformitäten in der betrachteten „künstlichen Welt“ selbst erzeugt, indem die Computerprogramme auf Repräsentationen operieren, für die diese Uniformitäten gelten. Ein realer Anwendungsbereich, in dem die normalerweise angenommenen Uniformitäten nicht oder nur teilweise gegeben sind, könnte durch die entsprechenden Simulationsprogramme vermutlich nur sehr inadäquat bewältigt werden. Wünschenswert wären daher gezieltere Untersuchungen zum adaptiven Charakter induktiver Heuristiken. Wenn gezeigt werden könnte, daß bestimmte Heuristiken sich an verschiedene Uniformitätsmuster in verschiedenen simulierten Welten angleichen können, dann wäre etwas Wesentliches für die Begründung solcher Heuristiken auch unabhängig von der realen Beschaffenheit der jeweiligen Welt gezeigt.

Aus der Tatsache, daß man ohne derartigen Annahmen offenbar kein wissenschaftliches Denken simulieren kann, kann gefolgert werden, daß wissenschaftliches Denken auf Uniformitätsannahmen beruht und daß es eine Aufgabe methodologischer Analysen sein sollte sie näher zu explizieren. Wir werden darauf noch zurückkommen.

## 6. Strukturalistische Analyse der Induktion

In diesem Abschnitt wollen wir die induktiven Anteile wissenschaftlicher Forschungsprozesse genauer lokalisieren und charakterisieren, indem wir von der strukturalistischen Wissenschaftskonzeption (**non-statement view**, Balzer, Moulines & Sneed, 1987; Stegmüller, 1979; siehe auch Gadanne, Kapitel 7 in diesem Band) und den Resultaten strukturalistischer Analysen tatsächlicher wissenschaftlicher Theorien und Forschungen (Westermann, 1987a; Westmeyer, 1989) ausgehen.

Die strukturalistische Wissenschaftskonzeption ermöglicht präzise und umfassende metatheoretische Analysen von Terminologien und Behauptungen, Verbindungen und Entwicklungen von wissenschaftlichen Theorien aus verschiedenen Fachgebieten, sie erlaubt und erfordert aber auch den Einbezug pragmatischer und konventionalistischer Aspekte, die den Wissenschaftsprozess entscheidend mitprägen. Damit bildet die strukturalistische Wissenschaftsauffassung tragfähige Brücken zwischen der **analytischen Wissenschaftsphilosophie** mit ihrem logisch-rekonstruktiven Grundanliegen, dem mehr methodologisch orientierten **Kritischen Rationalismus** und den wissenschaftshistorisch und -soziologisch ausgerichteten Ansätzen (Westermann, 1987b).

### 6.1 Induktion im Wissenschaftsprozess

Aus strukturalistischer Sicht ergeben sich einige Konsequenzen für Stellenwert und Problematik induktiver Argumente in empirischen Wissenschaften, die wir im folgenden thesenartig zusammenfassen wollen.

1. Maße für die globale induktive Bestätigung von **Theorien** sind nicht sinnvoll zu konzeptualisieren und zu verwenden, denn eine Theorie ist kein Aussagensystem mit unbegrenztem Geltungsanspruch, sondern ein epistemisches Instrument, dessen Elemente mit speziellen und jeweils begrenzten **intendierten Anwendungsbereichen** versehen werden.
2. Entscheidende Aspekte der empirischen Prüfung von Hypothesen, wie die Validitätssicherung und die signifikanzstatistische Auswertung, die traditionellerweise als induktiv bezeichnet werden, können weitgehend deduktiv rekonstruiert werden (Bredenkamp, 1980; Gadanne, 1984; Hager & Westermann, 1983; Erdfelder & Bredenkamp, Kapitel 14 in diesem Band; Hussy & Möller, Kapitel 11 in diesem Band).
3. Die Neuaufstellung oder Modifikation von Theorieelementen (**theoretische Fortschritte**) und die Ausdehnung von intendierten Anwendungsbereichen (**empirische Fortschritte**) beinhalten kein genuin induktives Denken, falls die resultierenden Hypothesen tatsächlich empirisch überprüft

werden sollen. Diese empirischen Hypothesen gehen zwar über das bisherige Wissen hinaus, ihr Wert hängt aber nicht so sehr davon ab, wie gut sie durch die bisher vorliegenden Erfahrungen gestützt werden. Er wird vielmehr vor allem durch den Informations- und Erkenntnisgewinn bestimmt, der mit einem nachfolgenden empirischen Untersuchungsergebnis verbunden ist.

4. Von Bedeutung ist induktives Denken hingegen bei der Formulierung und Bewertung von **empirischen Vermutungen** über die erfolgreiche Anwendbarkeit von Theorien auf empirische Systeme, die nicht empirisch überprüft werden können oder sollen.

Die vierte These soll im folgenden etwas näher begründet werden. Die zentrale Rolle in der strukturalistischen Theorienformulierung spielen die sog. **Theorieelemente**, die miteinander zu einem Theoriennetz verknüpft sind. Jedes Theorieelement besteht aus einem **formalen Kern** und einer Menge von **intendierten Anwendungen**. Diese intendierte Anwendungsmenge **I** ist ein pragmatisches Konzept und kann stets nur unscharf abgegrenzt werden. Sie enthält als Teilmengen vor allem die von der jeweiligen wissenschaftlichen Gemeinschaft akzeptierten paradigmatischen Anwendungen ( $I_0$ ) und die später hinzugekommenen erfolgreichen Anwendungen ( $I_0$ ). Außerdem gibt es „hypothetische“ Anwendungen ( $I_h$ ), die vorläufig in die Menge **I** aufgenommen worden sind, bei denen aber noch empirisch geprüft werden soll, ob sie tatsächlich erfolgreiche Anwendungen sind. Darüber hinaus kann in **I** aber auch eine Menge  $I_v$  von vermuteten Anwendungen enthalten sein.

Sie repräsentiert die festen Überzeugungen der Vertreter der Theorie darüber, auf welche anderen, ähnlichen empirischen Systeme das Theorieelement ebenfalls erfolgreich angewendet werden kann, ohne daß dies eigens empirisch überprüft wird. Diese Überzeugungen wollen wir als **empirische Vermutungen** bezeichnen.

Empirische Vermutungen können eindeutig als Ergebnis induktiver Argumente charakterisiert werden. Erstens sind sie insofern **gehaltserweiternd**, als sie über das akzeptierte empirisch fundierte Wissen hinausgehen. Zweitens werden sie aber nach Auffassung ihrer Vertreter durch Hintergrundwissen und bisherige Erfahrung stark **gestützt**.

Prinzipiell könnte man induktive Anwendungsvermutungen vermeiden, indem man die Gültigkeit einer Theorie strikt nur für genau die Fälle unterstellt, die erfolgreich empirisch überprüft worden sind. Tatsächlich sind induktive Geltungsbereichserweiterungen jedoch ein wesentliches Charakteristikum des üblichen Umgangs mit wissenschaftlichen Theorien, das es methodologisch zu analysieren gilt. Sie treten im Rahmen jeder wissenschaftlichen oder praktischen Anwendung einer Theorie **T** auf, bei der die Frage der Gültigkeit von **T** weder explizit überprüft noch auf spätere Untersuchungen verschoben wer-

den kann, sondern bei der die erfolgreiche Anwendbarkeit von T unterstellt werden muß, um andere Handlungsziele verfolgen zu können. Dies ist vor allem der Fall, wenn eine Theorie **T** als **Hilfstheorie** in einer Untersuchung zur empirischen Prüfung einer anderen Theorie T' verwendet wird oder wenn eine bewährte Theorie zur Phänomenerklärung und -vorhersage für andere Personen und Situationen herangezogen wird.

Die Frage der Übertragbarkeit von Untersuchungsergebnissen auf andere Personen und „natürliche“ Situationen ist gerade im Kontext **technologischer Forschungsprogramme** (siehe Herrmann, Kapitel 6 in diesem Band) von großer Bedeutung. Dies hat zu der bekannten Forderung geführt, wissenschaftliche Untersuchungen sollten nicht nur „intern“, sondern auch „extern“ valide sein (Campbell & Stanley, 1963; Cook & Campbell, 1979). Das Konzept der **externen Validität** ist aber in mehrfacher Hinsicht zu undifferenziert und kein sinnvolles generelles Gütemerkmal für empirische Untersuchungen (Gadene, 1984; Westermann, 1987a). In ihm werden die Frage der Güte einer hypothesenprüfenden Untersuchung mit der Frage der Gültigkeit der entsprechenden Theorie in einem nicht-untersuchten Kontext vermischt. Da grundlagenwissenschaftliche Theorien sich typischerweise auf kausale Zusammenhänge zwischen relativ wenigen Variablen beziehen, können sie nur durch Experimente in künstlich abgeschlossenen Situationen valide überprüft werden. Diese Experimente sind damit aber notwendigerweise extern invalide, weil ihre Ergebnisse nicht unbedingt auf „natürliche“ Situationen übertragen werden können.

## 6.2 Berechtigung empirischer Vermutungen

Wenn wir uns vor einem bestimmten Wissenshintergrund verschiedene mögliche empirische Vermutungen vorstellen, so können diese uns offensichtlich mehr oder minder begründet oder gewagt erscheinen. Wir wollen im folgenden von diesem Vorverständnis ausgehen und versuchen, Faktoren zu explizieren, von denen die Güte, Adäquatheit oder Berechtigung einer empirischen Vermutung abhängen könnte.

Allgemein gesagt soll die Berechtigung einer empirischen Vermutung um so höher sein, je mehr **gute Gründe** es für die Erwartung gibt, daß sie sich in einer hypothetisch möglichen Untersuchung bewähren würde oder bewährt hätte. Konkret können wir dann eine empirische Vermutung  $EV_x$  als berechtigter bezeichnen als eine andere empirische Vermutung  $EV_y$ , wenn alle Gründe, die für das Zutreffen von  $EV_y$  sprechen, auch für das Zutreffen von  $EV_x$  sprechen, und wenn es darüber hinaus mindestens einen Grund gibt, der für das Zutreffen von  $EV_x$ , nicht aber für das Zutreffen von  $EV_y$  spricht.

Wir können sicher nicht erwarten, zu einer vollständigen und endgültigen Liste von Determinanten und Korrelaten der Berechtigung einer empirischen Vermutung zu gelangen, vielmehr sind vielfältige Modifikationen, Ergänzungen und Ausdifferenzierungen möglich. Auch können wir die Wirkung dieser **Berechtigungs-faktoren** stets nur unter der Voraussetzung spezifizieren, daß sie nicht durch die gegensinnig wirkende Variation anderer Faktoren überlagert wird. Ohne Einschränkungen ist deshalb die Berechtigung von empirischen Vermutungen nur zu vergleichen, wenn sie sich in einem einzigen oder in mehreren gleichsinnig wirkenden relevanten Einflußfaktoren unterscheiden, während alle anderen bekannten und unbekanntem Faktoren gleich sind oder sich zumindest nur zufällig unterscheiden. Diese **Ceteris-paribus-Bedingung** (oder präziser: Ceteris-distributionibus-paribus-Klausel, Steyer, Kapitel **15** in diesem Band) verhindert zwar die vergleichende Bewertung beliebiger empirischer Vermutungen, wir können und wollen unsere Berechtigungs-faktoren aber auch nur im Sinne einer regulativen Idee verstehen: Unsere Überlegungen sollen Hinweise geben, welche Gesichtspunkte bei der Aufstellung einer konkreten empirischen Vermutung beachtet werden sollten, damit sie als möglichst berechtigt bezeichnet werden kann.

Aus den Überlegungen in den vorangegangenen Abschnitten ergibt sich, daß jedem induktiven Argument zumindest implizit spezielle **Uniformitätsannahmen** zugrundeliegen und daß eine adäquate wissenschaftliche Methodologie nicht ohne (implizite) Uniformitätsannahmen auskommt.

Deshalb kann jede Aussage über Kriterien für die Berechtigung einer empirischen Vermutung als Aussage über zugrundeliegende Vorstellungen von Uniformitäten im betrachteten Realitätsausschnitt interpretiert werden, und eine nach diesen Kriterien berechtigte empirische Vermutung kann in anderen „Welten“ inadäquat sein. Um nicht in infinite Begründungsregresse oder -zirkel zu gelangen, können wir die folgende Zusammenstellung von Berechtigungs-faktoren als Bestandteile einer **Definition** betrachten, durch die ein Uniformitätsmodell oder eine komparative Berechtigungsrelation zwischen empirischen Vermutungen charakterisiert wird, mit deren Hilfe wir unser Vorverständnis der unterschiedlichen Berechtigung oder Güte induktiver Geltungsbereichserweiterungen explizieren.

Um die prinzipiell mögliche Willkür einer derartigen Definition einzuschränken und um eine befriedigende Ähnlichkeit zwischen Explikat und Explikandum zu erreichen, wollen wir bei der Zusammenstellung der Berechtigungs-faktoren sowohl die tatsächlichen wissenschaftlichen Urteilsprozesse als auch ein möglichst breites Spektrum der wesentlichen philosophischen **Induktions-konzeptionen** berücksichtigen. Daraus folgt natürlich nicht, daß unser Ergebnis in irgendeiner Weise zwingend wäre. Vielmehr stellt es einen allgemeinen

konzeptuellen Rahmen dar, der im Lichte neuer Erkenntnisse oder kritischer Diskussion verworfen oder modifiziert werden kann.

Im folgenden wollen wir **vier Arten von Faktoren** beschreiben, die die Berechtigung einer induktiven empirischen Vermutung  $EV(T,a)$  über die erfolgreiche Anwendbarkeit eines Theorieelements  $T$  auf ein empirisches System  $a$  beeinflussen.

(1) Die Berechtigung einer induktiven empirischen Vermutung soll zunächst einmal um so höher sein, **je** größer die **Anzahl** der bisherigen erfolgreichen Anwendungen von  $T$  bzw. je größer ihre relative Häufigkeit in der Menge der bisherigen Anwendungen ist. Dies entspricht Reichenbachs und Carnaps enumerativen Induktionskonzeptionen (siehe S. 430f.).

(2) Popper hat betont, daß die Bewährung von Hypothesen zwar auch von der Zahl bestandener Prüfungen abhängt, viel stärker aber von ihrer Strenge (siehe S. 440ff.). Konkret soll deshalb eine induktive Ausdehnung des Anwendungsbereiches einer Theorie um so berechtigter sein, je „strenger“ ihre bisherigen erfolgreich bestandenen Anwendungsversuche gewesen sind.

Da aus strukturalistischer Sicht das wesentliche Ziel des Wissenschaftsprozesses nicht in der Falsifikation von Theorien besteht, sondern in der Identifikation und Abgrenzung ihrer erfolgreichen Anwendungsbereiche, dient es nicht dem Erkenntnisfortschritt, wenn wir nur einseitig nach strengen Prüfungen streben. Vielmehr müssen wir unseren Theorien faire Bewährungschancen geben und die Hypothesen über ihre empirische Anwendbarkeit auch mit einem gewissen **Wohlwollen** prüfen (Westermann, 1987b): Die Wahrscheinlichkeit, eine dieser Hypothesen abzulehnen, obwohl sie tatsächlich zutrifft, soll jeweils möglichst gering sein. Damit wird die Chance erhöht, tatsächliche erfolgreiche Anwendungen zu identifizieren, und die Basis für berechnete empirische Vermutungen wird vergrößert.

(3) Ähnlich wie Popper betont auch Cohen (1989), daß für die Induktion die Anzahl positiver Instanzen nur von sekundärer Wichtigkeit ist. Vielmehr hängt seiner Auffassung nach der Grad der induktiven Stützung einer Hypothese ganz wesentlich davon ab, wie stark sich die Beobachtungsevidenzen im Hinblick auf Faktoren unterscheiden, die auch als mögliche Ursachen für die beobachteten Phänomene in Frage kommen, d.h. inwieweit die bisherigen Prüfungen geeignet waren, Alternativerklärungen auszuschließen. Dementsprechend soll eine empirische Vermutung um so berechtigter sein, je größer die Zahl der Merkmale ist, in denen sich die bisherigen erfolgreichen Anwendungen von  $T$  unterscheiden und je größer die Unterschiede auf diesen divergierenden Merkmalen sind. Diese **Variabilitäten** beeinflussen die Berechtigung natürlich um so stärker, je relevanter sie sind, d.h. je stärker nach dem jeweiligen Hintergrundwissen die Erwartung berechtigt gewesen ist, daß die

divergierenden Merkmale die erfolgreiche Anwendbarkeit von T beeinflussen könnten.

Anders ausgedrückt wird die Berechtigung einer empirischen Vermutung erhöht, wenn theoriekonforme empirische Zusammenhänge über die Variation von wesentlichen Personen- und Untersuchungsmerkmalen konstant bleiben, d.h. wenn keine **Moderatoreffekte** bestehen oder wenn keine **disordinalen Interaktionen** (Bredenkamp, 1980; Hager & Westermann, 1983) zwischen den eigentlich interessierenden Ursachenvariablen und anderen Faktoren vorliegen.

(4) Da wir empirische Vermutungen als Ausdehnung des Anwendungsbereiches auf ähnliche Anwendungen charakterisiert haben, kann ihre Berechtigung nicht nur von Zahl, Strenge, Variabilität oder anderer Eigenschaften bisheriger Anwendungen bestimmt werden, sondern sie muß insbesondere auch von der **Ähnlichkeit** zwischen dem noch nicht untersuchten empirischen System a einerseits und den paradigmatischen und späteren erfolgreichen Anwendungen andererseits abhängen. Wir können hier keine detaillierten Ähnlichkeitsmodelle und -maße für intendierte Anwendungen einer Theorie entwickeln. Es genügt, davon auszugehen, daß eine empirische Vermutung generell um so berechtigter ist, je größer die Zahl der übereinstimmenden Merkmale zwischen der neuen Anwendung und bisherigen erfolgreichen Anwendungen von T ist, je geringer die Zahl der divergierenden Merkmale zwischen ihnen ist und je geringer die Unterschiede auf den nicht-übereinstimmenden Merkmalen sind.

Bei der genaueren Analyse der Ähnlichkeit zwischen den bisherigen und der vermuteten Anwendung muß natürlich auch wieder die **Relevanz** von Merkmalen und Unterschieden berücksichtigt werden. Wie Thagard und Nisbett (1982) betont haben (s. Abschn. 4), hängt unsere Bereitschaft zu induktiven Verallgemeinerungen besonders von unserem Hintergrundwissen über die übliche Variabilität der betrachteten Objekte hinsichtlich der betrachteten Merkmale ab. Im Falle der Einbeziehung eines nicht untersuchten Falls a in die Menge der intendierten Anwendungen des Theorieelements T ist insbesondere wichtig zu beurteilen, inwieweit die Merkmalsübereinstimmungen zwischen erfolgreichen und vermuteten Anwendungen dazu führen wurden, daß die theoriekonformen Variablenbeziehungen erhalten bleiben, und inwieweit entsprechende Merkmalsunterschiede zur Modifikation dieser Zusammenhänge führen könnten. Eine empirische Vermutung soll damit um so berechtigter sein, je stärker nach dem gegenwärtigen Hintergrundwissen die erfolgreiche Anwendbarkeit von T von der Konstanz der übereinstimmenden Merkmale abhängt und je weniger sie durch Variation der nicht-übereinstimmenden Merkmale beeinflusst wird. Damit ist die relevante Ähnlichkeit zwischen zwei Anwendungen kein „objektiver“ Sachverhalt, sondern einer, der von theore-

tischen Behauptungen abhängig ist, teilweise sogar von der Theorie, um deren vermutete Anwendbarkeit es bei den Ähnlichkeitsüberlegungen geht.

Die Ähnlichkeit von zwei Anwendungsbereichen kann also nur unter dem Vorbehalt der Richtigkeit der entsprechenden Theorien eingeschätzt werden. Der für die Berechtigung induktiver Vermutungen relevante Faktor der Ähnlichkeit setzt sich damit aus zwei Teilen zusammen: aus der allgemeinen Behauptung, daß die Berechtigung einer Anwendungsvermutung mit der Ähnlichkeit auf den relevanten Merkmalen ansteigt, und aus den speziellen Annahmen über die relevanten Merkmale. Aus diesem Grund kann im Falle einer nicht erfolgreichen empirischen Vermutung die allgemeine Ähnlichkeitsbehauptung und damit das ganze Uniformitätsmodell stets dadurch geschützt werden, daß man annimmt, daß die relevanten Merkmale falsch spezifiziert worden sind. Trotzdem werden (wie bei jedem Forschungsparadigma) anhaltende **Mißerfolge bei der Anwendung** dieses Modells seine Adaptation oder Aufgabe nahelegen.

Die Berechtigung einer empirischen Vermutung, die durch die in den vorangegangenen Absätzen angesprochenen Faktoren näher expliziert wurde, soll die auf unserem bisherigen Wissen basierende subjektive Erwartung ausdrücken, daß sie sich in einer hypothetischen Untersuchung bewähren würde. Sie ist deshalb nur für Anwendungsbehauptungen, die **nicht empirisch überprüft** werden, ein sinnvolles Bewertungskriterium. Empirische Hypothesen, die gezielt überprüft werden sollen, sind dagegen im allgemeinen am wertvollsten, wenn die Unsicherheit am größten ist, ob sie sich bewähren oder nicht, weil das Untersuchungsergebnis dann den höchsten **Informationswert** besitzt.

Außerdem hängt die Frage, ob eine empirische Vermutung als Grundlage für weitergehende Handlungen und Entscheidungen akzeptiert werden sollte, nicht nur von ihrer Berechtigung ab, sondern auch von den epistemischen und praktischen **Konsequenzen** der Akzeptanzentscheidung, d.h. von den erwarteten Nutzen und Kosten bei richtigen und falschen Entscheidungen über die empirische Vermutung.

Die identifizierten Berechtigungsfaktoren verdeutlichen das beständige Risiko, daß empirische Vermutungen über die erfolgreiche Anwendbarkeit von Theorien aufgrund späterer Erkenntnisse aufgegeben oder modifiziert werden müssen. Diese Unsicherheit und Vorläufigkeit des Wissens ist ein wesentliches Kennzeichen jeder empirischen Wissenschaft. Es unterscheidet sie von der deduktiven Logik und führt auch dazu, daß die Versuche der Philosophie, die (relative) Sicherheit induktiver Schlüsse formal-logisch zu begründen oder zu quantifizieren, so wenig Relevanz für die empirischen Wissenschaften haben. Überlegungen zur Berechtigung induktiv verallgemeinernder Aussagen in der Wissenschaft können nicht zu (absoluter) Sicherheit oder (hoher) Wahrscheinlichkeit für diese Generalisierungen führen, sondern sie können nur dazu ani-



mieren, sich bei der Formulierung empirischer Vermutungen insofern so rational wie möglich zu verhalten, als daß man bewußt die Faktoren beachtet, von denen es nach bisherigem Wissen abhängen könnte, ob diese Hypothesen sich empirisch bewähren werden.

## 7. Zusammenfassende Schlußbemerkungen

Zu Anfang dieser Arbeit haben wir zwei mit dem Induktionsproblem zusammenhängende Fragen formuliert. Sie betrafen die Rechtfertigung der **Ableitung** allgemeiner Aussagen und den Grad ihrer Stützung durch die zugrundeliegenden Beobachtungen (siehe S. 429). Nach den strukturalistischen Überlegungen im vorangegangenen Abschnitt finden sich im üblichen Umgang mit wissenschaftlichen Theorien nicht nur deduktive, sondern auch wesentliche induktive Argumentationen. Sie umfassen aber mehr als die einfachen Enumerationen und Extrapolationen, auf die sich der Großteil der philosophischen Induktionsdiskussion bezieht. Die induktiven Anteile der Wissenschaftsprozesse betreffen nach unserer Auffassung nicht die globale Bewertung von Theorien, sondern vor allem die Frage, über welche (angenommenerweise) irrelevanten Faktoren die Anwendungsbehauptung auf neue Fälle ausgedehnt werden kann. Damit stellt sich das Problem des (induktiven) Bestätigungsgrades von Theorien nicht mehr, so daß auch die in den Abschnitten 1 bis 3 angesprochenen Schwierigkeiten bei der Einführung von theoretisch angemessenen und praktisch verwendbaren Bestätigungsbegriffen und -maßen für uns keine unangenehmen Konsequenzen haben.

Auch aus strukturalistischer Sicht bestehen bleibt jedoch das Problem der Rechtfertigung induktiver Argumente. Die im Abschnitt 3 diskutierte Vermeidung jeder Art von Induktion durch eine rein deduktive Methodologie im Sinne von Popper erscheint nicht als befriedigende Antwort, vor allem weil auch sie verborgene induktive Annahmen enthält und keine begründeten Vorhersagen und Akzeptanzentscheidungen erlaubt. Einzig gangbar scheint der im Abschnitt 4 abgesteckte pragmatische Weg zu sein, induktive Argumente dadurch zu rechtfertigen, daß ihre **Nützlichkeit** bzw. Notwendigkeit zum Erreichen bestimmter wissenschaftlicher Ziele herausgestellt wird. Unterstützt wird diese Auffassung dadurch, daß aus kognitionspsychologischer Sicht (Abschnitt 5) induktives Denken keineswegs so irrational sein muß wie bloßes Raten, sondern als zielgerichtet, regelgeleitet und effektiv beschrieben werden kann.

Diese pragmatische Argumentation kann man als Rechtfertigung der Ableitung induktiver Anwendungsvermutungen ansehen. Die zweite Frage nach der Stützung solcher Anwendungsvermutungen konnten wir dadurch angehen, daß wir eine vorläufige Menge von Faktoren identifiziert haben, durch

die die Berechtigung einer empirischen Vermutung erhöht und die Adäquatheit induktiver wissenschaftlicher Argumente verbessert wird. Diese Kriterien sind auf Annahmen über die Uniformität des betrachteten Gegenstandsbezugs zurückzuführen und damit prinzipiell empirisch prüfbar, kritisierbar und revidierbar.

Wir sehen den Vorteil dieser pragmatischen Behandlung des Induktionsproblems darin, daß sie zu begründeten Handlungs- und Entscheidungskriterien in epistemisch unsicheren Situationen führt. Methodologische Positionen, die sich jeder Art von Induktion enthalten, können in diesen unvermeidbaren Situationen nicht weiterhelfen. Allgemein ist die Aufstellung von Normen, Präskriptionen oder Empfehlungen für bestimmte Situationen dann gerechtfertigt, wenn gute Gründe für die Erwartung genannt werden können, daß die entsprechenden Handlungen für die Erreichung der angestrebten Ziele hinreichend oder zumindest besonders nützlich oder effizient sind. Eine normative, präskriptive oder empfehlende Verwendung unserer Faktoren scheint uns vor allem deshalb gerechtfertigt zu sein, weil sie explizit aus anerkannten wissenschaftstheoretischen Positionen abgeleitet worden sind und damit die besten bekannten Mittel und Wege bezeichnen, die jeweils betonten epistemischen Ziele zu erreichen. Unter den jeweiligen erkenntnis- und entscheidungstheoretischen Prämissen können sie deshalb als rationale Handlungsgrundlage rekonstruiert werden. Außerdem haben sie sich zum Teil in Form von methodologischen Konventionen durchgesetzt und bewährt.

Um zu untersuchen, inwieweit sich mit Hilfe unserer Faktoren der tatsächliche Umgang mit empirischen Vermutungen adäquat beschreiben läßt, sind gezielte empirische Untersuchungen zur Beurteilung wissenschaftlicher Hypothesen und ihrer erfolgreichen Anwendbarkeit notwendig. Darüber hinaus ist eine Integration in die kognitionswissenschaftlichen Induktionsmodelle und Simulationsprogramme erstrebenswert. Notwendig ist aber vor allem eine präzisere Formulierung der Berechtigungsfaktoren. Dazu können wir uns z.B. psychologische Modellvorstellungen, insbesondere über die Abhängigkeit globaler Ähnlichkeitsbeurteilungen von spezifischen Merkmalsunterschieden (Tversky, 1977), zunutze machen (siehe im einzelnen Westermann, Heise & Gerjets, 1992).

## **Literatur**

- Adler, J. E. (1980). Criteria for good inductive logic. In L.J. Cohen & M. Hesse (Eds.), *Applications of inductive logic* (pp. 379-405). Oxford: Clarendon.
- Amico, R. P. (1986). On the vindication of deduction and induction. *Australasian Journal of Philosophy*, 64, 322-330.

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Balestra, D.J. (1979). Non-falsifiability: An inductivist perspective. *International Logic Review*, 10, 118-125.
- Balzer, W., Moulines, C. U. & Sneed, J. D. (1987). *An architectonic for science. The structuralist program*. Dordrecht: Reidel.
- Barker, S. & Achinstein, P. (1960). On the new riddle of induction. *Philosophical Review*, 69, 511-522.
- Bredenkamp, J. (1980). *Theorie und Planung psychologischer Experimente*. Darmstadt: Steinkopff.
- Campbell, D.T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp.413-463). La Salle, IL: Open Court.
- Campbell, D. T. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research in teaching. In N.L. Gage (Ed.), *Handbook of research in teaching* (pp. 171-246). Chicago: Rand McNally.
- Carnap, R. (1926). *Physikalische Begriffsbildung*. Karlsruhe: Braun.
- Carnap, R. (1946). Theory and prediction in science. *Science*, 104, 520-521.
- Carnap, R. (1950, 1962). *Logical foundations of probability*. (1st, 2nd ed.) Chicago, IL: University of Chicago Press.
- Carnap, R. (1952). *The continuum of inductive methods*. Chicago, IL: University of Chicago Press.
- Carnap, R. (1963). Replies and systematic expositions. In P.A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 859-1013). La Salle, IL: Open Court.
- Carnap, R. (1971). A basic system of inductive logic, part I. In R. Carnap & R.C. Jeffrey (Eds.), *Studies in inductive logic and probability*: Vol. I (pp. 33-165). Berkeley: University of California Press.
- Clement, J. (1988). Observed methods for generating analogies in scientific problem solving. *Cognitive Science*, 12, 563-586.
- Cohen, L.J. (1977). *The probable and the provable*. Oxford: Clarendon.
- Cohen, L.J. (1980). Inductive logic 1945-1977. In E. Agazzi (Ed.), *Modern logic - A survey* (pp. 353-375). Dordrecht: Reidel.
- Cohen, L.J. (1982). Intuition, induction, and the middle way. *Monist*, 65, 287-301.
- Cohen, L.J. (1989). *An introduction to the philosophy of induction and probability*. Oxford: Clarendon.
- Cohen, L. J. & Hesse, M. (Eds.) (1980). *Applications of inductive logic*. Oxford: Clarendon.
- Cohen, Y. (1979). A new view of *grue*. *Zeitschrift für allgemeine Wissenschaftstheorie*, 10, 244-252.
- Cook, T.D. & Campbell, D.T. (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Boston: Houghton Mifflin.
- Cox, J. W. R. (1986). Induction and disjunction. *Philosophical Papers*, 15, 89-95,

- De Olivera, M. B. (1985). The problem of induction: A new approach. **British Journal for the Philosophy of Science**, **36**, 129-145.
- Dreistadt, R. (1968). An analysis of the use of analogies and metaphors in science. **Journal of Psychology**, **68**, 97-116.
- Eddidin, A. (1984). Inductive reasoning and the uniformity of nature. **Journal of Philosophical Logic**, **13**, 285-302.
- Essler, W. K. (1973). **Wissenschaftstheorie: Band 3. Wahrscheinlichkeit und Induktion**. Freiburg: Alber.
- Essler, W. K. (1980). Induktion. In J. Speck (Hrsg.), **Handbuch wissenschaftstheoretischer Begriffe: Band 2** (S.297-307). Göttingen: Vandenhoeck & Ruprecht.
- Gadene, V. (1984). **Theorie und Erfahrung in der psychologischen Forschung**. Tübingen: Mohr.
- Gärdenfors, P. (1990). Induction, conceptual spaces and AI. **Philosophy of Science**, **17**, 78-95.
- Gaifman, H. (1985). On inductive support and some recent tricks. **Erkenntnis**, **22**, 5-21.
- Goodman, N. (1973). **Fact, fiction and forecast**. (3rd ed.) Indianapolis: Bobbs-Merrill.
- Grünbaum, A. (1963). **Philosophical problems of space and time**. New York: Knopf.
- Hager, W. & Westermann, R. (1983). Planung und Auswertung von Experimenten. In J. Bredenkamp & H. Feger (Hrsg.), Hypothesenprüfung. (**Enzyklopädie der Psychologie, Serie Forschungsmethoden der Psychologie: Band 5**) (S.24-238). Göttingen: Hogrefe.
- Hempel, C. G. (1943). A purely syntactical definition of confirmation. **Journal of Symbolic Logic**, **8**, 122-143.
- Hempel, C. G. (1945). Studies in the logic of confirmation. **Mind**, **54**, 1-26, 97-121.
- Hempel, C. G. (1965). **Aspects of scientific explanation and other essays in the philosophy of science**. New York: Free Press.
- Hempel, C.G. (1981). Turns in the evolution of the problem of induction. **Synthese**, **46**, 389-404.
- Hesse, M. (1966). **Models and analogies in science**. Notre Dame, IN: University of Notre Dame Press.
- Hesse, M. (1974). **The structure of scientific inference**. London: Macmillan.
- Hesse, M. (1980). What is the best way to assess evidential support for scientific theories?. In L.J. Cohen & M. Hesse (Eds.), **Applications of inductive logic** (pp.202-217). Oxford: Clarendon.
- Hintikka, J. (1966). A two-dimensional continuum of inductive methods. In J. Hintikka & P. Suppes (Eds.), **Aspects of inductive logic** (pp. 113-132). Amsterdam: North-Holland.
- Hintikka, J. & Niiniluoto, I. (1976). An axiomatic foundation for the logic of inductive generalization. In M. Przelecki, U. Szaniawski & R. Wojcicki (Eds.), **Formal methods of the methodology of science** (pp.57-81). Wrocław: Ossolineum.
- Holland, J. H., Holyoak, K. J., Nisbett, R. E. & Thagard, P. R. (1986). **Induction: Processes of inference, learning, and discovery**. Cambridge, MA: MIT-Press.

- Holyoak, K. J., Koh, K. & Nisbett, R.E. (1989). A theory of conditioning: Inductive learning within rule-based default hierarchies. *Psychological Review*, 96, 315-340.
- Holyoak, K. J. & Nisbett, R.E. (1988). Induction. In R. J. Sternberg & E.E. Smith (Eds.), *The psychology of human thought* (pp. 50-91).
- Hoppe, H. (1975). Goodmans Schein-Rätsel. *Zeitschrift für allgemeine Wissenschaftstheorie*, 6, 131-139.
- Hübner, K. (1980). Einige kritische Bemerkungen zum gegenwärtigen Popperianismus auf der Grundlage einer Theorie der Systemmengen. In G. Radnitzky & G. Andersson (Hrsg.), *Fortschritt und Rationalität der Wissenschaft (S. 275-286)*. Tübingen: Mohr.
- Hume, D. (1978). *Treatise of human nature*. (2nd ed.) Oxford: Clarendon (Original erschienen 1739).
- Jason, G. J. (1985). Two problems of induction. *Dialectica*, 39, 53-74.
- Jones, G.E. (1982). Vindication, Hume, and induction. *Canadian Journal of Philosophy*, 12, 119-129.
- Jones, G. & Perry, C. (1982). Popper, induction and falsification. *Erkenntnis*, 18, 97-104.
- Kemeny, J. G. (1963). Carnap's theory of probability and induction. In P.A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 711-738). La Salle, IL: Open Court.
- Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 148.
- Kronthaler, E. (1984). Zur kritischen Analyse des Induktionsproblems. *Zeitschrift für Philosophische Forschung*, 38, 278-291.
- Kuhn, T.S. (1981). *Die Struktur wissenschaftlicher Revolutionen*. (5. Aufl.) Frankfurt/Main: Suhrkamp (Original erschienen 1970: The structure of scientific revolutions).
- Kuipers, T. A. F. (1978). On the generalization of the continuum of inductive methods to universal hypotheses. *Synthese*, 37, 255-284.
- Kuipers, T.A.F. (1984). Two types of inductive analogy by similarity. *Erkenntnis*, 21, 63-87.
- Kulkarni, D. & Simon, H. A. (1988). The processes of scientific discovery: The strategy of experimentation. *Cognitive Science*, 12, 139-175.
- Kummer, W. (1982). Zur Interpretation der induktiven Logik. *Deutsche Zeitschrift für Philosophie*, 30, 603-613.
- Kutschera, F. von (1972). *Wissenschaftstheorie I, II*. München: Fink.
- Kyburg, H.E. (1961). *Probability and the logic of rational belief*. Middletown: Wesleyan University Press.
- Kyburg, H. E. (1964). Recent work in inductive logic. *American Philosophical Quarterly*, 1, 249-287.
- Kyburg, H. E. & Nagel, E. (Eds.) (1963). *Induction: Some current issues*. Middletown: Wesleyan University Press.
- Lakatos, I. (1968). Change in the problem of inductive logic. In I. Lakatos (Ed.), *The problem in inductive logic* (pp. 315-417). Amsterdam: North Holland.

- Lakatos, I. (1982). **Mathematik, empirische Wissenschaft und Erkenntnistheorie** (Philosophische Schriften: Band 2). Braunschweig: Vieweg.
- Lane, N. R. & Lane, S.A. (1984). Theory, observation and inductive learning. **Ratio**, 26, 140-151.
- Langley, P.A., Simon, H.A., Bradshaw, G.L. & Zytkow, J.M. (1987). **Scientific discovery. Computational explorations of the creative processes**. Cambridge, MA: MIT Press.
- Laudan, L. (1981). **Science and Hypothesis. Historical essays on scientific methodology**. Dordrecht: Reidel.
- Lenzen, W. (1974). **Theorien der Bestätigung wissenschaftlicher Hypothesen**. Stuttgart: Frommann-Holzboog.
- Levi, I. (1967). **Gambling with truth**. New York: Knopf.
- Levi, I. (1979). Inductive appraisal. **Current Research in Philosophy of Science**, 339-351.
- Levi, I. (1982). Self-profile. In R. Bogdan (Ed.), **Henry E. Kyburg jr. & Isaac Levi** (pp. 181-216). Dordrecht: Reidel.
- Levinson, A. (1974). Popper, Hume, and the traditional problem of induction. In A. Schilpp (Ed.), **The philosophy of Karl Popper** (pp.322-331). La Salle, IL: Open Court.
- Machina, K. E (1985). Induction and deduction revisited. **Nous**, 19, 571-578.
- Mackie, J.L. (1979). A defense of induction. In G. F. Macdonald (Ed.), **Perception and identity** (pp. 113-130). London: Macmillan.
- McClelland, J.L. & Rumelhart, D. E. (Eds.) (1986). **Parallel distributed processing: Vol. 2. Psychological and biological models**. Cambridge, MA: MIT Press.
- Miller, D. (1980). Can science do without induction?. In L.J. Cohen & M. Hesse (Eds.), **Applications of inductive Logic** (pp. 109-129). Oxford: Clarendon.
- Mura, A. (1990). When probabilistic support is inductive. **Philosophy of Science**, 17, 278-289.
- Newell, A. & Simon, H.A. (1972). **Human problem solving**. Englewood Cliffs, NJ: Prentice-Hall.
- Newell, A., Shaw, J.C. & Simon, H.A. (1958). Elements of a theory of human problem solving. **Psychological Review**, 61, 151-166.
- Niiniluoto, I. (1980). Analogy, transitivity, and the confirmation of theories. In L.J. Cohen & M. Hesse (Eds.), **Applications of inductive logic** (pp. 218-234). Oxford: Clarendon.
- Niiniluoto, I. (1981). Analogy and inductive logic. **Erkenntnis**, 16, 1-34.
- Niiniluoto, I. & Tuomela, R. (1973). **Theoretical concepts and hypothetico-inductive inference**. Dordrecht: Reidel.
- Nisbett, R.E., Krantz, D.H., Jepson, D. & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. **Psychological Review**, 90, 339-363.
- Pollock, J.L. (1984). A solution to the problem of induction. **Nous**, 18, 423-461.
- Popper, K. R. (1965). **Conjectures and refutations. The growth of scientific knowledge**. (2nd ed.) London: Routledge and Kegan Paul.

- Popper, K. R. (1973). **Objektive Erkenntnis. Ein evolutionärer Entwurf**: Hamburg: Hoffmann & Campe.
- Popper, K.R. (1974). Replies to my critics. In P.A. Schilpp (Ed.), **The philosophy of Karl Popper** (pp.961-1197). La Salle, IL: Open-Court.
- Popper, K.R. (1981). A proof of the impossibility of inductive probability. *Nature*, 302, 687-688.
- Popper, K. R. (1934, 1984). **Logik der Forschung**. (1., 8. Aufl.) Tübingen: Mohr.
- Popper, K.R. & Miller, D. (1983). A proof on the impossibility of inductive probability. *Nature*, 302, 687-688.
- Popper, K. & Miller, D. W. (1987). Why probabilistic support is not inductive. **Philosophical Transactions of the Royal Society of London**, Series A, 321, 569-591.
- Putnam, H. (1974). The „corroboration“ of theories. In P.A. Schilpp (Ed.), **The philosophy of Karl Popper** (pp.221-240). La Salle, IL: Open-Court.
- Reichenbach, H. (1935). **Wahrscheinlichkeitslehre. Eine Untersuchung über die logischen und mathematischen Grundlugen der Wahrscheinlichkeitsrechnung**. Leiden: Sijthoff.
- Reichenbach, H. (1938). **Experience and prediction. An analysis of the foundations and the structure of knowledge**. Chicago: University of Chicago Press.
- Rivadulla, A. (1987). Kritischer Realismus und Induktionsproblem. **Erkenntnis**, 26, 181-193.
- Rosenkrantz, R. D. (1982). Does the philosophy of induction rest on a mistake? **Journal of Philosophy**, 79, 78-97.
- Rumelhart, D.E. & McClelland, J.L. (Eds.) (1986). **Parallel distributed processing: Vol. 1. Foundations**. Cambridge, MA: MIT Press.
- Salmon, W. C. (1968). The justification of inductive rules of inference. In I. Lakatos (Ed.), **The problem of inductive logic** (pp.24-43). Amsterdam: North Holland.
- Salmon, W.C. (1978). Unfinished business: the problem of induction. **Philosophical Studies**, 33, 1-19.
- Salmon, W. C. (1979). **The foundations of scientific inference**. (5th ed.) Pittsburgh: University of Pittsburgh Press.
- Salmon, W.C. (1982). Further reflections. In R. McLaughlin (Ed.), **What? Where? When? Why?** (pp.231-280). Dordrecht: Reidel.
- Schlesinger, G. (1981). Strawson on induction. **Philosophia**, 10, 199-208.
- Shirley, E. S. (1981). An unnoticed flaw in Barkers and Achinstein's solution to Goodman's new riddle of induction. **Philosophy of Science**, 48, 611-617.
- Simon, H.A. (1966). Scientific discovery and the psychology of problem solving. In R. Colodny (Ed.), **Mind and Cosmos** (pp. 22-40). Pittsburgh: University of Pittsburgh Press.
- Simon, H. A. (1973). Does scientific discovery have a logic? **Philosophy of Science**, 40, 471-480.
- Spohn, W. (1981). Analogy and inductive logic: A note on Niiniluoto. **Erkenntnis**, 16, 35-52.

- Stegmüller, W. (1971). Das Problem der Induktion. Humes Herausforderung und moderne Antworten. In H. Lenk (Hrsg.), *Neue Aspekte der Wissenschaftstheorie* (S. 13-74). Braunschweig: Vieweg.
- Stegmüller, W. (1973). *Personelle und statistische Wahrscheinlichkeit: 1. Halbband. Personelle Wahrscheinlichkeit und rationale Entscheidung (Probleme und Resultate der Wissenschaftstheorie und analytischen Philosophie, Band 4)*. Berlin: Springer-Verlag.
- Stegmüller, W. (1978). *Hauptströmungen der Gegenwartsphilosophie*: Band I. (6. Aufl.) Stuttgart: Kröner.
- Stegmüller, W. (1979). *The structuralist view of theories*. Berlin: Springer-Verlag.
- Stemmer, N. (1975). The Goodman paradox. *Zeitschrift für allgemeine Wissenschaftstheorie*, 6, 340-354.
- Strawson, P.E. (1952). *Introduction to logical theory*. London: Methuen.
- Strube, G. (1990). Neokonnektionismus: Eine neue Basis für die Theorie und Modellierung menschlicher Kognition? *Psychologische Rundschau*, 41, 129-143.
- Suppe, F. (1977). *The structure of scientific theories* (2nd ed.). Urbana: University of Illinois Press.
- Tarski, A. (1977). *Einführung in die mathematische Logik*. (5.Aufl.) Göttingen: Vandenhoeck & Ruprecht.
- Thagard, P. (1990). The conceptual structure of the chemical revolution. *Philosophy of Science*, 57, 183-209.
- Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press.
- Thagard, P. & Nisbett, R. E. (1982). Variability and confirmation. *Philosophical Studies*, 42, 379-394.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- van Wright, G. H. (1957). *The logical problem of induction*. Oxford: Blackwell.
- Wertheimer, M. (1959). *Productive thinking*. (enlarged ed.) Westport, CT: Greenwood Press.
- Westermann, R. (1987a). *Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie*. Berlin: Springer-Verlag.
- Westermann, R. (1987b). Wissenschaftstheoretische Grundlagen der experimentellen Psychologie. In G. Lüer (Hrsg.), *Allgemeine Experimentelle Psychologie* (S. 5-42). Stuttgart: Fischer.
- Westermann, R., Heise, E. & Gerjets, P. (1992). The justification of empirical suppositions. A structuralist analysis of an inductive form of scientific reasoning. In H. Westmeyer (Ed.), *The structuralist program in psychology: Foundations and applications* (pp. 41-54). Seattle: Hogrefe and Huber.
- Westmeyer, H. (Ed.) (1989). *Psychological theories from a structuralist point of view*. Berlin: Springer-Verlag.
- Will, U. (1982). Eine pragmatische Rechtfertigung der Induktion. *Zeitschrift für Allgemeine Wissenschaftstheorie*, 13, 84-98.



**Teil D**  
**Forschungsschritte**

## Kapitel 11 Hypothesen

von Walter Hussy und Holger Möller

## Kapitel 12 Kontrolltechniken

von Jürgen Rehm und Fritz Strack

## Kapitel 13 Von Zahlzeichen zu Skalen

von Rainer Mausfeld

## Kapitel 14 Hypothesenprüfung

von Edgar Erdfelder und Jürgen Bredenkamp

## Kapitel 15 Stochastische Modelle

von Rolf Steyer

## Kapitel 16 Statistische Entscheidungen

von Axel Ostmann und Joachim Wutke

Teil D dieses Bandes ist am Ablauf des Forschungsprozesses bei der empirischen Bearbeitung einer Fragestellung orientiert. Einige markante Schritte werden herausgegriffen und mit ihren methodologischen Grundlagen und Problemen dargestellt.

Erörtert wird die Methodologie des Umgangs mit wissenschaftlichen Hypothesen (Kapitel 11), des Einsatzes von Techniken zur Kontrolle von Fehlern und Alternativverklärungen bei der Datenerhebung (Kapitel 12) und des Testens, Skalierens und Messens (Kapitel 13) mit der Frage, ob es in der Psychologie eigenständige Größen gibt, die durch Skalen numerisch darstellbar sind. Ein wichtiger Schritt ist auch der Übergang von inhaltlichen zu statistischen Hypothesen und deren Prüfung (Kapitel 14). Stochastische Modelle sind eine der Möglichkeiten, empirische und theoretische Begriffe explizit untereinander zu verknüpfen (Kapitel 15). Schließlich werden auch die äußerst difizilen und problematischen Grundlagen statistischer Entscheidungen im Forschungsprozeß vorgestellt und analysiert (Kapitel 16).

## Hypothesen

**Walter Hussy und Holger Möller**

### **1. Zur Definition, Klassifikation, Generierung und Funktion von Hypothesen**

#### **1.1 Wissenschaftliche und nicht-wissenschaftliche Hypothesen**

„Morgen wird es regnen.“ Dieser Satz stellt eine Aussage, eine Vorhersage bestimmter Ereignisse dar: Er drückt offenkundig eine **Vermutung** über das Wetter des nächsten Tages aus. Man könnte auch sagen, daß es sich um die **Hypothese** handelt, daß es am nächsten Tag regnen wird. Ist diese Hypothese nun wissenschaftlich, kann man sie als **wissenschaftliche Hypothesen** [WH bzw. WHn] bezeichnen? Diese vermeintlich leicht zu beantwortende Frage läßt sich jedoch losgelöst aus dem Kontext, in dem die jeweilige Hypothese aufgestellt wurde, nicht beantworten, wie die beiden folgenden Beispiele illustrieren:

Eventuell ist die Person, die diese Hypothese aufstellt, Teilnehmer einer Gartenparty und äußert die Vermutung zu Beginn eines Gesprächs, vielleicht um Kontakt zu einer anderen Person zu erhalten. Ganz offensichtlich sind es jedoch nicht diese Art von Hypothesen, die in der psychologischen Forschung eine noch näher zu bestimmende Rolle spielen.

Eventuell ist unsere Person aber auch ein Meteorologe, der zu dieser Vermutung im Rahmen seiner Forschungsaktivitäten kommt, aus denen er Vorhersagen für das Wetter des nächsten Tages ableiten kann. In diesem Falle wurde es sich um eine WH handeln. Worin mögen die Unterschiede beider Hypothesentypen, die doch in ihrer Aussage allein nicht differieren, begründet liegen? Was kennzeichnet eine WH und hebt sie von einfachen Aussagen ab? Wie läßt sich der Begriff der (wissenschaftlichen) Hypothese definieren?

## 1.2 Allgemeine Begriffsbestimmungen

Man geht häufig davon aus, daß wissenschaftliche Hypothesen theoretische Vermutungen bzw. Annahmen über die Zusammenhänge von interessierenden Sachverhalten sind oder auch vorläufige (provisorische) Antworten auf ein sich stellendes wissenschaftliches Problem beinhalten. Den WHn wird für gewöhnlich unterstellt, daß sie allgemeine Aussagen darstellen, die das Eintreten bestimmter Ereignisse, Erscheinungen oder Zusammenhänge vorhersagen und damit gleichzeitig das Eintreten anderer Ereignisse etc. ausschließen. Um von einer **wissenschaftlichen** Hypothese sprechen zu können, sollten zwei Merkmale erfüllt sein:

- (a) sie muß im Sinne einer synthetischen Aussage formuliert sein, also sich als falsch oder korrekt erweisen können („Wenn der Luftdruck fällt, regnet es!“; diese ausformulierte Hypothese liegt der Aussage des Meteorologen zugrunde, wenn er vermutet „Morgen wird es regnen“), und
- (b) sie muß überprüfbar sein (man muß prinzipiell feststellen können, ob sich der Luftdruck verändert und man muß beobachten können, ob es regnet oder nicht regnet).

Im Abschnitt 2.1 werden wir uns ausführlicher mit diesen Merkmalen beschäftigen.

WHn wird ein **vorläufiger** Charakter unterstellt, da sie im Rahmen der wissenschaftlichen Forschung sowohl verworfen als auch zunächst beibehalten bzw. weiterentwickelt werden können. Von den Konzepten, die in WHn auftreten, wird größtenteils angenommen, daß sie über die direkt beobachtbaren Sachverhalte hinausgehen, daß sie im Falle des Beibehaltens einer WH als Erklärungen dienen können, warum spezifische Sachverhalte aufgetreten sind bzw. nicht aufgetreten sind. Wir werden hierauf zurückkommen. Einschränkendere Definitionen betonen die „Beziehungsherstellung zwischen Determinanten und Resultanten“, die „in Form von Je-desto‘ oder ‚Wenn-dann‘-Ausagen“ Vorfindbar sind (z.B. Krapp, Hofer & Prell, 1982).

## 1.3 Arten von Hypothesen

Wissenschaftliche Hypothesen lassen sich nach zahlreichen Gesichtspunkten unterscheiden. Zur Illustration zunächst einige Beispiele:

- (1) „Kurt hat heute die dritte Aufgabe der vorliegenden Klausur intelligent gelöst.“
- (2) „In diesem Hörsaal gibt es männliche Personen, die einen HAWIE-IQ von mehr als 130 Punkten erreichen.“

- (3) „Alle Teilnehmer dieser Vorlesung Varianzanalyse haben einen HAWIE-IQ von mehr als 150 Punkten.“
- (4) „Frustrierte Personen reagieren zumeist aggressiv.“
- (5) „Die Mittelwerte der beiden Gruppen A und B unterscheiden sich.“
- (6) „Frustration erzeugt Aggression.“

Ein möglicherweise nützliches und sinnvolles Kriterium zur Differenzierung ist die Art der Einschränkungen, die den in Hypothesen angeführten Personenkreisen, Situationen, Aktivitäten, Zeiten etc. auferlegt werden. So wird z.B. in der o.g. Hypothese (3) für alle Personen einer spezifischen Lehrveranstaltung zu einer spezifischen Zeit eine Aussage getroffen (man könnte auch sagen, daß der Anwendungsbereich der Hypothese eine raum-zeitliche Beschränkung aufweist: beschränkt universelle Hypothese), während in Hypothese (6) weder Personen noch Orte oder Zeiten eingeschränkt werden: Hypothese (6) soll immer und für alle Personen in allen Situationen gelten (Universalhypothese oder unbeschränkte, universelle Hypothese). Groeben und Westmeyer (1981) haben so acht verschiedene Hypothesenarten differenziert, auf die wir kurz eingehen wollen. Die sehr unterschiedlichen Überprüfungsmöglichkeiten dieser Hypothesentypen werden uns erst später beschäftigen.

**Singuläre Hypothesen** haben sowohl hinsichtlich des Personenkreises, auf den sie sich beziehen, wie auch hinsichtlich der möglichen Situationen und den dort ausgeführten Aktivitäten Einschränkungen (siehe Beispiel (1) oben).

**Pseudo-singuläre (idiographische) Hypothesen** haben entgegen den singulären Hypothesen keine Einschränkungen in Raum und/oder Zeit: „Kurt ist intelligent.“

**Unbestimmte Existenzhypothesen** enthalten mindestens einen Existenzquantor ohne weitere Einschränkungen: „Es gibt ein psycho-physiologisches Korrelat der menschlichen Emotion.“

**Lokalisierende (bestimmte) Existenzhypothesen** beschränken ihre Existenzaussagen auf bestimmte Räume und/oder Zeiten (vgl. Beispiel (2) oben).

**Quasi-universelle Hypothesen** sind ähnlich wie (echt) universelle Hypothesen formuliert, beinhalten jedoch Einschränkungen, durch die die generellen Aussagen eher vage werden. Die quasi-universelle Hypothese soll nur mit einer (zumeist unbestimmten) Wahrscheinlichkeit gültig sein, Ausnahmen von den in der Hypothese aufgestellten Regeln sprechen nicht unbedingt gegen die Hypothese (siehe Beispiel (4) oben). Dieser Hypothesentyp ist in der Psychologie sehr häufig anzutreffen. Groeben und Westmeyer vermuten, daß Psychologen zwar einerseits generelle Aussagen von unbedingter Gültigkeit aufstellen möchten, andererseits jedoch die Generalität nicht einlösen können.

**Statistische Hypothesen** sind Hypothesen über Verteilungen von Werten bzw. über z.B. Mittelwerte, Varianzen, Korrelationen von Datenreihen (evtl. auch ohne Verteilungsannahmen; vgl. Beispiel (5) oben).

**Beschränkte universelle Hypothesen** beziehen sich einerseits auf alle Fälle einer Population von Individuen, führen aber Beschränkungen in Raum und Zeit ein (siehe Beispiel (3) oben). Groeben und Westmeyer versuchen die Generalisierungen der Sozialwissenschaften durch diese Hypothesenart zu rekonstruieren, die offenkundig eine universelle Gültigkeit beanspruchen, sich jedoch auf bestimmte „historisch-gesellschaftliche Rahmenbedingungen“ beschränken.

**Unbeschränkte universelle Hypothesen** beziehen sich auf alle Fälle einer bestimmten Art, haben keinerlei weitere Einschränkungen (vgl. Beispiel (6) oben).

Dazu noch einige Erläuterungen anhand des Meteorologenbeispiels, dem eine raumzeitliche Beschränkung anhaftet. Die vom Meteorologen implizit mitgedachte Ausformulierung „Wenn der Luftdruck fällt, regnet es!“, läßt verschiedene Einordnungen in das besprochene Klassifikationsschema zu: es handelt sich um eine (a) **quasi-universelle Hypothese**, wenn sie nur Wahrscheinlichkeitscharakter aufweist (wenn der Meteorologe unsicher ist); (b) **beschränkt universelle Hypothese**, wenn es räumliche und/oder zeitliche Einschränkungen gibt; (c) **unbeschränkt universelle Hypothese**, wenn keinerlei Einschränkungen in Ort, Zeit und Wahrscheinlichkeit mitgedacht sind.

Abschließend noch einige Bemerkungen zu den inhaltlichen Hypothesen aus unseren o. g. Beispielen. Vergleichen wir o. g. Hypothese (5) etwa mit Hypothese (4), so fällt auf, daß die erste eine **statistische** Hypothese ist (also Aussagen über statistische Parameter wie Mittelwerte, Korrelationen, Kovarianzen, Varianzen etc. beinhaltet), während die zweite eine **inhaltliche** Hypothese darstellt. Vergleichen wir weiterhin Hypothese (4) mit Hypothese (2), so fällt auf, daß sich Hypothese (2) auf konkret beobachtbare Dinge bezieht, Sachverhalte also, die empirisch beobachtbar sind, während Hypothese (4) sich auf Dinge bezieht, die eigentlich nicht empirisch beobachtbar sind. Die Begriffe, die in dieser Hypothese verwendet werden, sind eher Konstruktionen (theoretische Konstrukte), die als Erklärung für bestimmte Phänomene herangezogen werden.

Die aus Voruntersuchungen, eigenen Beobachtungen, Überlegungen bzw. aus Theorien abgeleiteten Vermutungen bezüglich des in Frage stehenden Untersuchungsgegenstandes bezeichnen wir als **Forschungshypothese**. . . . Der Forschungshypothese nachgeordnet ist die **operationale Hypothese**. Mit der operationalen Hypothese prognostiziert der Forscher den Ausgang einer konkreten Untersuchung (der natürlich im Einklang mit der allgemeinen Forschungshypothese stehen muß). (Bortz, 1984, S. 366)

Es erscheint uns sinnvoll, eine Trennung von theoretischen und empirischen Begriffen auch mit Blick auf die WHn durchzuführen. Daher wollen wir mit Hager (1984) eine solche Hypothese, die sich auf theoretische Begriffe und

Konstrukte bezieht, als **theoretisch-inhaltliche Hypothese** [TIH] bezeichnen. Hypothesen wie o.g. Hypothese (2) beziehen sich auf empirisch-beobachtbare Begriffe. Diese Art der Hypothese soll ebenfalls in Anlehnung an Hager (1984) als **empirisch-inhaltliche** Hypothese [EIH] bezeichnet werden. Die von den inhaltlichen Hypothesen abgegrenzten statistischen Hypothesen werden an dieser Stelle noch nicht weiter untergliedert. Von den inhaltlichen Hypothesen wollen wir annehmen, daß in ihnen im Regelfall psychologisch-inhaltliche Konzepte enthalten sind, daß statistische Konzepte wie Verteilungen, Mittelwerte o.ä. nicht in ihnen enthalten sind. Ausnahmen wären dann diejenigen psychologisch-inhaltlichen Hypothesen, die mathematische und/oder statistische Terme enthalten (z. B. Mathematische Modelle menschlichen Lernens etc.). Diese inhaltlichen Hypothesen lassen sich zwar ebenso unter o.g. Kategorien fassen, die dazu notwendigen Erläuterungen wurden uns jedoch zu weit vom eigentlichen Thema entfernen.

Die Unterscheidung in die zwei o. g. inhaltlichen Hypothesenebenen - je nach Art der verwendeten Begriffe - macht die eingangs genannten Äußerungen überdenkenswert: TIHn können das Eintreten bestimmter Ereignisse weder vorhersagen noch ausschließen, wenn diese Ereignisse beobachtbar sein sollen. Streng genommen haben TIHn keine direkte Verbindung mit beobachtbaren Sachverhalten, mit dem, was Forscher auch als Empirie bezeichnen. Dieses Problem wird uns später noch detaillierter beschäftigen, wenn wir die Frage beantworten wollen, wie WHn überprüft und gegebenenfalls als „falsch“ oder „richtig“ bezeichnet werden können.

## 1.4 Generierung von Hypothesen

Wie kommt ein Wissenschaftler eigentlich zu seinen Hypothesen? Eine allgemeingültige Anweisung dafür existiert nicht. Vielmehr muß man die wissenschaftliche Tätigkeit als Problemlösevorgang verstehen (vgl. z.B. Dörner, 1979; Hussy, 1983). Der wissenschaftlich tätige Psychologe interessiert sich für das **wie** und **warum** menschlichen Verhaltens und Erlebens. Er stellt Fragen und sucht nach Antworten. Hypothesen sind - wie gesehen - vorläufige Antworten auf solche Fragen, Sie sind Ergebnisse des Problemlösevorgangs (Lösungsmöglichkeiten), die überprüft werden müssen.

So kann man sich fragen, wie es kommt, daß kurzfristig nur eine begrenzte Zahl an Informationen behalten werden kann. Legt man einer Person eine siebenstellige Telefonnummer vor, so kann sie diese in der Regel für eine kurze Zeit behalten, etwa so lange, bis sie die Ziffern gewählt hat. Ist die Nummer deutlich länger, wird sie Schwierigkeiten bekommen. Ähnliche Schwierigkeiten wird sie haben, wenn die Telefonnummer nicht unmittelbar nach der Präsentation, sondern erst nach einer viertel Minute gewählt wird. Ein Wissen-

schaftler, der eine Vielzahl solcher und weiterer Beobachtungen angestellt hat, wird zu allgemeineren Vermutungen über das kurzfristige Behalten kommen, z. B. in Form der Hypothese: Wenn mehr als ca. sieben Informationseinheiten in unmittelbarer Abfolge präsentiert werden, dann kommt es bei der sich unmittelbar anschließenden Wiedergabe des Materials zu Fehlern. Der eigentliche Problemlösevorgang besteht hierbei in der Abstraktion der Gemeinsamkeiten der beobachteten Ereignisse. Es muß an dieser Stelle betont werden, daß Abstraktion hier im Sinne eines kognitiven Prozesses verstanden wird und keineswegs Bezug genommen wird auf die induktive Logik im Sinne von Carnap (1936; vgl. auch Westermann und Gerjets, Kapitel 10 in diesem Band). Dennoch bezeichnet man diese Art der Hypothesenfindung als **induktives Vorgehen**, weil von speziellen Einzelereignissen zu allgemeineren Vermutungen übergegangen wird.

Umgekehrt verhält es sich beim **deduktiven Vorgehen**. Hierbei werden aus einer vorliegenden allgemeinen Theorie spezielle (neue) Vermutungen abgeleitet. Ein gutes Beispiel dafür ist die Untersuchung von Lepper, Greene und Nisbett (1973), die die Selbstwahrnehmung (kognitive Dissonanztheorie) betrifft. Gemäß der Theorie wird zwischen zwei (oder mehreren) Gedanken, Erlebnissen etc. (sogenannten „kognitiven Elementen“) dann eine Dissonanz entstehen, wenn diese zueinander im Widerspruch stehen. Zentraler Gedanke der Theorie ist dann, daß der Mensch in solchen Fällen eine Tendenz verspürt, diese Dissonanz zu reduzieren. Er wird Maßnahmen ergreifen, die ihm eine solche Dissonanzreduktion ermöglichen, er wird Handlungen oder möglicherweise Umbewertungen seiner kognitiven Elemente vollführen.

In einer typischen Untersuchung zu dieser Theorie nehmen Personen an einem Experiment teil, in welchem recht stupide Aufgaben zu erledigen sind. Ein Teil von ihnen wird dafür gut, der andere Teil schlecht bezahlt. Die schlechtbezahlten Personen berichten hinterher, daß ihnen das Experiment gefallen hat, die gutbezahlten dagegen finden es langweilig. Gemäß der Theorie wurde man vermuten, daß bei den schlechtbezahlten Personen eine kognitive Dissonanz besteht - so wenig Geld für eine so langweilige Tätigkeit -, die dadurch beseitigt wird, daß man die Situation im Nachhinein als doch ganz interessant erlebt. Lepper et al. fragten sich nun, ob auch das Gegenteil zutrifft. Sie leiteten aus der Theorie die Hypothese ab, daß das (gute) Bezahlen einer kurzweiligen Tätigkeit dazu führt, daß diese Tätigkeit im Nachhinein als weniger kurzweilig erlebt wird. Tatsächlich bestätigten die Ergebnisse der Untersuchung diese Vermutung.

Gleichgültig ob man den induktiven oder deduktiven Weg beschreitet (in den meisten Fällen wird es ohnehin eine Mischform sein), profitiert der Vorgang der intentionalen Hypothesengenerierung - wie jeder Problemlöseprozeß - von einem umfangreichen, wohlstrukturierten, problembezogenen Faktenwis-



sen auf seiten des Wissenschaftlers. Allerdings kann auch der Zufall eine entscheidende Rolle spielen, wenngleich in der Literatur solche Fälle selten berichtet werden. So geht die Theorie des klassischen Konditionierens von Ivan Pawlow auf eine Zufallsentdeckung zurück. Er war ursprünglich - als Physiologe - an der Verdauung, speziell der Sekretion der Speicheldrüsen interessiert, etwa an der Frage, wie lange es bis zur Speichelsekretion dauert, wenn ein Hund gefüttert wird. Im Zuge seiner Untersuchungen fand er, daß die Tiere mit zunehmender Vertrautheit mit der Fütterungssituation (das gleiche Fressen, der gleiche Napf, der gleiche Pfleger) sogar schon Speichel produzierten, bevor sie das Fressen im Maul hatten. Das ging so weit, daß die Speichelproduktion schon beim Sehen des Pflegers begann. Er interessierte sich weiter für dieses überraschende Phänomen, das er psychische Sekretion nannte, da er vermutete, daß es auf die mentalen Aktivitäten der Tiere zurückzuführen sei und leistete mit seiner daraus entwickelten Theorie der klassischen Konditionierung einen wesentlichen, ursprünglich nicht geplanten Beitrag zu den psychologischen Lerntheorien.

## 1.5 Die Hypothese im Forschungsprozeß

Nicht nur das Generieren von Hypothesen ist als ein Vorgang des Problemlösens verstehbar. Auch die Entwicklung von Theorien (vgl. Kapitel 8) und die Aufstellung und Abarbeitung von Forschungsprogrammen (vgl. Kapitel 6) stellen Problemlöseprozesse dar. Sie repräsentieren die übergeordneten Ziele, zu deren Erreichung wissenschaftliche Hypothesen einen zentralen Beitrag leisten. Letztere bestimmen die Richtung der Forschungsarbeit und leiten das wissenschaftliche Arbeiten eines jeden Forschers (vgl. Bredenkamp, 1980; Gadenne, 1976, 1984 oder Hager & Westermann, 1983).

Wie bereits besprochen und in Abbildung 1 veranschaulicht, können WHn aus mehr oder minder elaborierten Theorien oder gar Theorienetzen abgeleitet sein (deduktives Vorgehen). In diesem Falle (den man wohl als Idealfall psychologischer Forschung bezeichnen müßte) wird die Bewertung der WHn auch von Bedeutung für die Bewertung der Theorien sein, aus denen sie abgeleitet wurden. Wenn WHn in größere Konzeptionen eingebettet sind, wie dieses bei Theorien und - verstärkt - bei Theorienetzen der Fall ist, können die Einzelerkenntnisse, die sich im Laufe des Überprüfungsprozesses der WHn ergeben, zu einem umfassenderen Erkenntnisfortschritt auf Theorieebene führen. Die besprochene Arbeit von Lepper et al. (1973) verdeutlicht diese Überlegung.

Gemäß Abbildung 1 führt der zweite Weg direkt von den Beobachtungen im Interessen- bzw. Problembereich zu den Hypothesen, etwa deshalb, weil noch keine theoretischen Vorstellungen zur Fragestellung vorliegen (induktives

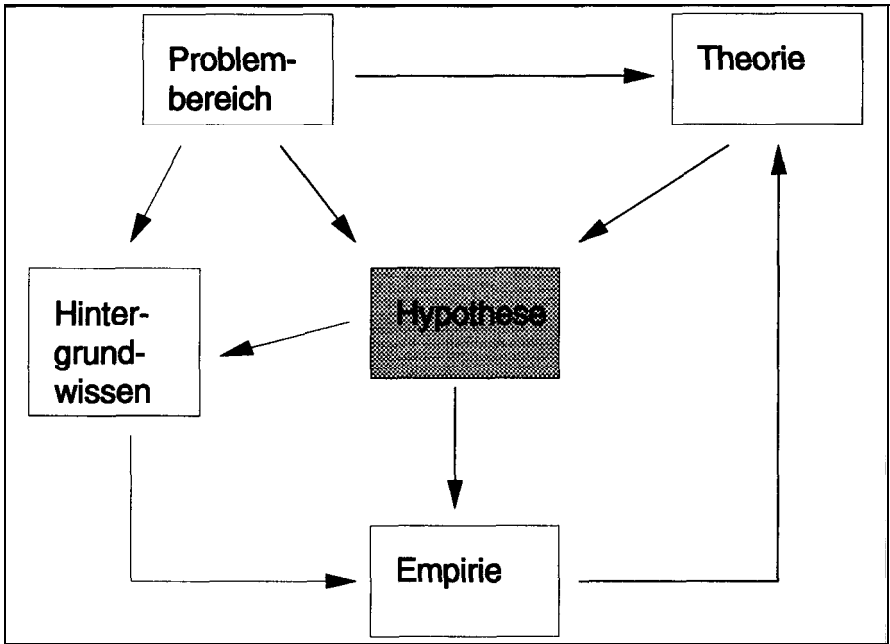


Abb. 1: Zusammenhänge von Theorie, Hypothese und Empirie

Vorgehen). In diesem Fall trägt die Hypothesenprüfung mit ihren Ergebnissen zur Schaffung einer soliden empirischen Datenbasis bei und bildet somit die Grundlage für die Entwicklung von Theorien. Vergewenärtigen wir uns das obengenannte Beispiel zum kurzfristigen Behalten. Aus einer Vielzahl von Einzelbeobachtungen zum fraglichen Gegenstandsbereich wurde eine WH gebildet und überprüft. Weitere diesbezügliche Untersuchungen (etwa zur Hypothese: „Wenn der Abruf von einmal präsentierten Informationen mehr als 15 Sekunden verzögert wird, dann kommt es zu Fehlern bei der Wiedergabe“) verbreitern die Ergebnisbasis und helfen beim Formulieren von Theorien. In diesem Beispiel führen Untersuchungen, die die Beibehaltung der Hypothesen stützen, zur Entwicklung des Konzepts des Kurzzeitgedächtnisses mit begrenzter Kapazität, sowohl bezüglich des Informationsumfangs als auch der Behaltensdauer.

Hypothesen haben eine weitere wichtige Funktion im Wissenschaftsprozess: Sie stellen die „theoretische Brille“ dar, durch die der Forscher seine empirische Arbeit sieht. Man kann die Auffassung vertreten, daß jede Beobachtung und Datenerhebung im Grunde genommen hypothesengeleitet erfolgt (wenngleich mit unterschiedlichem Explikationsgrad). Die Entscheidung des Forschers, an einer bestimmten Stichprobe von Versuchspersonen (VP bzw. VPn) ein spezifisches Treatment zu applizieren und genau definierte Maße empirisch

zu erheben, gründet sich immer auf mehr oder weniger klare Vorstellungen und Erwartungen, auf Vermutungen über Zusammenhänge der interessierenden Phänomene, die einerseits nichts anderes als Hypothesen sind (Hempel, 1974, S. 27), andererseits aber auch das Hintergrundwissen repräsentieren, welches durch die Fragestellung und die Hypothese (gegebenfalls auch durch die Theorie) aktiviert wird und auf die Gestaltung der empirischen Untersuchung zur Hypothesenprüfung - wie wir noch sehen werden - erheblichen Einfluß nimmt (vgl. Abbildung 1).

Generell kann man feststellen, daß Hypothesen im Rahmen des hypothetico-deduktiven Vorgehens das Bindeglied zwischen den Theorien und den empirischen Daten darstellen. Sie bilden die Grundlage für die Präzisierung der Fragestellung und führen zur Planung der Anlage der Untersuchung ebenso wie sie den Rückschluß von den statistischen Untersuchungsergebnissen auf die Ausgangstheorie (oder zu erstellende Theorie) bestimmen. Auf die dabei auftretenden Probleme werden wir im weiteren Verlauf des Textes eingehen. Sie gilt es in jedem Fall zu berücksichtigen, gleichgültig auf welcher Ebene eine Hypothese generiert wird (TIH oder EIH bzw. deduktiv oder induktiv), oder mit Hilfe welcher quantitativen Forschungsmethode (Experiment, Korrelationsstudie) sie überprüft wird.

## ***2. Zur Logik der Überprüfung wissenschaftlicher Hypothesen***

Da eine Vermutung über etwas „richtig“ oder „falsch“ sein kann, stellen sich an dieser Stelle mindestens zwei Probleme:

- (a) Läßt sich einer WH „Richtigkeit“ bzw. „Falschheit“ attestieren, „trifft die WH zu oder trifft sie nicht zu“?
- (b) Wenn ja, wie könnte der Überprüfungsprozeß aussehen bzw. wie lassen sich WHn überprüfen?

Unter (a) lassen sich die Vorbedingungen für die Überprüfbarkeit von Hypothesen zusammenfassen. Wir gehen im folgenden Abschnitt kurz darauf ein. Im Anschluß daran beschäftigen wir uns ausführlicher mit dem in (b) angesprochenen Überprüfungsprozeß, der eine Reihe von Ableitungsschritten umfaßt.

### **2.1 Widerspruchsfreiheit und Operationalisierbarkeit von Hypothesen**

Bevor man sich dem in der Regel nicht unerheblichen Aufwand unterzieht, eine WH empirisch zu überprüfen, sollte man sich vergewissern, daß einige Vorbedingungen eingehalten sind, deren Verletzung die Ergebnisse einer Un-

tersuchung wertlos machen bzw. eine sinnvolle Untersuchung erst überhaupt nicht ermöglichen. Es zählen dazu Überlegungen zur **Formulierung** von Hypothesen sowie zur prinzipiellen **Widerlegbarkeit** einer Hypothese und der **Operationalisierbarkeit** der darin enthaltenen Konzepte.

Der Art der Formulierung von Hypothesen wird generell ein großer Freiraum eingeräumt. Einem gewissen Standardformat entsprechen die „wenn,dann“-Hypothesen oder „je,desto“-Hypothesen (McGuigan, 1968, S. 313f.). Zum einen offenbaren diese Formulierungen in besonderer Weise den vorläufigen Charakter von Hypothesen und zum anderen differenzieren sie (bezogen auf die empirische Ebene) bereits zwischen unabhängigen und abhängigen Variablen (Kausalhypothesen) bzw. Prädiktor und Kriterium (Zusammenhangshypothesen). Wichtiger als diese sprachlichen Aspekte ist die Gefahr der widersprüchlichen Formulierung, wengleich diesem Problem selten viel Platz eingeräumt wird. Dieses mag u.a. damit zusammenhängen, daß bei einfachen Hypothesen ein Widerspruch in sich sofort auffällt. Allerdings muß bereits in solchen einfachen Fällen darauf geachtet werden, daß Widerspruchsfreiheit auch bzgl. der Zusatzannahmen (Hintergrundwissen, vgl. Abschnitt 2.2) besteht, welche bei der Operationalisierung der WH benötigt werden. Einen schwierigen Schritt stellt die Überprüfung der Widerspruchsfreiheit vor allem bei komplexen Hypothesen oder Hypothesengeflechten dar (vgl. z.B. Hussy und von Eye, 1988). Die in jüngerer Zeit erfolgte Entwicklung von Strukturgleichungsmodellen (z. B. EQS, Bentler, 1986; LISREL, Jöreskog und Sörbom, 1988) ermöglicht eine simultane Überprüfung mehrerer komplexer statistischer Hypothesen abgeleitet aus sogenannten Kausalmodellen. Neue und weiterführende Einsichten zur Lösung der Problematik der Überprüfung von Hypothesengeflechten scheinen daraus jedoch nicht zu resultieren (z. B. Möller, 1991).

Einige Überlegungen sollten auch zur prinzipiellen Widerlegbarkeit von Hypothesen angestellt werden. Daß die These „Wenn der Hahn kräht auf dem Mist, ändert sich das Wetter, oder es bleibt wie es ist“ nicht widerlegbar ist, liegt auf der Hand (vgl. Huber, 1987, S.53). Trifft eine solche Hypothese durch die Art ihrer Formulierung praktisch immer zu, so gibt es keinerlei empirische Sachverhalte, die zu den in der Hypothese behaupteten Sachverhalten in Widerspruch stehen können. Man sagt auch, daß die betreffende Hypothese keinen **empirischen Gehalt** hat. Der empirische Gehalt einer Hypothese wird häufig über die Anzahl von Falsifikatoren einer Hypothese erfaßt (je mehr mögliche Falsifikatoren eine Hypothese hat, desto höher ist ihr empirischer Gehalt; vgl. Opp, 1976).

Die prinzipielle Widerlegbarkeit ist auch dann gefährdet, wenn die in der Hypothese enthaltenen Begriffe nicht bzw. nicht eindeutig operationalisierbar sind. Die Hypothese „Erweitertes Bewußtsein erhöht die Fähigkeit zur De-

fokussierung der Aufmerksamkeit und mindert die Fähigkeit zur Fokussierung der Aufmerksamkeit“ ist solange prinzipiell nicht widerlegbar (und bedarf von daher keiner weiteren empirischen Überprüfung) als nicht eindeutig angegeben werden kann, was unter Bewußtsein verstanden und wie es erfaßt wird. Man spricht in diesem Fall auch davon, daß die Hypothese nicht mit empirischen Daten konfrontiert werden kann (vgl. Schulz, Muthig & Koeppeler, 1981, S. 30).

Abschließend sei darauf hingewiesen, daß natürlich auch die in der Hypothese gemachten Einschränkungen (bzw. das Fehlen solcher) und die logische (Re-)Konstruktion der Hypothese einen erheblichen Einfluß auf die prinzipielle Widerlegbarkeit (bzw. Belegbarkeit, Verifikation) der Hypothese haben: Je nach Art (vgl. Abschnitt 1.3) wird eine Hypothese z.B. eventuell überhaupt nicht widerlegbar (dieses gilt z.B. für die o. g. unbestimmten Existenzhypothesen) bzw. belegbar sein.

## 2.2 Die Überprüfung wissenschaftlicher Hypothesen

Nach diesen kurzen Vorbemerkungen zu den Vorbedingungen für die überprüfbarkeit wissenschaftlicher Hypothesen kommen wir nun zur Frage (b), wie nämlich der Überprüfungsprozeß aussieht. Diese Frage wollen wir zunächst unter Rückgriff auf Abschnitt 1.3 und die dort dargelegten Konzepte der verschiedenen Hypothesenarten nach Groeben und Westmeyer (1981) behandeln. Daß die Überprüfungen von Hypothesen immer nur unter Rückgriff auf ein spezifisches Hintergrundwissen möglich sind, soll zunächst weiter ausgeklammert bleiben.

Obleich sowohl singuläre als auch pseudo-singuläre Hypothesen als wissenschaftliche Hypothesen zu gelten haben, wollen wir Überprüfungsmöglichkeiten beider Arten hier nicht behandeln. Sie werden in der Psychologie zu meist als Beschreibungen (adverbialer bzw. adjektivischer Art) angewendet, es wird z.B. die Testleistung eines Probanden in einer bestimmten Aufgabe ausgedrückt.

Die unbestimmten Existenzhypothesen („Es gibt ein psycho-physiologisches Korrelat der menschlichen Emotion“) lassen sich prinzipiell bestätigen: Es reicht der Nachweis, daß die Hypothese bei einem einzigen Individuum bzw. einer einzigen Instanz gilt. Andererseits lassen sich unbestimmte Existenzhypothesen **nicht** widerlegen, da es niemals auszuschließen ist, daß irgendwann einmal eine positive Instanz gefunden wird.

Bestimmte Existenzhypothesen sind demgegenüber durchaus (zumindest prinzipiell) beleg- sowie widerlegbar. Durch die Einschränkungen in Raum und/oder Zeit ist eine Untersuchung aller Einheiten, aller Fälle möglich.

Anders verhält es sich mit den o. g. quasi-universellen Hypothesen, die in der Psychologie häufig vorliegen: Durch die Vagheit der eigentlich generellen Hypothese ist eine strenge Prüfung **nicht** möglich. Man wird auf Hilfskonstruktionen und Zusatzannahmen angewiesen sein, die uns im folgenden Abschnitt näher beschäftigen werden.

Beschränkte universelle Hypothesen lassen sich ähnlich den bestimmten Existenzhypothesen prinzipiell untersuchen, da Einschränkungen im Raum-Zeit-Bereich gemacht werden. Beispiele für solche Überprüfungen sind Meinungsumfragen, Einstellungsuntersuchungen etc., in denen die gesamte (endliche und abgeschlossene) Menge von Objekten untersucht wird.

Die unbestimmten universellen Hypothesen (die Allaussagen formuliert haben) sind (streng genommen) nicht zu belegen, sondern nur zu widerlegen. Da sie keinerlei Einschränkungen kennen, ist der Überprüfungsprozeß eigentlich nie abgeschlossen. Selbst wenn bislang bei allen Untersuchungen sich die Hypothese bewähren konnte, so ist prinzipiell nicht auszuschließen, daß nicht doch irgendwann eine negative Instanz auftritt. Das Auftreten einer solchen negativen Instanz reicht dann zur Widerlegung der Hypothese aus. Hier liegt insofern eine Umkehrung der unbestimmten Existenzhypothese vor.

Bei vielen WHn ist es (zumindest in der Psychologie) nicht immer leicht, sie als eine solche Art von Hypothese zu rekonstruieren, wie sie oben beschrieben werden. Daher läßt sich auch das Problem der Überprüfung einer WH immer nur unter dem Blickwinkel der spezifischen Rekonstruktion der betreffenden WH sehen („Diese Hypothese scheint die Struktur einer unbestimmten Existenzhypothese zu haben, und daher folgt für ihre Überprüfung, daß ...“). In der Psychologie finden wir sehr häufig die bereits erwähnten quasi-universellen Hypothesen (Wahrscheinlichkeitshypothesen). Da eine streng-logische Überprüfung nicht möglich ist, werden andere Wege beschritten werden müssen, über die wir nun berichten.

### ***2.2.1 Der Weg von den Inhalten zur Statistik***

Wir wollen zunächst die o.g. Hypothese (4) als Beispiel heranziehen: „Frustrierte Personen reagieren zumeist aggressiv.“ Diese Hypothese ist einerseits eine **inhaltliche** Hypothese, da sie auf inhaltliche Konzepte (Frustration, Aggression) abzielt. Andererseits ist sie eine **theoretisch-inhaltliche Hypothese**, da sich bei genauerer Betrachtung die o.g. inhaltlichen Konzepte als nicht beobachtbar (zumindest im wissenschaftlichen Sinne) herausstellen. Wollen wir das potentielle Zutreffen oder Nicht-Zutreffen dieser Hypothese möglichst stringent wissenschaftlich überprüfen, so werden wir die Hypothese mit der Realität, der Empirie, konfrontieren müssen. Dazu müssen wir vorher

darlegen, **was wir im Experiment** unter Frustration und Aggression verstehen wollen. Erst daran anschließend können wir die Hypothese überprüfen (z.B. durch ein entsprechend geplantes Experiment).

Zunächst sind die unabhängigen und abhängigen Variablen, die als theoretische Begriffe in der Hypothese enthalten sind, auf der Basis des vorliegenden Hintergrundwissens so festzulegen, daß sie einer Beobachtung, Messung und Erfassung zugänglich werden. Dieses Vorgehen wird als **Operationalisierung** bezeichnet (vgl. Hager, 1987, S. 53) und stellt einen der wichtigsten, aber auch schwierigsten Schritte in der psychologischen Forschung dar. Da die theoretischen Konzepte sich einer Beobachtung entziehen, müssen wir ihnen empirische Konzepte zuordnen.

Die Zuordnung von empirischen zu theoretischen Konzepten ist nicht möglich ohne (teils erhebliche) Reduktion des semantischen Gehaltes der theoretischen Konzepte, in einigen Fällen werden die vorgeschlagenen oder durchgeführten Operationalisierungen bei keinem zweiten Forscher Zustimmung finden. Dieses kann u. a. schon daraus resultieren, daß viele theoretische Konzepte einen nur sehr schwer zu definierenden semantischen Gehalt haben. Es scheint zwar in den meisten Fällen eine Art stiller Übereinkunft zwischen verschiedenen Forschern zu existieren, was sich theoretisch z.B. unter „Aggression“ verstehen läßt, wenn es dann jedoch daran geht, für ein konkretes Experiment Maßnahmen zu planen, die zu einer „Aggression“ der VPn führen sollen, findet man keinen Konsens.

Die Zuordnungsbeziehung sollte idealerweise wissenschaftlich exakt definierbar sein, jedoch fanden bislang vorgebrachte Lösungsvorschläge keine allgemeine Zustimmung: Als ein Beispiel dafür sei an die Vorschläge des logischen Empirismus via Carnap erinnert, die auf eine Definition der theoretischen Konzepte über beobachtbare Variablen hinausliefen, wobei die Zuordnung über sogenannte Reduktionssätze erfolgen sollte (vgl. u.a. Westermann, 1987a, S.11-12). Inwieweit aktuelle Ansätze aus dem Bereich der mathematischen Statistik (z.B. Einbezug von meßtheoretischen Modellen, damit möglicherweise präzisere Definition der Zuordnungsbeziehung von sogenannten latenten und manifesten Variablen über Strukturgleichungsmodelle wie z.B. LISREL etc.) hier eventuell Abhilfen schaffen können, muß an dieser Stelle offen bleiben (vgl. aber den Beitrag von Steyer in diesem Band, Kapitel 15). Bei letztgenannten Ansätzen scheint u.a. die Beziehung zwischen den theoretischen Konzepten und den latenten Variablen der Strukturmodelle unklar.

Die Operationalisierung erfolgt unter Bezugnahme auf das sogenannte Hintergrundwissen (Abbildung 1, vgl. u.a. Schulz, Muthig & Koepler, 1981), einem Wissensfundus, der dem Forscher zur Verfügung steht (oder zumindest stehen sollte). Dieses Hintergrundwissen enthält die zur Zeit in einem spezifischen Wissenschaftsbereich bekannten Gesetzmäßigkeiten, die zur Verfü-

gung stehenden technischen Voraussetzungen, weitere Vorannahmen, experimentelle Techniken, empirische Resultate etc., die in den Prozeß der Operationalisierung der theoretischen Variablen einmünden. Dieses Wissen ist selbst nicht Gegenstand der Überprüfung der WHn, es wird praktisch (ungeprüft) als zutreffend vorausgesetzt.

Damit ist einerseits ein praktischer Vorteil für jeden Forscher verbunden: Er muß nicht jedesmal erneut nachweisen, daß z.B. die Technik des „Freien Reproduzierens“ zur Operationalisierung der Behaltensleistung einer VP tauglich ist, sondern kann dieses Wissen, welches er aus anderen Studien, Untersuchungen etc. hat, ungeprüft auf seine Forschung anwenden. Andererseits ist es durchaus möglich, daß sich in dem Hintergrundwissen fehlerhafte Annahmen etc. befinden, die von dem Forscher jedoch ungeprüft übernommen werden und so potentiell zu falschen Bewertungen seiner WHn führen können.

Für die gewählten Operationalisierungen der theoretischen Begriffe werden (aus der TIH) empirisch-inhaltliche Hypothesen abgeleitet. Im einfachsten Falle jeweils nur einer Operationalisierung pro theoretischer Variablen resultiert auch nur eine inhaltliche Hypothese, die auf diese Operationalisierungen Bezug nimmt. Fehler, die in diesem frühen Stadium der Überprüfung der WH gemacht werden, lassen sich im folgenden kaum noch korrigieren. Daher muß dieser Schritt sorgfältig überdacht und durchgeführt werden.

Die dem Experiment und den dort realisierten empirischen Sachverhalten zugeordneten Hypothesen haben wir oben als empirisch-inhaltliche Hypothesen gekennzeichnet. Es sei hier darauf hingewiesen, daß für ein theoretisches Konzept durchaus verschiedene Operationalisierungen möglich sind und in der Praxis auch angewendet werden (vgl. z.B. für die „Behaltensleistung“ die experimentellen Erhebungsmethoden des „Freien Reproduzierens“, des „Wiedererkennens“ oder des „Wiedererlernens“). Daher ist es durchaus möglich, daß einer TIH in verschiedenen Experimenten unterschiedliche inhaltliche Hypothesen zugeordnet werden, die sich auf die unterschiedlichen Operationalisierungen beziehen. Die TIH kann so über verschiedene EIHn durchaus unterschiedliche Bewertungen erfahren. Es folgt hieraus unmittelbar die Anforderung an den Forscher, sowohl seine TIHn wie auch die gewählten EIHn im Rahmen des Forschungsberichtes explizit zu machen, damit dem Rezipienten des Berichtes ein kritisches Nachvollziehen ermöglicht wird.

In der TIH „Frustrierte Personen reagieren **zumeist** aggressiv“ taucht der Begriff „**zumeist**“ auf. Diesen müssen wir bei der Ableitung der EIH ebenso berücksichtigen, wie die Begriffe „Frustration“ und „Aggression“. Es sind nicht nur die Konzepte oder Variablen der TIH zu operationalisieren, um eine EIH abzuleiten. Die zwischen den theoretischen Begriffen bestehenden Relationen müssen ebenso auf die EIH übertragen werden. Das bedeutet an



dieser Stelle, daß eine Umschreibung und Klärung des Begriffes „zumeist“ erforderlich ist: Wir müssen darlegen, **was wir unter „zumeist“ verstehen wollen.**

So wäre ein Experiment mit zwei Gruppen denkbar, wobei in Gruppe 1 die VPn z.B. durch unausweichbaren Mißerfolg frustriert werden<sup>1</sup>, während dieses in der Gruppe 2 nicht der Fall ist. Es müßte ein Maß für Aggression erhoben werden<sup>2</sup> (Sprechgeschwindigkeit, Anzahl von Verbalinjurien pro Zeiteinheit ...), welches in Gruppe 1 höher liegen sollte als in Gruppe 2. Um den Begriff „zumeist“ in dem Experiment zu definieren, muß an dieser Stelle festgelegt werden, was „höher liegen“ bedeuten soll (alle VPn der Gruppe 1 haben höhere Werte als die der anderen Gruppe; von den 10 höchsten Werten kommen 75 % aus Gruppe 1 ...?).

Die EIH ist eine **inhaltliche** Hypothese, die sich auf die gewählten Operationalisierungen bezieht. In unserem Beispiel wurde die EIH bspw. lauten können: „Die Personen mit Mißerfolgserlebnissen zeigen in der sich anschließenden Diskussion durchschnittlich mehr Verbalinjurien als die Personen ohne induzierte Mißerfolgserlebnisse (wobei die Anzahl von Verbalinjurien innerhalb der Zeit von x Minuten von y unabhängigen und geschulten Beurteilern auf der Kategorienskala Z erhoben wird).“ Auf der Ebene dieser Hypothese werden empirische Daten gewonnen, die zur Beurteilung der Hypothese herangezogen werden sollen. Die Beurteilung der EIH anhand der Daten könnte nach vorher festgelegten, **inhaltlichen** Kriterien erfolgen. Leider liegen solche Kriterien (zumindest derzeit) kaum vor.

Die empirischen Daten psychologischer Experimente sind im allgemeinen fehlerbehaftet, d.h. es wird in aller Regel eine Diskrepanz zwischen den empirisch erhobenen Werteausprägungen und den zugrunde liegenden (wahren) Werten auftreten, die viele Gründe hat (z.B. könnte eine VP müde und unkonzentriert sein und wurde so ihre wirkliche Leistungsfähigkeit nicht zeigen können, sie wurde im psychologischen Test bzw. im psychologischen Experiment einen schlechteren Wert erhalten als unter normalen Bedingungen). Auf derartige Gründe wollen wir hier nicht näher eingehen, wohl aber auf eine Konsequenz, die daraus in den meisten psychologischen Untersuchungen gezogen wird.

Inhaltliche Hypothesen, die sich als quasi-universele Hypothesen rekonstruieren lassen, werden in der Psychologie zumeist **statistisch** überprüft. Haagen und Seifert sprechen demzufolge von einer „Operationalisierung der psychologischen Hypothese in statistische Konzepte“ (1979, S. 167). Ihnen werden **statistische Hypothesen** zugeordnet, die wir mit Hager (1987) als **statistische**

1 Operationalisierung des Begriffes „Frustration“

2 Operationalisierung des Begriffes „Aggression“

**Vorhersagen** [SV bzw. SVn] bezeichnen wollen, um in diesem Zusammenhang ihre Funktion im Rahmen des Überprüfungsprozesses einer inhaltlichen Hypothese anzudeuten. Über die SV wird dann in aller Regel mit Hilfe eines oder mehrerer Signifikanztests entschieden. Die in diesem Zusammenhang somit nachgeordneten Signifikanztests testen sehr spezifische statistische Hypothesen, die wir als **Testhypothesen** [ $H_0$ ,  $H_1$ ] bezeichnen wollen. Die Entscheidung über die SV erfolgt im günstigen Fall über einen Signifikanztest, dessen Null ( $H_0$ ) oder Alternativhypothese ( $H_1$ ) der interessierenden SV entspricht. Andernfalls sind mehrere Tests durchzuführen, die einzelnen Entscheidungen über die Null und/oder Alternativhypothesen der Tests sind über logische Verknüpfungen zur Entscheidung über die SV heranzuziehen.

Obgleich eine Vielzahl von SVn als Zuordnung zu den EIHN denkbar sind, werden in der Psychologie nahezu ausschließlich sogenannte Mittelwertshypothesen (unter der sogenannten Normalverteilungsannahme) statistisch überprüft (vgl. auch nächster Abschnitt).

### **2.2.2 Die Beziehung von inhaltlichen Hypothesen und statistischen Vorhersagen**

Wenn statistische Vorhersagen zur Überprüfung inhaltlicher Hypothesen herangezogen werden, wird man in der Regel die Beziehung beider Hypothesenarten näher zu betrachten haben, um die Art und Aussagekraft der Überprüfung beurteilen zu können. Blicken wir nochmals auf o.g. Beispiele (4) und (5), so wird wohl unmittelbar einsichtig, daß sich beide Hypothesen auf unterschiedliche Dinge und Sachverhalte beziehen. Inhaltliche Hypothesen [Beispiel (4)] haben zumeist (offene) Populationen von **Individuen** als Grundlage, die Gültigkeit der Hypothese wird für alle Individuen der Population behauptet. Diese Individuen sind die **Merkmalsträger**, eine Menge von Personen oder Tieren (VPn, Versuchstiere), denen durch die Hypothese eine spezifische Eigenschaft zugesprochen wird (z.B. aggressiv zu sein).

Statistische Vorhersagen [Beispiel (5)] beziehen sich auf statistische Konzepte, sie haben in diesem gewählten Kontext **Daten** zum Gegenstand, die (isoliert betrachtet) ohne inhaltliche Bedeutung sind und nur bestimmten Auflagen (z.B. normalverteilt zu sein) zu folgen haben. Diese Auflagen stehen in der Regel ohne Verbindung zur WHn und sind auch nur dann notwendig, wenn z.B. (theoretische) Modellparameter im Rahmen statistischer Modelle geschätzt werden sollen (sogenannte Reparametrisierungsbedingungen) oder um Testverteilungen mathematisch zu definieren, die einen (parametrischen) Signifikanztest überhaupt erst ermöglichen (sogenannte Verteilungs- und Varianzhomogenitätsannahmen etc.).

Aus den inhaltlich-psychologischen Hypothesen der hier beschriebenen Art werden normalerweise keine Kriterien hervorgehen, die eine spezifische Zuordnung statistischer Vorhersagen zu den inhaltlichen Hypothesen möglich machen würde. Solche inhaltlich-psychologische Aussagen lassen sich „unmöglich auf eine Aussage über Populationsmittelwerte“ reduzieren (Gadanne, 1984, S. 107), sie lassen sich allenfalls „auf dem Umweg über eine Populationsaussage“ testen (Gadanne, ebd.). Wir haben weiter oben davon gesprochen, daß die **Relationen**, die zwischen den Variablen der TIH behauptet werden, möglichst direkt auf die EIH übertragen werden sollten. Dieses ist zumeist relativ problemlos möglich („ist größer“, „ist gleich wie“, „erniedrigt die Werte“...). Die Relationen lassen sich ebenso auf die SV übertragen ( $>$ ,  $=$ ,  $A < B$ ). Gelingt eine solche Übertragung der Relationen, so spricht Hager (1984) von einer **adäquaten** Zuordnung der SV zur EIH.

Wie wir jedoch gesehen haben, ist eine direkte Übertragung der theoretischen oder empirischen **Variablen** der inhaltlichen Hypothesen auf statistische Variablen der SV nahezu unmöglich. Daraus folgt, daß wir die Überprüfung inhaltlicher Hypothesen durch eine Entscheidung über statistische Vorhersagen als sehr **indirekt** kennzeichnen wollen (z.B. Gadanne, 1984; Schulz, Mutzig & Koepler, 1981). Nicht nur die Beurteilung der TIH kann in verschiedenen Experimenten unterschiedlich sein in Abhängigkeit von der gewählten EIH, sondern auch die Bewertung einer konkreten EIH in Abhängigkeit von der gewählten Zuordnung der SV zu der EIH. In der Forschungspraxis wird diese Zuordnung zudem häufig durch die Verwendung spezifischer Signifikanztests determiniert (Testhypothese und statistische Vorhersage werden nicht unterschieden, die statistische Vorhersage wird nicht mit Bezugnahme auf die EIH kritisch reflektiert und dieser zugeordnet). Wir werden im Rahmen der Validitätsproblematik auf dieses Problem zurückkommen.

Diese nur sehr indirekte Beziehung beider Hypothesenarten wurde in der Vergangenheit des öfteren kritisiert, die Funktion der Statistik in der Psychologie in Frage gestellt. Auch Gadanne (1984) berührt diesen Punkt, wenn er schreibt:

Ungeklärt ist bereits, wie die in statistischen Tests geprüften Hypothesen mit den psychologischen Theorien, um deren Beurteilung es geht, logisch zusammenhängen. Die Kenntnis dieses Zusammenhanges ist Voraussetzung für die Beantwortung der Frage, welche Konsequenzen die Annahme oder Ablehnung einer statistischen Prüfungshypothese für die Beurteilung einer psychologischen Theorie haben sollte. (S. 104)

Zur Lösung dieses Problems wurde u.a. gefordert, daß zwischen EIH und SV eine (logische) **Implikationsbeziehung** [Wenn A, dann B; Aus A folgt B] bestehen (z.B. Bredenkamp, 1980), oder daß die SV folgerichtig aus der EIH abgeleitet sein sollte (Meehl, 1967).

Die EIH sollte über das Prinzip des *modus tollens* falsifizierbar werden, wenn sich die SV als falsch herausstellen sollte. Diese Falsifikation wurde sich jedoch logisch nur dann rechtfertigen lassen, wenn eine Implikationsbeziehung zwischen EIH und SV bestünde. Eine Bestätigung der EIH wäre danach allerdings nicht möglich bzw. allein mit logischen Argumenten nicht zu rechtfertigen. Leider scheint zumindest aus der Sicht der Forschungspraxis die Annahme einer solchen Beziehung zwischen EIH und SV kaum realistisch zu sein. So erfolgt bspw. die Ableitung und Überprüfung inhaltlicher Hypothesen immer auch unter Rückgriff auf das Hintergrundwissen (vgl. Abschnitt 2.2). Wenn überhaupt, so wäre nur die Verbindung aus inhaltlicher Hypothese und Hintergrundwissen (die logische Konjunktion) falsifizierbar, nicht die EIH allein. Überdies läßt sich die SV mit logischen Mitteln gar nicht falsifizieren, sondern es kann immer nur eine Entscheidung des Forschers über die Gültigkeit oder Ungültigkeit der SV getroffen werden. Da sich jedoch EIH und SV u.E. auf unterschiedlichen und nicht vergleichbaren Ebenen befinden, läßt sich selbst diese Entscheidung nicht direkt auf die EIH übertragen.

Eine Implikationsbeziehung wäre wohl nur dann möglich, wenn die psychologischen (inhaltlichen) Hypothesen anders aussähen als die empirisch-inhaltlichen Hypothesen (EIH), über die wir hier berichten. Die psychologischen Hypothesen [PH] sollten dann neben den inhaltlichen Aspekten auch Hinweise enthalten, die eine zweifelsfreie Übertragung auf die statistische Ebene ermöglichen wurden. In diese Form von PH werden also idealerweise statistische Konzepte einbezogen, sie stellen einen Bestandteil der PH dar (diese Auffassung der PH wird in diesem Band im Kapitel 14 von Erdfelder und Bredenkamp sowie im Kapitel 15 von Steyer vertreten). So gibt es in der Psychologie denn auch inhaltliche Hypothesen, die statistische Konzepte enthalten: Man denke nur an einige der Gesetzmäßigkeiten, die in der Psychophysik aufgestellt wurden. In den allermeisten Fällen sind jedoch Theorien und Hypothesen bei weitem nicht so exakt formuliert: Frustration schafft Aggression.

Aus Sicht der heutigen Forschungspraxis muß man u.E. befürchten, daß eine Forderung, PH „statistischer zu formulieren“, eher dazu führen wird, die PH dann gleich als SV<sub>n</sub> zu formulieren (Gleichsetzung von Inhalt und Statistik), um damit einigen Interpretationsproblemen aus dem Wege zu gehen. Dieses ließe sich fürwahr bei einigen Hypothesen der psychologischen Forschung bei oberflächlicher Betrachtung vermuten. Ein aufmerksames Lesen der als „Diskussion“ bzw. „Zusammenfassung“ ausgewiesenen Textstellen wissenschaftlicher Arbeiten offenbart jedoch fast ausnahmslos, daß die Autoren sich gerade nicht für die statistischen Parameter (und deren Schätzung) interessieren. Sie wollen in der Regel Aussagen über Verhaltensweisen, Wahrnehmungsleistungen, Denkvorgänge, Gedächtnisleistungen und dergl. ihrer VP<sub>n</sub> treffen.

Da wir die Beziehung zwischen inhaltlichen Hypothesen und statistischen Vorhersagen somit weiterhin als sehr indirekt kennzeichnen wollen und sich für diese Art psychologisch-inhaltlicher Hypothesen eine Implikationsbeziehung zur SV [wenn EIH, dann SV] in nahezu allen Fällen nicht ergeben wird, bliebe es zu erwähnen, daß der Signifikanztest **als Institution** im Rahmen psychologischer Forschung durchaus in Frage gestellt wird (z.B. Harnatt, 1975 oder Morrison & Henkel, 1970).

Wenn über SVn Entscheidungen getroffen werden sollen, ohne den Signifikanztest als (Hilfs)Kriterium zu nutzen, muß zunächst die Fehlerbehaftetheit psychologischer Daten ins Kalkül gezogen werden. Diese Entscheidungen sind wiederum mit spezifischen (zunächst aber unbekannt) **Fehlerrisiken** verbunden, welche wir jedoch kennen oder zumindest vorher abwägen sollten, bevor wir zu Entscheidungen kommen. Beide Problempunkte lassen sich jedoch im allgemeinen nicht aufgrund inhaltlicher Vorannahmen und Theorien lösen.

Die statistischen und mathematischen Theorien, die den bekannten Signifikanztests zugrunde liegen, ermöglichen dadurch, daß sie den Daten bestimmte Verteilungsfunktionen, Lokationsparameter, Streuungen usw. unterstellen, die Abschätzung der Fehlerrisiken, mit denen die Entscheidungen verbunden sind. Damit bietet der Signifikanztest Kriterien an, an denen wir uns orientieren können. Dadurch erhält er eine Rechtfertigung als Entscheidungskriterium in der psychologischen Forschung (vgl. auch Gadenne, 1984), wengleich an dieser Stelle keineswegs eine Institutionalisierung des Signifikanztests im Sinne einer „kochbuchartigen Standarddurchführung“ favorisiert werden soll.

Um die Abschätzung der Fehlerstreuungen und die Kalkulation der Fehlerrisiken zu erlangen, muß der Forscher allerdings bereit sein, den Daten ein spezifisches (mathematisch-statistisches) Datenmodell zu unterstellen und übernimmt damit ferner die Verpflichtung, für das Einhalten der Forderungen des adaptierten Datenmodelles Sorge zu tragen, es gegebenenfalls gegen ein eventuell geeigneter erscheinendes auszuwechseln. Statistische Entscheidungstheorien ermöglichen nur unter diesen Voraussetzungen die valide Abschätzung der Fehlerrisiken (der Entscheidungen).

Das Einhalten o. g. Verpflichtungen ist für den Forscher in einigen Fällen relativ leicht, in anderen Fällen kaum möglich. So ist es für den Forscher sehr schwierig, z.B. auf eine **Normalverteilung** der Daten (und noch dazu in einer meistens nicht bekannten Population) zu achten. Entspricht jedoch die Verteilung seiner Daten nicht der Verteilung, die das Modell des Signifikanztests voraussetzt, so werden die für den Signifikanztest als maximal tolerierbar geplanten Fehlerwahrscheinlichkeiten (die nominellen Werte) von den „realen“ Fehlerwahrscheinlichkeiten, die mit seinen Daten verbunden sind, abweichen.

Die Entscheidungsrationale unseres Forschers fußt in diesem Falle auf falschen oder nicht passenden Voraussetzungen, zumindest was die Verteilung der Daten anbelangt. Die Konsequenzen dieser Nichtbeachtung solcher Voraussetzungen des Signifikanztests wird allerdings je nach wissenschaftstheoretischer Einbettung anders bewertet, wobei diese Bewertungen mit einer durchaus unterschiedlichen Interpretation des Signifikanztests zusammenhängen (Westermann, 1987b; zusammenfassend Hager, 1987, S. 119-121).

Wenn wir nun an die unterschiedlichen Hypothesenebenen denken, können wir für jede Ebene (getrennt) mögliche Wahrscheinlichkeiten des Zutreffens der Bewertungen (TIH und EIH werden wir nicht weiter trennen) bzw. der Entscheidungen (SV und Signifikanztest) definieren, deren Zusammenhänge weitere Aufschlüsse über die Beziehungen von inhaltlichen und statistischen Hypothesen geben. Zur sprachlichen Vereinfachung wollen wir diese Wahrscheinlichkeiten im folgenden etwas ungenau als „Fehlerrisiken“ bezeichnen. Um die Beziehungen dieser Fehlerrisiken weiter zu untersuchen, übernehmen wir zunächst in Tabelle 2.1 bis Tabelle 2.3 die Definitionen von Westermann (1987a, S.38) und Hager (1987, S.77, 116 und 132). Im nächsten Abschnitt werden wir dann die Zusammenhänge der definierten Fehlerrisiken näher beleuchten.

Auf der (untersten) Ebene der Signifikanztests definieren sich mögliche Fehlerrisiken der Entscheidung zugunsten der  $H_0$  bzw.  $H_1$ -Hypothese wie in Tabelle 2.1 angegeben. Mit  $\alpha$  und  $\beta$  werden hier die Fehlerrisiken eines Signifikanztests bezeichnet. Es sind dieses die Wahrscheinlichkeiten von Fehlentscheidungen über die Testhypothesen des Signifikanztests.

Tabelle 2.1: Wahrscheinlichkeiten für richtige und falsche Entscheidungen in Signifikanztests

Entscheidung für	Wahr, aber unbekannter Sachverhalt	
	$H_0$ gilt	$H_1$ gilt
$H_0$	$1-\alpha$	$\beta$
$H_1$	$\alpha$	$1-\beta$

Auf der (nächsthöheren) Ebene der statistischen Vorhersage werden die Fehlerrisiken einer Entscheidung über die SV gemäß Tabelle 2.2 definiert. Mit  $\epsilon$  wird die Wahrscheinlichkeit einer fälschlichen Entscheidung zugunsten der SV, mit  $\phi$  die Wahrscheinlichkeit einer fälschlichen Entscheidung zuungunsten der SV bezeichnet.

Auf der (obersten) Ebene der EIH (und TIH) werden die Wahrscheinlichkeiten des Zutreffens der Bewertungen der inhaltlichen Hypothesen gemäß Tabelle 2.3 definiert. Mit  $f$  wird die Wahrscheinlichkeit bezeichnet, die EIH aufgrund der durchgeführten Untersuchung fälschlicherweise als nicht zutreffend zu charakterisieren, mit  $e$  die Wahrscheinlichkeit, die EIH fälschlicherweise als zutreffend zu bezeichnen.

Tabelle 2.2: Wahrscheinlichkeiten für richtige und falsche Entscheidungen über die statistische Vorhersage aufgrund eines oder mehrerer Signifikanztests

Entscheidung für	Wahrer, aber unbekannter Sachverhalt	
	SV gilt	SV gilt nicht
<b>SV</b>	$1-\varphi$	$\epsilon$
<b>¬SV</b>	$\varphi$	$1-e$

Tabelle 2.3: Wahrscheinlichkeiten für richtige und falsche Beurteilungen der empirisch-inhaltlichen Hypothese aufgrund der Entscheidungen über die SV sowie weiterer, inhaltlicher Kriterien

Beurteilung der EIH als	Wahrer, aber unbekannter Sachverhalt	
	EIH trifft zu	EIH trifft nicht zu
zutreffend	$1-f$	$e$
nicht zutreffend	$f$	$1-e$

Obleich die letztgenannten Wahrscheinlichkeiten  $e$  und  $f$  numerisch nicht bestimmbar sind, da ihre Ausprägungen zusätzlich noch von anderen Faktoren als nur den statistischen Entscheidungen abhängen (z. B. der internen Validität der Untersuchung), hängen sie doch mit den Fehlerrisiken  $\epsilon$  und  $\varphi$  und letztendlich mit  $\alpha$  und  $\beta$  zusammen (vgl. auch nächster Abschnitt). Eine Erhöhung der  **$\epsilon$ -Fehlerwahrscheinlichkeit** wird (andere Faktoren als konstant angenommen) in einer Erhöhung der  $e$ -Fehlerwahrscheinlichkeit resultieren und eine Erhöhung der  **$\varphi$ -Fehlerwahrscheinlichkeit** wird eine Erhöhung der  $f$ -Fehlerwahrscheinlichkeit nach sich ziehen. Die vier Wahrscheinlichkeiten der statistischen Ebene ( $\epsilon$ ,  $\varphi$ ,  $\alpha$  und  $\beta$ ) lassen sich (unter Vorbehalt, vgl. 3.3) numerisch bestimmen (vgl. dazu Hager, 1987). Der Signifikanztest macht die genauere Abschätzung von  $\alpha$  und  $\beta$  möglich (bei unexakten Hypothesen wie z.B. „ $\mu_1 > \mu_2$ “ eventuell als Maximalwahrscheinlichkeit; vgl. u. a. Hays, 1978). Damit kann es ebenso ermöglicht werden,  $\epsilon$  und  $\varphi$  zu bestimmen. Schließlich wird darüber eine gewisse Kontrolle der Fehlerwahrscheinlichkeiten  $e$  und  $f$  realisierbar.

### 2.2.3 Die Fehlerkontrolle von $\alpha$ und $\beta$

Wenn wir im folgenden einmal annehmen, wir hätten zwei Gruppen von VPn jeweils unterschiedlich frustriert (vgl. oben), die zweite nicht und die erste über irgendeine spezifische Technik im Rahmen der Untersuchung, so wurde gemäß der übergeordneten TIH „Frustrierte VPn reagieren aggressiver“ zu erwarten sein, daß die VPn der ersten Gruppe mit erhöhter Aggressivität reagieren. Wenn wir z.B. die „Anzahl von Verbalinjuriem gegen den Versuchsleiter innerhalb einer bestimmten Zeiteinheit“ [AVV] als Operationalisierung des Begriffes „Aggression“ definieren (vgl. Bredenkamp, 1980), sollte dieser Wert in der ersten Gruppe höher sein als in der zweiten (abgekürzt:  $AVV_1 > AVV_2$ ). Wir ordnen dieser EIH die statistische Vorhersage zu, daß der Mittelwert der abhängigen Variablen in der ersten Gruppe **größer** ist als in der zweiten:

$$\text{EIH: } AVV_1 > AVV_2 \Rightarrow \text{SV: } \mu_{AVV_1} > \mu_{AVV_2} \quad (2.1)$$

Das Zeichen „ $\Rightarrow$ “ stellt **keine logische Implikation** dar, sondern soll nur die o. g. Zuordnungsbeziehung verdeutlichen. über das mögliche Zutreffen oder Nicht-Zutreffen der SV läßt sich mit einem einfachen, gerichteten t-Test entscheiden, da diese SV der Alternativhypothese eines gerichteten t-Tests entspricht:

$$\text{SV: } \mu_{AVV_1} > \mu_{AVV_2} \equiv H_1: \mu_1 > \mu_2 \quad (\text{t-Test}) \quad (2.2)$$

Das Zeichen „ $\equiv$ “ deutet eine **Äquivalenzbeziehung** von SV und der Alternativhypothese des t-Tests an. Damit wird ebenso verdeutlicht, daß die SV zwar explizit nur eine Hypothese über zwei Mittelwerte darstellt, jedoch implizit auch spezifische weitere Annahmen (streng genommen damit weitere Hypothesen) macht, die nicht genannt werden (z.B. die Annahme der Normalverteilung etc.). Erst durch diese zusätzlichen Annahmen wird sie zu einer **testbaren** Hypothese. Diese zusätzlichen Annahmen folgen in den allermeisten Fällen nicht aus den inhaltlichen Theorien oder Hypothesen, sondern sind, wie bereits oben erwähnt, mehr technischer Natur.

Es läßt sich nun zeigen (vgl. Hager, 1987), daß die Fehlerwahrscheinlichkeit  $\varphi$  (fälschliche Ablehnung der SV) hier identisch ist mit der Fehlerwahrscheinlichkeit  $\beta$ , die sich aus dem gerichteten t-Test ergibt. Ferner ist die Fehlerwahrscheinlichkeit  $\epsilon$  (fälschliche Annahme der SV) numerisch gleich der Fehlerwahrscheinlichkeit  $\alpha$ . Wenn wir davon ausgehen, daß alle anderen Faktoren als konstant vorausgesetzt werden können, so resultieren daraus folgende Konsequenzen:



- (1) Ein **nicht-signifikantes** Ergebnis des t-Tests wurde die Ablehnung der  $H_1$ -Hypothese zur Folge haben. Daraus wurde die Entscheidung resultieren, die SV als nicht zutreffend zu bezeichnen. Eine solche Entscheidung kann (und wird im allgemeinen) die Beurteilung zur Folge haben, daß sich die EIH (zumindest in der vorliegenden Untersuchung) nicht bewähren konnte. Je größer also die Fehlerwahrscheinlichkeit  $\beta$  wird, desto größer wird damit auch die Wahrscheinlichkeit, daß wir die SV fälschlicherweise ablehnen (9). Dadurch wächst notgedrungen auch die Wahrscheinlichkeit  $f$  einer fälschlichen Beurteilung der EIH als „unzutreffend“ bzw. „nicht bewährt“.
- (2) Ein **signifikantes** Ergebnis des t-Tests wurde die Ablehnung der  $H_0$ -Hypothese zur Folge haben. Daraus würde die Entscheidung resultieren, die SV als zutreffend zu bezeichnen. Diese Entscheidung wiederum kann (und wird im allgemeinen) die Beurteilung zur Folge haben, daß die EIH sich (zumindest in der vorliegenden Untersuchung) bewähren konnte. Je größer also die Fehlerwahrscheinlichkeit  $\alpha$  wird, desto größer wird damit auch die Wahrscheinlichkeit, daß wir die SV fälschlicherweise annehmen (E). Dadurch wächst notgedrungen auch die Wahrscheinlichkeit  $e$  einer fälschlichen Beurteilung der EIH als „zutreffend“ bzw. „bewährt“.

Setzen wir nunmehr voraus, daß die Überprüfung der interessierenden inhaltlichen Hypothesen mit möglichst geringen Fehlerisiken erfolgen sollte (**strenge** und faire Prüfungen; vgl. Hager & Westermann, **1983**), so folgt daraus, daß sowohl eine Kontrolle des  $\alpha$  **wie auch** des  $\beta$ -Fehlers des statistischen Tests erfolgen muß. Auf das eher technische Problem, **wie** diese Kontrolle praktisch zu erfolgen hat, gehen wir in diesem Zusammenhang nicht weiter ein (vgl. dazu Hager, 1987). Nötig wird eine **Simultankontrolle** der Determinanten des Signifikanztests, die Hager (**1987**) als **Testplanung** beschrieben hat. Dabei kann sich der „Testplaner“ die funktionalen Zusammenhänge der Determinanten des Signifikanztests (die sogenannten Teststärkefunktionen; vgl. Cohen, 1977; Hager & Möller, 1986) zunutze machen, um eine Kontrolle der Einflußfaktoren des Signifikanztests zu erreichen.

Da die Verbindung der hier dargestellten inhaltlichen zu den statistischen Hypothesen als nur sehr „lose“ (Meehl, 1967) charakterisiert wurde, wollen wir im nächsten Abschnitt beschreiben, warum es in Anlehnung an die Forschungspraxis günstiger ist, wenn wir von einer **Bewertung** bzw. **Beurteilung** der inhaltlichen Hypothesen (u. a. aufgrund der signifikanztheoretischen Entscheidungen über statistische Hypothesen) sprechen, die zudem nicht ohne subjektive Anteile ist. Der Überprüfungsprozeß führt somit nicht zu einer Verifikation oder Falsifikation der inhaltlichen Hypothesen.

## 2.3 Der Weg von der Statistik zurück zu den Inhalten

Wir haben den Weg der Überprüfung inhaltlicher Hypothesen dadurch umschrieben, daß wir von TIHn ausgegangen sind, aus denen EIHn abgeleitet werden. Diesen EIHn werden SVn adäquat (und **suffizient**; vgl. Hager, 1987) zugeordnet, die im Anschluß via Signifikanztests überprüft werden, um zu einer Entscheidung über die **Testhypothesen** einerseits und die übergeordneten SVn andererseits zu gelangen. Der Begriff der Überprüfung der inhaltlichen Hypothesen (d.h. TIH und EIH) wurde zugunsten des Begriffes der Bewertung der inhaltlichen Hypothesen aufgegeben.

Wie läßt sich nun verfahren, wenn wir auf der untersten Ebene des Signifikanztests zu Ergebnissen gekommen sind? Können wir von signifikanten oder nicht-signifikanten Resultaten auf die Gültigkeit oder Ungültigkeit der inhaltlichen Hypothesen schließen? Wie läßt sich (aus dem Blickwinkel der Testhypothesen) der „Weg zurück“ beschreiben?

Die Frage nach dem Weg zurück ist gleichbedeutend mit der Frage nach einer Bewertung der wissenschaftlich-inhaltlichen Hypothesen (TIH und EIH). Die Antwort auf diese Frage läßt sich ohne Einbezug unterschiedlicher Güte- oder Validitätskriterien nicht geben. Wir haben die Betrachtung der unterschiedlichen Validitätskriterien an das Ende dieses Beitrages plaziert. In diesem Abschnitt sollen zuvor verschiedene Entscheidungs- und Bewertungsaspekte jeweils getrennt für die einzelnen Ebenen kurz skizziert werden.

### 2.3.1 Die Ebene des Signifikanztests

Die Entscheidung zugunsten der Alternativ bzw. Nullhypothese eines Signifikanztests ist mit den numerisch bestimmbaren Fehlerrisiken  $\alpha$  und  $\beta$  verbunden und ist als eine **konventionalistische** Entscheidung zu kennzeichnen: Unter jeder Testhypothese sind zunächst alle empirisch-statistischen Ergebnisse möglich, wenn auch unterschiedlich wahrscheinlich. Ein spezifisches empirisches Datum kann insofern nicht gegen oder für eine statistische Hypothese sprechen. Der Forscher wird sich dazu entschließen, eine der beiden sich widersprechenden Hypothesen anzunehmen. Er fällt damit eine Entscheidung.

### 2.3.2 Die Ebene der statistischen Vorhersage

Wenn die SV direkt in eine testbare statistische Hypothese eines Signifikanztests mündet, so ist die Entscheidung aus der untersten Ebene des Signifi-

kanztests mit der der nächsthöheren Ebene der SV identisch. Entspricht die SV der  $H_1$ -Hypothese des Signifikanztests, so führt ein signifikantes Resultat im Signifikanztest zur Annahme der SV, ein nicht-signifikantes Resultat wurde zur Ablehnung der SV führen. Wenn die SV mit einer  $H_0$ -Hypothese eines Signifikanztests identisch ist, wurde ein signifikantes Resultat im Test gerade zur Ablehnung der SV führen und umgekehrt. Die Entscheidung zugunsten bzw. zuungunsten der SV ist mit den Risikowahrscheinlichkeiten  $\epsilon$  bzw.  $\phi$  verbunden. Je nach Beziehung zwischen SV und Testhypothesen kann z.B.  $\epsilon$  der Fehlerwahrscheinlichkeit  $\alpha$  oder  $\beta$  entsprechen (Genaueres siehe Hager, 1987, S. 131-133).

Etwas komplizierter werden die Beziehungen, wenn **mehrere Signifikanztests** zur Entscheidung über eine **SV** herangezogen werden sollen. Die einzelnen Resultate der Signifikanztests lassen sich **logisch** miteinander verknüpfen. So könnte bspw. aus einer SV folgen, daß jeweils die Alternativhypothesen ( $H_{ik}$ ) von  $k = 3$  gerichteten t-Tests gelten sollten:

$$SV: \mu_1 > \mu_2 > \mu_3 > \mu_4 \equiv (H_{11}: \mu_1 > \mu_2) \wedge (H_{12}: \mu_2 > \mu_3) \wedge (H_{13}: \mu_3 > \mu_4) \quad (2.3)$$

In dem genannten Beispiel werden die drei t-Tests mit einer **Konjunktion** verbunden, d.h. es müßten sich in allen drei Tests signifikante Resultate einstellen, um den Forscher zu einer Entscheidung zugunsten der SV zu bewegen. Die Beziehungen der Fehlerrisiken  $\epsilon$  und  $\phi$  auf der einen und  $\alpha$  und  $\beta$  auf der anderen Seite werden etwas komplexer, einfache numerische Entsprechungen wie in dem Falle identischer SVn und Testhypothesen, sind nicht mehr ohne weiteres denkbar. Im übrigen ist es zudem möglich, daß die einzelnen Tests mit unterschiedlichen Fehlerrisiken  $a_i$  und  $\beta_i$  durchgeführt werden. Bei Hager (1987, S. 171-177) finden sich weitergehende Betrachtungen zur sogenannten **Kumulation** der Fehlerrisiken der statistischen Tests. In die Entscheidungen über die SV werden außerdem spezifische Validitätsaspekte einfließen, über die wir jedoch erst im Abschnitt 3 berichten.

### **2.3.3 Die Ebene der inhaltlichen Hypothesen**

Während wir auf den statistischen Ebenen (der Testhypothesen und der SV) von Entscheidungen gesprochen haben, werden wir bzgl. der Ebene der **EIH** nunmehr **von Beurteilungen** sprechen. Die Beurteilung der EIH „im Lichte der Daten“ (basierend auf den Entscheidungen über die SV) wird in Abhängigkeit von verschiedenen Gütekriterien der Untersuchung durchaus unterschiedlich ausfallen und nicht mehr relativ statisch bzw. automatisch wie noch z.B. im Signifikanztest ablaufen. Hier ist bspw. die Frage relevant, inwiefern die **Ableitungsvalidität der statistischen Hypothese** einerseits und die **statistische Validität** andererseits als hinreichend erfüllt angesehen werden kön-

nen. Wir werden auf diese Aspekte im nächsten Abschnitt zu sprechen kommen.

Für die Bewertung der inhaltlichen Hypothesen basierend auf den statistischen Testergebnissen wurden in der Vergangenheit unterschiedliche Vorschläge unterbreitet. Wir werden uns hier auf eine mögliche Konvention stützen, die von Hager (1984) als Erweiterung der Betrachtungen von Westermann und Hager (1982) eingeführt worden ist. Voraussetzung dieser Beurteilungen ist eine hinreichende Erfüllung der Validitätskriterien einerseits und eine entsprechend gelungene Kontrolle der Fehlerwahrscheinlichkeiten  $\alpha$ ,  $\beta$ ,  $\epsilon$  und  $\varphi$  andererseits.

Diese Beurteilung stützt sich nicht nur auf die Information, ob die Tests signifikant wurden, sondern gewichtet gleichzeitig Informationen über die eingetretenen statistischen Effekte (Effektstärken). Gestützt auf eine **Testplanung**, die vor dem Experiment stattfand und der Simultankontrolle der Determinanten des Signifikanztests dient (vgl. Hager, **1987**), läßt sich der **Beurteilungsraum der EIH** in vier bzw. drei Bereiche aufteilen (andere Kriterien werden u.a. bei Westermann & Hager, 1982, dargestellt). Danach wird die EIH als „bewährt“ oder „bedingt bewährt“ oder „bedingt nicht bewährt“ oder gar „nicht bewährt“ betrachtet, je nachdem, ob einerseits die relevante SV angenommen wurde, und andererseits ein vor dem Experiment geplanter statistischer Effekt über bzw. unterschritten worden ist. Die Beurteilung richtet sich im einfachsten Falle der Äquivalenz der SV mit einer Testhypothese danach, ob die SV der  $H_0$  oder der  $H_1$ -Hypothese entspricht und demzufolge der geplante statistische Effekt einen **Mindest-** oder **Maximaleffekt** darstellt. Für vertiefende Betrachtungen sei auf die Ausführungen von Hager (1987) verwiesen.

Der direkte Einbezug der von der Stichprobengröße bekanntermaßen unabhängigen Effektstärke läßt in engen Bahnen einen Einbezug inhaltlicher Überlegungen in den Signifikanztest zu (vgl. auch Hays, 1978). Der Forscher kann bspw. ein Testergebnis trotz signifikantem Ausgang auch als „nur bedingt für die EIH sprechend“ **interpretieren**, wenn die mit dem signifikanten Resultat verbundene Effektstärke ihm als zu niedrig erscheint, um eine **praktische Bedeutsamkeit** zu erhalten.

Die Bewertung der **TIH** aufgrund der bisherigen Überlegungen wird sich anschließend nahezu ausschließlich auf nicht-statistische Überlegungen stützen. Hier werden Bewertungen der Validitäten der Untersuchungen in hohem Ausmaß relevant: Ist die interne Validität der Untersuchung als hinreichend hoch anzusehen, wie sieht es mit der Sicherung der Populationsvalidität, der Situationsvalidität aus? Wurden die theoretischen Variablen annähernd valide operationalisiert (Variablenvalidität; vgl. Hager, 1984, S. 70), wurden die inhaltlichen Hypothesen folgerichtig abgeleitet etc.?

Die Fehlerwahrscheinlichkeiten  $e$  und  $f$ , die wir weiter oben dargestellt haben, werden in der Regel mit nicht-statistischen Überlegungen verbunden, so daß eine numerische Bestimmung von  $e$  und  $f$  ausgeschlossen sein wird. Des weiteren werden immer mehr subjektive Elemente Eingang in den Forschungsprozeß finden, je mehr wir uns wiederum den oberen, inhaltlichen Ebenen nähern. Exakt logisch begründbare Beurteilungskriterien lassen sich auf diesen Ebenen nicht mehr anführen. Muß bspw. ein Forscher davon ausgehen, daß aufgrund vielfältiger Störungen seines Experimentes die interne Validität als nicht gesichert angesehen werden muß, so kann er ebenfalls davon ausgehen, daß seine statistischen Ergebnisse keine Relevanz (zumindest für die TIH) haben werden. Dieses gilt selbst dann, wenn die statistische Validität der Untersuchung als sehr hoch einzuschätzen ist. Wir erkennen an diesem kleinen Beispiel, daß die Betrachtung der unterschiedlichen **Untersuchungvaliditäten** den eigentlich entscheidenden Anteil an der Beurteilung der inhaltlichen Hypothesen hat.

### ***3. Zur Validität der Überprüfung wissenschaftlicher Hypothesen***

Für eine wissenschaftliche Untersuchung werden unterschiedlichste Gütekriterien angegeben, nach denen die Gültigkeit der Untersuchung und deren Aussagen und Bewertungen beurteilt werden können. Je nach der spezifischen inhaltlichen Hypothese und je nach wissenschaftstheoretischer Orientierung des Forschers wird dabei dem einen Kriterium mehr, dem anderen weniger Bedeutung beizumessen sein. Die Gewichtung wird im Einzelfall zu erfolgen haben und kann nicht Gegenstand dieses Textes sein. Insoweit werden wir uns im folgenden nicht so sehr auf die Populationsvalidität, die ökologische Validität, die Situationsvalidität sowie die interne Validität (Ceteris-Paribus-Validität; vgl. Westermann, 1987a) konzentrieren (vgl. dazu z.B. die Abhandlungen von Bredenkamp, 1980 oder Hager & Westermann, 1983), sondern auf Gütekriterien eingehen, die sich direkt auf die Untersuchung und deren Hypothesenbewertungen beziehen. Wir werden dabei die Ausführungen von Hager (1984) aufnehmen bzw. ergänzen. Damit sollen jedoch andere Validitätsaspekte keineswegs als weniger bedeutsam eingeschätzt werden. So sind für eine intern nicht valide Untersuchung alle hier gemachten Überlegungen ohne Bedeutung. Es wäre ohne Relevanz, ob diese Untersuchung bspw. eine hohe Populationsvalidität besitzt etc., denn ohne Sicherung der internen Validität läßt sich nicht begründet darauf schließen, daß die Veränderung in den abhängigen Variablen, die sich in der Untersuchung eingestellt haben, überhaupt etwas mit den unabhängigen Variablen der Untersuchung zu tun haben. Dieses gilt selbst dann, wenn die unabhängigen und abhängigen Variablen der Untersuchung optimal operationalisiert worden sind, dieses gilt ebenfalls selbst dann, wenn die durchgeführten statistischen Tests die bestmöglichen Tests dar-

stellen und die erhobenen Daten eine vollkommene Reliabilität aufweisen wurden.

### 3.1 Die Zuordnung der empirisch-inhaltlichen zu den theoretisch-inhaltlichen Hypothesen

Wir haben erwähnt, daß die theoretischen Variablen der TIH operationalisiert werden, um empirische Variablen konkret zu untersuchen. Die Zuordnung der empirischen Variablen zu den theoretischen Begriffen kann im Experiment durchaus unterschiedlich erfolgen, so daß wir die Güte dieser Operationalisierung durch den Begriff der Variablenvalidität erfassen wollen. Diese kann sowohl für die unabhängigen als auch für die abhängigen Variablen der Untersuchung beeinträchtigt sein. So stellt sich bspw. das Problem, ob das „Auszählen von Verbalinjurien innerhalb einer bestimmten Zeit“ inhaltlich betrachtet tatsächlich etwas von dem intendierten Konstrukt der Aggression abzubilden vermag. Weiter kann es durchaus fraglich erscheinen, ob die experimentellen Maßnahmen, die die VPn frustrieren sollen, tatsächlich etwas ausgelöst haben, was sich im Sinne des Konstruktes interpretieren ließe.

Wichtig ist, daß allein die Bedeutungsreduktion des theoretischen Begriffes auf der experimentellen Ebene noch nicht ein Indiz für mangelnde Variablenvalidität darstellt, denn wir werden uns im Experiment immer auf bestimmte Facetten des allgemeineren Begriffes einschränken müssen. Da wir jedoch die Möglichkeit besitzen, mehrere Untersuchungen durchzuführen, können wir in diesen dann jeweils andere Operationalisierungen i.S. der konzeptuellen Replikation (Bredenkamp, 1980) durchführen. So können wir aus der Feststellung, daß für bestimmte empirische Interpretationen (Operationalisierungen) des theoretischen Begriffes die Hypothesen nicht anwendbar sind, wichtige Erkenntnisse gewinnen. Eine zu geringe oder gar fehlende Variablenvalidität wird jedoch die Validität der gesamten Untersuchung infrage stellen. Leider läßt sich in der Forschungspraxis zumeist nicht eindeutig feststellen, in welchem Umfang die Variablenvalidität eingeschränkt worden ist, es lassen sich keine Werte, keine Validitätskoeffizienten o.ä. angeben, wie man dieses z.B. im Rahmen der Testkonstruktion psychologischer Tests oft versucht. Die Variablenvalidität hängt direkt mit dem Operationalisierungsproblem zusammen, über das wir weiter oben bereits berichtet haben.

Neben dem Problem der Sicherung der Variablenvalidität bleibt zu klären, ob wir bei der Übertragung der **Relationen** zwischen den theoretischen Begriffen auf die empirische Ebene Fehler gemacht haben. Die **Ableitungsvalidität der empirisch-inhaltlichen Hypothese** [AblVal(EIH)] kann beeinträchtigt sein

(vgl. Hager, 1984, S. 63). Einen solchen Fehler wurden wir bspw. dann begehen, wenn in der TIH ausgesagt wurde, daß „A *höher ist als* B“, wir als EIH jedoch „A\* *ist ungleich* B“<sup>3</sup> schließen würden<sup>3</sup>.

In Abhebung zu der bislang benutzten Definition wollen wir an dieser Stelle den Begriff der Operationalisierung in einer erweiterten Fassung vorstellen: Da die Güte der Operationalisierung von der Variablenvalidität **und** der Abl-Val(EIH) abhängt, wurde u. E. „Operationalisieren“ demzufolge bedeuten, sowohl die Konzepte wie auch die Relationen zwischen diesen Konzepten einer Beobachtung, Messung und Erfassung zugänglich zu machen.

### 3.2 Die Zuordnung der statistischen Vorhersagen zu den empirisch-inhaltlichen Hypothesen

Der EIH wird eine SV zugeordnet. Die **Übersetzung** der empirischen Variablen in statistische scheint (nach unseren vorangegangenen Ausführungen) nicht direkt möglich, nur eine **Zuordnung** statistischer Konzepte zu empirischen Variablen erscheint realistisch. Die Güte dieser Zuordnungsbeziehung hängt u. a. vom Skalenniveau der erhobenen Daten ab. Es stellt sich die Frage, ob alle **relevanten** Informationen, die die EIH betreffen, sich auf der statistischen Ebene (in einer entsprechend veränderten Form) wiederfinden, Ist dieses der Fall, so können wir mit Hager (1987) von einer **suffizienten** Zuordnung sprechen.

So könnte eine Einschränkung einer suffizienten Zuordnung vorliegen, wenn in der SV eine Ranghypothese vertreten wird, obgleich sich aus den Daten Mittelwertsinformationen gewinnen lassen. Die Reduktion auf eine Ranghypothese bringt einen Informationsverlust mit sich, da die Abstandsinformationen der Rohdaten verloren gehen.

Werden die Relationen zwischen den Variablen der EIH korrekt auf die statistische Ebene übertragen, haben wir die Zuordnung als **adäquat** bezeichnet. Faßt man beide Kriterien zusammen, so lassen sich Beeinträchtigungen als Störungen der **Ableitungsvalidität der statistischen Vorhersagen** [Abl-Val(SV)] bezeichnen. So wurde bspw. eine Beeinträchtigung der AblVal(SV) vorliegen, wenn einer EIH „A\* ist größer als B“ die SV zugeordnet werden wurde, daß „ $\mu_A \neq \mu_B$ “.

<sup>3</sup> Mit dem hochgestellten \* soll verdeutlicht werden, daß es sich um die Operationalisierung der theoretischen Variablen handelt.

### 3.3 Die Zuordnung der Signifikanztests zu den statistischen Vorhersagen

Eine Beeinträchtigung der **Ableitungsvalidität der Testhypothesen** [AblVal( $H_0, H_1$ )] liegt vor, wenn Signifikanztests herangezogen werden, die zur Prüfung der SV nur eingeschränkt geeignet sind. Gehen wir vereinfachend davon aus, daß Mittelwerte eine suffiziente Zuordnung statistischer zu empirischen Konzepten darstellen, dann wurde im folgenden Beispiel eine Beeinträchtigung der AblVal( $H_0, H_1$ ) vorliegen:

Ein Forscher bildet vier Gruppen von VPn und äußert die Vermutung, daß die unterschiedlichen Treatments, die den Gruppen appliziert werden, dazu führen, daß die Arbeitszufriedenheit, die über einen spezifischen Fragebogen erfaßt wurde [AZ], über die vier Gruppen ansteigt und in der vierten Gruppe die höchste AZ vorliegen sollte. Das Ansteigen sollte von Gruppe zu Gruppe vorliegen. Er ordnet dieser EIH folgende SV zu:  $\mu_1 < \mu_2 < \mu_3 < \mu_4$ . Diese SV kann man als adäquate Zuordnung bezeichnen. Der Forscher testet diese Hypothese im folgenden mit einer einfaktoriellem Varianzanalyse und bewertet ein signifikantes Ergebnis als für seine EIH sprechend. Da die einfaktoriellem Varianzanalyse als Testhypothese jedoch die  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$  gegen die unspezifische  $H_1: \mu_i \neq \mu_j$  für mindestens ein Paar mit  $i \neq j$  testet, liegt eine Störung der AblVal( $H_0, H_1$ ) vor.

Die Kriterien einer adäquaten und suffizienten Zuordnung verschiedener Hypothesenarten zueinander lassen sich ebenso auf die Beziehung von Testhypothesen und SV anwenden. In dem o. g. Beispiel kann man davon sprechen, daß weder eine adäquate, noch eine **suffiziente Ableitung** der Testhypothesen vorliegt: Die Signifikanz des varianzanalytischen Tests spricht allein noch nicht für die vorangestellte SV. Hätte der Forscher im Anschluß an die Varianzanalyse sogenannte (gerichtete) Post hoc-Vergleiche durchgeführt, so wäre zwar die Beeinträchtigung der AblVal( $H_0, H_1$ ) zumindest geringer, ließe sich jedoch aufgrund der wahrscheinlich vergrößerten Fehlerrisiken für  $\epsilon$  und/oder  $\phi$  (die mit den unterschiedlichen Teststärken der verwendeten Testverfahren zusammenhängen) immer noch nicht ausschließen.

Da es u. E. in Einzelfällen denkbar erscheint, daß ein Forscher irgendeinen leider vollkommen ungeeigneten Signifikanztest als Überprüfung seiner SV heranzieht, wird man in solchen Fällen kaum noch von einer Ableitungsvalidität sprechen können. In diesem sicherlich extremen Fall wurde zwischen der SV und der/den Testhypothese(n) keine Beziehung bestehen und die Fehlerwahrscheinlichkeiten  $\epsilon$  und  $\phi$  wären demzufolge numerisch **nicht** bestimmbar.



### 3.4 Die Validität des statistischen Schlusses

Schließlich bleibt noch anzumerken, daß die Verwendung von Signifikanztests auf spezifischen Voraussetzungen beruht, deren Erfüllung bzw. Nichterfüllung zu einer Beeinträchtigung der **statistischen Validität**<sup>4</sup> per se führen können. Während die Fehlerwahrscheinlichkeiten  $\epsilon$  und  $\phi$  mit der  $\text{AbVal}(H_0, H_1)$  zusammenhängen, bewirkt eine Einschränkung der statistischen Validität (meistens) unkontrollierte Veränderungen der Fehlerwahrscheinlichkeiten für  $\alpha$  und  $\beta$ . Die tatsächlichen Stichprobenkennwerteverteilungen der Tests weichen von den tabellierten ab, die tatsächlichen Fehlerraten erster und zweiter Art können (teils erheblich) von den nominellen, die ein Forscher sich vorgibt, differieren. Die statistische Validität hängt zusätzlich u. a. mit dem Skalenniveau der Daten und einigen anderen Faktoren zusammen, auf die wir in diesem Zusammenhang nicht näher eingehen können (vgl. u.a. Bredenkamp, 1980; Hager & Westermann, 1983 oder Hager, 1987).

## 4. Abschließende Bemerkungen

Der Prozeß der Beurteilung wissenschaftlicher Hypothesen, ihre Überprüfung, hängt von zahlreichen Bedingungen ab und enthält an vielen Stellen subjektive Komponenten. Die Einteilung der Hypothesen in vier mögliche Ebenen mag für einige Zwecke zu undifferenziert sein, kann jedoch spezifische Probleme der Forschungspraxis verdeutlichen. Sobald ein Forscher zum Zwecke der Bewertung inhaltlicher (und theoretischer) Hypothesen Untersuchungen anstellt, wird er Entscheidungen fallen müssen, die richtig, aber auch falsch, optimal, aber auch suboptimal sein können.

Die unterschiedlichen Validitätsaspekte wissenschaftlicher Untersuchungen behandeln letztendlich diese potentiellen Fehlermöglichkeiten und versuchen sie an zahlreichen Stellen sogar zu quantifizieren. Ein reines „Schielen“ nach signifikanten Ergebnissen vermag wohl den Anforderungen der heutigen psychologischen Forschung kaum noch gerecht zu werden, zumal ein Rücktransfer der (signifikanztheoretischen) „Wahrheit“ auf die inhaltlichen Hypothesen u.E. nicht möglich ist. Die Beziehung zwischen den hier beschriebenen inhaltlichen und den statistischen Hypothesen ist dafür bei weitem zu indirekt, als daß eine solche Übertragung gelingen könnte.

Andererseits wäre wohl auch ein Umformen der inhaltlichen Hypothesen gemäß den Erfordernissen der Statistik und des Signifikanztests nicht das, was u. E. eine Lösung erbringen könnte. Schließlich stellt sich die Psychologie als inhaltliche und naturwissenschaftliche Disziplin dar, die den Methodenkanon

---

4 Validität des statistischen Schlusses

der Statistik, der Mathematik, eigentlich nur als Hilfswissenschaft nutzt. Die wissenschaftlich sicherlich wichtigen Fragen eines Mathematikers sind eben nicht die inhaltlich relevanten Fragen eines Psychologen. Daraus folgt u.E. konsequent die Unmöglichkeit, valide psychologische Untersuchungen dadurch durchführen zu wollen, indem man nur die statistischen Entscheidungsprozesse optimiert. Sicherlich ist letzteres eine wichtige Perspektive psychologischer Forschung, die endgültigen Bewertungen der inhaltlichen Hypothesen können jedoch immer nur unter Bezugnahme auf nicht-statistische Kriterien gemacht werden, die sich (leider?) auch nicht (noch nicht?) quantifizieren lassen.

## **Literatur**

- Bentler, P. M. (1986). ***EQS - Ein Ansatz zur Analyse von Strukturgleichungsmodellen für normal- bzw. nichtnormalverteilte quantitative Variablen***. In C.Möbus & W. Schneider (Hrsg.), *Strukturmodelle für Längsschnittdaten und Zeitreihen* (S. 27-56). Bern: Huber.
- Bortz, J. (1984). ***Lehrbuch der empirischen Forschung***. Berlin: Springer.
- Bredenkamp, J. (1980). ***Theorie und Planung psychologischer Experimente***. Darmstadt: Steinkopff.
- Carnap, R. (1936). Testability and meaning. ***Philosophy of Science*, 3**, 419-471.
- Cohen, J. (1977). ***Statistical power analysis for the behavioral sciences*** (2nd ed.). Hillsdale, NJ: Erlbaum
- Dörner, D. (1979). ***Problemlösen als Informationsverarbeitung*** (2. Auflage). Stuttgart: Kohlhammer.
- Gadenne, V. (1976). ***Die Gültigkeit psychologischer Untersuchungen***. Stuttgart: Kohlhammer.
- Gadenne, V. (1984). ***Theorie und Erfahrung in der psychologischen Forschung***. Tübingen: Mohr.
- Groeben, N. & Westmeyer, H. (1981). ***Kriterien psychologischer Forschung*** (2. Aufl.). München: Juventa.
- Haagen, K. & Seifert, H. G. (1979). ***Methoden der Statistik für Psychologen (Band 2)***. Stuttgart: Kohlhammer.
- Hager, W. (1984). Aspekte eines deduktiven Forschungsansatzes in der empirischen Pädagogik: Fragen der Hypothesenvalidität, der Untersuchungsplanung und der Hypothesenbeurteilung. ***Zeitschrift für Empirische Pädagogik und Pädagogische Psychologie*, 8**, 55-75.
- Hager, W. (1987). ***Grundlagen einer Versuchsplanung zur Prüfung empirischer Hypothesen in der Psychologie***. In G. Lüer (Hrsg.), *Allgemeine experimentelle Psychologie* (Kap. 3, S. 43-264). Stuttgart: Fischer.
- Hager, W. & Möller, H. (1986). Tables and procedures for the determination of power and sample sizes in univariate and multivariate analyses of variance and regression. ***Biometrical Journal*, 28**, 643-667.

- Hager, W. & Westermann, R. (1983). Planung und Auswertung von Experimenten. In J. Bredenkamp & H. Feger (Hrsg.), **Hypothesenprüfung. Enzyklopädie der Psychologie, Forschungsmethoden der Psychologie**, Bd. 5 (S. 24-238). Göttingen: Hogrefe.
- Harnatt, J. (1975). Der Signifikanztest in kritischer Betrachtung. **Psychologische Beiträge**, 17, 595-612.
- Hays, W. L. (1978). **Statistics for the social sciences** (2nd ed.). New York: Holt, Rinehart & Winston.
- Hempel, C. G. (1974). **Philosophie der Naturwissenschaften**. München: DTV.
- Huber, O. (1987). **Das psychologische Experiment**. Stuttgart: Huber.
- Hussy, W. (1983). Komplexe menschliche Informationsverarbeitung: das SPIV-Modell. **Sprache & Kognition**, 2, 47-62.
- Hussy, W. & Eye, A. von (1988). **On cognitive operators in information processing and their effects on short-term memory performance in different age groups**. In F.E. Weinert & M. Perlmutter (Eds.), *Memory development: Universal changes and their individual differences* (pp.275-291). Hillsdale, NJ: Erlbaum.
- Jöreskog, K. G. & Sörbom, D. (1988). **LISREL 7 - a guide to the program and applications**. Chicago: Scientific Software Inc.
- Krapp, A., Hofer, M. & Prell, S. (1982). *Forschungs-Wörterbuch*. München: Urban & Schwarzenberg.
- Lepper, M. R., Greene, D. & Nisbett, R.E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the 'overjustification' hypothesis. **Journal of Personality and Social Psychology**, 28, 129-137.
- McGuigan, F. J. (1968). **Experimental psychology** (2nd Ed.). Englewood Cliffs, Ca.: Prentice-Hall.
- Meehl, P.E. (1967). Theory testing in psychology and physics: A methodological paradox. **Philosophy of Science**, 34, 103-115.
- Möller, H. (1991). **Möglichkeiten und Grenzen Linearer Strukturmodelle zur Parametrisierung ereigniskorrelierter Potentiale**. Eine Untersuchung am Beispiel von emotional bedeutsamem Reizmaterial. Trier: WVT.
- Morrison, D.E. & Henkel, R.E. (Eds.). (1970). **The significance test controversy**. Chicago: Aldine.
- Neyman, J. & Pearson, E.S. (1933). The testing of statistical hypotheses in relation to probabilities a priori. **Proceedings of the Cambridge Philosophical Society**, 29, 492-510.
- Opp, K. D. (1976). **Methodologie der Sozialwissenschaften** (3. Aufl.). Reinbek bei Hamburg: Rowohlt.
- Schulz, T., Muthig, I? & Koepller, K. (1980). **Theorie, Experiment und Versuchsplanung in der Psychologie**. Stuttgart: Kohlhammer.
- Westermann, R. (1987a). **Wissenschaftstheoretische Grundlagen der experimentellen Psychologie**. In G.Lüer (Hrsg.), *Allgemeine experimentelle Psychologie* (Kap. 2, S. 5-42). Stuttgart: Fischer.
- Westermann, R. (1987b). **Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie**. Berlin: Springer.

## Kontrolltechniken

***Jürgen Rehm und Fritz Strack***

“In our judgment the controlled experiment still remains the single most powerful tool for enabling investigators to disentangle just what causes what, and is the standard against which other research methods continue to be measured.”

Aronson, Ellsworth, Carlsmith und Gonzales, 1990, p. XVII

Ein Kapitel über Kontrolltechniken bei psychologischen Erhebungen in einem Teilband der Enzyklopädie der Psychologie sollte Redundanz zu bereits erschienenen und thematisch ähnlich gelagerten Beiträgen der Enzyklopädie und des Handbuchs der Psychologie vermeiden (dies gilt insbesondere hinsichtlich der Arbeiten von Bredenkamp, 1969, 1983; Feger, 1983; sowie Hager & Westermann, 1983) und versuchen, eine neue Perspektive einzunehmen.

Der vorliegende Beitrag hat dieses Problem im Einklang mit dem Titel des Bandes durch eine Fokussierung auf methodologische Probleme (unter Vernachlässigung von statistischen Einzelfragen) zu lösen versucht. Im einzelnen geht es um die Frage, wie psychologische Hypothesen oder Theorien geprüft und gleichzeitig Alternativerklärungen ausgeschlossen oder unwahrscheinlich gemacht werden können (vgl. Mill, 1886/1968 zu diesem Kriterium der Kausalität). Dabei steht das Problem von „Fehlereinflüssen“ und ihrer Kontrolle durch geeignete Verfahren im Vordergrund.

Bei der Beantwortung dieser Frage haben wir das Experiment als Forschungsmethode der Wahl ins Zentrum der Überlegungen gestellt, da das Experiment mit seinen Kontrollmöglichkeiten nach Ansicht maßgeblicher Autoren (z.B. Aronson, Ellsworth, Carlsmith & Gonzales, 1990) als methodologisches Ideal auch dann zur Richtschnur werden sollte, wenn ein Forscher sich nach Prüfung aller Alternativen für nicht-experimentelle Versuchspläne und damit für andere Kontrolltechniken entscheidet.

Dieser Standpunkt wird im folgenden an einem psychologischen Forschungsbeispiel entwickelt. Ausgehend von einem konkreten Problem werden unter-

schiedliche Umsetzungen in Versuchspläne entwickelt. Dabei reicht die Logik der Argumentation über die strikt experimentelle Lösung unter Anwendung der Kontrolltechnik der Randomisierung bis hin zu nicht-experimentellen Lösungsversuchen mit gezielten Kontrolltechniken wie Parallelisierung. Den Abschluß des Kapitels bilden die Darstellung von Kontrolltechniken mit Hilfe statistischer Prozeduren sowie Überlegungen zur Ethik des Experiments. Diese Abfolge wurde lediglich aus didaktischen Gründen gewählt und impliziert keineswegs, daß sich psychologische Forschung induktiv vom Phänomen zu dessen Erklärung hin entwickelt.

## 1. Beobachtung und Schlußfolgerung

Das folgende Beispiel soll dazu dienen, die Gültigkeit von aus Alltagsbeobachtungen hergeleiteten psychologischen Schlüssen zu untersuchen. Angenommen, eine Kindergärtnerin stellt auf der Grundlage vieler Beobachtungen fest, daß Kinder, die sich viele gewalttätige Sendungen im Fernsehen anschauen, häufiger aggressive Verhaltensweisen zeigen als andere Kinder. Sie kommt daraufhin zu dem Schluß,<sup>1</sup> daß gewalttätige Sendungen aggressiv „machen“; mit anderen Worten: daß Sendungen dieser Art Aggressivität verursachen.

Dieses Beispiel beschreibt eine typische Alltagssituation, in der auf der Grundlage von Beobachtungen Schlüsse über zugrundeliegende psychologische Zusammenhänge gezogen werden. Diese Art der alltagspsychologischen Erkenntnisgewinnung hat mit der Methodologie der wissenschaftlichen Psychologie eines gemeinsam: empirische Beobachtungen sind der Prüfstein für Hypothesen bzw. Theorien. Der wesentliche Unterschied zwischen Alltags-erkenntnis und Wissenschaft besteht darin, daß in der wissenschaftlichen Psychologie die Beobachtungen **systematischer** und **kontrollierter** sein müssen, weil sonst die resultierenden Schlüsse nicht hinreichend gerechtfertigt sind.

Die Kindergärtnerin im Beispiel hat implizit oder explizit bereits zwei Gruppen mit unterschiedlichen Sehgewohnheiten verglichen: Kinder, die häufig Sendungen gewalttätigen Inhalts rezipieren und Kinder, die dies nicht oder vergleichsweise selten tun. Diese Gruppen unterschieden sich in spezifischen aggressiven Verhaltensweisen, so wie diese von der Kindergärtnerin definiert wurden. Der gezogene Schluß war dagegen allgemeiner Natur und nicht mehr auf die beobachteten Gruppen beschränkt: Gewalttätige Sendungen verursachen aggressive Handlungen.

<sup>1</sup> Auf die Problematik von induktiven Schlüssen aus Beobachtungen soll hier nicht weiter eingegangen werden (vgl. dazu allgemeiner Chalmers, 1978; hinsichtlich von Folgerungen für die Psychologie s. Gadenne, 1984; sowie Westermann & Gerjets, Kapitel 10 in diesem Band).

Die gezogene Schlußfolgerung ist jedoch keineswegs zwingend, denn zwei alternative Schlüsse wären gleichermaßen mit den Beobachtungen vereinbar gewesen:

- a) Aggressive Kinder schauen sich lieber gewalttätige Sendungen an als andere Kinder.
- b) Kinder, die von ihren Eltern vernachlässigt werden, entwickeln zum einen aggressive Tendenzen und bevorzugen zum anderen Sendungen mit gewalttätigen Inhalten.

Bei der ersten Alternativerklärung handelt es sich um die Umkehrung der Kausalrichtung. Nicht der Fernsehkonsum ist die Ursache für die Aggression, sondern die Aggression ist die Ursache für den Fernsehkonsum. Die zweite Alternativerklärung beinhaltet eine dritte Ursache (den Erziehungsstil der Eltern), die sowohl für den Fernsehkonsum wie auch für die Aggression verantwortlich ist.

Schon an diesem Beispiel wird deutlich, daß sich auf der alleinigen Grundlage von unsystematischen Beobachtungen keine Entscheidungen über die Gültigkeit von alternativen Erklärungsmöglichkeiten treffen lassen. Durch einfache Beobachtung und darauf basierende Schlüsse ist es offensichtlich **nicht** möglich, zu eindeutigen Kausalerklärungen zu gelangen. Ziel wissenschaftlicher Psychologie ist es jedoch, zu gültigen, wahren oder zutreffenden Aussagen zu gelangen.'

Aus den angeführten Gründen liefert die unsystematische Beobachtung eine unzureichende empirische Grundlage für die wissenschaftliche Psychologie. Vielmehr ist es notwendig, Beobachtungen hier im vorhinein so zu planen und anzulegen, daß möglichst eindeutige Schlüsse erreicht werden können. Im folgenden sollen verschiedene Techniken vorgestellt werden, die dieses Ziel erreichen helfen. Dabei wird der unserer Meinung nach wichtigste Zugang am ausführlichsten vorgestellt: das psychologische Experiment mit Randomisierung.

## 2. ***Das psychologische Experiment als Idealtypus kontrollierter Beobachtung***

### 2.1 Experiment und Kontrolle durch Randomisierung

Das Experiment gilt nicht nur in der Psychologie, sondern auch in anderen Naturwissenschaften als Methode der Wahl (allerdings z.T. mit anderen Kontrollmechanismen; s. u.). Die experimentelle Methode kann als **planmäßige**

2 Diese Aussage basiert bereits auf einer wissenschaftstheoretischen Entscheidung, die hier nicht weiter ausgeführt werden muß - vgl. dazu bspw. Gadenne, 1984, bzw. Kapitel 7 in diesem Band.

**Manipulation von Variablen zum Zwecke der Beobachtung unter kontrollierten Randbedingungen** definiert werden (zu dieser Definition s. auch Hummell, 1980, S.210; Merz, 1987<sup>2</sup>, S. 538; Metzger, 1952, S. 143; oder Traxel, 1964, S. 91) und ist in der Psychologie durch drei Merkmale gekennzeichnet:

- a) **die Herstellung der Untersuchungsbedingungen;**
- b) **die systematische Variation (Manipulation) der zu untersuchenden Ursache;**
- c) **die randomisierte (zufallsgesteuerte) Zuteilung der Versuchspersonen zu den einzelnen Bedingungen.**

Um aus Beobachtungen eindeutige Schlußfolgerungen zu ziehen, ist das dritte Merkmal, die **Kontrolle durch Randomisierung**, von größter Bedeutung (zu den Grundlagen der Kontrolle durch Randomisierung s. R.A. Fisher, 1930/1960, pp. 18ff.; vgl. auch Bredenkamp, 1969, S.232ff., Aronson et al., 1990, sowie die dort angeführte Literatur).

Hätte die Kindergärtnerin im Eingangsbeispiel zur Prüfung ihrer Vermutung des kausalen Einflusses gewalttätiger Fernsehsendungen auf aggressives Verhalten die experimentelle Vorgehensweise gewählt (einige praktische Aspekte der Realisierung sollen zunächst vernachlässigt werden), so wären die Kinder per Zufall entweder (1) einer Bedingung mit Präsentation von gewalttätigen Filmen zugeteilt worden oder (2) einer sogenannten Kontrollbedingung ohne eine solche Präsentation. Im Anschluß daran wären in beiden Gruppen aggressive Handlungen gemessen worden.

Inwiefern können die alternativen Erklärungsmöglichkeiten (Umkehrung der Kausalrichtung bzw. gemeinsame dritte Ursache, s.o.) durch die beschriebene Vorgehensweise ausgeräumt werden? Zunächst kann bei einer zufälligen Zuteilung (**Randomisierung**) angenommen werden, daß sich die Kinder hinsichtlich von Aggressivität als Persönlichkeitseigenschaft zwischen den Bedingungen nicht überzufällig unterscheiden. Zwar werden in beiden Experimentalbedingungen Kinder unterschiedlicher Aggressivität vertreten sein, eine **systematische** Überrepräsentation von aggressiven Kindern in einer Gruppe ist jedoch auf der Grundlage von wahrscheinlichkeitstheoretischen Überlegungen nicht zu erwarten. Aus diesem Grund kann ein beobachteter Unterschied in der Häufigkeit aggressiven Verhaltens nicht auf Unterschiede in aggressiven Dispositionen zurückgeführt werden, die möglicherweise zu einer selektiven Auswahl von Fernsehprogrammen führen. Der beobachtete Verhaltensunterschied kann also nicht durch die Umkehrung der Kausalrichtung erklärt werden.

Bei der Konstruktion der Alternativerklärung wurde die dispositionale Ursache als bekannt vorausgesetzt. In solchen Fällen kann ein entsprechender Einfluß auch dadurch kontrolliert werden, daß die dispositionale Aggression vor

der Durchführung des Experiments erfaßt wird und dann auf der Grundlage des Meßergebnisses eine gleiche Verteilung dieses Merkmals über beide Experimentalbedingungen sichergestellt wird („Parallelisierung“, s. auch Abschnitt 3.2 unten). Werden die Teilnehmer an einer Untersuchung jedoch nach Zufallskriterien den einzelnen Bedingungen zugewiesen, so ist sichergestellt, daß die Experimentalbedingungen nicht nur hinsichtlich bekannter Persönlichkeitsmerkmale vergleichbar sind, sondern auch hinsichtlich anderer, in ihrem Einfluß unbekannter, Charakteristika. Durch die Randomisierung finden sich beispielsweise auch Kinder aus unterschiedlichen Elternhäusern in jeder Gruppe. Somit wird weiterhin ausgeschlossen, daß diese oder andere Drittvariablen das beobachtete Verhalten in Abhängigkeit von der Experimentalbedingung unterschiedlich beeinflussen.

An diesem Beispiel wird deutlich, daß durch Randomisierung **alle Eigenschaften sowie Voraussetzungen, die Versuchspersonen von vornherein mitbringen** als Alternativerklärung für die Ergebnisse ausgeschlossen werden können. Aronson et al. (1990, p. 18) sprechen hier von einem „**great equalizer**“.<sup>3</sup> Die Eigenschaft der Nivellierung aller Gruppenunterschiede durch Randomisierung wird immer wieder als entscheidender Vorteil der experimentellen Methode angeführt (z. B. Campbell & Stanley, 1966), weil Störfaktoren ansonsten ausschließlich gezielt durch den Experimentator ausgeschaltet werden könnten, sei es durch Aufnahme als unabhängige Variable in den Versuchsplan oder durch Parallelisierung (vgl. Aronson et al., 1990, pp. 18ff. sowie p. 148ff.; s. auch 3.2). Die **gezielte Ausschaltung von Störfaktoren scheitert** in der Praxis jedoch an zwei Hindernissen:

- a) Da die Anzahl der Störfaktoren, die durch Aufnahme oder Parallelisierung gezielt kontrollierbar sind, numerisch eng begrenzt ist, können nicht alle bekannten Störfaktoren auf diese Weise kontrolliert werden.
- b) Selbst wenn eine solche Kontrolle möglich wäre, müßten alle Störfaktoren dem Experimentator a priori bekannt sein, was unmöglich erscheint, da die Anzahl potentieller Störfaktoren unendlich ist (Fisher, 1960, p. 18).

Aus diesem Grund gilt die Randomisierung als einzige Methode, die geeignet ist, den Einfluß bekannter und unbekannter Störfaktoren auszuschalten. Einschränkung bleibt allerdings zu beachten, daß Gruppenunterschiede nicht absolut ausgeschlossen werden können, sondern lediglich im Rahmen wahr-scheinlichkeitstheoretisch definierter Grenzen.

Das Problem von Unterschieden zwischen Gruppen, die trotz Randomisierung bestehen bleiben, kann an einem typischen Beispiel aus der Medizinpsy-

<sup>3</sup> Es sei angemerkt, daß die Randomisierung mit den genannten Folgen nur Grundlage für Experimente mit Menschen, Pflanzen und Tieren ist. In der Physik oder Chemie kann auf Randomisierung verzichtet werden, weil die dort untersuchten Teilmengen von Stoffen als a priori gleichartig vorausgesetzt werden können.



chologie aufgezeigt werden. Nehmen wir an, durch ein kontrolliertes Experiment soll die Wirksamkeit eines neuen Medikaments in Verbindung mit Psychotherapie zur Heilung von endogenen Depressionen untersucht werden. Aus praktischen Gründen (begrenzte Zeitdauer des Experiments, ausreichende Anzahl an Personen mit entsprechenden Merkmalen) nehmen mehrere Kliniken mit einer jeweils niedrigen Patientenzahl an der Untersuchung teil. Eine Randomisierung über alle Personen (ungeachtet der Kliniken) kann dazu führen, daß sich beispielsweise alle drei Patienten einer Klinik in der Kontrollgruppe wiederfinden. Üblicherweise unterscheiden sich Kliniken und ihr Klientel jedoch beträchtlich, so daß speziell bei kleinen Stichproben hier die Gefahr besteht, daß unkontrollierte Faktoren die abhängige Variable (mit)beeinflussen.

Um dieser Gefahr einer zufälligen Ungleichverteilung prognostisch wichtiger Faktoren zu entgehen, wurden Fortentwicklungen und Adaptationen des Randomisierungsverfahren wie stratifizierte, parallelisierte oder adaptive Randomisierung entwickelt (s. Friedman, Furberg & DeMets, 1985, insbesondere Kapitel 5, als Überblick; s. Abel, 1987; Birkett, 1985; Wahrendorf, Blettner & Edler, 1985, als methodische Studien zu verschiedenen Einzelverfahren). Die entwickelten Verfahren gehen davon aus, daß nicht alle Mitglieder einer Stichprobe auf die Versuchsgruppen per Zufall verteilt werden, sondern daß zunächst Unterstichproben (z.B. anhand von Geschlecht und Alter oder - im Falle des gerade ausgeführten Beispiels - von Klinikzugehörigkeit) gebildet werden und dann die Randomisierung auf die Versuchsgruppen *innerhalb* dieser Unterstichproben erfolgt. Da Verfahren dieser Art auf den gleichen Prinzipien wie die unbedingte Randomisierung basieren, ergeben sich hinsichtlich unserer methodologischen Schlußfolgerungen keine Änderungen.

Die Kontrolltechnik der Randomisierung rechtfertigt außerdem die Verwendung der üblicherweise zur Analyse benutzten statistischen Verfahren (vgl. dazu Edgington, 1966, oder Bredenkamp, 1972). Psychologische Hypothesen beziehen sich im allgemeinen auf „offene Populationen“, die nicht raum-zeitlich begrenzt sind. Dies gilt auch für die Fragestellung unseres Eingangsbeispiels. Hier lautet die Hypothese, daß das Sehen von gewalttätigen Sendungen bei Personen eines gewissen Alters zu aggressiven Handlungen führt. Eine solche Hypothese auf Kinder eines bestimmten Kindergartens an einem bestimmten Ort zu einer bestimmten Zeit einschränken zu wollen, wäre theoretisch schwer zu rechtfertigen und in der psychologischen Forschung unüblich. Diese spezifische Eigenart psychologischer Theoriebildung ist jedoch insofern problematisch, als zur Anwendung von Signifikanztests die übliche Rechtfertigung in Frage gestellt wird. Diese Rechtfertigung basiert auf Zufallsauswahl aus Populationen, und eine solche Zufallsauswahl ist aus offenen Populationen nicht möglich.

Allerdings ist es möglich, die **Anwendung von Signifikanztests über das Randomisierungsmodell zu rechtfertigen**. In diesem Modell ist kein statistischer Schluß auf eine Population vorgesehen. Vielmehr wird geprüft, ob die unabhängigen Variablen **in der Stichprobe** einen überzufälligen Effekt haben. Ein solches Modell ist nur im Rahmen einer deduktiven Methodologie anwendbar (vgl. Bredenkamp, 1972, 1980; Gadenne, 1984), in der davon ausgegangen wird, daß eine allgemeine Hypothese an **jeder** Stichprobe überprüfbar sein muß.

Entsprechende statistische Tests werden **Randomisierungstests** genannt (vgl. Edgington, 1969, 1980). In verschiedenen Simulationsstudien hat sich gezeigt, daß die Verteilungen von Randomisierungstests mit den entsprechenden Verteilungen von parametrischen Tests übereinstimmen, so daß diese bei der Auswertung von Experimenten als Randomisierungstests interpretiert werden können (vgl. Bredenkamp, 1980; Gadenne, 1984).

Bei allen positiven Eigenschaften der Randomisierung sind jedoch auch Kosten in Kauf zu nehmen, die vor allem darin bestehen, daß auf diese Weise gebildete Gruppen oft heterogener sind als natürliche Gruppen. „Heterogenität“ bezieht sich hier auf Faktoren, die neben der systematisch variierten, postulierten Ursache ebenfalls auf die abhängige Variable einwirken. Dadurch erhöht sich die Variabilität innerhalb der Versuchsbedingungen, was mit einer Erhöhung des sogenannten **Zufallsfehlers** (random error) einhergeht. Da experimentell ermittelte Ergebnisse üblicherweise mit statistischen Verfahren geprüft werden, bei denen Abweichungen zwischen den Gruppen (= systematische Varianz) mit Unterschieden innerhalb der Gruppen verglichen werden (Varianzanalyse, vgl. Fisher, 1925; Winer, 1971; Box, Hunter & Hunter, 1978; Wilcox, 1987), kann ein hoher Zufallsfehler dazu führen, daß Unterschiede zwischen Gruppen unentdeckt bleiben oder aber große Stichproben verwendet werden müssen. In diesem Zusammenhang sind modifizierte Randomisierungspläne wie z.B. Zufallszuteilung von parallelisierten Paarlingen auf die Gruppen oft wesentlich ökonomischer. „Student“ hatte bereits 1931 (p. 405) bei seiner Evaluation einer großangelegten Studie zum Einfluß verschiedener Milcharten auf das Größenwachstum festgestellt, daß bei anderer Versuchsplanung (randomisierte Zwillingstudie statt interindividuellem Randomisierungsdesign) die Gesamtstichprobengröße bei gleicher Effizienz von 20.000 auf 100 Personen (= 50 Zwillingspaare) hätte gesenkt werden können (wobei diese Anzahl von Zwillingen wahrscheinlich schon unsystematisch in der Stichprobe vorhanden waren; vgl. p. 405).

Zusammenfassend lassen sich Experimente als geplante Beobachtungen unter kontrollierten Randbedingungen beschreiben. Sie sind insbesondere dadurch gekennzeichnet, daß eine zufallsgesteuerte Zuteilung der Versuchspersonen auf die unterschiedlichen (variierten) Bedingungen erfolgt. Durch diese Ran-

domisierung werden kausale Schlußfolgerungen ermöglicht, da mögliche Alternativerklärungen (Umkehrungen der Kausalrichtung, Erklärungen durch dritte Variablen) ausgeschlossen werden können. Weiterhin erlaubt die Randomisierung eine Anwendung statistischer Testverfahren zur Auswertung von Experimenten. Die Kosten der Randomisierung bestehen in einer Erhöhung des Zufallsfehlers.

## 2.2 Zur Rolle von systematischen Störfaktoren

Ein mögliches Problem bei der Durchführung von Experimenten sind systematische Fehler oder Störfaktoren, die nicht durch Merkmale der Versuchsperson bedingt sind. Von systematischen Störeinflüssen wird gesprochen, wenn sich die untersuchten Bedingungen nicht nur durch die variierte unabhängige Variable, sondern durch andere Determinanten der Experimentalsituation unterscheiden. Im angeführten Beispiel wäre dies dann der Fall, wenn die Versuchsleiterin in der Gruppe mit gewalttätigen Filmen barsch, streng und ungerecht reagieren würde, in der anderen Gruppe dagegen freundlich, zuvorkommend und ausgleichend. In einem solchen Fall wären Unterschiede zwischen den Bedingungen nicht zweifelsfrei auf die variierte Variable (= gewalttätige Filme) zurückzuführen, sondern könnten ebenso durch das unterschiedliche Verhalten der Versuchsleiterin hervorgerufen sein. Versuchsleitereffekte, die oft aus dem Wissen eines Experimentators über die Art der experimentellen Intervention und ihre theoretische Bedeutung resultieren (vgl. Rosenthal, 1966, 1967, 1969), sowie andere systematische Fehler, die durch den spezifischen „Aufforderungscharakter“ einer Experimentalsituation entstehen (vgl. Orne, 1962, 1969; Aronson et al., 1990, pp.292ff.), müssen bei der Planung von Experimenten in Betracht gezogen werden. Dazu sind oftmals Vorversuche notwendig. Besteht der Verdacht, daß das Wissen des Versuchsleiters über die experimentelle Beeinflussung das Verhalten bestimmen können, dann sollte der Experimentator hinsichtlich der jeweiligen Versuchsbedingung „blind“ sein.

Eine Vorkehrung gegen systematische Fehler, bei der sowohl dem Versuchsleiter wie auch der Versuchsperson jegliches Wissen über die Art der konkreten Beeinflussung vorenthalten wird, ist der sogenannte **Doppel-Blind-Versuch**, der in der medizinischen Forschung eine wichtige Rolle spielt, gerade bei Prüfung der Wirkung von Pharmaka im Vergleich zu Scheinpräparaten (Placebos). Viele Versuchspersonen reagieren auf Präparate ohne aktive Substanzen und zeigen dabei so unerwartete Effekte wie das Schrumpfen von Tumoren (Klopfer, 1957). Insgesamt geht man davon aus, daß Placebos im Durchschnitt 30-40 Prozent der Wirkkraft von üblichen Medikamenten entwickeln (Price, 1984). Die Wirkung wird dabei auf das Bewußtsein der Ver-

suchspersonen zurückgeführt, sich in einer therapeutischen Situation zu befinden, in der bestimmte Medikamente vom Arzt verordnet wurden und deshalb entsprechende Wirkungen erwartet werden können.<sup>4</sup>

Die Auswirkungen des Placebo-Effekts hat in der Medizin zu einem Standard geführt, daß die Wirksamkeit von Medikamenten immer mit einer Placebo-Medikation im Doppel-Blind-Versuch verglichen werden muß, bei dem weder die Versuchspersonen noch die Versuchsleiter die jeweilige Versuchsbedingung (Medikament oder Placebo) kennen.

An dieser Stelle sollen auch diejenigen Grenzen experimenteller Kontrollmöglichkeiten erwähnt werden, die mit der jeweiligen Intervention und dem gewählten Versuchsplan verbunden sind. Ein Experiment wird in der Regel mit dem Ziel durchgeführt, eine bestimmte Theorie oder Hypothese zu testen. Dabei werden aber zwangsläufig Bedingungen geschaffen, die aus einem anderen theoretischen Blickwinkel in anderer Weise interpretiert werden können (s. Gadenne, 1976, S. 31ff., für entsprechende Beispiele aus der Psychologie). Mit anderen Worten: Auch bei gewissenhaftester Durchführung von Experimenten können diejenigen Alternativerklärungen nicht eliminiert werden, die der verwendeten Intervention aus einer anderen theoretischen Perspektive eine neue Interpretation zumessen. Somit muß die Aussage, daß Randomisierung alle möglichen Alternativerklärungen ausschließt, relativiert werden. Durch Randomisierung werden lediglich solche Alternativerklärungen ausgeschlossen, die nicht durch die gewählte Manipulation der unabhängigen Variablen selbst entstanden sind. Darin ist jedoch kein Nachteil der experimentellen Methode zu sehen, da durch die Randomisierung immerhin mehr Alternativerklärungen ausgeschlossen werden können als durch jedes andere Verfahren.

### 2.3 Kennzeichen eines guten Experiments - experimenteller Realismus

Wie wir bisher ausgeführt haben, sind Experimente Beobachtungen unter kontrollierten Randbedingungen, bei denen eine oder mehrere Variablen systematisch variiert werden (= unabhängige **Variablen**), um ihre Effekte auf die sogenannten **abhängigen Variablen** zu messen. In den seltensten Fällen ist das Interesse dabei aber ausschließlich auf die manipulierten oder gemessenen Variablen gerichtet. Vielmehr steht dahinter im allgemeinen eine Hypothese oder Theorie, die allgemeinere Aussagen beinhaltet (zum Problem der Hypothesen und ihrer Prüfung s. Hussy und Möller - Kapitel 11 -, zum Stellenwert von

<sup>4</sup> In ähnlicher Weise wirken könnte das Bewußtsein, sich an einem psychologischen Experiment zu beteiligen, insbesondere dann, wenn die Hypothesen des Experiments bekannt sind oder als bekannt erscheinen.

Theorien s. Gadenne, Kapitel 7 in diesem Band). Dies soll anhand des bisher verwendeten Beispiels erläutert werden.

Nehmen wir an, statt einer realistischen Filmdarstellung von Gewalt sehen die Kinder einen Zeichentrickfilm gewalttätigen Inhalts. Die zu prüfende Hypothese beinhaltet jedoch keine Differenzierung zwischen den unterschiedlichen Arten des Stimulusmaterials und impliziert, daß jegliche Darstellung von Gewalt in Filmen zu aggressivem Verhalten führt. Zeichentrickfilme werden lediglich als eine mögliche experimentelle Realisierung der theoretischen Einflußvariable gewählt. Die Prüfung der Hypothese wird nun in starkem Maße davon bestimmt, ob und inwieweit die von der Hypothese postulierten Konstrukte tatsächlich im Experiment umgesetzt werden konnten. Dies gilt für die abhängige Variable ebenso wie für die unabhängige Variable.

Welche Realisierungen im einzelnen gewählt werden, wird vor allem von Eigenschaften der Untersuchungsteilnehmer bestimmt. Realisierungsziel sollte es sein, Situationen zu schaffen, die bei den Versuchspersonen ein **Höchstmaß an Involviertheit** auslösen. Dazu muß ein Mindestmaß an Aufmerksamkeit und Interesse geweckt werden. Teilnehmer an der Aggressionsstudie sollten somit für den Fernsehfilm interessiert werden, so daß die Aufmerksamkeit auf das Stimulusmaterial gerichtet ist und nicht auf Aspekte, die für die Untersuchung bedeutungslos sind und die Wirkung der unabhängigen Variablen abschwächen oder gar verfälschen. Aus denselben Gründen sollte der Inhalt der Darstellungen für die gewählte Teilnehmergruppe möglichst verständlich präsentiert sein. Allgemein läßt sich sagen, daß sämtliche experimentellen Realisierungen sich an der Situation der Versuchspersonen orientieren müssen.

Dies gilt jedoch nicht nur für die Versuchspersonen, die der Experimentalbedingung zugeteilt wurden, sondern in gleichem Maße für die Kontrollgruppe. Im gewählten Beispiel gilt es, eine Kontrollbedingung herzustellen, in der sich die Versuchssituation von der Situation der Experimentalgruppe ausschließlich durch die Abwesenheit der unabhängigen Variablen unterscheidet. Idealerweise sollten die Kinder der Kontrollgruppe einen Zeichentrickfilm sehen, der keine Gewalt enthält, aber ansonsten keine Unterschiede zum Experimentalfilm aufweist. Vor allem sollte der Kontrollfilm die Aufmerksamkeit und das Interesse der Kinder in ähnlicher Weise ansprechen. Falls den Kindern in der Kontrollbedingung zum Beispiel ein Landschaftsfilm gezeigt wurde, könnte es sein, daß die Kinder Langeweile verspüren und ihr Aktivitätsniveau sich aus diesem Grund reduziert. Unterschiede zwischen den Gruppen könnten dann ebenso auf die hemmende Wirkung der Kontrollsituation zurückgeführt werden. Das Beispiel macht deutlich, daß Schlußfolgerungen aus experimentell induzierten Unterschieden erschwert werden, wenn sich die Realisierung von Experimental- und Kontrollbedingung in mehreren Details unterscheiden.

Wie bereits erwähnt, gelten für die Messung der abhängigen Variablen ähnliche Überlegungen. Es versteht sich von selbst, daß die Messung der Aggressivität bei Kindern sich von der Aggressivitätsmessung bei Erwachsenen unterscheidet. Ein Fragebogen zur Selbsteinschätzung des Verhaltens wird bei Kindern weniger geeignet sein und entsprechende Antworten werden durch die experimentelle Variable anders beeinflußt werden als bei Erwachsenen. Deshalb wäre im Beispielerperiment eine Realisierung vorzuziehen, bei der Interaktionen der Kinder aufgezeichnet und dann von geeigneten Personen (z. B. von Peers oder Kindergärtnerinnen) bewertet werden.

Zusammenfassend gilt, daß die experimentelle Realisierung sämtlicher Variablen wesentlich von der ins Auge gefaßten Versuchspopulation abhängig ist. Es soll eine Situation geschaffen werden, in der die postulierten Mechanismen wirken können, die aber hinsichtlich alternativer Wirkmechanismen kontrolliert ist.

Dazu ist es allerdings nicht notwendig, Realisierungen zu finden, die den Abläufen der Welt außerhalb des Laboratoriums möglichst genau entsprechen. Es gilt vielmehr, die theoretischen Konstrukte adäquat umzusetzen und eine Situation zu schaffen, welche die Operationalisierung für die Versuchsperson plausibel und interessant einbettet. Dieses Ziel wurde von Aronson und Carlsmith (1968) als „**experimenteller** Realismus“ gekennzeichnet.

Ob experimenteller Realismus dadurch erreicht wird, daß das Experiment in natürlicher Umgebung stattfindet oder ob völlig artifizielle Situationen geschaffen werden, ist unerheblich. Ein gutes Beispiel einer artifiziellen Situation mit hohem experimentellen Realismus sind die Milgram-Experimente (1963, 1973), in denen Versuchspersonen als Lehrer in vorgeblichen Lernexperimenten ihren „Schülern“ Stromstöße als Bestrafung für Fehler austeilen mußten. Bekanntlich erteilten die Versuchspersonen auf Anweisungen des Versuchsleiters Stromstöße selbst dann, wenn sie annehmen mußten, daß diese gefährlich oder gar lebensbedrohlich für die „Schüler“ waren. Der geschilderte Versuchsaufbau wurde in verschiedenen Teilerperimenten variiert und diente zur Prüfung von Determinanten gehorsamen Verhaltens.

Obwohl die Situation in bezug auf die Alltagsrealität völlig artifiziell war, beschrieben die meisten Teilnehmer das Experiment als äußerst einschneidend und emotional stark involvierend.<sup>5</sup> Die Ergebnisse des Milgram-Experiments waren denn auch von höchster gesellschaftlicher Relevanz. Die Bedeutung der Erkenntnisse bezog sich jedoch nicht auf die Rolle der Bestrafung bei Lern-

<sup>5</sup> Zur ethischen Diskussion des Milgram-Experiments vgl. Baumrind, 1964, 1985; Milgram, 1964; zur Ethik des Experiments allgemein vgl. Aronson et al., 1990, pp. 83ff., sowie Abschnitt 6 weiter unten.

aufgaben (dies wurde durch simple Analogie zur Alltagsrealität nahegelegt), sondern auf Autoritätssituationen im allgemeinen.

An diesem Beispiel und aus den vorangegangenen Überlegungen wird deutlich, daß die Oberflächenähnlichkeit einer Experimentalsituation zu einer natürlichen Situation weder eine notwendige noch eine hinreichende Voraussetzung für hohe Aussagekraft der Ergebnisse darstellt. Insofern hat die oft beschworene „ökologische Validität“ eines Experiments keineswegs die ihr oft zugemessene Bedeutung (z.B. bei Dörner und Lantermann, 1991). Vielmehr kommt es darauf an, die experimentelle Studie so zu gestalten, daß die zugrundeliegende Theorie einem möglichst aussagekräftigen Test unterzogen wird. Denn es ist die Theorie, die zur Erklärung von natürlichen Situationen herangezogen wird und nicht der Analogieschluß vom Experiment auf die Realität.

Ein letztes Problem bei der Umsetzung von Hypothesen in experimentelle Verfahren betrifft die **Standardisierung** der Realisierung von Variablen in verschiedenen Studien. In der Forderung nach Standardisierung verbirgt sich das Postulat, theoretische Konstrukte immer auf gleiche Weise empirisch zu realisieren. Dadurch soll die Vergleichbarkeit von Ergebnissen verschiedener Studien verbessert werden. Eine solche Standardisierung ist in verschiedenen Situationen unabdingbar (z.B. Medikamentenstudien), hat aber auch - gerade in der Sozialpsychologie und der kulturell vergleichenden Psychologie (vgl. Aronson et al., 1990, pp.226ff.) - schwerwiegende Nachteile.

Beispielsweise wurde ein in den fünfziger Jahren verwendetes Verfahren zur Auslösung von Schamgefühlen durch Vorlesen bestimmter Worte (vgl. Aronson & Mills, 1959), die damals als anzüglich galten, heute eher Heiterkeit als Scham hervorrufen. Standardisierung hat also dort ihre Grenzen, wo aus historischen, kulturellen, populationspezifischen oder anderen Gründen die Bedeutung des Stimulus eine Veränderung erfahren könnte. Anders ausgedrückt ist es oft nicht sinnvoll, die Standardisierung auf der Stimulusebene zu fordern und deshalb analoge Stimulussituationen herzustellen. Sinnvoller ist es dagegen, die Vergleichbarkeit der subjektiven Erfahrungen anzustreben, die durch die Stimulussituation ausgelöst werden sollen. Dazu ist es oft nötig, unterschiedliche Realisierungen zu wählen, die dem jeweiligen Kontext und der Population der Untersuchungsteilnehmer angemessen sind.

Darüber hinaus werfen standardisierte Vorgehensweisen ein Interpretationsproblem auf. Wenn ein theoretisches Konstrukt wiederholt mit nur einer Realisierung experimentell geprüft wurde, besteht die Möglichkeit, daß der ermittelte Effekt auf die spezifische Realisierung, nicht jedoch auf das dahinter vermutete Konstrukt zurückzuführen ist. Übersetzt man das theoretische Konstrukt dagegen auf verschiedene Arten in experimentelle Vorgehensweisen und erhält immer den gleichen Effekt, so ist dessen Unabhängigkeit von der

jeweiligen Realisierung belegt, und das Vertrauen in die theoretische Kausalbeziehung gestärkt.

Die letzten beiden Punkte lassen sich folgendermaßen zusammenfassen: Experimente sollten hohen experimentellen Realismus anstreben, während Alltagsnähe kein Ziel an sich darstellt. Standardisierung in der Realisierung theoretischer Konstrukte hat neben dem Vorteil der Vergleichbarkeit auch potentielle Nachteile, die im Einzelfall jeweils gegeneinander abzuwägen sind.

## 2.4 Probleme bei der Realisierung der abhängigen Variablen: Verhaltensmessung und Kontexteinflüsse

Bisher wurden allgemeine Prinzipien der experimentellen Vorgehensweise dargestellt. Dabei konzentrierten sich die Argumente auf Kontrolle von systematischen Einflüssen durch Randomisierung sowie auf die Herstellung von experimentellem Realismus, d.h. auf das Schaffen einer geeigneten Umsetzung der theoretischen Konstrukte, die für die Versuchspersonen involvierend und interessant ist.

Nachdem die unabhängigen Variablen zur Wirkung gebracht wurden, ist die adäquate Messung der erzielten Effekte erforderlich. Mit der Wirkungsmessung beschäftigen sich die nächsten Abschnitte, welche die Themen „Verhaltensmessung“ und „verbale Maße“ (Fragebogen- bzw. Interviewerhebungen) zum Thema haben.

**Verhaltensmessung:** Die Messung von Verhalten ist als abhängige Variable im allgemeinen aufwendiger als die Erhebung verbaler Maße und wird - zum Teil aus diesem Grunde - seltener angewandt. Bei Verwendung von Verhaltensmessungen in einem Experiment werden diese meist so in den Ablauf eingebaut, daß die Versuchspersonen sich gar nicht darüber bewußt sind, an einem Experiment teilzunehmen. Beispielsweise könnte aggressives Verhalten in einem Kindergarten gemessen werden, indem die Reaktionen der Kinder auf ein Kasperltheater für die Kinder nicht sichtbar aufgezeichnet und dann von KindergärtnerInnen in Hinblick auf Aggressivität beurteilt werden (vgl. Metz, 1989). In einem solchen Versuchsplan ist es zusätzlich möglich, die Verhaltensäußerungen blind - d.h. ohne Kenntnis der Versuchsbedingung - beurteilen zu lassen.

Die Messung von Verhalten ohne entsprechendes Bewußtsein der Versuchspersonen vermeidet das Problem der **Reaktivität**. Unter Reaktivität versteht man das Phänomen, daß die bloße Messung der abhängigen Variablen mit der unabhängigen Variablen oder anderen Ereignissen während des Experiments interagiert und so Wirkungen hervorruft, die ansonsten nicht aufgetreten wä-



ren (vgl. Aronson et al., 1990, p.267; oder Cook & Campbell, 1979). In unserem Beispiel zur Wirkung von Fernsehen auf Aggressivität wäre Reaktivität dann gegeben, wenn die Messung von Aggressivität so erfolgte, daß dadurch die Wirkungen der Filme mit gewalttätigen Inhalten verändert wurden. Nehmen wir beispielsweise an, solche Filme führten normalerweise zu aggressiven Verhaltensweisen. Die Messung der Aggressivität wird aber im Experiment so auffällig betrieben (z.B. durch das Filmen von allen aggressiven Akten, und nur von diesen, mit einer Videokamera), daß die Kinder gerade aus diesem Grund aggressive Handlungen unterlassen.

Eine anderes Beispiel für die nicht-reaktive Erhebung von aggressivem Verhalten findet sich bei Rehm, Steinleitner und Lilli (1987). Dort wurde die Wirkung von De-Individuation auf Aggression dadurch überprüft, daß als Studienreferendare angekündigte Versuchsleiter dem Turnunterricht beiwohnten, mit plausibler Begründung die unabhängige Variable durch Trikotvergabe bei Handballspielen variierten und die abhängige Variable als Anzahl von Fouls protokollierten. Diese Art der Messung aggressiven Verhaltens war für die Versuchspersonen unauffällig und von daher weitgehend „nonreaktiv“, da von Referendaren ohnehin Aufzeichnung über den Unterrichtsverlauf erwartet werden. In diesem Fall war daher auszuschließen, daß Messung selbst den Gegenstand wesentlich veränderte.

Zusammenfassend kann festgehalten werden, daß Verhaltensmaße dazu dienen können, Probleme der Reaktivität auszuschalten (vgl. dazu auch den Überblick von Webb, Campbell, Schwartz & Sechrest, 1966). Dies ist bei besonderen Vorkehrungen zwar auch durch verbale Maße möglich (s.u.); Verhaltensmaße sind aber oft die einfachere und elegantere Wahl.

Ein weiterer Vorteil von Verhaltensmaßen wird darin gesehen, daß ein **gefordertes Verhalten** Versuchspersonen normalerweise **stärker involviert** als verbale Äußerungen. Weiter oben wurde die Wichtigkeit von Involviertheit der Versuchspersonen in Experimenten bereits angesprochen.

Kritik an Verhaltensmaßen wurde vor allem wegen des Auftretens eines größeren Zufallsfehlers geäußert. Dieser ergibt sich auch daraus, daß sich beobachtete Verhaltensweisen oft einer eindeutigen Interpretation und Klassifikation entziehen, was die Zuverlässigkeit der Messung stärker beeinträchtigt als bei vielen Fragebogenmaßen.

Die Wahl eines Typs der abhängigen Variablen schließt jedoch die Verwendung des anderen keineswegs aus: häufig ergänzen sich Verhaltensmessungen und verbale Maße. Dies ist der Fall, wenn zusätzlich zur Antwort auf eine Frage die benötigte Antwortzeit gemessen wird (vgl. Fazio, 1990) oder wenn nach der Verhaltensmessung ein Fragebogen mit dem Ziel ausgegeben wird, Mehrdeutigkeiten bei der Verhaltensinterpretation und -klassifikation zu ver-

meiden (vgl. dazu Webb et al., 1966, S. 1). Bei der Bestimmung der abhängigen Variablen ist deshalb die Alternative „Verhaltensmessung vs. Fragebogenerhebung“ keinesfalls eine Grundsatzentscheidung. Vielmehr geht es darum, im Einzelfall die postulierten theoretischen Konstrukte möglichst eindeutig zu erfassen.

**Kontexteinflüsse:** Entscheidet sich der Forscher dazu, die Antwort der Teilnehmer auf eine oder mehrere Fragen als abhängige Variable zu verwenden, so ist es notwendig, sich der möglichen Einflüsse von Faktoren bewußt zu sein, die nicht nur durch den Inhalt der Frage bestimmt werden, sondern durch den Kontext, in dem die Frage gestellt wird. Solche Kontext-Einflüsse können die Antwort in systematischer Weise verzerren und so das Ergebnis einer Untersuchung bestimmen.

Um herauszufinden, wie sich Kontextfaktoren auf die Beantwortung von Fragen in standardisierten Situationen auswirken, ist es wichtig, die psychologischen Prozesse zu verstehen, die der Fragebeantwortung zugrunde liegen. Mit diesem Ziel wurde in letzter Zeit von kognitiv orientierten (Sozial-)Psychologen und Umfrageforschern eine große Zahl von Untersuchungen durchgeführt (als Überblick, vgl. Bless, Strack & Schwarz, im Druck; Schwarz, Strack & Hippler, 1991; Strack, im Druck).

Diese Studien basieren auf der Annahme, daß Befragte eine Reihe von Aufgaben lösen müssen (vgl. u.a. Strack & Martin, 1987). Sie müssen zunächst die gestellte Frage verstehen. Durch semantische und pragmatische Interpretation versucht der Befragte, die intendierte Bedeutung herauszufinden. Dabei spielen Aspekte des Fragenkontextes eine entscheidende Rolle. So kommt es darauf an, welche speziellen Inhalte durch eine frühere Fragen aktiviert wurden und deshalb leicht aus dem Gedächtnis abrufbar sind (vgl. Nebel, Strack & Schwarz, 1989; Strack, Schwarz & Wänke, 1991). Darüber hinaus können Befragte aus dem Kontext erschließen, welche von mehreren möglichen semantischen Bedeutungen vom Fragesteller gemeint ist. Als „cues“ für derartige Schlußfolgerungen können sowohl vorangehende Fragen (vgl. Schwarz, Strack & Mai, 1991; Strack, Martin & Schwarz, 1988) als auch die verwendete Antwortskala (vgl. Schwarz, Strack, Müller & Chassein, 1988) dienen.

So fanden Strack, Martin und Schwarz (1988), daß Informationen über spezielle Lebensbereiche (in diesem Fall über romantische Beziehungen zum anderen Geschlecht), die durch eine Vorfrage aktiviert wurden, dann nicht in die Beurteilung der allgemeinen Lebenszufriedenheit einbezogen wurden, wenn beide Fragen als zusammengehörig dargestellt wurden. Dieser Befund (vgl. auch Schwarz, Strack & Mai, 1991) legt nahe, daß Personen auch in standardisierten Frage- und Antwortsituationen Regeln befolgen, die in natürlichen Konversationen verhaltensbestimmend sind (vgl. Grice, 1975).

In einer Studie zum Einfluß der vorgegebenen Antwortskala auf die Interpretation der Frage ließen Schwarz et al. (1988) Befragte berichten, wie oft sie sich „ärgern“. Um diese Frage zu beantworten, muß der Befragte herausfinden, was der Fragesteller mit „sich ärgern“ meint. Geht es um unbedeutende Ärgernisse des Alltags oder um schwerwiegendere Sachverhalte? In einem solchen Fall kann die Antwortskala zur Interpretation herangezogen werden. So fanden Schwarz et al. (1988), daß Befragte an extremere Beispiele von Ärger-Erfahrungen dachten, wenn die Skala niedrige Antwortvorgaben für die Häufigkeitsurteile enthielten („weniger als einmal alle drei Monate“ bis „mehrmals pro Jahr“), als wenn die Skala hohe Vorgaben enthielt („weniger als einmal pro Woche“ bis „mehrmals täglich“).

Wenn die Versuchspersonen die Frage nun (korrekt oder inkorrekt - vom Standpunkt des Versuchsleiters aus betrachtet) interpretiert haben, müssen sie eine Antwort generieren. Sofern nicht bereits zu einem früheren Zeitpunkt ein entsprechendes Urteil gebildet wurde, das die Versuchspersonen zum Befragungszeitpunkt aus dem Gedächtnis abrufen können (vgl. Lingle & Ostrom, 1979), muß die Urteilsbildung in der aktuellen Situation erfolgen. Dann müssen relevante Informationen aus dem Gedächtnis abgerufen werden, die das Generieren eines Urteils zum Befragungszeitpunkt ermöglichen. Solche Informationen können das eigene Verhalten betreffen, wenn z.B. nach der Häufigkeit des eigenen Fernsehkonsums gefragt wird und auf dieser Grundlage Urteile über die Zufriedenheit mit bestimmten Lebensaspekten gebildet werden (vgl. Schwarz, Hippler, Deutsch & Strack, 1985). Solche Informationen können Erinnerungen an spezifische hedonisch relevante Ereignisse sein, die dann die Beurteilung des eigenen Wohlbefindens beeinflussen (vgl. Strack, Schwarz & Gschneidinger, 1985). Dabei fungiert die Zugänglichkeit („accessibility“) und die Verfügbarkeit („availability“) von Informationen als wichtige Urteilsdeterminante. Die Zugänglichkeit bestimmt, ob spezifische Inhalte in die Urteilsbildung eingehen. Ein Befragter, dem eine oder mehrere Episoden aggressiven Handelns zum Urteilszeitpunkt einfallen, wird den Handelnden für aggressiver halten als ein Befragter, der keine entsprechenden Inhalte aus dem Gedächtnis abrufen kann. In ähnlicher Weise kann aber auch der Prozeß der Abrufung von Verhaltensepisoden das Urteil bestimmen. Wenn es für den Befragten leicht (oder schwierig) ist, bestimmte Inhalte abzurufen, dann kann die empfundene Leichtigkeit oder Schwierigkeit zur Urteilsgrundlage werden (vgl. Salancik & Conway, 1975; Schwarz, Bless, Strack, Klumpp, Rittenauer-Schatka & Simons, 1991). Jemand, der ohne große Anstrengung eine aggressive Verhaltensepisode einer Person erinnert, wird dem anderen eher die Eigenschaft „aggressiv“ zuschreiben, als einer, der dazu lange nachdenken muß (vgl. Tversky & Kahneman, 1973). Beide Aspekte werden durch die vorherige Verwendung der jeweiligen Informationen beeinflusst. Das heißt, eine vorangehende Frage, zu deren Beantwortung eine bestimmte In-

formation aktiviert werden muß, beeinflusst die Leichtigkeit der anschließenden Abrufung und damit Urteile, bei denen die Verfügbarkeit der Information als Heuristik herangezogen wird.

Nachdem Informationen aus dem Gedächtnis abgerufen sind, muß der Befragte entscheiden, wie diese Informationen bei der Urteilsbildung verwendet werden sollen. Dabei kommen zwei Möglichkeiten in Betracht. Zum einen können die Informationen als Urteilsgrundlage wirken und das Urteil in die Richtung ihrer Implikationen beeinflussen (Assimilationseffekt). Zum anderen können Informationen als Vergleichsstandard in die Urteilsbildung eingehen und das Urteil in die entgegengesetzte Richtung beeinflussen (Kontrasteffekt). Die Verwendung der Information wird vor allem durch ihre Ähnlichkeit mit dem Urteilsgegenstand auf der Urteilsdimension beeinflusst (vgl. Herr, Sherman & Fazio, 1983). Aber auch Ähnlichkeit auf einer peripheren Dimension kann urteilsbestimmend sein. Eine solche Determinante ist die zeitliche Distanz. Informationen, die aus der entfernten Vergangenheit kommen, werden eher zum Vergleichsstandard und führen zu Kontrasteffekten bei der Beurteilung eines aktuellen Sachverhalts als Informationen, die der Gegenwart entstammen. Letztere führen zu einem Assimilationseffekt und beeinflussen das Urteil in Richtung ihrer Implikationen (vgl. Strack et al., 1985). Ähnliches gilt für Erwartungen, die in der Zukunft angesiedelt sind (vgl. Strack, Schwarz & Nebel, 1990). Angewandt auf das Beispiel des vorliegenden Kapitels wurde man vorhersagen, daß extrem aggressive Verhaltensepisoden aus der entfernten Vergangenheit einer Zielperson dazu führen, daß gegenwärtiges Verhalten als weniger aggressiv eingeschätzt wird, während Verhaltensepisoden aus der Gegenwart zu einer Erhöhung des Aggressionsurteils führen.

Da nicht anzunehmen ist, daß Befragte ihr Urteil in den Kategorien im Gedächtnis repräsentiert haben, die vom Fragesteller zur Beantwortung vorgegeben wurden (z. B., als „+3“ auf einer bipolaren 7-Punkte-Skala), ist es nötig, die Antwort in ein vorgegebenes Antwortformat zu „übersetzen“. Dabei geht es vor allem darum, die Antwortskala an einem Urteilsgegenstand zu „verankern“. Meistens wird einer der Extremwerte der Skala einem möglichen Urteilsgegenstand zugewiesen. Wird beispielsweise ein bestimmtes aggressives Verhalten mit einem Skalenwert verknüpft, dann werden Verhaltensweisen, die sich durch weniger Aggressivität davon unterscheiden, mit einem kleineren Skalenwert versehen. Wäre dagegen ein aggressives Verhalten anderer Ausprägung mit demselben Skalenwert verbunden worden, dann hätte sich auch die Zuordnung der übrigen Skalenpunkte verändert. Eine solche Verankerung der Antwortskala läßt sich durch die Veränderung der „Perspektive“ der Versuchspersonen beeinflussen. Zahlreiche Studien zu diesem Problembereich wurden von Upshaw und Ostrom durchgeführt (vgl. Ostrom & Upshaw, 1968; s. a. Parducci, 1965).

Schließlich ist der soziale Kontext bei der Urteilsabgabe von Bedeutung. Dies gilt vor allem dann, wenn eine Antwort offen und nicht anonym angegeben werden muß. Unter solchen Bedingungen ist es besonders wahrscheinlich, daß Einflüsse „sozialer Erwünschtheit“ (vgl. Crowne & Marlowe, 1964) das Urteil bestimmen. Befragte möchten einen guten Eindruck erwecken oder zumindest einen negativen Eindruck vermeiden. In den meisten Fällen bewirkt eine offene Antwortabgabe eine positivere Selbstbeschreibung (vgl. aber Strack, Schwarz, Chassein, Kern & Wagner, 1990, für die Beschreibung von Situationen mit gegenteiligen Effekten).

In den vorangegangenen Abschnitten konnten nur einige ausgewählte Einflüsse des Kontexts auf die Beantwortung von Fragen dargestellt werden, die im Verlauf der Befragung - von der Interpretation der Frage bis zur Antwortabgabe - zum Tragen kommen. Im Rahmen des vorliegenden Kapitels ist eine auch nur annähernd vollständige Darstellung ausgeschlossen. Deshalb konnte das Ziel dieser Ausführungen lediglich sein, die Sensibilität des Forschers für Einflüsse dieser Art zu schärfen. Bleibt die Frage nach den Konsequenzen. Einerseits erscheint es erstrebenswert, ungewünschte Kontexteinflüsse zu eliminieren. Andererseits gibt es keine „kontextfreien“ sozialwissenschaftlichen Untersuchungen. Aber dennoch wird sich ein Forscher bemühen, Einflüsse, von denen er weiß, daß sie die Ausprägung der abhängigen Variablen bestimmen und möglicherweise die Wirkung der experimentellen Manipulation überdecken, zu vermeiden. Falls dies nicht möglich ist, muß sichergestellt sein, daß sich die experimentellen Bedingungen nicht systematisch hinsichtlich der Kontextvariablen unterscheiden. Darüber hinaus können jedoch Kontexteinflüsse als Forschungsinstrument zur Geltung kommen. Wenn zum Beispiel die Aktivierung von Gedankeninhalten durch eine Vorfrage nicht als Störfaktor betrachtet wird, sondern als Möglichkeit, experimentell nicht manipulierbare Einflüsse der experimentellen Erforschung zugänglich zu machen (als Beispiel vgl. Schwarz & Strack, 1981; Strack, Schwarz, Weidner, Hippler & Schwarz, 1983), erscheint es nicht als notwendige Konsequenz, Einflüsse dieser Art in jedem Fall eliminieren zu wollen. Vielmehr wird es sinnvoll, sie gezielt herbeizuführen und im Dienste von verschiedenen Fragestellungen fruchtbar zu machen.

## 2.5 Zur Frage von intraindividuellen vs. interindividuellen Versuchsplänen

Im vorangegangenen Abschnitt wurde in rudimentärer Weise versucht, eine Theorie von Kontexteinflüssen bei der Beantwortung von Fragen vorzustellen. Die Komplexität der dabei gezeigten möglichen Urteilsprozesse macht es unmöglich, einfache allgemeine Ratschläge im Sinne von Kochrezepten abzulei-

ten. Dies gilt auch hinsichtlich der Entscheidung für eine interindividuelle oder eine intraindividuelle Vorgehensweise. Bei **interindividuellen Versuchsplänen** werden die Teilnehmer unterschiedlichen Experimentalbedingungen zugewiesen und die gemessenen Werte werden zwischen den Gruppen verglichen. Bei **intraindividuellen Plänen** werden die Versuchspersonen einer wiederholten Messung unterzogen, wobei die experimentelle Beeinflussung durch einen Vorher-Nachhervergleich der Meßwerte innerhalb der Versuchspersonen bestimmt wird.

Handelt es sich um reaktive Erhebungen (z.B. Befragungen), muß jede Messung als möglicher Einflußfaktor (= treatment) behandelt und als Determinante von nachfolgenden Messungen ernstgenommen werden (s. dazu bereits Campbell & Stanley, 1966). So fanden Mark, Sinclair und Wellens (1991) in einer neueren Untersuchung, daß sich die Beantwortung von Becks Depressionsinventar (BDI) auf nachfolgende Berichte über die eigene Stimmungslage auswirkt. Bei Befragten mit hoher Depressivität führte die Beantwortung des BDI zu negativeren Stimmungsberichten, während Teilnehmer mit niedriger oder mittlerer Depressionsausprägung nach der Bearbeitung des BDI eine positivere Stimmung berichteten. Die Beantwortung eines bestimmten Fragebogens kann auch den Zusammenhang zu anderen Maßen erhöhen. So fanden Knowles und Mitarbeiter (als Überblick s. Knowles et al., im Druck), daß sich im Laufe der Beantwortung eines Persönlichkeitsfragebogens die Korrelation der Beantwortung eines Items mit Gesamtscore erhöht. Dalbert, K. Strack, Boesken und Schwenkmezger (1991) fanden eine Zunahme des Zusammenhangs zwischen Zustandsangst und den Werten auf ungünstigen Dimensionen der Krankheitsbewältigung als Funktion des Bearbeitungs eines Coping-Fragebogens.

Will man derartige Einflüsse vermeiden, sollte interindividuellen Versuchsplänen der Vorzug gegeben werden. Dies gilt vor allem dann, wenn ein gegenseitiger Einfluß der einzelnen Messungen zu vermuten ist und eine explizite Prüfung möglicher Effekte nicht vorgesehen ist (z. B. durch einen Vier-Gruppen-Solomon-Versuchsplan, vgl. Solomon, 1949). Auf der anderen Seite können Erkenntnisse der kognitiven Psychologie zur gezielten Beeinflussung der Antworten durch vorherige Messung zur Aufklärung von Kausalbeziehungen nutzbar gemacht werden (z.B. Schwarz & Strack, 1981). Weiterhin darf nicht übersehen werden, daß Versuchspläne mit Meßwiederholung ein geringeres Maß an Zufallsfehler aufweisen als interindividuelle Pläne und auf diese Weise Hypothesen mit kleinen Effekten leichter geprüft werden können (vgl. bspw. Aronson et al., 1990, pp. 143ff., s. also pp. 138ff.). Schließlich erlauben intraindividuelle Versuchspläne unter bestimmten Umständen (Veränderungen bei abhängigen **und** unabhängigen Variablen) die statistische Kontrolle von nicht gemessenen zeitkonstanten Faktoren wie genetischen Faktoren (Rehm, Armingier & Kohlmeier, 1992).

### **3. Alternativen zum randomisierten Experiment**

Bisher wurden das Experiment als Methode herausgestellt, die in besonderer Weise geeignet ist, eindeutige Schlußfolgerungen über Ursache-Wirkungs-Beziehungen zuzulassen, sowie Merkmale eines guten Experiments erläutert. In diesem Abschnitt sollen einige nichtexperimentelle Herangehensweisen besprochen werden. Die bislang dargestellten Prinzipien werden dabei als (allgemeiner) Maßstab zur Beurteilung dieser Alternativen benutzt (Punkt 3.1). In der Folge werden dann drei spezifische Versuchspläne und die damit verbundenen Möglichkeiten der Kontrolle behandelt: Kohortenstudie, Fall-Kontroll-Studie (3.2) und Einzelfallstudie (3.3).

#### **3.1 Allgemeine Grundsätze zur Erstellung von nicht-experimentellen Forschungsplänen**

Wenn das Experiment als Methode der Wahl anzusehen ist, stellt sich die Frage nach den Bedingungen, unter denen alternative Beobachtungsmethoden in Betracht kommen. Einige Voraussetzungen werden im folgenden aufgeführt.

Eine Bedingung für die Verwendung nicht-experimenteller Versuchspläne liegt vor, wenn das Erhebungsziel keine kausale Fragestellung beinhaltet. Dies trifft immer dann zu, wenn es lediglich um das Aufzeigen der Existenz bestimmter Phänomene geht. Ein Beispiel hierfür wäre ein Forschungsprogramm, mit dem geprüft werden soll, ob es Alkoholiker gibt, die nach einem Entzug lernen können, Alkohol kontrolliert zu konsumieren. Nicht kausaler Art wäre auch eine Fragestellung zur Inzidenz und Prävalenz von psychischen Erkrankungen in der Allgemeinbevölkerung oder eine Fragestellung zum non-verbalen Ausdruck verschiedener Emotionen in unterschiedlichen Kulturkreisen. Die aufgeführten Forschungsfragen haben gemeinsam, daß Ursache-Wirkungs-Zusammenhänge zu ihrer Beantwortung keine Rolle spielen.

Nicht-experimentelle Verfahren sind auch dann angezeigt, wenn die Zuteilung auf Bedingungsvariationen unmöglich oder unethisch ist. Ein Beispiel für die Unmöglichkeit der Zufallszuweisung zu Bedingungen sind alle angeborenen Eigenschaften von Personen. Geschlecht, Körpergröße oder Hautfarbe einer Person sind experimentell nicht manipulierbar und Personen können angeborenen Merkmalen nicht nach Zufallsprinzipien zugeordnet werden. Dennoch haben Bedingungen dieser Art eine große Bedeutung als Kausalfaktoren in psychologischen und sozialwissenschaftlichen Erklärungen. Psychologische Theorien definieren allerdings in den seltensten Fällen solche Eigenschaften als alleinige Ursache, sondern meist als Wechselwirkung mit spezifischen Sozialisationsmechanismen. So geht man beispielsweise davon aus, daß Jungen

in unserer Gesellschaft lernen, ihre Aggressivität eher non-verbal zu zeigen, während Mädchen eher zu verbalen Aggressionen sozialisiert werden (Eagly & Steffen, 1986). Auch solche Sozialisationsmechanismen sind experimentell oft schwer prüfbar, weil sie über einen langen Zeitraum hinweg wirksam werden, während Experimente immer zeitlich eng begrenzt sind.<sup>6</sup> Wie die experimentelle Entwicklungspsychologie gezeigt hat, ist es dagegen sehr wohl möglich, einzelne Kausalfragestellungen der Sozialisation einer experimentellen Prüfung zu unterziehen. Das Lernen von aggressivem Verhalten durch Imitation (Bandura, Ross & Ross, 1963) oder die Rolle von Modellbelohnung oder -bestrafung auf das Imitationslernen (Bandura, 1965) sind Beispiele.

Neben der naturgegebenen Unmöglichkeit einer Bedingungsvariation bestehen oft ethische Bedenken (s. auch Abschnitt 6 unten). Diese treten insbesondere dann auf, wenn aus der systematischen Variation der unabhängigen Variablen gesundheitliche oder psychische Schäden entstehen. Wollte man beispielsweise experimentell zeigen, daß Streß zu akutem Myokardinfarkt führt, so wäre es ethisch nicht vertretbar, eine zufällig ausgewählte Gruppe absichtsvoll streßhaften Situationen zuzuweisen. Zur Beantwortung von solchen Fragestellungen werden heutzutage üblicherweise Fall-Kontroll-Verfahren eingesetzt (vgl. 3.2). Eine mögliche experimentelle Alternative wäre in diesem Fall der Tierversuch, der jedoch neue Probleme (Übertragbarkeit, Ethik bei Tierversuchen) aufwirft.

Somit bleibt festzuhalten, daß Situationen existieren, in denen experimentelle Methoden nicht angewandt werden können. Sollen dennoch kausale Wirkungsmechanismen untersucht werden, kann man die Prinzipien des Experiments (s. Abschnitt 2) als Standard heranziehen, um die Tauglichkeit von alternativen Vorgehensweisen zu bewerten und eine entsprechende Auswahl zu treffen. Diese Vorgehensweise soll im folgenden erläutert werden (die Ausführungen bauen auf Campbell & Stanley, 1966, sowie auf Cook & Campbell, 1979).

Kehren wir zu unserem ursprünglichen Beispiel zurück und nehmen an, die Kindergärtnerin hätte ihre Beobachtungen von Kindern, die nach dem Sehen von gewalttätigen Comic Strips aggressiv reagierten, nur an **einem** Tag durchgeführt. Dies wäre eine sehr schwache Grundlage für eine Generalisierung, weil nicht ausgeschlossen werden kann, daß eine Besonderheit des Erhebungszeitraums für den ermittelten Zusammenhang verantwortlich ist. Etwas trag-

6 Die prinzipiell mögliche Dauer von Experimenten sollte aber nicht zu niedrig veranschlagt werden. Aus der epidemiologischen Forschung sind Arzneimittelversuche bekannt, die auf einen Zeitraum von fünf bis zehn Jahre angesetzt sind. Ein Beispiel ist die sogenannte Physicians' Health Study zum Einfluß von Aspirin und Beta-Karotin auf Herz-Kreislauf-Mortalität (vgl. Hennekens & Buring, 1987, pp. 182ff.). In dieser Studie wurden 22'071 Ärzte den vier Befragungen eines 2-faktoriellen Versuchsplanes mit je zwei Stufen (Gabe von Aspirin und Beta-Karotin jeweils vs. Placebo) randomisiert zugeteilt und mußten über mehrere Jahre in regelmäßigen Abständen die den Bedingungen entsprechende Medikation einnehmen.



fähiger wäre die empirische Basis, wenn das Zeigen von gewalttätigen Filmen an verschiedenen Tagen zu verschiedenen Zeiten wiederholt wurde, und in jedem Fall aggressivere Handlungsweisen nach den Filmen resultierten (im Vergleich zu dem Verhalten vor der Darbietung der Filme). In einem solchen Fall wäre zum einen die Herstellung der Untersuchungsbedingungen durch den Experimentator gegeben und zum anderen ein gewisses Maß an Kontrolle über externe Einflüsse, da durch Replikation des Effektes unter verschiedenen Bedingungen viele potentielle Alternativerklärungen unplausibel werden. Einige Alternativerklärungen können jedoch nur experimentell ausgeräumt werden. Dies gilt für die Möglichkeit, daß nicht der gewalttätige Film die Aggressionen hervorgerufen hat, sondern die damit einhergehende Unterbrechung des normalen Spielalltags zu Frustrationen geführt hat, die ihrerseits die Aggressionen verursacht haben. Hätte man zwei randomisierte Gruppen gebildet und nur einer Gruppe gewalttätige Filme gezeigt und der anderen Gruppe gewaltfreie Filme, so hätte diese Alternativerklärung eliminiert werden können.

Die bisherigen Ausführungen gelten auch für Langzeitbeobachtungen (vgl. dazu auch die Ausführungen zur Kohortenstudie in 3.2). Nehmen wir an, eine psychologische Theorie ginge davon aus, daß gewalttätige Sendungen nicht unmittelbar zu aggressiven Handlungen führen, sondern daß erst die Kumulation gewalttätiger Sendungen allmählich ein Aggressionspotential aufbaut. Falls sich diese Kumulation erst über Jahre hinweg auswirkt, ist eine experimentelle Prüfung schon aus Zeitgründen problematisch. Auch in diesem Fall bieten sich wiederholte Messungen von Ursache (Sehen gewalttätiger Sendungen) und postuliertem Effekt (Aggressivität) ohne Bedingungsvariation als Alternative an. Falls die Langzeithypothese zutrifft, muß für jeden Teilnehmer ein bestimmtes Datenmuster entstehen, das empirisch geprüft werden kann. Als Beispiel sei hier die Forschung von Lefkowitz, Eron, Walder und Huesmann (1972) genannt, die zu einem ersten Zeitpunkt die durchschnittliche Gewalttätigkeit in den beliebtesten Sendungen von Drittklässlern sowie die im Klassenzimmer gezeigte Aggressivität ermittelten. Es zeigte sich der erwartete Zusammenhang in Form von Korrelationen zwischen Expertenratings der Gewalttätigkeit der Fernsehsendungen und Aggressivitätsbeurteilungen der Mitschüler. Zehn Jahre später wurden die gleichen Variablen wieder erhoben. Es zeigte sich, daß der Zusammenhang zwischen Fernsehgewalt und Aggressivität zugenommen hatte, während kein Zusammenhang zwischen aggressivem Verhalten zum ersten Zeitpunkt und der Gewalttätigkeit der Sendungen bestand, die nach zehn Jahren gesehen wurden.

Eine solche Datenstruktur<sup>7</sup> legt nahe, daß die gewalttätigen Sendungen die Ursache von aggressiven Handlungen darstellen. Insofern erlaubt uns das ge-

<sup>7</sup> Die Ausführungen über die Datenstruktur anhand von Korrelationskoeffizienten dienen lediglich zur Illustration des Versuchsplanes und seiner Möglichkeiten. Inzwischen würde man

schilderte Design (sog. „Cross-lagged panel design“) ein größeres Vertrauen in die postulierte Kausalrichtung, als dies bei Vorliegen einer nur korrelativen Beziehung der Fall gewesen wäre. Durch einen solchen Versuchsplan kann allerdings nicht ausgeschlossen werden, daß eine dritte Variable für das gesamte Datenmuster verantwortlich ist.

Noch mehr Kontrolle bieten nicht-experimentelle Verfahren, die sich „natürlich“ auftretende Variationen zunutze machen. Ein Beispiel beruht auf einer Besonderheit bei der Verbreitung des Fernsehens in den USA. Dort wurde das Fernsehen nicht abrupt, sondern in Stufen eingeführt, da eine staatliche Behörde ein Einfrieren der Sendelizenzen zwischen 1949 und 1952 verordnete. Dies führte dazu, daß viele Gemeinden in den USA drei Jahre früher als andere über Fernsehen verfügten. Trifft der postulierte Einfluß des Fernsehens zu (z.B. in Form eines Anwachsens aggressiver Übergriffe durch Jugendbanden laut Statistik), so müßte dieser Effekt in den „eingefrorenen“ Gemeinden drei Jahre später sichtbar werden als in den „nicht eingefrorenen“ Gemeinden. Eine solche Annahme läßt sich in ein statistisches Model überführen und inferenzstatistisch als unterbrochener Zeitreihenversuchsplan mit wechselnden Replikationen testen. Hennigan, Del Rosario, Heath, Cook, Wharton und Calder (1982) prüften mit diesem Erhebungsplan verschiedene Hypothesen über die Wirkung des Fernsehens auf verschiedene Formen der Kriminalität. Die Autoren kamen zu dem Ergebnis, daß Fernsehkonsum tatsächlich einen Einfluß auf eine bestimmte Kriminalitätsart hat, nämlich Diebstahl. Im einzelnen zeigte sich, daß die Häufigkeit von Diebstählen jeweils nach einer fixen Latenzzeit im Anschluß an die Einführung des Fernsehens anstieg und daß dieser Effekt in den „eingefrorenen“ Gemeinden drei Jahre später auftrat als in den „nicht eingefrorenen“ Gemeinden.

Die Autoren teilen in diesem Zusammenhang nicht die Skepsis von Frey und Frenz (1982), die in der Ausrichtung an Gütekriterien des Experiments die Ursache für eine inhaltliche Entleerung der psychologischen Feldforschung sehen. Wir meinen dagegen, daß die Orientierung am Experiment dazu beiträgt, in natürlichen Situationen Bedingungen aufzusuchen und vielleicht zu verändern, die eine besonders diagnostische Prüfung theoretischer Annahmen zulassen.

Im Rahmen eines überblicksartikels kann auf Einzelfragen der Auswahl und Bewertung der vielen möglichen speziellen Versuchspläne für unsere Beispielfragestellung nicht eingegangen werden. (An dieser Stelle sei auf Campbell & Stanley, 1966, sowie Cook & Campbell, 1979, verwiesen, die von ihrer Aktualität nichts eingebüßt haben.) Wichtiger erscheint dagegen der Hinweis auf

---

ein solches „Cross-lagged panel design“ mit statistisch adäquateren Methoden auswerten (vgl. Rogosa, 1980). Entsprechende Reanalysen der Daten von Lefkowitz et al. legten aber keine unterschiedlichen Schlußfolgerungen nahe.

die allgemeinen Prinzipien, mit denen eine Auswahl des geeignetsten Versuchsplans möglich ist. Die Richtlinie lautet: ***Falls Experimente zur Beantwortung kausaler Fragestellungen nicht durchführbar sind, sollten diejenigen Alternativen bevorzugt werden, die den Prinzipien des Experiments hinsichtlich Kontrolle durch Randomisierung am nächsten kommen.***

### 3.2 Kohortenstudie und Fall-Kontroll-Studie

Im bisher verwandten Beispiel bildete Aggressivität oder aggressives Verhalten die abhängige Variable. Bei solchen Verhaltensweise geht man - zumindest bei kurzfristiger situativer „Erzeugung“ ohne langfristige Folgen für die Versuchspersonen - davon aus, daß entsprechende Experimente ethisch vertretbar sind. Anders stellt sich die Situation dar, wenn als abhängige Variablen Krankheiten wie akuter Myokardinfarkt, Brustkrebs oder Alkoholismus untersucht werden.

Getestet werden soll beispielsweise die Hypothese, daß sogenannte kritische Lebensereignisse ohne Vorhandensein von sozialen Netzwerken zur Bewältigung zu Brustkrebs führen. Präziser, da es sich hier um eine unvollständige Hypothese handelt (Gadonne, 1990): daß innerhalb einer Gruppe von Personen, die kritische Lebensereignisse erfahren haben, die Neuerkrankungsrate an Brustkrebs erhöht ist, wenn zur Bewältigung der kritischen Ereignisse keine sozialen Netzwerke zur Verfügung stehen.

Sicherlich besteht in diesem Falle Übereinstimmung, daß Experimente am Menschen zur Prüfung dieser Fragestellung aus ethischen und praktischen Gründen ausscheiden. Als nichtexperimentelle Alternative bieten sich zwei Versuchspläne an. Zum einen kann eine Stichprobe ausgewählt werden, die jeweils nach bestimmten Zeitintervallen hinsichtlich kritischer Lebensereignisse und nach dem Auftreten bestimmter Krankheiten befragt wird (sog. **Kohortenstudie**). Bei hinreichend großer Stichprobe und hinreichend langem Untersuchungszeitraum (das 10-Jahres-Mortalitätsrisiko für Brustkrebs lag Ende der 80er Jahre für die 40-45jährigen im Bereich von 400 pro 100.000; Angabe aus Häußler, Rehm, Naß & Kohlmeier, 1990) kann die o. g. Hypothese einer Prüfung unterzogen werden. Dabei sollte bedacht werden, daß die Stichprobengröße für bekannte Kohortenstudien in den USA zur Prüfung von Ursache-Wirkungs-Zusammenhängen im Krebsbereich 100'000 Personen übersteigt (vgl. Hennekens & Buring, 1987, pp. 22ff.) und auch Untersuchungszeiten von über 10 Jahren die Regel sind. Da die Stärke des Einflusses von kritischen Lebensereignissen auf die Inzidenz von Brustkrebs als eher gering einzuschätzen ist (zu einer Übersicht der Risiken für Brustkrebs vgl. Kohlmeier, Rehm & Hoffmeister, 1990; vgl. auch Geyer, 1991), dürfte die genannte Stichprobengröße eher eine vorsichtige Schätzung darstellen.

Für methodologische Überlegungen bedeutet dies, daß Kohortenstudien für die Prüfung von psychologischen Hypothesen im Bereich der allgemeinen Epidemiologie wohl in den allermeisten Fällen als Methode der Wahl aus Kosten- und Praktikabilitätsgründen ausscheiden. In anderen Bereichen, wenn die untersuchten Merkmale ubiquitär und intervallskaliert (z.B. Größe, Intelligenz) oder relativ häufig in den untersuchten Populationen auffindbar sind (z. B. psychische Störungen mit 21 % in der Allgemeinbevölkerung, vgl. Fichter, 1990, S. 48), bilden Kohortenversuchspläne durchaus ein wichtiges Untersuchungsinstrument (z. B. in der Entwicklungspsychologie oder in der psychiatrischen Epidemiologie).

Die zweite Alternative zum kontrollierten Experiment im Bereich der Versuchsplanung zur Prüfung psychologischer Einflüsse auf degenerative Krankheiten ist die **Fall-Kontroll-Studie**. Auf diese Art Versuchsplan wird hier ausführlich eingegangen, da Fall-Kontroll-Studien keineswegs auf den genannten Bereich beschränkt sind, sondern zu Recht in viele andere Bereiche übertragen werden (als Beispiel für die Übertragung der Methode auf den Bereich Straßenverkehrsunfall s. Fahrenkrug & Rehm, 1992).

Wollte man die Hypothese über den Einfluß von nicht bewältigten kritischen Lebensereignissen auf Brustkrebs mit Hilfe einer klassischen Fall-Kontroll-Studie (vgl. dazu Rothman, 1986, als Einführung; s.a. Schlesselman, 1982) prüfen, wurde man folgendermaßen vorgehen:

- 1) Definition der Fälle ( Beispiel: alle neu auftretenden Fälle von Brustkrebs vom 1.1.1992 - 30.6.1992 in Berlin, die in Krankenhäusern nach ICD diagnostiziert werden);
- 2) Definition der Kontrollgruppe (Dabei sollten die Kontrollpersonen aus der gleichen Population stammen wie die Fälle und diesen mit Ausnahme hinsichtlich der postulierten Einflußfaktoren möglichst ähnlich sein; Beispiel: Frauen gleichen Alters, die vom Einwohnermeldeamt im gleichen Stadtbezirk nach einem Zufallsverfahren ausgewählt werden);
- 3) Befragungen sämtlicher Teilnehmer hinsichtlich der postulierten Einflußfaktoren;
- 4) Auswertung hinsichtlich des Einflusses der postulierten Faktoren auf die abhängige Variable, die der Einteilung in Fälle und Kontrollen zugrundelag.

Wo liegen die methodologischen Probleme der genannten Vorgehensweise? Zunächst in Schritt 2: Wie kann gewährleistet werden, daß die Kontrollpersonen tatsächlich der gleichen Population entstammen wie die Fallpersonen? Um dies sicherzustellen, werden oft Schichtungskriterien festgelegt, die außer den postulierten Kausalfaktoren mit der abhängigen Variablen in Verbindung stehen. Für jede Person in der Fallgruppe werden dann eine oder mehrere Kontrollpersonen gezogen, die die gleiche Konfiguration von Ausprägungen

hinsichtlich der gewählten Schichtungskriterien aufweisen (sog. **Parallelisierung** oder „**Matching**“; vgl. Schritt 2 oben). Dieses Verfahren bietet allerdings nicht die Kontrolle der Randomisierung, weil nicht ausgeschlossen werden kann, daß sich Fall- und Kontrollgruppe in anderen als den kontrollierten Faktoren, die auf die abhängige Variable Einfluß haben, unterscheiden.\* Im vorliegenden Fall ist es jedoch die Möglichkeit, die am meisten Kontrolle möglicher weiterer Einflußfaktoren bietet.

Methodologisch kann das grundlegende Problem anders umschrieben werden: Es gibt kein Verfahren mit Garantie für das Zutreffen der grundlegenden Annahme, daß Fall- und Kontrollgruppe tatsächlich bis auf die unterschiedliche Exposition gegenüber den postulierten Einflußfaktoren vergleichbar wären. Da die Fallgruppe qua definitionem systematisch rekrutiert wird, kann es keine stichprobentheoretischen Verfahren geben, die Kontrollgruppe aus der gleichen Population zu gewinnen.

Eine weitere Problematik von Fall-Kontroll-Studien liegt im dritten Schritt des Verfahrens begründet. Die Befragung über die mögliche Ursache kann aus Gründen des Versuchsplans erst nach dem Eintreten der Konsequenz erfolgen. Deshalb kann nicht ausgeschlossen werden, daß eine umgekehrte Kausalrichtung vorliegt und die Konsequenz die Angaben hinsichtlich der Ursachen determiniert. So wird eine Krankheit wie Brustkrebs sicherlich kognitive Prozesse hinsichtlich möglicher Ursachen in Gang setzen. Es kann durchaus sein, daß diese Suche nach möglichen Ursachen durch das Stellen der Frage nach kritischen Lebensereignissen in eine bestimmte Richtung gelenkt wird mit dem Ergebnis, daß die untersuchten Frauen sich tatsächlich an mehr derartige Ereignisse erinnern (vgl. dazu Geyer, 1991).

Die skizzierte Problematik hat dazu geführt, daß zum einen versucht wird, die möglichen Ursachen mit Tricks vor Eintreffen des Ereignisses zu erheben. Im Falle des Brustkrebs bedeutet dies, daß Frauen, die ins Krankenhaus gingen, um möglichen Brustkrebs abklären zu lassen, vor der eigentlichen Diagnose befragt wurden (für eine Darstellung der unterschiedlichen Vorgehensweisen, vgl. Geyer, 1991). Erst nach Vorliegen des Befunds der medizinischen Untersuchung wurden diese Frauen dann in Fall- und Kontrollgruppe eingeteilt. Auch dieses Design hat Probleme, da sich herausgestellt hat, daß Frauen

---

8 In diesem Punkt weisen die Autoren den Standpunkt von Huber (1987, p. 95) energisch zurück, demnach generell Parallelisieren „bei kleinen Stichproben besser als das Randomisieren“ sei. Wenn man wegen der Stichprobengröße Extremgruppen hinsichtlich einer bekannten Variablen befürchtet, bieten sich zweistufige Verfahren wie z.B. das Bilden von Paarlingen anhand dieser Variablen mit anschließender randomisierter Zuteilung auf Experimental- und Kontrollgruppe an (s. dazu auch die ausführliche Diskussion entsprechender Verfahren in Abschnitt 2.1). In jedem Fall sollte man aber - wenn möglich - Verfahren mit Randomisierung den Vorzug vor Parallelisieren geben, weil nur so unbekannte Einflußfaktoren konstant gehalten werden können.

das Untersuchungsergebnis mit relativ hoher Sicherheit vorhersagen können. Außerdem ist das Verfahren nicht auf andere unabhängige Variablen generalisierbar.

Als weitere Problemlösestrategie wird versucht, durch kognitive Prozesse nicht veränderbare Indikatoren der unabhängigen Variablen zu finden (wie z.B. biochemische Parameter). Leider ist eine solche Strategie hinsichtlich psychisch relevanter Variablen begrenzt.

### 3.3 Einzelfallstudie

Die eingangs geschilderte Situation, in der ein Laie oder Wissenschaftler ohne Verwendung von natürlicher oder induzierter Bedingungsvariation das Verhalten von Kindern einmal beobachtet und dabei zu Einsichten über den Zusammenhang von Fernsehen und aggressivem Verhalten gelangt, ist in der Terminologie von Campbell und Stanley (1966) eine „one-shot case study“. So sehr ein solches Vorgehen von einem Gefühl der unmittelbaren Evidenz begleitet sein mag (vgl. Chapman & Chapman, 1967), so wenig spricht aus methodologischer Perspektive für die kasuistische Methode.

Kasuistisches Vorgehen ist am wenigsten geeignet, allgemeine (d.h. personen- und situationsübergreifende) theoretische Zusammenhänge empirisch zu überprüfen, da die wichtigsten methodologischen Prinzipien einer solchen Prüfung (insbesondere Kontrolle, s. o.) verletzt werden. Deshalb haben Fallstudien vor allem eine heuristische Funktion im Vorfeld experimenteller Untersuchungen. Darüber hinaus gilt die Durchführung einer Fallstudie als indiziert (vgl. Aronson et al., 1990), wenn wegen der Natur der Dinge ein experimentelles oder quasi-experimentelles Vorgehen nicht möglich ist. Einmalige oder seltene Ereignisse können hier als Beispiele angeführt werden.

So ist ein Forscher, der zum spezifischen Verständnis des Unglücks von Tschernobyl beitragen möchte, mit einem bislang historisch einmaligen Fall konfrontiert, der eine Bedingungsvariation oder -beobachtung ausschließt. Bei seiner Arbeit wird es jedoch darauf ankommen, experimentell geprüftes theoretisches Wissen auf den konkreten Fall anzuwenden. Andererseits wird jede Einsicht, die sich aus kasuistischer Betrachtung ergibt, einer experimentellen Prüfung zu unterziehen sein, bevor sie als triftige Evidenz für einen allgemeinen Kausalzusammenhang gelten kann. **Somit beinhaltet kasuistisches Vorgehen in erster Linie Theorieanwendung und weniger Theorieprüfung.**

**Beispiele kasuistischer Vorgehensweise:** Im Bereich der Gedächtnisforschung hat Neisser (1982) Fallbeschreibungen von Personen und Situationen gesammelt, die auf unterschiedliche Weise einmalig sind oder waren. Als weitere

Beispiele für kasuistische Forschung sollen zwei sozialpsychologische Studien herangezogen werden.

Inspiziert durch die Idee, daß Denk- und Urteilsprozesse durch kognitive Dissonanz (vgl. Festinger, 1957) beeinflußt werden, haben Festinger, Riecken und Schachter (1956) Mitglieder einer Sekte teilnehmend beobachtet, die den Weltuntergang für unmittelbar bevorstehend hielten und erwarteten, zu einem bestimmten Zeitpunkt von außerirdischen Wesen zu ihrer Rettung abgeholt zu werden. Für Festinger und seine Mitarbeiter war es interessant herauszufinden, wie sich die Überzeugungssysteme der Sektenmitglieder nach dem zu erwartenden Fehlschlag der Vorhersage verändern wurden. Wie Festinger, Riecken und Schachter eindrucksvoll beschrieben, haben die Anhänger der Sekte nicht das Ausbleiben der Rettung, sondern das des Weltuntergangs als Zeichen Gottes interpretiert. Dieses Verhalten stand in Einklang mit den ursprünglichen Vermutungen der Forscher, die aus den Kernhypothesen der Theorie der kognitiven Dissonanz abgeleitet wurden. Die theoretische Weiterentwicklung dieser Theorie beruhte in der Folge vorwiegend auf umfangreichen experimentellen Studien (vgl. z.B. Festinger & Carlsmith, 1959), in denen Bedingungen hergestellt wurden, die zur Prüfung der unterschiedlichen Aspekte der Theorie besonders diagnostisch waren.

Eine weitere Fallstudie wurde über die unterschiedliche Wahrnehmung des Verlaufs eines Football-Spiels durch die Fans der beiden Mannschaften durchgeführt (Hastorf & Cantril, 1954). Die Autoren stellten fest, daß ein und dasselbe soziale Ereignis in Abhängigkeit von der Präferenz einer der Mannschaften völlig unterschiedlich „gesehen“ wurde. Diese Beobachtung wurde zum Ausgangspunkt eines umfassenden Forschungsprogramms, das mit dem Schlagwort „social perception“ gekennzeichnet wurde und eine Vielzahl von kontrollierten Experimentalstudien beinhaltet (als neuerer Überblick vgl. Zebrowitz, 1990).

**Fallstudie vs. Experiment: Computer-Kasuistik als Alternative?** In den bisherigen Ausführungen wurde die wissenschaftliche Bedeutung von Fallstudien aus methodologischen Gründen hinter die des Experiments gestellt, wobei die Annahme zugrundelag, daß ein experimentelles Verfahren *ceteris paribus* einer kasuistischen Vorgehensweise vorzuziehen ist. Allerdings wurde von einigen Autoren auch die entgegengesetzte Position vertreten. In jüngster Zeit haben Dörner und Lantermann (1991) das Experiment als Königsweg der psychologischen Forschung in Frage gestellt und einen Standpunkt vertreten, der als „Computer-Kasuistik“ gekennzeichnet werden kann.

Ausgangspunkt ihrer Ausführungen ist die Beobachtung der Autoren, daß menschliches Verhalten einerseits multikausal determiniert ist, sich aber andererseits im Experiment nur eine begrenzte Zahl von Bedingungen realisieren lassen, die kein wirklichkeitsgetreues Abbild der natürlichen Situation darstel-

len. Um die „multifaktorielle Bedingtheit psychischen Geschehens“ in der Forschung dennoch zu berücksichtigen, schlagen Dörner und Lantermann vor, Versuchspersonen an einem Computerspiel teilnehmen zu lassen, in dem das Programm sowohl nach bestimmten Regeln als auch zufallsgeneriert auf das Spielverhalten der Versuchsperson reagiert. Nach der Charakterisierung der Autoren handelt es sich um ein „dynamisches Problem mit hoher Unbestimmtheit, multiplen Zielen und mit Zeitdruck“ (S.40), das vielen Alltagssituationen entspreche. In dem geschaffenen Szenario werden unterschiedliche Variablen beobachtet (Art der Entscheidung, Zielerreichung, Reaktionszeit etc.), die dann wiederum mit Persönlichkeitseigenschaften des Spielers in Beziehung gesetzt werden können. Wie von den Autoren berechnet, beinhaltet das beschriebene Spiel 16'384 Bedingungskombinationen, deren Untersuchung in einem faktoriellen Versuchsplan über 80'000 Versuchspersonen erforderlich machte. Da natürliche Situationen - so die Ausführungen von Dörner und Lantermann - zumindest so komplex wie die computergenerierte Spielsituation seien, könnte die experimentelle Methode zur Erklärung menschlichen Verhaltens nicht fruchtbar herangezogen werden. „Experimentalizisten“ (Dörner und Lantermanns Bezeichnung für Forscher, die dennoch am Experiment festhalten) hätten angesichts der aufgezeigten Probleme verschiedene Immunisierungs-Strategien gegen Selbstzweifel“ (S. 49) entwickelt, die zu einer verengten Weltsicht führten. Als Konsequenz verlören diese die Komplexität und Interdependenz psychischer Prozesse aus dem Auge, ignorierten die „Totalität der Person“ und machten sich auf diese Weise einer „experimentell induzierte[n] Ent-Personalisierung der Versuchsperson“ (S. 51f.) schuldig.

Nach diesen schwerwiegenden Vorwürfen an die Adresse der Experimentalforscher bieten Dörner und Lantermann als Alternative ihren computer-kasuistischen Ansatz an, der die empirische Psychologie aus der Sackgasse herausführen soll, in die sie durch die experimentelle Methode hineingeraten sei. Als neue Methode der Wahl wird eine Variante der Einzelfallbeobachtung empfohlen. Dabei soll das Verhalten der Versuchsperson beim Computerspiel nicht unter einer isolierten Bedingungsvariation, sondern in seiner gesamten Komplexität sorgfältig beobachtet und protokolliert werden. Der Forscher ist angehalten, das Verhalten der Versuchsperson zu „verstehen“ und auf diese Weise Hypothesen über die jeweilige Verhaltensdetermination zu gewinnen. Die gewonnenen Hypothesen sollen dann in ein Computermodell einfließen, dessen Qualität sich an der Güte der Simulation des realen Verhaltens bemißt. Mit dieser Vorgehensweise werde das „langwierige und mühselige Experimentieren mit vollständiger Bedingungsvariation ersetzt“. Das computer-kasuistische Verfahren basiere „auf der genauen Betrachtung des Einzelfalls, ohne aber auf die Verallgemeinerung zu verzichten“ (S. 56).

Um eine Würdigung dieses neuerlichen Plädoyers für die Fallstudie vornehmen zu können, ist es notwendig, die Vor- und Nachteile des beschriebenen



Verfahrens im Vergleich zum Experiment unter die Lupe zu nehmen. Was die Vorteile der computer-kasuistischen Methode angeht, so müssen diese wohl vor allem im Kontrast zu den vorgeblichen Mängeln des Experiments gesucht werden, da die hinlänglich bekannten methodologischen Nachteile der „one-shot case study“ (Campbell & Stanley, 1966; Cook & Campbell, 1979) bei der Hypothesenprüfung von Dörner und Lantermann überhaupt nicht angesprochen werden. Sicher hat auch das computer-kasuistische Verfahren einen heuristischen Wert bei der „Erfindung“ von Hypothesen. Zu deren Prüfung ist es jedoch weit weniger geeignet als das Experiment. Insofern scheint die Überlegenheit des methodischen Vorschlags fast ausschließlich aus den geschilderten Nachteilen des Experiments zu resultieren, welche die methodologischen Schwächen kasuistischer Verfahren aufzuwiegen scheinen.

Aus der Sicht der Autoren besteht das Grundübel des Experiments darin, daß die Realität nicht in einem multifaktoriellen Versuchsplan abgebildet werden kann. Sicher ließe sich dasselbe Argument auch auf das computerkasuistische Vorgehen übertragen (auch 16'384 Bedingungskombinationen umfassen sicher nicht sämtliche möglichen Handlungsdeterminanten in einer bestimmten Situation, geschweige denn in einem Situationstyp), wichtiger ist jedoch die Feststellung, daß die von Dörner und Lantermann vorgebrachte Kritik einem grundlegenden Mißverständnis der experimentellen Methode entspringt. Denn weder in den Naturwissenschaften noch in der Psychologie ist das Ziel des experimentellen Vorgehens, eine bestimmte natürliche Situation möglichst wirklichkeitsgetreu zu simulieren, um dann die experimentellen Ergebnisse auf entsprechende natürliche Situationen übertragen zu können. „ökologische Validität“ („mundane realism“ im Sinne von Aronson et al., 1990) ist für die Güte eines Experiments ohne jede Bedeutung („experimenteller Realismus“ dagegen sehr, wie wir an anderer Stelle in diesem Kapitel argumentiert haben - vgl. 2.3; s.a. Turner, 1981, für die Ausführung einer ähnlichen Argumentationslinie).

Dieses Mißverständnis beruht auf der unzutreffenden Annahme, Experimente hätten die Aufgabe, die Dynamik von Einzelfällen zu erklären. In Wirklichkeit geht es im Experiment darum, Hypothesen und Theorien zu prüfen, die dann zum Verständnis von **vielen** Einzelfällen angewandt werden können. Theorien sind deshalb möglichst allgemein und können nicht nur auf einen bestimmten Anwendungsfall beschränkt sein (vgl. Gadenne, 1984).

Da Theorien wegen ihres allgemeinen Charakters nicht situationsgebunden sind, kommt es bei ihrer experimentellen Prüfung auch nicht darauf an, eine bestimmte natürliche Situation im Experiment zu simulieren. Wichtig ist lediglich, daß die in der Theorie enthaltenen Variablen adäquat realisiert werden und der Einfluß der übrigen Situations- oder Personenvariablen kontrolliert wird. Dabei werden nicht alle in realen Situationen auftretenden Variablen

explizit miteinbezogen werden können. Vielmehr wird eine idealisierte Situation geschaffen, um jeweils als wichtig angesehene Hypothesen aus einer Theorie zu prüfen. In theoriegeleiteten Forschungsprogrammen (vgl. Abschnitt 4) werden dies üblicherweise zunächst die Kernhypothesen der Theorie sein; nach deren Bewährung wird dann versucht, zunehmend komplexere Bedingungsgefüge einer Prüfung zu unterziehen.

In jedem Fall sind Experimente aber hinsichtlich der Anzahl der untersuchten Variablen begrenzt und damit im Vergleich zu realen Einzelsituationen meist reduziert. Dies entspricht aber den psychologischen Theorien, die ja gerade versuchen, die Komplexität von Einzelsituationen auf die entscheidenden Ursache-Wirkungs-Beziehungen zu reduzieren und dadurch viele Einzelsituationen zu erklären.

Oberstes Ziel einer theoretisch-experimentell orientierten psychologischen Wissenschaft ist es demnach, allgemeine Erkenntnisse zu gewinnen, die ein besseres Verständnis menschlichen Denkens, Fühlens und Verhaltens ermöglichen und nicht, eine Datenbank von kasuistisch gewonnenen Computer-Simulationen für möglichst viele verschiedene Handlungssituationen und Persönlichkeiten aufzubauen. Aus diesem Grund erscheint die von Dörner und Lantermann (1991) vorgeschlagene methodologische Alternative gegenüber dem Experiment keine Vorteile zu bieten und mit denselben Nachteilen behaftet zu sein, die alle Verfahren ohne Kontrollmöglichkeit auszeichnen.

#### **4. Forschungsprogramme und ihr Einfluß auf die Planung psychologischer Untersuchungen**

In den bisherigen Ausführungen wurde im allgemeinen implizit so vorgegangen, als ob komplexe kausale Fragestellungen wie der Einfluß von gewalttätigen Sendungen auf Aggression mit einer einzigen Untersuchung und einem Versuchsplan zu beantworten seien. Dies ist natürlich nicht der Fall. In der Praxis werden komplexere Fragestellungen im Rahmen von umfassenderen **Forschungsprogrammen** (vgl. Herrmann, 1976; sowie Herrmann, Kapitel 6 in diesem Band) zu beantworten gesucht. Innerhalb solcher Forschungsprogramme werden dann Einzelfragestellungen aufgeworfen und mit verschiedenen Methoden bearbeitet.

Für das Beispiel des vorliegenden Kapitels bedeutet dies, daß bei einer singulären Bestätigung des Einflusses von gewalttätigen Sendungen auf kindliche Aggression die entsprechende Fragestellung keineswegs ad acta gelegt werden kann. Vielmehr sollte als nächstes geprüft werden, welche psychologischen Mechanismen dabei zur Wirkung kommen. Welche Hinweisreize sind innerhalb gewalttätiger Sendungen entscheidend? Welche Rolle spielen kognitive

Enkodierprozesse? Gibt es Interaktionen der Wirkung von gewalttätigen Filmen mit Persönlichkeitsvariablen? Solche und andere Fragestellungen sind als Teile eines Forschungsprogramms „Gewalttätige Sendungen und Aggression bei Kindern“ vorstellbar (für eine erste inhaltliche Beantwortung der aufgeworfenen Frage s. Josephson, 1987).

Bei genauerem Betrachten solcher Forschungsprogramme wird ein weiterer Grund offensichtlich, warum das Experiment eine überragende Rolle in der psychologischen Forschung spielt. Selbst bei Problemstellungen, die per se nicht experimentell zugänglich scheinen (wie Sozialisationsfragestellungen) tauchen im Verlauf des Forschungsprozesses immer wieder Fragen nach spezifischen Mechanismen auf, die am besten im Laborexperiment beantwortet werden können.

Aronson, Brewer und Carlsmith (1985, s.a. Aronson et al., 1990) plädieren deshalb dafür, daß sich bei der Ausarbeitung von Forschungsprogrammen verschiedene Methoden ergänzen sollen (s. a. Nowack, 1991). Dabei erlauben Laborexperimente die Beantwortung von Einzelfragestellungen beispielsweise zu kognitiven Mechanismen in wesentlich kontrollierterer Art und Weise. Feldexperimente sollten gewählt werden, wenn es darum geht, die Wirksamkeit von bekannten Kausalfaktoren in Interaktion mit natürlichen Hintergrundvariablen zu testen. Quasi-experimentelle Versuchspläne sollten zur Anwendung kommen, wenn kausale Fragestellungen geklärt werden sollen, aber Experimente nicht durchführbar erscheinen. Andere Felduntersuchungen schließlich sollten eingeplant werden, wenn bestimmte Phänomene erst beschrieben und klassifiziert werden müssen sowie im Verlauf der Hypothesengenerierung. Zahlreiche Einzelbeispiele von gelungenen Interaktionen verschiedener Forschungsmethoden finden sich in Aronson et al. (1985, pp.483f.).

## **5. Statistische Kontrolltechniken**

Im letzten Abschnitt zu Kontrolltechniken soll auf sogenannte statistische Kausalanalysen eingegangen werden. Die moderne Statistik bietet dem Psychologen heute ein Instrumentarium, das sehr diagnostische Analysen erlaubt (vgl. die in Abschnitt 3.1 beschriebene Studie von Hennigan et al., 1982). Als positive Folge kann der inhaltlich orientierte Forscher über eine größere Variationsbreite bei der Realisierung seiner Variablen verfügen, da die spätere Auswertung nur wenige Restriktionen für die Versuchsplanung auferlegt. Insbesondere gibt es verbesserte Verfahren für die Auswertung nicht intervallskalierter abhängiger Variablen (loglineare Analyse - Kennedy, 1983; logistische Regression - Hosmer & Lemeshow, 1989; oder Modellbildung mit explizit vorgeschaltetem Meßmodell inklusive Fehlerannahmen, vgl. Arminger & Rehm, 1989).

Die negative Folge dieser Entwicklungen besteht allerdings darin, daß offensichtlich der Anschein erzeugt wird, statistische Verfahren wie sogenannte Kausalanalysen könnten experimentelle Versuchspläne ersetzen. Dabei sind statistische Verfahren und Versuchsplanung zum Zweck der Bestimmung von Kausalität keineswegs gleichwertig; denn kein noch so gutes Verfahren kann Designfehler a posteriori ausmerzen und aus einer korrelativen Querschnittserhebung eindeutig Kausalität bestimmen.

Dieser Gedankengang soll anhand eines Beispiels mit dem Programmpaket LISREL (Jöreskog & Sörbom, 1989) aufgezeigt werden. Die gleiche Argumentation gilt analog auch für andere Programmpakete wie EQS (Bentler, 1989) oder EZPATH (Steiger, 1989). Um Mißverständnisse zu vermeiden: Programmpakete wie LISREL stellen bei sachgemäßer Anwendung eine große Bereicherung für die Sozialwissenschaften dar und ermöglichen z.B. die Auswertung von Versuchsplänen mit wiederholten Messungen, die bisher nicht adäquat ausgewertet werden konnten (s. Arminger & Müller, 1990). Es geht hier lediglich um die Fehlinterpretation von Partialkoeffizienten als belegte oder gar bewiesene Kausalität.

LISREL erlaubt die Anpassung von komplexen Kausalmodellen an multivariate Daten. Als Datengrundlage dienen in der Regel Korrelations-, Kovarianz- oder Momentenmatrizen. Für beobachtete und latente Variablen kann ein Kausalsystem formuliert werden, wobei im Fall des Einschlusses von latenten Variablen gemessene Variablen als Indikatoren für die latenten Variablen spezifiziert werden müssen (vgl. Stelzl, 1983; ausführliche Beschreibungen in Jöreskog & Sörbom, 1989, oder Hodapp, 1984). Ein Modell kann gerichtete (z. B. gewalttätige Sendungen beeinflussen aggressive Handlungen) und/oder ungerichtete Beziehungen (z. B. zwischen psychischen und somatischen Krankheiten besteht ein Zusammenhang) zwischen Variablen enthalten. Diese Beziehungen müssen vom Benutzer in Form linearer Gleichungen angegeben werden; üblich ist außerdem die graphische Veranschaulichung in Form von Pfaddiagrammen. LISREL und andere Programme schätzen dann simultan Gewichtungszahlen für die einzelnen Pfade und - sofern das Kausalmodell restriktiv genug ist - einen Signifikanztest für die Modellanpassung.

Die Pfadkoeffizienten werden aber nur dann erwartungstreu und konsistent geschätzt, wenn die zugrundegelegte Kausalstruktur korrekt spezifiziert wurde und außerdem vollständig ist. Vollständigkeit bedeutet in diesem Zusammenhang, daß keine weiteren Variablen existieren, die auf die abhängigen Variablen im Gleichungssystem einwirken. Beide Voraussetzungen sind in der Psychologie selten erfüllt. Psychologische Theorien sind fast nie vollständig in dem Sinn, daß sie **alle möglichen Einflußfaktoren** auf die untersuchten Phänomene spezifizieren. Vielmehr haben wir es in der Psychologie im allgemeinen mit sog. unvollständigen Hypothesen zu tun (vgl. Gadenne, 1984; 1990).

Daraus ergibt sich die Frage, inwieweit bei Verletzungen der Voraussetzungen mit Programmpaketen wie LISREL noch Kausalstrukturen entdeckt bzw. geprüft werden können. Die Antwort ist eher skeptisch. Die Beziehungen zwischen verschiedenen Variablen, so wie sie in Zusammenhängigkeitsmatrizen vorliegen, sind meist mit unterschiedlichen, sich von der Kausalitätsrichtung her widersprechenden Kausalsystemen vereinbar (ausführlicher in Stelzl, 1983, S. 317ff.). Mit anderen Worten: Es läßt sich mit Hilfe von LISREL in der Regel nicht entscheiden, ob eine Variable A auf die Variable B wirkt oder umgekehrt.

Ferner können Regeln angegeben werden, um aus Kausalbeziehungen äquivalente Systeme hinsichtlich der empirischen Anpassung zu erzeugen (vgl. Stelzl, 1986). Mit anderen Worten: Es gibt Regeln, um aus einem empirisch den Daten angepaßten System von Kausalbeziehungen ein neues System zu erzeugen, das andere Kausalbeziehungen behauptet, aber hinsichtlich der vorliegenden Daten eine ebenso gute Anpassung zeigt.

Was bedeutet diese Argumentation hinsichtlich unserer Ausgangsfragestellung? Zum einen kann LISREL nicht zur **Entdeckung** von Kausalstrukturen verwendet werden. Zum anderen können auch bei der **Prüfung** von Kausalstrukturen in vielen Fällen keine eindeutigen Ergebnisse in dem Sinne erzielt werden, daß eine Kausalstruktur angenommen werden kann, während alle anderen Strukturen verworfen werden müssen. Es ist allerdings möglich, Aussagen von explizit ausgearbeiteten Theorien oder Hypothesenbündeln hinsichtlich ihrer Vereinbarkeit mit vorliegenden empirischen Daten zu beurteilen.

Insgesamt läßt sich sagen, daß die statistische Kausalanalyse zwar gegenüber konventionellen Zusammenhangsmaßen zahlreiche Vorteile aufweist, das experimentelle Vorgehen hinsichtlich des Aufzeigens von Kausalität aber nicht ersetzen kann.

## **6. Anmerkungen zur Ethik psychologischer Untersuchungen**

### 6.1 Vorbemerkung

Empirische Untersuchungen mit Menschen werfen ethische Fragen auf. Die Beantwortung dieser Fragen beruht auf Wertentscheidungen, die vom einzelnen, der Gemeinschaft der Wissenschaftler oder der Gesellschaft in unterschiedlicher Weise getroffen werden.

Auf der einen Seite dient Forschung dem Ziel des Erkenntnisfortschritts, der in aufgeklärten Gesellschaften als wertvoll erachtet wird. Auf der anderen Seite kann jedoch dem Erkenntnisgegenstand durch die Forschungsaktivität

Schaden zugefügt werden. Die physische und psychische Unversehrtheit der Person ist aber ebenfalls ein Grundwert moderner Gesellschaften. Dieser Konflikt zwischen unterschiedlichen Wertsystemen manifestiert sich nicht nur in der psychologischen Forschung, sondern in gleichem oder stärkerem Maße in der Medizin oder den Sozialwissenschaften im allgemeinen (vgl. z.B. Sieber, 1982a, b; Textor, 1989).

In diesem Abschnitt kann es jedoch weder darum gehen, eine allgemeine Ethik der Forschung darzustellen oder auch nur die ethische Problematik des Forschens mit Menschen in sämtlichen Facetten aufzuzeigen. Für eine ausführliche Darstellung der Problemlage sei auf Keith-Spiegel und Koocher (1985) sowie auf Schuler (1980) verwiesen. Eine Diskussion aus unterschiedlichen Wissenschaftsperspektiven findet sich in einem von Kruse und Kumpf (1981) vorgelegten Sammelband. Möglich im vorliegenden Rahmen ist lediglich, einen Blick auf solche ethischen Probleme zu werfen, die sich dem experimentell arbeitenden Psychologen stellen und einige pragmatische Vorschläge zur Minderung des Konflikts zu unterbreiten.

## 6.2 Mögliche Schädigungen

Der Schaden, den Teilnehmer an empirischen Untersuchungen erleiden können, läßt sich in drei Kategorien fassen. Zum einen können die Versuchspersonen **physischen und psychischen Streß** erleben. Das zweite Problem besteht in der **Täuschung** der Teilnehmer. Das dritte Problem ist eine mögliche **Verletzung der Anonymität**.

**Physischer und psychischer Streß:** Obwohl **physischer Streß** in psychologischen Untersuchungen selten auftritt, existieren einige experimentelle Untersuchungen, in denen Schmerz und körperliches Unwohlbefinden zu Erkenntniszwecken herbeigeführt wird. Als ein Beispiel können Arbeiten von Lanzetta, Kleck und Mitarbeitern (z.B. Lanzetta, Cartwright-Smith & Kleck, 1976) dienen, in der die Autoren herausfinden wollten, ob die willkürliche Unterdrückung des Ausdrucks einer Empfindung deren Intensität verringert. Dazu erhielten Versuchspersonen einen elektrischen Schock mit der Instruktion, sich die Verabreichung des Schocks nicht anmerken zu lassen. Ein weiteres Beispiel sind Experimente, in denen physisches Unwohlbefinden oder Schmerz induziert wird, indem die Versuchspersonen zum Beispiel veranlaßt wurden, ihre Hände für längere Zeit in Eiswasser zu halten (vgl. Rollman, 1983) oder indem unangenehme Geräusche über Kopfhörer eingespielt wurden (z.B. Ross, Rodin & Zimbardo, 1969). Mit solchen Prozeduren wurden Stimmungseinflüsse bzw. die Wirkung von Bestrafung auf die Lernleistung untersucht.

Weit häufiger werden Versuchspersonen **psychischem Streß** ausgesetzt. Dies kann in der Auslösung von negativen Emotionen wie Angst (z.B. Schachter, 1959), Schuld (z.B. Milgram, 1963), Scham (z.B. Sarnoff & Zimbardo, 1961), Ärger (z.B. Ax, 1953), Ekel (z.B. Rozin & Fallort, 1987) bestehen oder in der Induktion von Frustration durch die Rückmeldung eines vermeintlichen Mißerfolgs in einer Leistungsaufgabe (z.B. Bohner, Bless, Schwarz & Strack, 1988). Bisweilen wird psychisches Unwohlbefinden bewirkt, indem Person zu einem Handeln verleitet werden, das ihren eigenen Überzeugungen oder Wertvorstellungen widerspricht (vgl. Festinger, 1957; Festinger & Carlsmith, 1959). Dabei wird manchmal ein negatives Selbstbild aktiviert, indem eigenes Verhalten, das durch subtile experimentelle Manipulationen ausgelöst wurde, eigenen Persönlichkeitseigenschaften zugeschrieben wird („fundamental attribution error“, ROSS, 1977). So mögen Teilnehmer, die vom Versuchsleiter ohne augenfälligen Anreiz dazu aufgefordert wurden, gegenüber einer anderen Versuchsperson zu lügen, sich selbst als unehrlich erleben. In der Theorie der kognitiven Dissonanz (Festinger, 1957) ist es aber gerade dieses unangenehme Gefühl, das als motivationale Ausgangsbedingung von Einstellungsänderungen relevant wird.

Betrachtet man die Induktion von physischem und psychischen Streß unter forschungsethischer Perspektive, so besteht Konsens darüber, daß eine längerfristige Schädigung von Versuchspersonen nicht zulässig ist. Die Beweislast liegt dabei beim Forscher und bisweilen wird es nötig sein, längerfristige Auswirkungen empirisch auszuschließen (z.B. Milgram, 1964). Daneben gilt es vor allem, die Intensität und Dauer der Einwirkung zu berücksichtigen sowie die Freiwilligkeit der Teilnahme an der Untersuchung. Darüber hinaus sollte der Versuchsperson die Möglichkeit eingeräumt werden, zu jedem Zeitpunkt aus dem Versuch auszusteigen.

Ein Spezialproblem stellen randomisierte Studien in der Psychotherapieforschung dar (Pritsch & Noack, 1991). Hier ist es unter bestimmten Umständen möglich, daß die randomisierte Zuteilung zu einer Vergleichsgruppe dazu führt, daß physische und/oder psychische Schmerzen nicht bestmöglich gelindert werden. Das ethische Problem stellt sich hier nicht für die Versuchspersonen, die in Kenntnis der beschriebenen Sachlage einer Teilnahme zugestimmt hatten, sondern für den Forscher, ob die Durchführung eines solchen Experiments insgesamt ethisch vertretbar ist. Auch diese Entscheidung beinhaltet im Kern eine Abwägung von Kollektivinteressen (Chance auf bessere Behandlung für die Allgemeinheit in der Zukunft) mit Individualinteressen (der teilnehmenden Versuchspersonen). Falls diese Entscheidung gegen die Durchführung eines randomisierten Experiments fällt (und damit gegen das methodologische Optimum), finden sich in Pritsch und Noack (1991, S. 43ff.) alternative Versuchspläne (s. a. Zelen, 1990).

**Täuschung der Versuchsteilnehmer:** Ein weiteres forschungsethisches Problem besteht in der Täuschung von Versuchspersonen (vgl. dazu auch Irle, 1979). Sicherlich wäre es einfacher und ethisch weniger problematisch, die Teilnehmer an Experimenten von Anfang an über die Hypothesen einer Studie und den Sinn bevorstehender experimenteller Interventionen zu informieren. Allerdings ist es für die Validität von Untersuchungen oftmals von Bedeutung, daß den Teilnehmern die Fragestellung verborgen bleibt. Die Notwendigkeit dieser Vorkehrung wird vor allem in der medizinisch-pharmakologischen Forschung deutlich, in der es dem Standard des Experimentierens entspricht, Scheinpräparate (Placebos) zu verwenden und nicht nur den Versuchspersonen, sondern auch dem Versuchsleiter Informationen über die Wirksamkeit der jeweiligen Droge vorzuenthalten („Doppel-Blind-Versuch“, s. a. Abschnitt 2.2 oben). Auf diese Weise sollen mögliche Einflüsse, die auf das Wissen um die Wirkung eines Präparats zurückzuführen sind, kontrolliert und eliminiert werden.

In der experimentellen sozialwissenschaftlichen und psychologischen Forschung spielt die Täuschung der Teilnehmer als methodologisches Instrument eine ebenso bedeutende Rolle. So mag ein Gedächtnisforscher, der herausfinden will, wie gut bestimmte Informationen erinnert werden, wenn eine Versuchsperson nicht explizit zu deren Enkodierung aufgefordert wird, den Teilnehmern mitteilen, bei der Experimente gehe es um die Erforschung der Eindrucksbildung (und nicht um eine Gedächtnisstudie). Ein Sozialpsychologe wie Stanley Milgram (1963), der an der Dynamik des Gehorsams gegenüber Autoritäten interessiert ist, mag seine Studie als Untersuchung zum Einfluß von Bestrafung auf das Lernverhalten beschreiben. Eine ähnliche „Tarnung“ des Untersuchungszwecks mag von einem Aggressionsforscher (z. B. Buss, 1966) verwendet werden.

Diese Beispiele illustrieren, daß die Enthüllung des tatsächlichen Untersuchungsziels oft eine Beeinträchtigung der Validität darstellen würde. So wird eine Versuchsperson, die das tatsächliche Ziel des Gedächtnisexperimentes kennt, die dargebotenen Informationen eher enkodieren und abspeichern als eine Person, der die Fragestellung nicht bekannt ist. Ein Teilnehmer in Milgrams oder Buss' Untersuchung wurde weniger Gehorsam und geringere Feindseligkeit zeigen, wenn er wußte, daß seine Autoritätsgläubigkeit bzw. Aggressivität auf dem Prüfstand steht. In einer neueren Untersuchung haben Larsen und Simnett (1991) die Wirkung des sogenannten Velden-Verfahrens zur Induktion von Stimmungen einer Metaanalyse unterzogen. Dabei haben die Autoren zwischen Untersuchungen unterschieden, in denen die Teilnehmer über das Ziel des Verfahrens unterrichtet wurden und Studien, in denen dies nicht der Fall war. Larsen und Simnett fanden, daß die Kenntnis des Zieles der Velden-Prozedur die Selbst-Beschreibungen der eigenen Stimmung beeinflusste. Versuchspersonen, die über den Sinn der Intervention Bescheid



wußten, schrieben sich in höherem Maße die ausgelöste Emotion zu als Teilnehmer, die keine entsprechenden Informationen erhalten hatten.

Diese Überlegungen und Befunde zeigen, daß die Kenntnis der Untersuchungsziele die Ergebnisse in systematischer Weise verzerren kann. Aus diesem Grund haben sich Forscher auf unterschiedliche Weise bemüht, das Erkenntnisziel einer Studie oder die Absicht einer bestimmten Intervention zu verbergen. Dabei reichte das Spektrum vom Vorenthalten der entsprechenden Information bis zur aktiven Irreführung.

Betrachtet man die Täuschung von Versuchspersonen unter forschungsethischer Perspektive, so stellt sich unter anderem die Frage nach möglicher Schädigung der Versuchspersonen. Bisher haben empirische Untersuchungen allerdings keinerlei Evidenz für Schädigungen in Folge von Täuschungen erbracht. Vielmehr bringen Versuchspersonen, die im nachhinein über eine Täuschung informiert wurden, in der Regel Verständnis für diese Maßnahme auf, wenn der Versuchsleiter ihre Notwendigkeit erläutert (vgl. Christensen, 1988). Darüber hinaus liegen Ergebnisse vor, denen zufolge Studien als interessanter eingeschätzt wurden, wenn diese mit einer Täuschung verbunden waren.

Der ethisch problematischste Aspekt von Täuschungen ist die Beschränkung der Entscheidungsfreiheit, die darin besteht, daß Personen durch unzutreffende Informationen oder ohne zureichendes Wissen zur Teilnahme an einer Untersuchung veranlaßt werden. Dieses Dilemma liegt in der Natur des Vorgehens und läßt sich nur im nachhinein durch entsprechende Aufklärung kompensieren. In jedem Fall wird der Forscher sicherstellen müssen, daß sich die Versuchsperson während der Untersuchung nicht getäuscht *fühlt*. Das heißt, der Ablauf des Experiments muß den Erwartungen entsprechen, die zu Beginn geweckt wurden. Dies ist besonders dann der Fall, wenn die Teilnahme mit physischem oder psychischem Streß verbunden ist. Allerdings werden Abweichungen von den ursprünglichen Erwartungen oft durch einen „unvorhergesehenen“ Zwischenfall oder durch eine Einflußnahme bewirkt, die vermeintlich außerhalb des eigentlichen Experiments erfolgt. Auch ein solches Vorgehen, bei dem sich die Teilnehmer ihrer Rolle als Versuchsperson nicht bewußt sind, ist forschungsethisches problematisch.

Wie bereits erwähnt, ist ein Forscher, der die Versuchspersonen über den Zweck der Untersuchung im Unklaren läßt, verpflichtet, die Teilnehmer nach Abschluß des Experiments über das tatsächliche Ziel der Studie zu informieren. Diese „Aufklärung“ soll so erfolgen, daß mögliche Effekte der Manipulation rückgängig gemacht werden. Dieses Desiderat ist allerdings nicht so einfach zu verwirklichen, wie man vielleicht vermuten möchte. So haben Untersuchungen zur Wirksamkeit von verschiedenen Aufklärungsverfahren (z. B., Ross, Lepper & Hubbard, 1975) gezeigt, daß ein im Experiment geschaffener Eindruck oftmals auch nach einer Aufklärung andauert. Versuchs-

teilnehmer, die positive oder negative Rückmeldung über eine ungewöhnliche Fähigkeit erhielten, waren auch dann noch von der Information beeinflusst, wenn sie wußten, daß die Rückmeldung nach einem Zufallsprinzip erfolgte und in keinerlei Zusammenhang zu der tatsächlichen Leistung der Teilnehmer stand. Allerdings konnten ROSS et al. (1975) nachweisen, daß der Effekt der Manipulation dann erfolgreich beseitigt werden konnte, wenn die Versuchspersonen einer „Prozeß-Aufklärung“ („process debriefing“) unterzogen wurden. Unter dieser Bedingung bestand die Aufklärung nicht nur in Informationen über die Art der Irreführung, sondern auch über die typischen Auswirkungen dieser Beeinflussung auf die Urteilsbildung.

Während die postexperimentelle Aufklärung an der tatsächlichen Beeinflussung orientiert sein sollte, ergeben sich unter bestimmten Umständen Konflikte mit dem ethischen Ziel der Minimierung von Unwohlbefinden. So wird abzuwägen sein, ob eine durch fingierte Erfolgsrückmeldung induzierte positive Stimmung einer Person durch Information über den fiktiven Charakter der Rückmeldung neutralisiert oder gar in eine negative Stimmungslage transformiert werden sollte. Dieses Beispiel illustriert die oftmals fehlende Eindeutigkeit bei forschungsethischen Entscheidungen. Obwohl in vielen Fällen Konsens über die Unzulässigkeit von bestimmten Praktiken besteht, müssen Entscheidungen oft von Fall zu Fall getroffen werden. Auch wenn sich die Forscher an denselben ethischen Richtlinien orientieren, die von professionellen Vereinigungen wie der American Psychological Association oder den Psychologenvereinigungen in der Bundesrepublik Deutschland kodifiziert wurden (eine Übersicht findet sich bei Schuler, 1980), können im Einzelfall unterschiedliche Entscheidungen resultieren. Somit bleibt die Verantwortung letztlich beim jeweiligen Forscher.

**Datenschutz:** Ein dritter Problembereich betrifft den Datenschutz. Teilnehmer an wissenschaftlichen Untersuchungen müssen darauf vertrauen können, daß Aufzeichnungen über die Manifestation ihres Verhaltens oder die Äußerung von Meinungen anonym verarbeitet und gespeichert sowie vertraulich behandelt werden. Da die Daten in der experimentellen Forschung normalerweise ohne Identifikation erhoben und abgespeichert werden, stellt sich das Problem des Datenschutzes in geringerer Weise als etwa in der empirischen Sozialforschung, wo repräsentative Wiederholungsbefragungen eine wichtige Rolle spielen.

Dennoch ist es auch in der experimentellen Forschung bisweilen notwendig, Datenprotokolle so zu kennzeichnen, daß sie mit anderen Aufzeichnungen derselben Person zusammengeführt werden können. Dies ist vor allem dann der Fall, wenn eine Studie aus mehreren Phasen besteht oder wenn zu Beginn des Semesters Merkmale von potentiellen Untersuchungsteilnehmern erhoben werden, um zu einem späteren Zeitpunkt mit experimentellen Ergebnissen

verknüpft zu werden. Dabei wird oft anstelle des Namens ein Code verwendet, der von den Teilnehmern selbst generiert wird und vom Versuchsleiter nicht entschlüsselt werden kann. Allerdings ist dieses Vorgehen ausgeschlossen, wenn Teilnehmer nach den erhobenen Merkmalen selektiert werden sollen. Dann müssen gesonderte Vorkehrungen getroffen werden, um die Anonymität der Individualdaten zu sichern. Eine eingehendere Darstellung von Anonymisierungsverfahren findet sich bei Dittrich und Schlörer (1987).

### 6.3 Durchführungsregeln für experimentelle Studien

Um die Durchführung von Experimenten in ethisch relevanten Aspekten zu vereinheitlichen, haben einige Forschungsinstitute Regeln erarbeitet, in denen allgemeine Prinzipien forschungspraktisch konkretisiert wurden. Eine solche Konkretisierung wurde von der Forschungsgruppe „Intention und Handeln“ am Max-Planck-Institut für psychologische Forschung erstellt (Gollwitzer, 1991). In diesen Richtlinien werden die folgenden drei forschungsethischen Grundregeln auf die Experimentalsituation angewandt: a) die Freiwilligkeit der Teilnahme, b) die vertrauliche Behandlung der erhobenen Daten, c) die Durchsichtigkeit von Ziel und Zweck der Studie.

Um die Freiwilligkeit der Teilnahme zu sichern, wird ein Anwerbungsverfahren durchgeführt, das den Versuchspersonen die Möglichkeit gibt, zu jedem Zeitpunkt - das heißt auch vor Beginn der eigentlichen Untersuchung - die Teilnahme ohne Angabe von Gründen zu verweigern.

Zur Wahrung der Anonymität der Individualdaten werden bei der Erhebung keinerlei personenidentifizierende Informationen erfaßt. Sämtliche Namenslisten, aus denen hervorgeht, wer an welcher Studie teilgenommen hat, werden nach Abschluß der Untersuchung vernichtet. Sollten Nachbefragungen notwendig sein, erfolgt die Zuordnung der Protokolle durch einen von der Versuchsperson gewählten Code, den die Versuchsleiter nicht entschlüsseln können. Hat eine Versuchsperson ihren Code vergessen, wird sie von der Untersuchung ausgeschlossen.

Die Durchsichtigkeit einer Studie wird so weit als möglich gewährleistet, indem die Teilnehmer zunächst über den konkreten Ablauf des Experiments und die zu bewältigenden Aufgaben unterrichtet werden. Nach Abschluß der Studie erfolgt eine umfangreiche Aufklärung, bei der die Versuchspersonen über die in Frage stehenden Hypothesen sowie über deren theoretischen Hintergrund informiert werden. Die Teilnehmer erfahren dabei, welcher Experimentalbedingung sie zugeordnet waren und welche weiteren Bedingungen in der Studie implementiert wurden. Bei Studien, in denen die Versuchspersonen getäuscht wurden, erfolgt die Aufklärung so, daß die Teilnehmer selbst die

Irreführung entdecken. Auf diese Weise soll sichergestellt werden, daß die Aufklärung verstanden wird und sich die Versuchspersonen nicht als leicht täuschbar erleben. Bevor eine Täuschung eingesetzt wird, muß a) geprüft werden, ob die Täuschung gerechtfertigt und notwendig ist; b) festgestellt werden, ob alternative Täuschungsmöglichkeiten existieren, die weniger schwerwiegend sind und eine wirksamere Aufklärung ermöglichen; c) gesichert sein, daß die verwendete Täuschung mit Sicherheit aufgeklärt werden kann und daß mögliche negative emotionalen Empfindungen ruckgängig gemacht werden können.

Auch wenn die Verallgemeinerung von konkreten Verfahrensrichtlinien dieser Art von der jeweiligen Fragestellung bestimmt wird, so liefern sie doch eine nützliche Orientierung zur praktischen Umsetzung von forschungsethischen Prinzipien.

## Literatur

- Abel, U. (1987). Modified replacement randomization. *Statistics in Medicine*, **6**, 127-135.
- Arminger, G. & Müller, E. (1990). *Lineare Modelle zur Analyse von Paneldaten*. Opladen: Westdeutscher Verlag.
- Arminger, G. & Rehm, J. (1989). Models for analyzing non-metric dependent variables: Applications to nutritional epidemiology. In L. Kohlmeier & E. Helsing (Eds.), *Epidemiology, nutrition, and health*. Proceedings of the 1st Berlin Meeting on Nutritional Epidemiology (pp. 67-78). London: Smith-Gordon, Nishimura.
- Aronson, E. & Carlsmith, J.M. (1968<sup>2</sup>). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology*, (Vol. 2, pp. 1-79). Reading, MA: Addison-Wesley.
- Aronson, E., Brewer, M. & Carlsmith, J.M. (1985<sup>3</sup>). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 1, pp. 441-486). New York: Random House.
- Aronson, E., Ellsworth, P. C., Carlsmith, J.M. & Gonzales, M. H. (1990). *Methods of research in social psychology*. New York: Mc Graw-Hill.
- Aronson, E. & Mills, J. (1959) The effects of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology*, **59**, 177-181
- Ax, A. F. (1953). The physiological differentiation between fear and anger in humans. *Psychosomatic Medicine*, **15**, 433-442.
- Bandura, A. (1965). Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology*, **1**, 589-595.
- Bandura, A., Ross, D. & Ross, S.A. (1963). Imitation of filmmediated aggressive models. *Journal of Abnormal and Social Psychology*, **66**, 3-11.

- Bentler, **P.M.** (1989). **EQS Structural equations program manual**. Los Angeles: BMDP Statistical Software.
- Bless, H., Strack, F. & Schwarz, N. (im Druck). The informative function of research procedures: Bias and the logic of conversation. **European Journal of Social Psychology**.
- Bohner, G., Bless, H., Schwarz, N. & Strack, F. (1988). When do events trigger attributions? The impact of valence and subjective probability. **European Journal of Social Psychology**, **18**, 335-345.
- Box, G. E., Hunter, W. G. & Hunter, J.S. (1978). **Statistics for Experimenters**. New York: Wiley.
- Bredenkamp, J. (1969). Experiment und Feldexperiment. In C.F. Graumann (Ed.), **Handbuch der Psychologie in 12 Bänden. Band 7: Sozialpsychologie** (1. Halbband, pp. 332-374). Göttingen: Hogrefe.
- Bredenkamp, J. (1972). **Der Signifikanztest in der psychologischen Forschung**. Frankfurt/Main: Akademische Verlagsgesellschaft.
- Bredenkamp, J. (1980). **Theorie und Planung psychologischer Experimente**. Darmstadt: Steinkopff.
- Bredenkamp, J. (1983). Übersicht. In J. Bredenkamp & H. Feger (Eds.), **Hypothesenprüfung. Enzyklopädie der Psychologie** (Serie 1, Bd. 5., pp. 1-23). Göttingen: Hogrefe.
- Buss, A. H. (1966). The effect of harm on subsequent aggression. **Journal of Experimental Research in Personality**, **1**, 249-255.
- Campbell, D. T. & Stanley, J. C. (1966). **Experimental and quasi-experimental designs for research**. Chicago: McNally.
- Cannell, C. F. & Kahn, R. L. (1968<sup>2</sup>). Interviewing. In G. Lindzey & E. Aronson (Eds.), **The handbook of social psychology** (Vol. 2, pp. 526-595). Reading, MA: Addison-Wesley.
- Chalmers, A. F. (1978). **What is this thing called science?** Queenstand: University of Queenstand Press.
- Chapman, L.J. & Chapman, J.P. (1967). Genesis of popular but erroneous diagnostic observations. **Journal of Abnormal Psychology**, **72**, 193-204.
- Christensen, L. (1988). Deception in psychological research: When is its use justified? **Personality and Social Psychology Bulletin**, **14**, 664-675.
- Cook, T. D. & Campbell, D.T. (1979). **Quasi-experimentation: Design and analysis issues for field settings**. Boston: Houghton Mifflin.
- Crowne, D. & Marlowe, D. (1964). **The approval motive**. New York: Wiley.
- Dalbert, C., Strack, K., Boesken, W.H. & Schwenkmezger, P. (1991). Beeinflußt die Auseinandersetzung mit Bewältigungsverhalten das Angsterleben? **Zeitschrift für Klinische Psychologie**, **20**, 65-74.
- Dittrich, K. & Schlörner, J. (1987). Anonymisierung von Forschungsdaten und Identifikation anonymisierter Datensätze. **Datenschutz und Datensicherung**, **1**, pp. 30ff.
- Dörner, D. & Lantermann, E.D. (1991). Experiment und Empirie in der Psychologie. In K. Grawe, R. Hänni, N. Semmer & F. Tschan (Eds.), **über die richtige Art Psychologie zu betreiben**. Göttingen: Hogrefe.

- Edgington, E. S. (1966). Statistical inference and nonrandom samples. *Psychological Bulletin*, 66, 485-487.
- Edgington, E. S. (1969). *Statistical inference: The distribution-free approach*. New York: McGraw-Hill.
- Edgington, E. S. (1980). *Randomization tests*. New York: Marcel Dekker.
- Fahrenkrug, H. & Rehm, J. (1992). *Sozialepidemiologie alkoholbedingter Straßenverkehrsunfälle junger Erwachsener*. Lausanne: ISPA.
- Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M.S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 74-97). Newbury Park: Sage.
- Feger, H. (1983). Planung und Bewertung von wissenschaftlichen Untersuchungen. In H. Feger & I. Bredenkamp (Eds.), *Datenerhebung. Enzyklopädie der Psychologie* (Serie 1, Bd. 2., pp. 1-75). Göttingen: Hogrefe.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- Festinger, L. & Carlsmith, J.M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58, 203-210.
- Festinger, L., Riecken, H. W. & Schachter, S. (1956). *When prophecy fails*. Minneapolis: University of Minnesota Press.
- Fichter, M. M. (1990). *Verlauf psychischer Erkrankungen in der Bevölkerung*. Berlin: Springer.
- Fisher, R.A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Fisher, R.A. (1930/1960). *The design of experiments*. Edinburgh: Oliver and Boyd.
- Friedman, L.M., Furberg, C.D. & Demets, D. L. (1985<sup>2</sup>). *Fundamentals of clinical trials*. Littleton, MA: PSG Publishing Company.
- Frey, S. & Frenz, H.-G. (1982). Experiment und Quasi-Experiment im Feld. In J.-L. Patry (Ed.), *Feldforschung. Methoden und Probleme sozialwissenschaftlicher Forschung unter natürlichen Bedingungen* (pp. 229-258). Bern: Huber.
- Gadenne, V. (1976). *Die Gültigkeit psychologischer Untersuchungen*. Stuttgart: Kohlhammer.
- Gadenne, V. (1984). *Theorie und Erfahrung in der psychologischen Forschung*. Tübingen: Mohr.
- Gadenne, V. (1990). Unvollständige Erklärungen. In M. Sukale (Ed.), *Sprache, Theorie und Wirklichkeit* (pp. 263-287). Frankfurt/Main: Lang.
- Geyer, S. (1991). *Lebensverändernde Ereignisse und Brustkrebs*. Bern: Huber.
- Gollwitzer, P.M. (1991). *Ethische Richtlinien für die Durchführung von sozial- und motivationspsychologischen Studien am Max-Planck-Institut für psychologische Forschung*. Unveröffentlichtes Manuskript. München: Max-Planck-Institut für psychologische Forschung
- Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics*. 3: *Speech acts* (pp.41-58). New York: Academic.

- Häubler, A., Rehm, J., Naß, E. & Kohlmeier, L. (1990). Ernährung, Krankheit, Gesundheit - Wechselwirkungen. **Bundesgesundheitsblatt**, **33**, 94-96.
- Hager, W. & Westermann, R. (1983). Planung und Auswertung von Experimenten. In J. Bredekamp & H. Feger (Eds.), **Hypothesenprüfung. Enzyklopädie der Psychologie** (Serie 1, Bd. 5., pp. 24-238). Göttingen: Hogrefe.
- Hastorf, A.H. & Cantril, H. (1954). They saw a game: A case study. **Journal of Abnormal and Social Psychology**, **49**, 129-134.
- Hennekens, C. H. & Buring, I. E. (1984). **Epidemiology in medicine**. Boston: Little, Brown and Company.
- Hennigan, K.M., Del Rosario, M. L., Heath, L., Cook, T. D., Wharton, J. D. & Calder, B. J. (1982). Impact of the introduction of television on crime in the United States: Empirical findings and theoretical implications. **Journal of Personality and Social Psychology**, **42**, 461-477.
- Herr, P.M., Sherman, S. J. & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. **Journal of Experimental Social Psychology**, **19**, 323-340.
- Herrmann, T. (1976). **Die Psychologie und ihre Forschungsprogramme**. Göttingen: Hogrefe.
- Hodapp, V. (1984). **Analyse linearer Kausalmodelle**. Bern: Huber.
- Hosmer, D. W. & Lemeshow, S. (1989). **Applied logistic regression**. New York: Wiley.
- Huber, O. (1987). **Das psychologische Experiment: eine Einführung**. Bern: Huber.
- Hummell, H. J. (1980). Experiment in den Sozialwissenschaften. In J. Speck (Ed.), **Handbuch wissenschaftstheoretischer Begriffe** (Bd. 1, pp.210-216). Göttingen: Vandenhoeck und Ruprecht.
- Irle, M. (1979). Das Instrument der „Täuschung“ in der Verhaltens- und Sozialwissenschaftlichen Forschung. **Zeitschrift für Sozialpsychologie**, **10**, 305-330.
- Jöreskog, K. G. & Sörbom, D. (1989). **LISREL 7. User's reference guide**. Mooresville, IN: Scientific Software.
- Josephson, W. L. (1987). Television violence and children's aggression: Testing the priming, social script, and disinhibition predictions. **Journal of Personality and Social Psychology**, **53**, 882-890.
- Keith-Spiegel, P. & Kocher, G. P. (1985). **Ethics in psychology**. Hillsdale: Erlbaum.
- Kennedy, J. J. (1983). **Analyzing qualitative data. Introductory log-linear analysis for behavioral research**. New York: Praeger.
- Klopfer, B. (1957). Psychological variables in human cancer. **Journal of Projective Techniques**, **21**, 331-340.
- Knowles, E., Coker, M.C., Cook, D. A., Diercks, S. R., Irwin, M.E., Lundeen, E. J., Neville, J. W. & Sibicky, M.E. (im Druck). Order effects within personality measures. In N. Schwarz & S. Sudman (Eds.), **Context effects in social and psychological research**. New York: Springer.
- Kohlmeier, L., Rehm, J. & Hoffmeister, H., (1990). Lifestyle and trends in worldwide breast cancer rates. In D. L. Davis & D. G. Hoel (Eds.), **Trends in cancer mortality in industrial countries**. Annals of the New York Academy of Sciences (Vol.609, pp.259-268). New York: New York Academy of Sciences.

- Kruse, L. & Kumpf, M. (Eds.) (1981). **Psychologische Grundlagenforschung: Ethik und Recht**. Bern: Huber.
- Lanzetta, J.T., Cartwright-Smith, J. & Kleck, R.E. (1976). Effects of nonverbal dissimulation on emotional experience and autonomic arousal. *Journal of Personality and Social Psychology*, **33**, 354-370.
- Larsen, R. J. & Sinnett, L.M. (1991). Meta-analysis of experimental manipulations: Some factors affecting the Velten mood induction procedure. *Personality and Social Psychology Bulletin*, **17**, 323-334.
- Lefkowitz, M.M., Eron, L. D., Walder, L. O. & Huesman, L. R. (1972). Television violence and child aggression: A follow up study. In G.A. Comstock & E.A. Rubenstein (Eds.), **Television and social behavior** (Vol. 3: Television and adolescent aggressiveness). Washington, DC: United States Government Printing Office.
- Lingle, J. H. & Ostrom, T. M. (1979). Retrieval selectivity in memory-based impression judgments. *Journal of Personality and Social Psychology*, **37**, 180-194.
- Mark, M.M., Sinclair, R. C. & Wellens, T. R. (1991). The effect of completing the Beck Depression Inventory on self-reported mood state: Contrast and assimilation. *Personality and Social Psychology Bulletin*, **17**, 457-465.
- Merz, F. (1987<sup>2</sup>). Experiment. In A. Wilhelm, H. J. Eysenck & R. Meili (Eds.), **Lexikon der Psychologie** (Bd. 1, pp. 538-544). Freiburg: Herder.
- Metz, G. (1989). **Deindividuation und geschlechtsspezifisches Aggressionsverhalten in Gruppen**. Mannheim: Unveröffentlichte Diplomarbeit.
- Metzger, W. (1952). Das Experiment in der Psychologie *Studium Generale*, **5**, 142-163
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, **67**, 371-378.
- Milgram, S. (1964). Issues in the study of obedience: A reply to Baumrind. *American Psychologist*, **19**, 848-852.
- Milgram, S. (1973). **Obedience to authority: An experimental view**. New York: Harper and Row.
- Mill, J. S. (1968/1886). **System der deduktiven und induktiven Logik**. Band 3. Aalen: Scientia. (Neudruck der Ausgabe Leipzig 1886.)
- Nebel, A., Strack, F. & Schwarz, N. (1989). Tests als Treatment. Wie die psychologische Messung ihren Gegenstand verändert. *Diagnostica*, **35**, 191-200.
- Neisser, U. (1982). **Memory observed. Remembering in natural contexts**. San Francisco: Freeman.
- Nowack, W. (1991). **Labor „vs.“ Feld = Interne „vs.“ Externe Validität?** Vortrag auf dem 7. Hamburger Symposium zur Methodologie der Sozialpsychologie 11.1. - 12.1.1991.
- Orne, M. (1962). On the social psychology of the psychological experiment. *American Psychologist*, **17**, 776-783.
- Orne, M. (1969). Demand characteristics and the concept of quasi-experimental control. In R. Rosenthal & R. Rosnow (Eds.), **Artifact in behavioral research** (pp. 143-179). New York: Academic.



- Ostrom, T. M. & Upshaw, H. S. (1968). Psychological perspective and attitude Change. In A. G. Greenwald, T. C. Brook & T. M. Ostrom (Eds.), **Psychological foundations of attitudes** (pp. 217-242). New York: Academic.
- Parducci, A. (1965). Category judgment: A range-frequency model. **Psychological Review**, *72*, 407-418.
- Pocock, S. J. (1983). **Clinical trials**. Chichester: Wiley.
- Price, L. (1984). Art, science, faith, and medicine: The implications of the placebo effect. **Sociology of Health and Illness**, *6*, 61-73.
- Pritsch, M. & Noack, H. (1991). Möglichkeiten und Grenzen randomisierter Studien in der Psychotherapieforschung. In: BMFT (Ed.), **Nachlese zum 4. Statuskolloquium am 18./19. April in Tübingen Bereich: Therapie und Rückfallprophylaxe chronischer psychischer Erkrankungen im Erwachsenenalter** (pp. 36-47). Bonn
- Rehm, J., Arminger, G. & Kohlmeier, L. (1992). Using follow-up data to avoid omitted variable bias: An application to cardiovascular epidemiology. **Statistics in Medicine**, *11*, 1195-1208.
- Rehm, J., Steinleitner, M. & Lilli, W. (1987). Wearing uniforms and aggression - a field experiment. **European Journal of Social Psychology**, *17*, 357-360.
- Rollman, G.B. (1983). Measurement of experimental pain in chronic pain patients: Methodological and individual factors. In R. Melzack (Ed.). **Pain measurement and assessment** (pp.251-258). New York: Raven Press.
- Rogosa, D. (1980). A critique of cross-lagged correlations. **Psychological Bulletin**, *88*, 245-258.
- Rosenthal, R. (1966). **Experimenter effects in behavioral research**. New York: Appleton-Century-Crofts.
- Rosenthal, R. (1967). Covert communication in the psychology experiment. **Psychological Bulletin**, *67*, 356-367.
- Rosenthal, R. (1969). Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal & R. Rosnow (Eds.), **Artifact in behavioral research** (pp. 181-277). New York: Academic.
- Ross, L.D. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), **Advances in Experimental Social Psychology** (Vol. 10). New York: Academic Press.
- Ross, L. D., Lepper, M. R. & Hubbard, M. (1975). Perseverance in self perception and social perception: Biased attributional processes in the debriefing paradigm. **Journal of Personality and Social Psychology**, *32*, 880-892.
- Ross, L. D., Rodin, J. & Zimbardo, P. G. (1969). Toward an attribution therapy: the reduction of fear through induced cognitive-emotional misattribution. **Journal of Experimental Social Psychology**, *12*, 279-288.
- Rozin, P. & Fallon, A.E. (1987). A perspective on disgust. **Psychological Review**, *94*, 23-41.
- Rothman, K. (1986). **Modern epidemiology**. Boston: Little, Brown & Company.
- Salancik, G.R. & Conway, M. (1975). Attitude inferences from salient and relevant cognitive content about behavior. **Journal of Personality and Social Psychology**, *32*, 829-840.

- Sarnoff, I. & Zimbardo, P.G. (1961). Anxiety, fear, and social affiliation. *Journal of Abnormal and Social Psychology*, **62**, 356-363.
- Schachter, S. (1959). *The psychology of affiliation*. Stanford: Stanford University Press.
- Schlesselman, J. J. (1982). *Case-Control Studies. Design, Conduct, Analysis*. New York: Oxford University Press.
- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H. & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, **61**, 195-202.
- Schwarz, N., Hippler, H. J., Deutsch, B. & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly*, **49**, 388-495.
- Schwarz, N. & Strack, F. (1981). Manipulating salience: Causal assessment in natural settings. *Personality and Social Psychology Bulletin*, **7**, 554-558.
- Schwarz, N., Strack, F. & Hippler, H.J. (1991). Kognitionspsychologie und Umfrageforschung: Themen und Befunde eines interdisziplinären Forschungsgebietes. *Psychologische Rundschau*, **42**, 175-186.
- Schwarz, N., Strack, F., Müller, G. & Chassein, B. (1988). The range of response alternatives may determine the meaning of the question: Further evidence on informative functions of response alternatives. *Social Cognition*, **6**, 107-117.
- Schuler, H. (1980). *Ethische Probleme psychologischer Forschung*. Göttingen: Hogrefe.
- Sieber, J. E. (1982a). *The ethics of social research: Survey and experiments*. New York: Springer Verlag.
- Sieber, J. E. (1982b). *The ethics of social research: Fieldwork regulations, and publication*. New York: Springer Verlag.
- Soloman, R.L. (1949). An extension of control group design. *Psychological Bulletin*, **46**, 137-150.
- Steiger, J.H. (1989). *EzPATH. Causal Modeling*. Evanston, 11: SYSTAT.
- Stelzl, I. (1983). Zur Uneindeutigkeit von LISREL-Lösungen: Überlegungen und Beispiele. *Psychologische Beiträge*, **25**, 315-335.
- Stelzl, I. (1986). Changing causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research*, **21**, 309-331.
- Strack, F. & Martin, L.L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude Surveys. In H. J. Hippler, N. Schwarz & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 123-148). New York: Springer.
- Strack, F., Martin, L.L. & Schwarz, N. (1988). Priming and communication: Social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology*, **18**, 429-442.
- Strack, F., Schwarz, N. & Gschneidinger, E. (1985). Happiness and reminiscing: The role of time perspective, affect, and mode of thinking. *Journal of Personality and Social Psychology*, **49**, 1460-1469.

- Strack, F., Schwarz, N. & Nebel, A. (1990). **Happiness and anticipating the future: Affective and cognitive** consequences. Paper presented at the General Meeting of the European Association of Experimental Social Psychology, Budapest, Hungary.
- Strack, F., Schwarz, N., Chassein, B., Kern, D. & Wagner, D. (1990). The salience of comparison Standards and the activation of social norms: Consequences for judgments of happiness and their communication. **British Journal of Social Psychology**, 29, 303-314.
- Strack, F., Schwarz, N., Weidner, R., Hippler, G. & Schwarz, R. (1983). Politische Einschüchterung als sozialpsychologisches Problem. In J. Haisch (Ed.), **Ange wandte Sozialpsychologie** (pp. 195-210). Bern: Huber.
- „Student“ (1931). The Lanarkshire milk experiment. **Biometrika**, 23, 398-406.
- Textor, M. R. (1989). Experimente am Menschen - ethische und juristische Aspekte. **Medizin Mensch Gesellschaft**, 14, 208-213.
- Traxel, W. (1964). **Einführung in die Methodik der Psychologie**. Bern: Huber.
- Turner, J. C. (1981). Some considerations in generalizing experimental social psychology. In G. M. Stephenson & J. M. Davis (Eds.), **Progress in applied social psychology** (Vol. 1, pp. 3-34). New York: Wiley.
- Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. **Cognitive Psychology**, 4, 207-232.
- Wahrendorf, J., Blettner, M. & Edler, L. (1985). An investigation of the adequacy of a randomization. **Controlled Clinical Trials**, 6, 249-258.
- Webb, E. J., Campbell, D.T., Schwartz, D. & Sechrest, L. (1966). **Unobtrusive measures: Nonreactive research in the social sciences**. Chicago: Rand-McNally.
- Wilcox, R. R. (1987). New designs in analysis of variance. **Annual Review of Psychology**, 38, 29-60.
- Winer, B. J. (1971<sup>2</sup>). **Statistical principles in experimental design**. New York: Mc Graw Hill.
- Zebrowitz, L. A. (1990). **Social perception**. Pacific Grove, Cal. : Brooks/Cole Publishing Co.
- Zelen, M. (1990). Randomized consent designs for clinical trials: An update. **Statistics in Medicine**, 9, 645-656.

## Von Zahlzeichen zu Skalen

***Rainer Mausfeld***

### ***1. Der Zahlgebrauch in der Psychologie***

Testen, Skalieren, Messen: drei Bezeichnungen in der Psychologie für einen Sachverhalt, den der Physiker Messung nennt oder den er als ‚meßbar machen‘ umschreibt, falls er erst die Grundlagen der Messung bereitstellt. Die begriffliche Vielfalt in der Psychologie ist nicht Resultat einer logisch-systematischen Einteilung unterschiedlicher Meßoperationen im Bereich des Psychischen, sondern vorrangig historische Konsequenz unterschiedlicher Perspektiven, aus denen heraus in der Psychologie das Meßproblem betrachtet wurde.

In der diagnostischen oder psychometrischen Tradition stehen der unterschiedliche Ausprägungsgrad von Eigenschaften verschiedener Individuen, wie beispielsweise Intelligenz, Konzentrationsfähigkeit und Neurotizismus, im Mittelpunkt des Interesses. Diese vorrangig sozialtechnologisch orientierte Tradition versteht unter dem Begriff des Messens die Erstellung geeigneter Indizes innerhalb statistischer Vorhersagetechniken; die Bewertung dieser Indizes ergibt sich dabei eher aus der pragmatischen Brauchbarkeit denn aus einem Theoriegefüge.

In der psychophysikalischen Tradition werden allgemeine Prinzipien der Abhängigkeit ‚subjektiver Dimensionen‘, wie Lautstärke, Helligkeit und Farbton, von den ihnen zugrunde liegenden physikalischen Größen untersucht; diese Abhängigkeiten sucht man durch Skalen zu modellieren und die Beziehungen zwischen verschiedenen psychologischen ‚Dimensionen‘ zu charakterisieren. Mit dem Ziel, die Natur und Struktur perzeptueller Codes zu erfassen, orientiert sich hier der Meßbegriff an dem Vorbild der Physik.

In der meßtheoretischen Tradition schließlich werden abstrakt die Bedingungen behandelt, unter denen eine physikalische oder psychologische Eigenschaft durch Zahlen im Sinne der Mathematik charakterisiert werden kann. Hier sucht man für psychologische Größen den Größenbegriff in ähnlich präziser Weise zu fassen, wie die Physik dies für ihren Größenbegriff tut, und

zugleich zu zeigen, daß psychologische Relationen strukturell so reichhaltig sein können, daß sie sich durch numerische Relationen symbolisch repräsentieren lassen.

In der Geschichte des Gebrauchs von Zahlen lassen bis heute - stark vereinfachend - zwei Pole bestimmen, die in gewisser Weise die historische Entwicklung der Arithmetik umspannen. An dem einen Pol finden wir den konkreten Umgang mit individuellen Zahlzeichen; dieser steht in engem Zusammenhang mit Alltagsproblemen, etwa des Handwerks, des Warentausches oder konkreter Maßtechniken. Der operative Umgang des Zählens und Rechnens differenzierte sich allmählich in vielfältiger Weise aus und führte, wie bei den Babyloniern, zu hochentwickelten Rechentechniken und Regelsystemen für das Operieren mit Zahlzeichen. Dieser konkret-operativen Umgangsweise steht als anderer Pol eine abstrakt-theoretische Auffassung von Zahlen gegenüber. Historisch findet sich diese Haltung zuerst bei den Griechen, die eine abstrakte Definition der Zahl suchten und erkannten, daß der Begriff der Zahl dem der Größe untergeordnet ist. Zahlen stellen hier eigenständige theoretische Entitäten dar, die ihre Bedeutung durch eine umfassende abstrakte Theorie erhalten. Verschiedene Zahlzeichen können dabei dieselbe Zahl darstellen. Dieser hochgradig theoretische Gebrauch des Konzeptes der Zahl ist kennzeichnend für die Mathematik.

Beide Pole des Zahlgebrauchs spielten bei der Entwicklung der neuzeitlichen Naturwissenschaft eine wichtige Rolle; die moderne Naturwissenschaft verdankt ihre theoretischen Erfolge jedoch vorrangig der Anbindung ihrer Theorieentwicklung an die Mathematik. Erst mit dem theoretischen Konzept der Zahl, wie es die Mathematik entwickelte, konnte die Idee einer umfassenden Quantifizierung ihre Faszination als Leitbild der Theorieentwicklung entfalten. Ob man, wie Cusanus, das Wort ‚mens‘ mit dem Wort ‚mensura‘ in Beziehung setzt oder wie Keppler der Überzeugung ist, wie sich das Auge auf das Erkennen von Farben und das Ohren auf das Erkennen von Tönen richte, so richte sich der menschliche Geist auf die Erkenntnis des Quantitativen: in dem Maße, wie sich in der neuzeitlichen Wissenschaft das Streben nach Naturerkenntnis mit dem Bedürfnis nach Gewißheit dieser Erkenntnis verbindet, läßt sich ein theoretischer Erkenntnisfortschritt nur in Anbindung an die Mathematik erreichen.

Die Erfolge, welche das quantitativ-theoretische Denken für die Theorieentwicklung der Physik verzeichnen konnte, trugen dazu bei, daß auch die Psychologie ihre Entwicklung als eigenständige, empirisch orientierte Wissenschaft an die Bedeutung band, die sie dem Quantitativen in ihrer Theorieentwicklung zu geben vermochte. In jenen Bereichen der Psychologie, deren Entwicklung durch theoretische Intentionen geleitet war, wie die Fechnersche Psychophysik, wird man eine solche Haltung, in geeigneter Präzisierung,

durchaus als berechtigt ansehen können. Doch folgte die Psychologie auch in jenen Bereichen diesem Leitbild, wo sie sich für die Bewältigung von Alltagsproblemen mit einer operativen Verwendung von Zahlzeichen hätte zufrieden geben können. Ohne Beachtung der schon in ihrer Geschichte angelegten Heterogenität der Psychologie wird man die Rolle des Quantitativen in der Psychologie nicht diskutieren können. Theoretische Intentionen auf der einen Seite und konkrete, mit Alltagsproblemen verbundene Zielsetzungen auf der anderen charakterisieren in der Psychologie zwei sich parallel entwickelnde Wissenschaftsbereiche, die trotz mannigfacher Querbezüge von gänzlich unterschiedlichen Prinzipien und Erkenntniszielen geprägt sind. Als Stellvertreter für diese beiden Wurzeln der Psychologie seien Fechner und Binet genannt. Dem Quantitativen kommt nun in beiden - hier schematisch gegenübergestellten - Bereichen eine jeweils andere Bedeutung zu: Sucht man eine Theorie über die Natur und Funktionsweisen des menschlichen Geistes zu erstellen, so kann man - wie Fechner - fragen, inwieweit der Aufbau einer solchen Theorie auf der Entwicklung eines eigenständigen Konzeptes psychologischer Größen beruht und ob in dieser Theorie Konzepte, die sich in Analogie zur Physik als subjektive Dimensionen beschreiben lassen, eine Rolle spielen. Zielt man jedoch aus einer sozialtechnologischen Perspektive auf die Erfassung und Beherrschung von Zusammenhängen in komplexen Systemen, so möchte man quantitative Indizes gewinnen, die sich zur Lösung des jeweiligen Problems als besonders geeignet erweisen. Nicht nur der Erklärungs-begriff ist hier ein anderer, sondern auch die im Vordergrund stehenden Aspekte des Konzeptes der Zahl. Die sich in einer naturwissenschaftlich orientierten Psychologie möglicherweise stellende Frage, ob ihre theoretischen Strukturen ähnlich denen der Physik zu eigenständigen Größen, Dimensionen und Skalen führen, ob also die Zahlzeichen durch den strukturellen Gehalt der Theorie als Zahlen aufgefaßt werden können, ist für eine sozialtechnologisch orientierte Psychologie unerheblich (vgl. Herrmann, in diesem Band, Kap. 6.5.3). Die funktionale Beherrschung eines Systems setzt weder eine substantielle, an einem kausalen Erklärungs-begriff orientierte Theoriebildung noch einen eigenständigen Größenbegriff voraus. Die empirische Struktur der Indizes, wie man sie durch Tests, Rating-Skalen u.ä. gewinnt, kann so reichhaltig sein, daß sie die jeweils erforderlichen Einsichten in die Funktionszusammenhänge eines komplexen Systems bereitstellen kann. Daß derartige Indizes, d.h. Zahlzeichen, innerhalb komplexer Rechenprozeduren, etwa mathematisch-statistischer Schätz- und Prüfverfahren, wieder als Zahlen im theoretischen Sinne (in der Regel als reelle Zahlen) aufgefaßt werden, rechtfertigt nicht, diese Indizes als Messung einer psychologischen Größe, d.h. als Skala, zu betrachten - es sei denn man liberalisiert die Begriffe der Größe, Dimension und Skala in entsprechender Weise; mit dieser Liberalisierung verliert jedoch das Problem der Quantifizierbarkeit psychologischer Größen sein theoretisches Interesse. (Die Interpretation von Parametern in einem wahrscheinlichkeitstheoretischen Modell als Skalenni-

veau ist Folge einer solcher Bedeutungsliberalisierung.) Das naturwissenschaftliche Ideal einer Quantifizierung bindet das Quantitative an den strukturellen Gehalt einer substantiellen Theorie. Statistische Theorien können aber kein Ersatz für inhaltliche Theorien sein; losgelöst von diesen können sie zu einer Quantifizierung nicht beitragen.

Dieses Kapitel befaßt sich mit der Frage, ob es eigenständige Größen in der Psychologie gibt, die sich durch Skalen numerisch darstellen lassen; es diskutiert einige der mit ihr verbundenen Konzepte und Probleme. Nicht zuletzt aus den genannten Gründen soll dies exemplarisch am Beispiel der Psychophysik erfolgen: Die Psychophysik ist zur Untersuchung einer solchen Frage in einer vergleichsweise glücklichen Lage, da sie über Bereiche mit reichhaltigen theoretischen Strukturen verfügt; auch lassen hier die Möglichkeiten einer präzisen physikalischen Beschreibung des Reizes sowie seine hohe experimentelle Kontrollierbarkeit es am ehesten als sinnvoll erscheinen, den Versuch einer Definition psychologischer Größen zu unternehmen und die Beziehungen dieser Größen zueinander zu studieren. Die Frage der Meßbarkeit und Quantifizierung, wie sie sich hier stellt, läßt zudem die Bezüge zum allgemeinen Problem der Rolle des Quantitativen in den Naturwissenschaften deutlich erkennen.

Innerhalb der Psychologie wird man beim Konzept der Skala überwiegend an Skalen aus dem Bereich der Sozialpsychologie und Diagnostik denken: Aggressions- und Vorurteilsskalen sind ebenso geläufig wie Intelligenz- und Neurotizismusskalen. Der Skalenbegriff spielt also gerade in jenen Bereichen der Psychologie eine wichtige Rolle, die sich mit der Entwicklung einer Methodologie zur Erfassung komplexer Funktionszusammenhängen beschäftigen. Doch hängt dieser Skalenbegriff nur noch metaphorisch mit dem Skalenbegriff zusammen, aus dem die Frage der Meßbarkeit psychologischer Größen ihre theoretische Bedeutung bezieht. Für die Einstellungsmessung und Diagnostik ist dies freilich keine Einschränkung, da die Erreichbarkeit ihrer jeweiligen Ziele in keiner Weise davon abhängt, daß es einen eigenständigen psychologischen Größenbegriff gibt; daher ist es schwer verständlich, warum sich hier der Blick oftmals auf Probleme einengt, die mit einer Rechtfertigung der Meßbarkeit verbunden sind. Bei der Behandlung gesellschaftlicher Probleme oder als interessant erachteter Alltagsfragen haben numerische Indizes ganz unterschiedlicher Art seit jeher eine große Rolle gespielt (s. Gould, 1981). Die Psychologie hat für die Behandlung dieser Probleme ein umfangreiches Instrumentarium und eine eigene Methodologie hervorgebracht, deren pragmatischer Nutzen jenseits der metatheoretischen Frage einer Meßbarkeit des Psychischen außer Zweifel steht. Doch wirft die Diskrepanz zwischen Methodologie und theoretischem Substrat, zu der Suppes (1962, S.260) bemerkt, „It is a paradox of scientific method that the branches of empirical science that have the least theoretical developments often have the

most sophisticated methods of evaluating evidence“, wiederum Fragen auf. Insbesondere birgt sie einerseits die Gefahr, daß in den Teilen der Psychologie, die ihre psychologisch-theoretischen Ansprüche reduzieren und ihren Gegenstandsbereich unter dem Postulat der unmittelbaren Anwendbarkeit verkürzen, eine Funktionalisierung der Theorieentwicklung für externe Zwecke (Böhme et al., 1974) überhand nimmt und daß andererseits bei einer Behandlung des Meßbarkeitsproblems Fragen nach der Inkongruenz von theoretischem Verständnis und sozialtechnischer Brauchbarkeit als ungerechtfertigt erscheinen und die Methodologien selbst als Lösung des Meßbarkeitsproblems betrachtet werden.

Im folgenden untersuchen wir zunächst den Begriff der psychologischen Größe und beschäftigen uns mit der Frage, ob es bereits a priori Einwände gegen eine theoriebezogene Quantifizierung des Psychischen gibt (die weiterführende Frage, ob die Theorieentwicklung in der Psychologie tatsächlich zur Entwicklung eigenständiger fundamentaler Skalen geführt hat, werden wir hier nicht behandeln können; vgl. jedoch Mausfeld, in diesem Band, Kap.4). Sodann werden wir einige jüngere Entwicklungen, die mit dem Ziel der Skalenkonstruktion verbunden sind, diskutieren. Die verschiedenen Arten der Skalenkonstruktion können nach vielerlei Gesichtspunkten klassifiziert werden, beispielsweise nach der Art der sich aus unterschiedlichen psychologischen Relationen ergebenden Datenstruktur (Coombs, 1964; Roskam, 1983). Da in diesem Kapitel metatheoretische Betrachtungen im Vordergrund stehen sollen, scheint eine abstraktere und zugleich gröbere Klassifikation geeigneter zu sein, in der sich grundsätzliche theoretische Perspektiven der Skalenkonstruktion ausdrücken. Schließlich werden wir uns der Frage zuwenden, inwieweit man sicherstellen kann, daß numerische Relationen und Aussagen in sinnvoller Weise an psychologische angebunden sind.

## ***2. Qualität und Quantität: Ist Psychisches meßbar?***

„Quantität“ und „Messung“, auf diesen Begriffen, die im Erkenntnisprozeß eine besondere Wirkungskraft zu entfalten scheinen, gründet sich wesentlich die epistemologische Dignität der Naturwissenschaften, voran der Physik. Ihren Erkenntnisidealen folgend bemühte sich auch eine empirisch orientierte Psychologie mit der Eliminierung alles philosophisch Spekulativen um den Nachweis der prinzipiellen Möglichkeit einer Quantifizierung des Psychischen.

Die klassische Grenzziehung von Qualität und Quantität war in der Physik schon früh durch eine fortlaufende Neubestimmung dieser Begriffe aufgehoben worden. Die Physik löste sich von der unmittelbaren Erfahrung, qualitative Unterschiede suchte man auf Differenzen des Quantitativen zurückzuführen, sei es der geometrischen Struktur oder der Zahl. Mit der pythagorä-



schen Idee, die Einheit der Welt in einer abstrakten Harmonie zu suchen, bildet sich ein neues Erkenntnisideal aus. Es kehrt wieder im neuzeitlichen Gedanken einer *Mathesis Universalis* einer allgemeinen Theorie der Größen und Größenverhältnisse, in der die sich ausdifferenzierenden und zunehmend von der Welt unmittelbarer Erfahrungen sich lösenden Wissenschaften eine methodische Einheit finden sollen. Seitdem prägte es die Entwicklung der neuzeitlichen Naturwissenschaften, voran der Physik, auf deren wichtigstes Charakteristikum es hinweist: ihre Anbindung an die Mathematik.

Zu einer *Mathesis Universalis* zählen all jene Unterfangen, wo, so Descartes, „nach Ordnung und Maß geforscht wird, und daß es hierbei gar nicht darauf ankommt, ob man dieses Maß nun in den Zahlen, oder in den Figuren oder den Gestirnen oder den Tönen oder in irgendeinem anderen Gegenstand zu suchen hat“ (zit. nach Mittelstraß & Schroeder-Heister, 1986). Der für diese formale Einheit zu entrichtende Preis war freilich ein Wachsen der Kluft zwischen alltäglicher Erfahrungswelt und Welt der Physik. Eben dieses Ideal einer theoretischen Quantifizierung, das sich in der Physikgeschichte allmählich entfaltete, tritt mit G.Th. Fechner weitgehend unvermittelt in die Psychologiegeschichte ein (daß Fechner den Neologismus *Psychophysik* für sein Forschungsprogramm bildet, ist ein Indiz für diese Diskontinuität) und stellt fortan die Richtschnur eines sich naturwissenschaftlich begreifenden psychologischen Denkens dar. Hingegen fand das ursprüngliche *Warum*, Fechners philosophische Perspektive nämlich, der zählende und messende Gott der Weisheit Salomons habe auch die psycho-physische Einheit nach Zahl und Maß geordnet, ihres metaphysischen Gehaltes wegen in der sich rein empirisch verstehenden Psychologie keinen Platz mehr. Fechner ging es nämlich, vor dem Hintergrund des Leib-Seele-Problems, noch im eigentlichen Sinne um ‚Empfindungsmessung‘.

Hat nun das Fechnersche Forschungsprogramm in seinen modernen Varianten den Anspruch eingelöst, psychologische Größen in einem theoretisch fruchtbaren Sinne zu quantifizieren und der Psychologie durch so erstellte Skalen ein theoretisches Fundament bereitzustellen? Allein schon die Auseinandersetzung mit dieser Fragen läßt hoffen, daß dadurch der Gegenstandsbereich der Psychologie selbst erhellt wird.

Zwei Problemkreise, zwei Begriffskomplexe berühren sich in der Frage nach der Metrisierbarkeit psychologischer Größen. Da ist zunächst der Begriff der psychologischen Größe oder Eigenschaft. Zweitens der Begriff der Metrisierbarkeit und mit ihm verbunden die Begriffe ‚Zahl‘, ‚Größe‘ und ‚Quantität‘. Welches Verhältnis haben Größen, seien es nun Länge, Gewicht oder Fechners Konzept der Stärke der Empfindung, und Zahlen zueinander? Warum lassen sich die aus der Wirklichkeit abstrahierten Größen durch mathematische Entitäten so erfolgreich beschreiben?

Werfen wir einen kurzen Blick auf das Verhältnis von Zahlen und empirisch interpretierten Größen. Diese Beziehung ist für unser Verständnis von Wissenschaft grundlegend, steht doch „am Anfang der Mathematik und am Anfang der Philosophie und am Anfang des abendländischen Denkens diese Begeisterung für die Zahl“ (Martin, 1956, S. 16). Im griechischen Sinne waren Zahlen nur das, was in heutiger Terminologie positive ganze Zahlen sind. Allgemeiner als der Begriff der Zahl war der der Größe. Als Größe wurde alles aufgefaßt, was in Bestandteile zerlegbar ist, wobei die Teile einer Größe für sich bestehen können und selbst wieder Größen sind, wie etwa die Teile einer Länge wieder Längen sind. Diese beiden Konzepte, Zahl und Größe, welche entscheidende Wurzeln des naturwissenschaftlich-mathematischen Denkens bildeten, aus denen heraus Mathematik und Physik gleichermaßen erwachsen, ließen sich durch das Konzept der Proportion (als einem Verhältnis zweier Größen) in eleganter Weise miteinander in Beziehung setzen, freilich nur unter der stillschweigenden Voraussetzung, daß je zwei Größen kommensurabel sind, d.h. ein gemeinsames Maß haben. Mit der Erschütterung, welche die Entdeckung der Inkommensurabilität auslöste, mit dem Nachweis von Größen nämlich, die keine durch ganze Zahlen ausdrückbare Proportion haben, zerbrach auch die ursprünglich angenommene Beziehung von Zahl und Größe. So finden wir bei Euklid Zahlen und Größen streng geschieden. Es war der von Eudoxos entwickelte Kalkül mit Größenverhältnissen und keineswegs das griechische Konzept der Zahl, das zur Entwicklung der so mächtigen Theorie der reellen Zahlen führte (zwischen der klassischen Proportionslehre und der Theorie der Automorphismengruppe extensiver Strukturen besteht eine enge Beziehung, wie Krull, 1960, gezeigt hat). Eine der heute üblichen Präzisierungen des Konzepts der reellen Zahlen, nämlich die von Dedekind, knüpft unmittelbar an das 5. Buch Euklids an, das die Darstellung der Eudoxischen Größenlehre zum Inhalt hat. Angesichts dieser Entwicklung sieht man, daß unser heutiges Konzept der Zahl, das ja allem Quantifizieren und Messen unterliegt, historisch selbst aus der Behandlung geometrischer und physikalischer Größen erwachsen ist. Die reellen Zahlen sind nicht etwas in der Mathematik Vorgefundenes, sondern eine Abstraktionsleistung des menschlichen Geistes, konstruiert zu dem Zwecke, Größen gedanklich handhabbarer zu machen. Die reellen Zahlen stellen gleichsam eine universelle Größenstruktur dar, durch welche die formale Struktur aller eindimensionalen Größen erfaßt wird (vgl. Bourbaki, 1966, S. 12). Interessanterweise knüpft auch in der abstrakten Meßtheorie Niederée (1987, 1992a,b) an die Proportionslehre an; er entwickelt eine modelltheoretische Verallgemeinerung von ihr, die er zur Grundlegung und Verallgemeinerung meßtheoretischer Konzepte, wie sie im folgenden beschrieben werden, heranzieht. Anders als der herkömmliche Repräsentationsansatz, doch diesen gleichsam fundierend, geht Niederée nicht von einem vorgegebenen Zahlbereich (wie dem der reellen Zahlen) als Grundmengen der repräsentierenden numerischen Struktur aus.

Statt dessen geht sein Typenansatz von einem abstrakten Konzept des iterativen **Meßverfahrens** aus, und ‚Zahlen‘ im Sinne von ‚Meßwerten‘ werden sodann im Kontext der jeweiligen qualitativen Struktur auf theoretischer Ebene aus dem betreffenden Meßverfahren heraus gleichsam konstruiert. Dieser Ansatz ermöglicht zum einen eine befriedigendere Behandlung einer Reihe grundsätzlicher Fragen (z.B. zur ‚Konstruierbarkeit‘ von Skalen und zur Rolle Archimedischer Axiome) als bisher möglich. Darüber hinaus erweitert sich der ‚meßtheoretische Horizont‘, da im Prinzip auch Zahlbereiche denkbar erscheinen, welche wesentlich anderer Natur sind als der Bereich der reellen Zahlen (der ja seinem Ursprung nach primär auf elementare physikalische und geometrische Probleme zugeschnitten ist).

Meßtheoretische Betrachtungen im eigentlichen Sinne wurden durch erkenntnistheoretische Fragen der Physik veranlaßt. Man wandte sich in abstrakter Weise der Größenmessung selbst zu, um die erkenntnistheoretischen und formalen Grundlagen zu legen, durch die jede Art der Größenmessung ihre Rechtfertigung finden soll. Diesem Ziel widmete 1887 Helmholtz seine Schrift **Zählen und Messen erkenntnistheoretisch betrachtet**. „Was“, so Helmholtz, „ist der objective Sinn davon, dass wir Verhältnisse reeller Objecte durch benannte Zahlen als Größen ausdrücken, und unter welchen Bedingungen können wir dies thun?“ Auf der Suche nach einer Antwort stellte Helmholtz - hierbei zugleich eine Präzisierung des traditionellen Begriffs der extensiven Größe anstrebend - der arithmetischen Addition eine physikalische Verknüpfung, der arithmetischen Gleichheit ein Verfahren der „physikalischen Vergleichung“ gegenüber. Eine Größenmessung für die betrachteten Objekte ist nun, Helmholtz zufolge, genau dann gegeben, wenn die den Objekten zugeordneten Zahlen die physikalische Gleichheit respektieren und wenn die arithmetische Addition mit der physikalischen Addition verträglich ist. Das durch den Zahlgebrauch als Quantitativ Aufscheinende ist nichts anderes als das Spiegelbild der qualitativen Relationen zwischen den Objekten: eine folgenreiche Relativierung (oder genauer eine Neufassung) der traditionellen Dichotomie von Qualität und Quantität. Diese Behandlung von Größen wurde in einer einflußreichen Arbeit von Hölder (1901) mathematisch weitergeführt und erweitert; damit waren die Grundlagen der abstrakten Meßtheorie gelegt, die später im Scott-Suppes-Paradigma zur Standardtheorie der **fundamentalen Messung** in Psychologie und Ökonomie wurden.

Helmholtz zufolge scheint jedes fundamentale Messen extensives Messen zu sein und erfordert folglich eine additionsanaloge Verknüpfung zwischen den empirischen Entitäten. Unter **fundamentalem Messen** wird, nach einer von dem Physiker N. R. Campbell in den 20er Jahren getroffenen Unterscheidung, ein Messen verstanden, daß im Gegensatz zur **sog. abgeleiteten Messung** grundlegenden Charakter hat und nicht selbst wiederum auf einer Messung beruht. In modernen Worten: Fundamentale Größenmessung liegt dann vor,

wenn die Zahlen den qualitativ-strukturellen Gehalt der physikalischen Relationen und Operationen treu widerspiegeln. Mit der Anbindung des fundamentalen an das extensive Messen schien a priori eine fundamentale Messung psychologischer ‚Größen‘ ausgeschlossen zu sein, denn was ließe sich unter einer Addition im Bereiche des Psychischen vorstellen? Fechners Lösungsvorschlag, eine Empfindung mit Hilfe gerade bemerkbarer Unterschiede als eine „Summierung eines Soundsovielmal des Gleichen“ zu bestimmen, hatte mit der extensiven Messung über direkte, assoziative und monotone Verknüpfungsoperationen wenig gemein und ließ sich nicht als ein fundamentales Messen betrachtet. v. Kries (1882) hatte schon eingewandt, das Fechnersche Vorgehen beruhe nicht auf eigentlicher Messung, sondern lediglich auf einer willkürlichen Konvention. Bis zur Mitte dieses Jahrhunderts blieb die Auffassung, fundamentales Messen in der Psychologie sei a priori nicht möglich, vorherrschend. Obwohl der Psychophysiker S. S. Stevens im Zusammenhang mit neuartigen Konstruktionsprinzipien psychophysikalischer Skalen wesentliche Anstöße gab, aus operationalistischer Perspektive über diese Auffassung hinaus zu denken, fand sie ihren deutlichsten Ausdruck 1940 in einem entsprechenden Urteil einer eigens zur Untersuchung dieser Frage eingesetzten Kommission der British Royal Society (vgl. Narens & Luce, 1986). Doch zeigte sich alsbald, daß der tatsächliche Umfang einer allgemeinen Theorie der in einem fundamentalen Sinne meßbaren Größen sehr viel weiter war. Hatte sich schon Fechner durch Bernoullis Überlegungen zur Nutzentheorie leiten lassen, so stellten auch hier wieder Betrachtungen zur Nutzentheorie entscheidende Impulse für eine geeignete Verallgemeinerung des Konzepts fundamentaler Messung bereit (Ramsey, 1931; von Neumann & Morgenstern, 1947, Debreu, 1959) und führten schließlich zum Konzept **additiv-verbundener Messung**. Hinzu kam die modelltheoretische Ausarbeitung des Gedankens ‚Messen als numerische Repräsentation qualitativer empirischer Strukturen‘ durch den mathematischen Logiker Dana Scott und den Philosophen, Physiker und mathematischen Psychologen Patrick Suppes im Jahre 1958. Diese löste eine Fülle weiterer Untersuchungen aus, durch die sich erneut und in neuem Lichte die Frage psychologischer Messung stellte (Krantz, Luce, Suppes & Tversky, 1971). Mit der Abstrakten Meßtheorie konstituierte sich ein sich rasch selbstständiges Forschungsfeld - von Cliff (1992) als „one of the most important developments of scientific psychology“ bezeichnet -, dessen weitreichende Bedeutung für eine naturwissenschaftlich orientierte Psychologie erst zukünftige Entwicklungen erschließen werden.

Die Ausarbeitung des Konzepts der additiv-verbundenen Messung (s. Krantz et al., 1971, Kap. 5) bildete den wichtigsten Ertrag dieser Untersuchungen. Statt sich wie im extensiven Fall auf eine explizite und empirisch interpretierte Verknüpfungsoperation zu beziehen, geht man bei **additive-verbundenen Strukturen** von einer Produktstruktur aus, auf der lediglich die Relation einer

schwachen Ordnung definiert ist. Empirisch entsprechen Produktstrukturen Attributen, die durch mindestens zwei unabhängige Komponenten bestimmt werden. Die entscheidenden strukturellen Annahmen, die in Form sog. Axiome formuliert werden, beziehen sich auf eine Zerlegbarkeit der Produktstruktur in geordnete Teilstrukturen sowie auf einen *trade off* bzw. auf eine Kompensierbarkeit zwischen den Teilstrukturen. Ein psychophysikalisches Beispiel für eine solche additiv-verbundene Struktur ist die binaurale Lautstärke, bei der eine Versuchsperson (Vp) Töne, die sich jeweils aus einer Reizkomponente des linken und rechten Ohres zusammensetzen, der Lautstärke nach ordnet (Levelt, Riemersma & Bunt, 1972). Für additiv-verbundene Strukturen läßt sich dann die Repräsentierbarkeit in einer über der geordneten additiven Gruppe der reellen Zahlen definierbaren numerischen Struktur beweisen (wobei sich dies unter geeigneten Verallgemeinerungen des Scott-Suppes-Ansatzes ebenfalls als Homomorphismus formulieren läßt). Daß dabei ‚lediglich‘ aus Ordnungsinformationen eine additive Repräsentation mit Intervallskaleninvarianz erhalten wird, ist weniger erstaunlich, wenn man sich verdeutlicht, daß in den Anforderungen an diese Ordnung gerade die Strukturrestriktionen verborgen sind, die einer abstrakten additiven Operation entsprechen. Additiv-verbundene Strukturen sind gleichsam „krypto-extensive“ Strukturen (Niederée, 1992a), die ohne explizite Angabe einer empirisch interpretierbaren Verknüpfung eine fundamentale Messung in gleich strengem Sinne erlauben wie die extensiven Strukturen.

Es lassen sich weitere Arten fundamentaler Messung formulieren, die mit den additiv-verbundenen Strukturen verwandt sind und ebenfalls keine extensive Verknüpfung voraussetzen. Die wichtigsten sind Differenzenstrukturen und Bisymmetriestrukturen. Erstere beziehen sich auf eine Ordnung auf Intervallen, durch die sich implizit eine additionsanaloge Verknüpfung von Intervallen formulieren läßt. Für eine Anwendung in der auditiven Psychophysik siehe Schneider, Parker & Stein (1974). Bisymmetriestrukturen, die zuerst von Pfanzagl (vgl. 1971) formal studiert wurden, beziehen sich auf eine Operation  $\circ$ , die hinsichtlich einer Ordnung monoton ist, doch weder positiv ist (d.h.  $a \circ b > a$ ) noch assoziativ oder kommutativ, jedoch eine die Axiome der Assoziativität und Kommutativität abschwächende Bedingung erfüllt, nämlich die Bisymmetrie:  $(a \circ b) \circ (c \circ d) \sim (a \circ c) \circ (b \circ d)$ . Das Studium solcher Strukturen leitet sich aus der Untersuchung algebraischer Eigenschaften von Mittelwerten von Paaren reeller Zahlen ab. Empirisch lassen sich derartige Strukturen mit den Bisektionsmethoden der Psychophysik in Beziehung setzen: Interpretiert man  $a \circ b$  als die Operation der Mittenbildung zwischen den Reizen  $a$  und  $b$ , so ist infolge von Hysterese-Effekten i.d.R. die Kommutativität nicht erfüllt, d.h.  $(a \circ b) \neq (b \circ a)$ , wohl aber die Bisymmetrie. Das Bisektionsverfahren wurde erstmals von Plateau (1872) in der Psychophysik angewandt, um eine Skala von Grautönen zu erstellen.

Für die Psychologie bedeuten die formalen Resultate zu neuartigen Strukturen fundamentaler Messung dies: Der aprioristische Einwand, fundamentales Messen in der Psychologie sei nicht möglich, da ihre Größen nicht-extensiver Natur seien, ist beseitigt. Oftmals ist der qualitative Gehalt einer empirischen Struktur so reichhaltig, daß er auch ohne additive Verknüpfung eine Repräsentation durch die Struktur der reellen Zahlen, d.h. ein fundamentales Messen gestattet. Der Größenbegriff erweist sich somit als sehr viel komplexer und umfassender, als es eine Theorie physikalischer Größen zunächst nahelegt. Konzeptuell scheint folglich der grundsätzlichen Möglichkeit einer quantitativen Erfassung psychologischer Größen nichts im Wege zu stehen. Damit ist freilich nur ein einzelner Aspekt psychologischer Größenmessung geklärt, der sich auf ein begriffslogisches Problem bezieht. Alle weiterführenden konzeptuellen Fragen, etwa welchen Status die psychologischen Relationen haben und was die eigentlichen Objekte einer psychologischen Messung sind, werden dadurch in keiner Weise berührt. Fechner wollte Empfindungen messen und sah folglich diese als eigentliche Objekte der Messung an. Die ‚Empfindung‘ ist verlorengegangen und an ihre Stelle sind die ‚Reize‘ zusammen mit bestimmten Urteilsaufgaben getreten; doch was ist dann der Gegenstand der Messung? Zudem bleibt zu zeigen, daß numerischen Skalen bei der Entwicklung psychologischer und psychophysikalischer Theorien tatsächlich eine Bedeutung zukommt (vgl. Mausfeld, in diesem Band, Kap.4). Die Theorieentwicklung in Naturwissenschaften wie der Chemie und Biologie hat nämlich nicht zu einer eigenständigen Konstruktion fundamentaler Skalen über die physikalischen hinaus geführt; eine naturwissenschaftliche Theoriebildung muß weder notwendig auf einer eigenen Art fundamentaler Messung beruhen, noch setzt sie numerische Skalen voraus.

Wie bedeutend oder unbedeutend sich eigene Skalen auch für die Theoriebildung in der Psychologie erweisen mögen, es bleibt Verdienst der Psychophysik, die Beschäftigung mit diesen abstrakten Fragen der fundamentalen Meßbarkeit jenseits des Konzepts der extensiven Messung ausgelöst zu haben. Dadurch wurde in der Folge ein sehr viel tieferes Verständnis und zugleich eine Neubestimmung der klassischen Dichotomie von Qualität und Quantität erzielt, in der prinzipiell einer fundamentalen Messung psychologischer Größen nichts entgegensteht.

### ***3. Arten der Skalenkonstruktion***

Im folgenden Abschnitt werden typische Vorgehensweisen bei der Konstruktion von Skalen betrachtet; dadurch wollen wir Aufschlüsse darüber gewinnen, was die Objekte einer solchen Messung sein können und welche Rolle die erhaltene Skala für die Theoriebildung haben könnte. Für derartige Be-

trachtungen bietet die Psychophysik einen besonders geeigneten Rahmen: Beispielhaft lassen sich hier sowohl die formalen Techniken wie auch die Anbindung formaler Konzepte an solche der substantiellen Theoriebildung in präziser Weise behandeln; zudem bildet die Psychophysik den ideengeschichtlichen Ursprung, von dem andere Bereiche der Psychologie, wie die Entscheidungstheorie, die Diagnostik und die Sozialpsychologie, entscheidende Anstöße und Intuitionen der Skalenkonstruktion bezogen. Die Möglichkeit einer präzisen physikalischen Beschreibung des zu skalierenden Attributes in der Psychophysik läßt Grundzüge und Probleme der Skalenkonstruktion deutlicher als in anderen Bereichen hervortreten. Was sich bei einer solchen Behandlung an Einsichten gewinnen läßt, kann zudem in wesentlichen Teilen in andere mit einer Skalenkonstruktion befaßte Bereiche der Psychologie übertragen werden. Insbesondere gilt dies für die unterschiedlichen Weisen, wie sich der Begriff eines Fehlers konzipieren läßt. Der nun folgende Abschnitt stellt weder eine Einführung noch eine Übersicht über Methoden der Skalierung darstellt; hierzu sei auf Luce & Galanter (1963a,b), Roskam (1983), Tack (1983), Falmagne (1985), Luce & Krumhansl (1988) und Drösler (1989) verwiesen.

Verschiedene Arten der Skalenkonstruktion können nach unterschiedlichen Gesichtspunkten klassifiziert werden, etwa danach, ob sie auf ‚direkten‘ oder ‚indirekten‘ Schätzmethoden beruhen, nach der Art der Urteilsaufgabe, danach, ob sie unter dem Aspekt der Identifikation, Klassifikation, Entdeckbarkeit oder Diskriminierbarkeit von Reizen betrachtet werden, nach der Art der Vorstellungen über die interne Reizrepräsentation (**random vs. constant utility model**) oder nach der Dimensionalität der Repräsentation. Diese und andere Klassifikationen kreuzen sich in vielfältiger Weise, und eine einheitliche Systematik psychophysikalischer Verfahren gibt es nicht. Für unsere Zwecke scheint folgende abstrakte Einteilung zweckmäßig zu sein:

- a) lokale vs. globale Verfahren,
- b) probabilistische vs. deterministische Verfahren,
- c) Verfahren, die meßtheoretisch formuliert sind vs. solche, für die dies nicht gilt.

Die Einteilung ‚lokal vs. global‘ bezieht sich darauf, ob sich die Gesamtskala aus Diskriminationsurteilen in den jeweiligen Umgebungen ihrer Elemente zusammengesetzt oder ob die strukturellen Annahmen, auf denen die Skala beruht, sich auf den Gesamtbereich beziehen. Diese traditionelle Einteilung erscheint bei genauerer begrifflicher Bestimmung der beiden Unterscheidungsmerkmale als nur schwer präzisierbar; tatsächlich wird sich, wie wir noch sehen werden, bei probabilistischen Verfahren diese Grenzziehung verwischen.

Die Einteilung ‚probabilistisch vs. deterministisch‘ unterscheidet Verfahren, deren Grundkonzepte eine Behandlung von Zufallsfluktuationen erlauben von solchen, die dies nicht zulassen.

Die dritte Unterscheidung schließlich bezieht sich darauf, ob sich ein Verfahren so darstellen läßt, daß sich die Skala aus der Gültigkeit qualitativer Bedingungen als numerische Repräsentation im Sinne der Meßtheorie ergibt.

Verdeutlichen wir die Klassifikation anhand einiger Beispiele. Zu den deterministischen, lokalen Verfahren gehören sowohl die Fechnerschen Verfahren, sofern die Schwelle deterministisch bestimmt wird, als auch die Semiordnungen. Zu den deterministischen, globalen Verfahren zählen die auf dem algebraischen additiv-verbundenen Messen beruhenden Vorgehensweisen, ferner das genannte Bisektionsverfahren, Stevens' indirekte Verfahren sowie die nichtmetrische multidimensionale Skalierung. Zu den probabilistischen, lokalen Verfahren sind alle Verfahren zu rechnen, die zur Bestimmung von Diskriminationswahrscheinlichkeiten das Konzept der psychometrischen Funktion (oder ein analoges Konzept) heranziehen, zu den probabilistischen, globalen Verfahren hingegen jene, bei denen eine additiv-verbundene Struktur auf Wahrscheinlichkeiten oder Zufallsvariablen formalisiert wird, sowie probabilistische Varianten der Stevensschen Verfahren und der multidimensionalen Skalierung. Die dritte Kategorie der Klassifikation kreuzt sich mit den andern, da es in jeder der Teilkategorien meßtheoretisch begründete Verfahren gibt. Als Beispiele seien die Krantz-Shepard-Reformulierung der Stevensschen Verfahren genannt, die meßtheoretische Darstellung der multidimensionalen Skalierung und die Schwellenrepräsentationen. Wir werden hierauf noch zurückkommen. Das Attribut ‚meßtheoretisch‘ deutet hauptsächlich eine bestimmte metatheoretische Haltung an; der formale Kern eines Verfahren kann nämlich in unterschiedliche Sichtweisen eingebettet sein, wie die Kontroverse über **conjoint measurement vs. functional measurement** (s. Anderson, 1982) illustriert.

### 3.1 Globales, deterministisches Vorgehen 1:

#### Die Stevensschen Verfahren

Zu globalen, deterministischen Verfahren sind Ratingskalen-Methoden sowie die ‚direkten‘ Skaliervorgänge von Stevens zu rechnen. Von diesen sind die wichtigsten die Größenschätzung (**magnitude estimation**) und der Intermodalvergleich (**cross modality matching**). Da auch die Größenschätzung als ein Vergleich zweier Bereiche betrachtet werden kann, nämlich dem eines physikalischen mit einem Zahlenkontinuum, kann sie als Spezialfall des **cross modality matching** betrachtet werden. Statt der zuvor betrachteten Ordnungsre-



lation liegt diesen Methoden die Relation einer Korrespondenz oder Gleichheit zugrunde.

Bei der Methode der Größenschätzung ist es Aufgabe der Vp, jedem dargebotenen Reiz ein Zahlzeichen zuzuordnen, so daß die Zahlen den durch die Reize hervorgerufenen Empfindungen proportional sind; dabei wird zuvor ein Bezugsreiz durch eine willkürliche numerische Zuordnung ausgezeichnet, um die Response-Skala zu verankern. Bei den Herstellungsmethoden soll zu einem dargebotenen Reiz ein anderer von der Vp eingestellt werden, der eine ‚halbe oder doppelt so starke Empfindung‘ hervorruft wie der Ausgangsreiz. Bei dem Intermodalvergleich schließlich werden zwei Modalitäten zunächst willkürlich miteinander verankert, indem ein Reiz aus einer Modalität vorgegeben wird, der als ‚gleich‘ zu einem bestimmten Reiz einer anderen Modalität definiert wird. Mit dieser Verankerung hat die Vp dann zu beliebigen anderen Reizen der ersten Modalität solche der zweiten anzugeben oder zu adjustieren, die dem Reiz der ersten Modalität wiederum ‚gleich‘ sind. Bereits hier stellen sich eine Reihe konzeptueller Probleme im Zusammenhang mit der Frage, welche psychologischen Prozesse sich in den durch diese Methoden erhobenen Daten widerspiegeln könnten: Verwenden die Vpn Halbierungsinstruktionen so, als würden sie numerische Meßwerte halbieren? Bezeichnen die Zahlen-Äußerungen der Vpn tatsächlich Zahlen, die der Forscher als numerische Skalen im Sinne der Meßtheorie interpretieren kann? Wie läßt sich Stevens' Behauptung, durch Größenschätzung gewonnene Skalen seien Verhältnisskalen, rechtfertigen? Welche interne Empfindungsarithmetik wird bei diesen Verfahren postuliert? Diese Beispiele konzeptueller Fragen deuten bereits an, daß sich hinter der scheinbaren Einfachheit dieser Verfahren zahlreiche metatheoretische Probleme verbergen. Ausführliche theoretische Behandlungen der Stevensschen Methoden finden sich in Luce & Galanter (1963b), Krantz (1972a, b) sowie Baird & Noma (1978), der Kategorienmethoden bei Restle & Greeno (1970, S. 149ff.).

Stevens hatte vor den Hintergrund operationalistischer und behavioristischer Denkweisen nach psychophysikalischen Methoden gesucht, die lediglich auf Reiz-Reaktions-Beziehungen basieren. (Dennoch redet er weiterhin von ‚Empfindungen‘ und ‚Empfindungsskalen‘.) Stevens hielt zudem Skalierverfahren für ungeeignet, die, wie die Fechnerschen Methoden, allein im Bereich der Unsicherheit zwischen Reizen anwendbar sind. Denn diese klassischen Skalierverfahren, die Stevens irreführend „indirekt“ nannte und denen er seine „direkten“ Methoden gegenüberstellte, erlauben eine Skalenkonstruktion nur dort, wo eine genaue Unterscheidung zwischen den Reizen nicht möglich ist; aus der Perspektive deterministischer Modelle macht sich dies als Inkonsistenz der Urteile bemerkbar. Der Fechnersche (wie auch der Thurstonesche) Ansatz beruhe darauf, „to transform variabilities, discriminial dispersions, or confusions into units of measure“, und behaupte, „that all we can know about

magnitude is what confusion tells us“ (Stevens, 1957, S. 154). Die Möglichkeit, innerhalb probabilistischer Modelle eine psychophysikalische Skala erstellen zu können, wurde von Stevens also ebenso in Abrede gestellt wie die Fechnersche Annahme, daß jeder ebenmerkliche Zuwachs subjektiv gleich groß ist, daß also gleichen Reizverhältnissen gleiche Empfindungsdifferenzen entsprechen.

Diese Kritik des Fechnerschen Ansatzes findet sich bereits in klassischen Arbeiten des vergangenen Jahrhunderts. Techniken einer ‚globalen‘ Skalenkonstruktion, die nicht auf Urteilen über ‚lokale‘ Diskriminierbarkeit beruhen und im Stevensschen Sinne „direkt“ sind, wurden von Merkel (1888) entwickelt, etwa die „Methode der doppelten Reize“. Brentano (1874/1924, S. 99f.) vertrat die Auffassung, daß empirische Daten eher nahelegen anzunehmen, daß gleichen Reizverhältnissen gleiche Empfindungsverhältnisse entsprechen. Plateau (1872, S. 383) hatte diese Annahme als eine sinnvolle a priori Annahme diskutiert, die er dadurch gestützt sah, daß Helligkeitsverhältnisse unter Beleuchtungsänderung konstant bleiben. Diese Annahme führt, worauf Fechner (1877, S.24ff.) hinweist, auf das von Plateau (1872) postulierte Potenzgesetz zwischen Reiz  $x$  und Empfindung  $\Psi(x)$ :

$$\Psi(x) = c \cdot x^r \quad (c, r > 0). \quad (3.1)$$

Empirisch konnte für viele eindimensionale Reizkontinua, die sog. prothetischen Kontinua (Stevens, 1957) gezeigt werden, daß sich die mittlere Größenschätzung  $Y(x)$  durch eine Potenzfunktion der physikalischen Reizintensität gut approximieren läßt und daß sich zudem im Intermodalvergleich der Exponent aus dem Verhältnis der unabhängig gewonnenen Exponenten für die beiden einzelnen Dimensionen vorhersagen läßt, d. h. mit  $\Psi_1(x) = c \cdot x^r$  und  $\Psi_2(y) = c' \cdot y^{r'}$  ergibt sich aus dem Vergleich  $\Psi_1(x) = \Psi_2(y)$ , daß  $y = (c/c')^{1/r'} \cdot x^{r/r'}$ . (Prothetische Kontinua sind solche, die sich, wie etwa Lautstärke und Helligkeit, nur auf die Intensität beziehen, während die sog. metathetischen Kontinua auf Qualitäten wie Tonhöhe, Farbton oder Geruch Bezug nehmen.)

Die Stevensschen Verfahren können als deterministisch betrachtet werden, da die Urteile der  $V_p$  unmittelbar als Skalenwerte der psychophysikalischen Skala angesehen werden und üblicherweise das Fehlerproblem nicht explizit gestellt wird; vielmehr wird hier als ‚Fehler‘ indirekt das definiert, was durch Mittelung über  $V_{pn}$  und durch eine Kleinste-Quadrate-Anpassung einer Geraden in log-log-Koordinaten eliminiert wird. Die psychologische Skala ergibt sich als eine Transformation der physikalischen Skala, zu der sie gleichsam ein internes Analogon ist. Die Fehlerbehandlung folgt daher einer physikalischen Fehlerbehandlung, ohne daß freilich wie in der Physik Kriterien darüber vorliegen, was als akzeptabel anzusehen ist und was nicht. Auf diesen Punkt werden wir später zurückkommen.

Die Methoden der Größenschätzung und des Intermodalvergleichs lassen sich meßtheoretisch reformulieren, so daß sich die Skala als Konsequenz der Gültigkeit qualitativer Bedingungen ergibt: In der Krantz-Shepard-Theorie (Krantz, 1972a; s.a. Tack, 1983), die davon ausgeht, daß alle Urteile implizit oder explizit relational sind, wurden die idealisierten empirischen Regularitäten, die sich in empirischen Studien mit diesen Methoden zeigten, herangezogen, um aus ihnen die Existenz und die Art der numerischen Repräsentation abzuleiten. Luce (1990) erhält die Potenzfunktion als Repräsentation in nicht-relationaler Weise durch die Forderung, daß die Korrespondenz- bzw. Gleichheitsrelation invariant unter bestimmten Paaren von Translationen der beiden beteiligten physikalischen Strukturen ist. Qualitative Bedingungen für die Größenschätzung und den Intermodalvergleich, deren Gültigkeit Voraussetzung für den Erhalt einer Verhältnisskala ist, formuliert Narens (1993).

Die Theorien von Krantz-Shepard und Luce sind der Natur ihrer Grundkonzepte nach deterministisch und liefern keine Kriterien zur Bewertung von Zufallsfluktuationen der Gleichheitsurteile. Während bei den klassischen Stevensschen Verfahren der ‚Fehler‘ in geometrisch-numerischer Weise als Abweichung von der postulierten numerischen Funktion definiert wird, werden beim meßtheoretischen Vorgehen als ‚Fehler‘ alle auftretenden Arten von Verletzungen der qualitativen Bedingungen angesehen.

Ein wichtiger Unterschied in der Forschungsperspektive zwischen den Stevenschen Verfahren und ihren meßtheoretischen Reformulierungen sei abschließend noch einmal hervorgehoben: Es charakterisiert die meßtheoretische Perspektive, daß sie das qualitative Testen eines Modells in den Mittelpunkt stellt und eine konkrete numerische Skala nur als Nebenprodukt der Modellgültigkeit betrachtet. Nicht die numerische Skala als solche, sondern der strukturelle Gehalt des Modelles steht hier im Vordergrund.

### 3.2 Globales, deterministisches Vorgehen II: Die additiv-verbundene Messung

Das Vorgehen bei der additiv-verbundenen Messung wollen wir schematisch an einem idealisierten Beispiel der Untersuchung binauraler Lautstärke betrachten. Dabei werden dem linken und rechten Ohr simultan Töne bestimmter Intensität dargeboten. Die Menge der Reize für das linke Ohr sei mit  $A = \{a, b, c, \dots\}$  bezeichnet, die für das rechte Ohr mit  $P = \{p, q, r, \dots\}$ , wobei die Elemente jeweils 1000 Hz-Sinustönen darstellen, die sich nur in der Intensität unterscheiden. Aufgabe der  $V_p$  ist es, (etwa in einem Paarvergleichsdesign) anzugeben, welche von zwei dargebotenen Kombinationen (a,p), (b,q) lauter ist. Die Urteile haben also die Form  $(a,p) > (b,q)$ . Der  $V_p$  werden alle Paare

an Kombinationen aus  $A \times P$  dargeboten, und aus diesen paarweisen Urteilen wird eine verbundene und transitive Ordnungsrelation auf  $A \times P$  erstellt. Für die erhaltene Ordnung auf  $A \times P$  kann sodann überprüft werden, ob sie die qualitativen Bedingungen einer additiv-verbundenen Struktur erfüllt. Wenn diese Bedingungen, die oft als Axiome bezeichnet werden, erfüllt sind, so folgt mathematisch, daß es eine additive Repräsentation der Daten gibt, d.h. reellwertige Funktionen  $f$  auf  $A$  und  $g$  auf  $P$  mit

$$,p) \succcurlyeq (b,q) \Leftrightarrow f(a) + g(p) \geq f(b) + g(q). \quad (3.2)$$

Unter den Axiomen einer additiv-verbundenen Struktur spielt die Bedingung der Doppelkürzbarkeit eine wichtige Rolle; diese Bedingung lautet:

Wenn für beliebige  $a,b,c \in A$  und  $p,q,r \in P$  gilt  $(c,p) \succcurlyeq (b,r)$  und  $(a,r) \succcurlyeq (c,q)$ , so muß auch  $(a,p) \succcurlyeq (b,q)$  gelten.

Die Wahrscheinlichkeit, daß eine Ordnung, welche die Unabhängigkeit und die Doppelkürzbarkeit erfüllt, keine additive Repräsentation hat, ist nämlich sehr gering (Arbuckle & Larimer, 1976; McClelland, 1977), so daß diese beiden Bedingungen bereits eine starke Modellrestriktion darstellen.

Zumeist will man sich nicht mit dem Nachweis der Existenz einer additiven Repräsentation begnügen, sondern eine solche konkret erstellen. Erfüllen die Daten das Modell einer additiv-verbundenen Struktur fehlerfrei, so kann die Bestimmung einer Skala über das Lösen eines entsprechenden Systems linearer Ungleichungen erfolgen, wodurch sich der gesamte Lösungsraum, d.h. die Menge der möglichen additiven Repräsentationen, bestimmen läßt (McClelland & Coombs, 1975). Die Größe des Lösungsraumes hängt von der Größe der Grundmengen ab und konvergiert mit steigender Anzahl von Ungleichungen gegen die durch Intervallskalentransformationen beschriebene Lösungsmenge, d.h. alle Skalen der Lösungsmenge gehen durch Intervallskalen-Transformationen auseinander hervor. Bei einem größeren Lösungsraum kann man nach gewissen Kriterien eine repräsentative Lösung auswählen.

Erweisen sich die Daten hinsichtlich des obigen Modells als fehlerhaft, so sind zwei Betrachtungsweisen möglich: Entweder man geht zu einem probabilistischen Modell über, worauf wir noch zu sprechen kommen werden, oder man behandelt den Fehler in geometrisch-statistischer Weise direkt auf der Ebene der numerischen Skala. Dies kann durch Techniken der isotonen Regression (s. Robertson et al., 1988) geschehen, durch die sich additive Repräsentationen  $f$  und  $g$  bestimmen lassen, die in einem bestimmten Sinne, nämlich der Minimierung des sog. Stress-Index, möglichst nahe an den erhaltenen Daten liegen. Hat man durch ein geeignetes Verfahren Skalen  $f$  und  $g$  bestimmt, so betrachtet man sie als psychophysikalische Skalen für das Attribut ‚binaurale Lautstärke‘.

In einem weiteren Schritt kann man die Frage nach der Beziehung dieser Skalen zu der zugrunde liegenden physikalischen Skala, hier etwa der des Schalldrucks oder der Schallintensität der verwendeten Reize, stellen. So untersuchten Levelt, Riemersma & Bunt (1972), ob sich die Skalen  $f$  und  $g$  als Potenzfunktionen der physikalischen Skala ergeben.

Das obige Modell einer additiv-verbundenen Struktur ist in rein algebraischer Weise formuliert und läßt die Zufallsschwankungen empirischer Daten außer acht. Will man diese unmittelbar in die Formulierung des Modells mit einbeziehen, so kann man zu analogen probabilistischen Modellen übergehen, die wir als globale, probabilistische Verfahren klassifizieren.

### 3.3 Lokales, deterministisches Vorgehen

Als Beispiel wollen wir Verfahren und Repräsentationen anführen, denen ein nicht-probabilistisches Schwellenkonzept zugrunde liegt.

Betrachten wir zunächst die vierstellige Relation  $\succcurlyeq$  in einer additiv-verbundenen Struktur  $(A \times P, \succcurlyeq)$ ; diese wird ebenso als transitiv angenommen wie die entsprechenden zweistelligen Relationen  $\succcurlyeq$  einer einfachen Ordnungsstruktur  $(A, \succcurlyeq)$  oder einer extensiven Struktur  $(A, 0, \succcurlyeq)$ . In empirischen Daten können aber Intransitivitäten der folgenden Arten auftreten: (i)  $a \sim b$  und  $b \sim c$ , jedoch  $a > c$ , (ii)  $a > b$  und  $b > c$ , jedoch  $c > a$ . Will man diese Intransitivitäten nicht lediglich auf Seiten der numerischen Repräsentation durch einen numerisch-geometrisch inspirierten Fehlerbegriff behandeln, sondern zu einem Bestandteil der qualitativen Charakterisierung des Modells machen, lassen sich verschiedene Wege beschreiten. Man kann die relationale Struktur in Termini von Wahrscheinlichkeiten oder von geeigneten Statistiken von Zufallsvariablen beschreiben oder man kann die Intransitivitäten, insbesondere des Typs (ii), als Hinweise auf eine ‚Mehrdimensionalität‘ der Struktur betrachten und zu entsprechenden Formulierungen der Struktur übergehen; drittens schließlich kann man die Anforderungen an die Ordnungsrelation selbst modifizieren, indem man diese, wie im Fall der Semiordnungen, in eine transitive strenge Ordnung und eine intransitive und symmetrische Indifferenzrelation zerlegt, wobei ein Indifferenzintervall niemals größer als ein Ordnungsintervall sein kann. Im letzten Fall erhält man die übliche numerische Repräsentation, nur daß nun die Indifferenzrelation einem Schwellenbereich entspricht. Eine formale meßtheoretische Behandlung aller drei Vorgehensweise findet sich bei Suppes et al. (1989, Kap. 16 & 17). Ersetzt man etwa bei der additiv-verbundenen Messung die Relation  $\succcurlyeq$  durch eine Semiordnung  $>_m$ , so erhält man folgende Repräsentation:

$$(a,p) >_m (b,q) \Leftrightarrow f(a) + g(p) \succcurlyeq f(b) + g(q) + \delta \quad (\delta > 0).$$

Ein besonders interessanter Fall ergibt sich, wenn auf einer extensiven Struktur  $(A, 0, \geq)$  zusätzlich eine psychologische Ordnungsrelation  $>_m$  definiert wird durch:

$$a >_m b \Leftrightarrow \text{„a ist **merklich** größer (z.B. heller, lauter, länger, schwerer) als b“}.$$

Diese Relation  $>_m$ , die als asymmetrisch, transitiv und irreflexiv angenommen wird, erlaubt für jede additive Repräsentation  $f$  der extensiven Struktur eine Schwellenrepräsentation der Art

$$a >_m b \Leftrightarrow f(a) > f(b) + \Delta_f(b). \quad (3.3)$$

Die Gültigkeit des Webersehen Gesetzes läßt sich nun durch bestimmte qualitative Bedingungen charakterisieren (s. Narens, 1985, S. 89), so insbesondere durch die Additivität des qualitativen Schwellenoperators  $J$ , wobei

$$J(a) = \sup\{b \mid b \circ a \not>_m a\}.$$

$J(a)$  bezeichnet das größte physikalische ‚Inkrement‘, daß zu  $a$  ‚hinzuaddiert‘ von  $a$  nicht merklich unterscheidbar ist. Additivität bedeutet dann gerade:

$$J(a \circ b) = J(a) + J(b).$$

Hieraus folgt die Additivität des entsprechenden numerischen Terms 4; es gilt dann nämlich  $\Delta_f(a) = k \cdot f(a)$  ( $k > 0$ ). Die Schwellenrepräsentation (3) geht in diesem Fall über in

$$a >_m b \Leftrightarrow f(a) > f(b) + k \cdot f(b) \text{ bzw. } f(b)(1 + k).$$

Mit  $u = \log f$  und  $d = \log(1 + k)$  erhält man dann die Fechnersche Repräsentation

$$a >_m b \Leftrightarrow u(a) > u(b) + d,$$

in der alle ebenmerklichen Unterschiede denselben numerischen Wert erhalten. Diese Repräsentation mit konstantem Schwellenwert ist ein formales Analogon zu Fechners Annahme, daß gleichen Reizverhältnissen gleiche Empfindungsdifferenzen entsprechen.

Ausführliche Behandlungen von Schwellenrepräsentationen finden sich bei Luce & Galanter (1963a, S.201ff.), Pfanzagl (1971, S. 176ff.) und Suppes et al. (1989, S. 353ff.).

### 3.4 Lokales, probabilistisches Vorgehen

Die bisher betrachteten Modelle gehen davon aus, daß die Beziehungen zwischen Reiz, ‚Empfindung‘ und Urteil in den betrachteten Ausschnitten idealiter deterministisch sind. Inkonsistenzen zwischen Daten und Theorie, insbesondere Urteilsfluktuationen, werden also im Sinne externer Fehlerkonzepte behandelt. Probabilistische Modelle nehmen demgegenüber gerade Charakteristiken der Variabilität zum Ausgangspunkt. ‚Fehler‘ im Sinne dieser Variabilität wird hier als wesentlicher Theoriebestandteil, d.h. als theorieintern, angesehen. Bei einem solchen Vorgehen gibt es konzeptuell wie formal sehr unterschiedliche Wege, den Begriff der Variabilität in einem Modell zu fassen und Skalen als Parameter probabilistischer Modells zu erhalten. Beispielsweise kann man die interne Repräsentation eines physikalischen Reizes selbst als eine Zufallsvariable auffassen und erhält dann die sog. **random utility-Modelle**, die ihren Namen analogen nutzentheoretischen Modellen verdanken und von denen das Thurstone-Modell das bekannteste ist. Die ihnen zugrunde liegenden abstrakten Prinzipien lassen sich zur Formulierung einer modelltheoretischen ‚Probabilisierung‘ von Konzepten der abstrakten Meßtheorie heranziehen (Heyer & Niederée, 1992). Den **constant utility-Modellen** in der Psychophysik liegt demgegenüber das Bild zugrunde, daß Reizen in deterministischer Weise jeweils eine Empfindung zugeordnet ist, während das Urteil von Personen Fluktuationen unterworfen ist. Beide Klassen von Modellen können jedoch auch unabhängig von diesen Bildern rein formal betrachtet und verglichen werden (siehe die ausführlichen Darstellungen bei Luce & Suppes, 1965, Colonius, 1984, sowie Falmagne, 1985).

Tatsächlich nahm bereits Fechner (1860, I, S. 104ff.) an, daß sich die Werte bei der Schwellenbestimmung als eine Zufallsvariable auffassen und bestimmen lassen. Diese Idee findet in Thurstones Modell ihre Weiterführung und explizite formale Ausgestaltung: Der durch einen Reiz ausgelöste interne Effekt kann durch eine Zahl beschrieben werden, die jedoch von Darbietung zu Darbietung variieren kann. Die interne Repräsentation eines Reizes  $x$  ist also durch Zufallsvariablen  $U_x$ , welche die Verteilung der zugehörigen „Diskriminanzprozesse“ darstellt, beschreibbar. Diese Zufallsvariablen sollen, so nimmt Thurstone an, Normalverteilungen sein. Die Entscheidungsregel lautet somit: Ein Reiz  $x$  wird von einer  $V_p$  genau dann als ‚größer‘ als ein Reiz  $y$  beurteilt, wenn der erhaltene Wert der Zufallsvariablen  $U_x$  größer als der Wert von  $U_y$ . Die Wahrscheinlichkeit  $p(x,y)$  für dieses Ereignis ist also

$$p(x,y) = \Pr[U_x \geq U_y] = \Pr[U_x - U_y \geq 0] = \Phi[(\mu_x - \mu_y)/\sigma_{x-y}], \quad (3.4)$$

wobei  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung ist. Der „modale Diskriminanzprozeß“  $\mu_x$  bzw.  $\mu_x$  kann somit als Skalenwert einer internen Skala  $S$  betrachtet werden und die Varianz als Maß für die Größe der Unter-

schiedsschwelle. Die klassische Frage nach der Beziehung zwischen physikalischem Reiz und Empfindung übersetzt sich dann in die Frage nach der Abhängigkeit der entsprechenden Erwartungswerte einer Repräsentation von den physikalischen Reizwerten. Zwischen dem Erwartungswert dieser Zufallsvariablen und ihrer Varianz gibt es jedoch keine zwingende Beziehung. Thurstone wollte, wie später Stevens, sein Modell nicht an die Fechnersche Annahme binden, daß alle Unterschiedsschwellen durch einen konstanten Wert repräsentiert werden. Lediglich in dem als „Case V“ bezeichneten Spezialfall seines Modells nimmt er alle Varianzen als gleich an; in diesem Fall vereinfacht sich das Modell zu  $p(x,y) = \Phi[S(x) - S(y)]$ . Modelle der Art

$$p(x,y) = D[u(x) - u(y)] \quad (3.5)$$

mit beliebiger Verteilungsfunktionen  $D$  werden als Fechner-Modelle bezeichnet (Luce & Suppes, 1965, S. 334; Falmagne, 1985. Die Verteilungsfunktion  $D$  hat hier die Rolle der psychometrischen Funktion, durch die Differenzen auf der psychophysikalischen Skala in Dominanzwahrscheinlichkeiten übersetzt werden.

Mit dieser Modellgleichung bietet sich eine Möglichkeit, einen probabilistischen Ansatz mit dem algebraischen Konzept der additiv-verbundenen Messung zu verknüpfen, indem man definiert:

$$(x,y) \geq (v,w) \Leftrightarrow p(x,y) \geq p(v,w).$$

Hierbei ist vorauszusetzen, daß die Dominanzwahrscheinlichkeiten stets von 0 und 1 verschieden sind. Die Zufallsfluktuationen werden hier gleichsam zum strukturellen Kern der Theorie, indem man als eine Skala gerade das definiert, was den Fehler regularisiert (ein ähnlicher Gedanke findet sich in Levine, 1970). Dies wird besonders deutlich, wenn man beachtet, daß hier zu einer gegebenen Familie von reell-wertigen Zufallsvariablen eine Transformation gesucht wird, die die Zufallsvariablen in solche überführt, die bis auf Verschiebungen um eine additive Konstante identisch verteilt sind. Diese Transformation der physikalischen Reizskala ergibt dann gerade die gesuchte psychophysikalische Skala  $u$ , welche dann die Eigenschaft hat, die psychometrischen Funktionen zu parallelisieren. An diesem Beispiel wird bereits die Problematik der Unterscheidung ‚lokal‘ vs. ‚global‘ erkennbar.

Läßt sich in einem Fechner-Modell  $D$  wiederum als Differenzverteilung zweier unabhängiger und identisch verteilter, stetiger Zufallsvariablen schreiben, so erhält man ein **random utility-Modell**. Wählt man für  $D$  die logistische Funktion, so ist für Paarvergleichsmodelle die Verteilung der Diskriminanzprozesse nicht mehr eindeutig festgelegt (vgl. Luce & Krumhansl, 1989, S. 27). Eine empirische Entscheidung für eine bestimmte Verteilungsfunktion  $D$  läßt sich in der Regel nicht treffen, doch läßt sich in Analogie zu bestimmten



neurophysiologischen Vorstellungen über die Art der Aggregation unabhängiger und identisch verteilter Signale die Verteilungswahl aus entsprechenden Grenzwertsätzen motivieren (Wandell & Luce, 1978; vgl. Tack, 1983). Weitere Aspekte von Thurstone-Modellen werden in Luce (1977) und Suppes et al. (1989, S. 421ff.) behandelt.

An dieser Stelle müssen wir auf die schon angedeuteten Probleme der Unterscheidung lokaler und globaler Verfahren zurückkommen. Das oben beschriebene Vorgehen, die Skala über die psychometrische Funktion zu bestimmen, zeigt bereits, wie im probabilistischen Fall die Grenze zwischen ‚lokal‘ und ‚global‘ verschwimmt. Die psychometrische Funktion wird nämlich idealisiert als eine Funktion angesehen, die über dem gesamten betrachteten, d.h. über dem globalen, Reizbereich stets ungleich 0 oder 1 ist; folglich könnte theoretisch die Skalenkonstruktion auf einem fest gewählten Reiz basieren, durch dessen Vergleich mit allen anderen Reizen sich die globale Skala ergäbe. Die psychometrische Funktion gibt gerade an, wie sich das ‚Lokale‘ zu ‚Globalem‘ verbindet. Auf einer so weitgehenden Idealisierung wird man aber die Skalenkonstruktion nicht gründen wollen, und so nimmt man an, daß jeder Reiz der Skala Vergleichsreiz für die Reize seiner Umgebung ist, d.h., daß es eine Familie psychometrischer Funktionen gibt. Nun benötigt man aber eine (meist nicht explizit ausgesprochene) Miniatur-Theorie darüber, wie Schwellenverhalten mit ‚globalen‘ Urteilen zusammenhängt. Diese Miniatur-Theorie ist im Falle der durch Gleichung (5) beschriebenen Fechner-Modelle in der Forderung versteckt, daß auf der gesuchten Skala  $u$  alle psychometrischen Funktionen parallel sind, d.h. durch Verschiebung auf der Abszisse auseinander hervorgehen. Der empirische Gehalt dieser postulierten theoretischen Beziehung zwischen einer stückweise aus lokalen Urteilen erstellten Skala und Urteilen über globale Skalenbereiche wird an folgendem Beispiel deutlich: Eine Skala, die nach Fechners Vorschlag durch Kumulieren von Unterschiedsschwellen (über der absoluten Schwelle) erzeugt wird, führt (unter bestimmten Annahmen) zu einer logarithmischen psychophysikalischen Funktion. Die durch Logarithmieren der physikalischen Werte bestimmte psychophysikalische Skala hat die Eigenschaft, daß physikalisch gleichen Reizverhältnissen gleiche ‚Empfindungsdifferenzen‘ entsprechen. Die in der Akustik verwendete Dezibel-Skala (dB) der ‚Lautstärke‘ ist eine solche Skala, da sie als Logarithmus des physikalischen Schalldrucks (bezogen auf den Schwellenwert) definiert ist. Gleiche Abstände auf der Dezibel-Skala sollen also perzeptuell auch gleich erscheinen. Ein 1000 Hz-Sinuston von 50 dB wird aber nicht als halb so laut empfunden wie ein solcher von 100 dB. Die postulierte Beziehung zwischen lokalem und globalem Urteilsverhalten ist hier also nicht erfüllt und ‚lokaler‘ und ‚globaler‘ ‚Empfindungs‘-Begriff fallen auseinander.

Auf eine weitere Schwierigkeit in der Unterscheidung von ‚global‘ und ‚lokal‘ werden wir am Ende des folgenden Abschnitts zu sprechen kommen.

### 3.5 Globale, probabilistische Modelle

Diese Kategorie soll an zwei Ansätzen von Falmagne veranschaulicht werden, die versuchen, im Rahmen der oben behandelten additiv-verbundenen Strukturen Skalen unter Berücksichtigung von Zufallsfluktuationen zu konstruieren.

Die erste Möglichkeit basiert darauf, ein algebraisches Meßkonzept dadurch zu probabilisieren, daß die zugrunde gelegten Relationen unter Bezugnahme auf Urteilswahrscheinlichkeiten definiert werden. Anders als im Fall der Fechner-Struktur gehen wir dabei nicht von einfachen Dominanzwahrscheinlichkeiten aus, sondern von einer vierstelligen Dominanzrelation. Dabei wird die Wahrscheinlichkeit, die Reizkombination  $(a,p)$  als größer zu beurteilen als die Kombination  $(b,q)$ , mit  $p[(a,p),(b,q)]$  bezeichnet. Eine Modellgleichung ergibt sich durch folgende Forderung:

$$p[(a,p),(b,q)] \leq 0.5 \Leftrightarrow f(a) + g(p) \leq f(b) + g(q). \quad (3.6)$$

Die Relation  $\langle A \times P, \succ_{0.5} \rangle$  läßt sich nun definieren durch

$$(a,p) \succ_{0.5} (b,q) \Leftrightarrow p[(a,p),(b,q)] > 0.5.$$

Auf diese Weise können alle Axiome der Struktur in Aussagen über Wahrscheinlichkeiten übersetzt werden, etwa die Transitivität der Ordnung in die **sof. schwache stochastische Transitivität**. Diese Übersetzung der probabilistischen in eine algebraische Struktur ist jedoch sehr grob; zumeist nimmt man darüber hinaus an oder wünscht, daß Unterschiede in den Dominanzwahrscheinlichkeiten Differenzen in den Skalenwerten widerspiegeln. Man kann daher obige Modellgleichung verschärfen zu

$$p[(a,p),(b,q)] = F[f(a) + g(p), f(b) + g(q)] \quad (3.7)$$

wobei  $F$  streng monoton steigend in der ersten und streng monoton fallend in der zweiten Variablen ist. Ein spezieller Fall hiervon ist wiederum

$$p[(a,p),(b,q)] = D[f(a) + g(p) - f(b) - g(q)] \quad (3.8)$$

mit  $D$  streng monoton steigend. Mit dieser Modellgleichung ist für den Fall einer Komponentenstruktur  $A \times P$  der Anschluß an die Modellgleichung (3.5) der Fechner-Modelle hergestellt.

Für eine detaillierte Behandlung dieser Klasse von Modellen sowie von Bedingungen probabilistischer Konsistenz siehe Luce & Suppes (1965), Falmagne (1979), Roberts (1979), Colonius (1984) sowie Suppes et al. (1989, S.388ff.).

In dem zweiten, von Falmagne (1976) vorgeschlagenen Ansatz zur Behandlung von Zufallsfluktuationen werden statt Urteilen der Art  $(a,p) \succ (b,q)$ , Gleichheitsurteile der Art  $(a,p) \sim (b,q)$  betrachtet. Hier wäre etwa an Ein-

Stellungsmethoden zu denken, bei welchen die von der  $V_p$  manipulierte Variable als Zufallsvariable aufgefaßt wird: Wird bei vorgegebenen  $a, p$  und  $q$  ein Gleichheitsurteil durch Einstellen eines Wertes  $b$  abgegeben, so sei die Verteilung der entsprechenden Zufallsvariablen mit  $B(a, p; q)$  bezeichnet. (Hierbei wird implizit jede Reizkomponente mit ihrem zugeordneten reellen physikalischen Skalenwert identifiziert.) Postuliert man die Gültigkeit einer additiven Kompositionsregel für die Effekte der Variablen  $a, p$  und  $q$  und nimmt man zudem ein Fehlermodell mit (symmetrischer) Fehlerverteilung  $\epsilon(a, p; q)$  an, so gelangt man zu der Modellgleichung

$$f(\mathbf{B}(a, p; q)) = f(a) + g(p) - g(q) + \epsilon(a, p; q). \quad (3.9)$$

Den Anschluß an die Modellgleichung (3.2) für die additiv-verbundene Messung gewinnt man, wenn man bei dem betrachteten Gleichheitsurteil als Wert für  $b$  eine Statistik der Zufallsvariablen  $B(a, p; q)$  wählt, die invariant unter monotonen Transformationen ist. Falmagne (1976) benutzte zur Formulierung seines „random conjoint measurement“ den Median  $Md$ . Da für jede streng monotone Transformation  $h$  gilt

$$h(Md[\mathbf{B}(a, p; q)]) = Md[h(\mathbf{B}(a, p; q))]$$

und da für die als symmetrisch angenommene Fehlerverteilung

$$Md[\epsilon(a, p; q)] = 0$$

gilt, erhält man aus (9)

$$Md[\mathbf{B}(a, p; q)] = f^{-1}[f(a) + g(p) - g(q)]. \quad (3.10)$$

Mit  $b = Md[\mathbf{B}(a, p; q)]$  ergibt sich dann gerade Gleichung (3.2).

Die Bedingung (3.10) führt zu einer Reihe von qualitativen testbaren Bedingungen, die sich nur auf die experimentell erhobenen Werte  $Md[\mathbf{B}(a, p; q)]$  beziehen. So folgt aus (3.10) etwa die Bedingung

Diese Bedingung ist das Analogon zur Doppelkürzbarkeitsbedingung, und sie läßt sich durch einen Median-Test statistisch prüfen, wie Falmagne (1976) zeigt. (Gigerenzer und Strube (1983) untersuchten mit diesem Modell noch einmal die Hypothese binauraler Additivität und fanden sie, anders als Levell et al. (1972), nicht bestätigt.)

Die beiden Ansätze von Falmagne lassen sich formal wieder in Beziehung setzen. Unter naheliegenden Zusatzannahmen ergibt sich nämlich  $(a, p) \succcurlyeq_{0,5} (\mathbf{B}(a, p; q), q)$ .

Warum nun, und damit kommen wir noch einmal auf die Problematik ‚lokal‘ vs. ‚global‘ zurück, haben wir diese beiden Zugangsweisen als globale klassi-

fiziert? Offensichtlich wird bei diesen Verfahren die Existenz einer Skala durch strukturelle Bedingungen garantiert, die sich auf den gesamten betrachteten Reizbereich beziehen: nämlich gerade durch die Ordnungsrestriktionen einer additiv-verbundenen Struktur. Besonders deutlich wird dies bei der erstgenannten Zugangsweise: Mit der durch die Bedingung

$$(a,p) \geq (b,q) \Leftrightarrow p[(a,x),(b,q)] > 0.5.$$

festgelegten Klassifikation der Dominanzwahrscheinlichkeiten wird die psychometrische Funktion gleichsam wieder deterministisch betrachtet, und die Modellrestriktionen beziehen sich lediglich auf die Ordnung der so klassifizierten Dominanzwahrscheinlichkeiten. Man könnte nun einwenden, daß die erhaltene Skala dennoch auf lokalen Informationen beruhe, da die zugrunde liegenden Urteile Diskriminations- bzw. Dominanzurteile sind und somit in einem Konfusionsbereich liegen müssen. Wir wollen daher zwischen den Konzepten ‚lokal‘ und ‚Diskriminationsurteile‘ folgende Unterscheidung treffen: Als lokale Eigenschaften einer Skala seien solche bezeichnet, die sich auf einzelne Elemente der Skala und deren Umgebung im Sinne eines Konfusionsbereiches beziehen. Diskriminationsurteile seien abstrakt durch eine Unterscheidbarkeitsrelation  $>$  beschrieben; diese kann sich wiederum unmittelbar auf die Elemente der Skala beziehen, d.h. eine zweistellige Relation der Art  $a > b$  sein, oder sie kann, wie bei additiv-verbundenen Strukturen, höherstufig sein, also  $(a,x) > (b,y)$ . Würde man eine Skala erstellen wollen, deren Elemente Paare der Art  $(a,x)$  sind, d.h. ist man nicht an den Elementen der vierstelligen Relation selbst interessiert, so könnte man Diskriminationsurteile der Art  $(a,x) > (b,y)$  wieder als ‚lokal‘ bezeichnen. Die üblichen psychophysikalischen Skalen beziehen sich aber, veranlaßt durch die korrespondierenden eindimensionalen physikalischen Größen, gerade auf die einzelnen Reizkomponenten  $a, b, \dots$  der Relation. Folglich ist die probabilistische additiv-verbundene Messung als globales Verfahren anzusehen, das auf komplexeren Diskriminationsurteilen basiert.

Diese Beispiele mögen für eine Veranschaulichung unterschiedlicher Arten der Skalenkonstruktion in der Psychophysik genügen. Weiterführende Darstellungen finden sich in Luce & Galanter (1963a,b), Bock & Jones (1968), Krantz (1972b), Holman & Marley (1974), Luce & Green (1974), Luce (1977) sowie Laming (1986).

### 3.6 Vergleichende Zusammenfassung

Die zuvor genannten meßtheoretischen Zugangsweisen wollen wir noch einmal kurz unter dem Aspekt der Beziehung von theoretischer Struktur und empirischen Daten betrachten.

Eine Auffassungsweise ist, wesentliche algebraische Bedingungen, die eine Klasse psychophysikalischer Struktur charakterisieren, unmittelbar als empirische Hypothesen aufzufassen, indem man die theoretischen Relationen als empirische ansieht. Durch Beschränkung auf ordinale Daten können bei einer geeigneten Verteilung der Datenpunkte Zufallsfluktuationen unbemerkt bleiben in dem Sinne, daß keine Inkonsistenzen zwischen Daten und Theorie auftreten. Treten Inkonsistenzen auf, so gibt es zwei prinzipielle Strategien zu ihrer Behandlung: Man kann sie auf Seiten der numerischer Repräsentation durch ein globales Verletzungsmaß wie den oben genannten Stress-Index ‚bewerten‘. Oder man sucht Verletzungen auf der Ebene qualitativer (meist universeller) Axiome (oder auch deren Folgerungen) zu beschreiben. Dieses Vorgehen verknüpft sich häufig mit dem Anliegen einer Diagnose systematisch-strukturellen Fehlanpassung zwischen Daten und Theorie. (Die Begrenztheit eines solchen diagnostischen Vorgehens zeigt sich u.a. darin, daß eine bestimmte Fehlanpassung zwischen Daten und Theorie bei äquivalenten Axiomensystemen gleichsam auf verschiedene Axiome projiziert werden kann!) Für die Diagnose und Bewertung der Modellgüte gibt das Auszählen von Axiomverletzungen - ebenso wie der Stress-Index - allenfalls vage Kriterien an die Hand.

Ein anderer Weg, die Beziehung einer meßtheoretischen Struktur zu ihrem empirischen Gegenstandsbereich zu behandeln, liegt darin, bestimmte Arten von strukturellen ‚Abweichungen‘ als theoretisch interessant zu betrachten und sie bereits innerhalb der deterministischen Theorie bei der Formulierung der Axiome zu berücksichtigen. Beispielsweise kann die Ordnungsrelation der Struktur zu einer Semiordnung abgeschwächt werden, um die Intransitivität von Indifferenzen als Theoriebestandteil zu berücksichtigen. Dieses Vorgehen führt zu den oben genannten Schwellenrepräsentationen (vgl. Gleichung (3.3)). Oder die Struktur selbst wird als Struktur mit maximalen Elementen angesehen; bestimmte existentielle Axiome können etwa dadurch verletzt sein, daß maximale Elemente vorliegen, was zu einer entsprechend verfeinerten Axiomatisierung Anlaß geben kann.

Einen Übergang von der deterministischen zu einer probabilistischen Betrachtungsweise ergibt sich, wenn man die Schwelle mittels der psychometrischen Funktion über ein Wahrscheinlichkeitskriterium definiert. Dies entspricht der allgemeinen Strategie deterministische Relationen unter Bezug auf Wahrscheinlichkeiten zu definieren, die sie etwa dem in den Gleichungen (3.6-3.8) beschriebenen Vorgehen von Falmagne (1979) zugrunde liegt. Die Anbindung an die Daten erfolgt hier über relative Häufigkeiten ordinaler Urteile.

Eine andersartige Stochastisierung der Struktur liegt vor, wenn ihre Elemente selbst, wie bei den **random utility-Modellen**, als Zufallsvariablen aufgefaßt werden, über deren Parameter sich wiederum eine Skala definieren läßt. Die

algebraische Theorie wird gleichsam probabilistisch ‚verwackelt‘, wodurch einige der algebraischen Bedingungen zu statistisch testbaren Hypothesen werden (vgl. Iverson, 1991).

Ein sehr allgemeiner Ansatz, auf theoretischem Wege zu einer Probabilisierung deterministischer Meßstrukturen zu gelangen, wurde von Heyer & Niederée (1992; vgl. Luce & Narens, 1993) entwickelt. Er beruht auf dem Konzept ‚probabilistischer Mischungen‘ qualitativer deterministischer Meßstrukturen (z. B. additiv-verbundener Strukturen). Dieser Ansatz verallgemeinert in natürlicher Weise das aus der Nutzentheorie bekannte Konzept des **binary choice system induced by ranking** (vgl. Fishburn, 1992). Für diesen Fall ist bekannt, daß sich für endliche Trägermengen fester Größe die resultierenden ‚Mischungen‘ durch ein endliches System linearer Ungleichungen charakterisieren lassen. Heyer und Niederée zeigen, daß ein entsprechendes Resultat auch im allgemeinen Fall gilt, und setzen das genannte Konzept zu einer Verallgemeinerung des **random utility-Konzeptes** in Beziehung.

Mit diesen Vorgehensweisen ist ein Spektrum theoretischer Konzeptualisierungen von Zufallsfluktuationen skizziert, zwischen dessen Extrempunkten sich die verschiedenen Möglichkeiten, den Zufallsfehler bereits bei der Formulierung der theoretischen Struktur zu berücksichtigen, spannen.

Im folgenden Abschnitt werden wir das Problem der Bestimmung eines ‚Fehler‘-Konzeptes, das zumeist implizit die bisherigen Abschnitte zur Skalenkonstruktion durchzog, aus allgemeinerer Perspektive behandeln.

#### **4. Anmerkungen zum Fehlerbegriff**

Im letzten, eher technischen, Abschnitt haben wir idealtypisch zwei Zugangsweisen zur Behandlung des ‚Fehlerproblems‘ an die Unterscheidung von deterministischen und probabilistischen Modellen geknüpft. Dieser Gegenüberstellung entspricht eine Unterscheidung von ‚Fehlerkonzepten‘, die sich bezogen auf die Theorie als intern bzw. extern auffassen lassen. Einige Überlegungen grundsätzlicher Art seien nun nachgetragen. Dabei werden wir das Augenmerk auch auf die ideengeschichtliche Entwicklung entsprechender Konzepte richten.

Beginnen wir mit theorieexternen Fehlerkonzepten. Das sich dabei ergebende Problem, den ‚eigentlichen‘ Meßwert von störenden ‚Nebeneffekten‘ zu bereinigen, stellte sich nicht erst mit der psychologischen Skalenkonstruktion. Jede Messung, auch die physikalische, ist mit ihm konfrontiert. Was aber als eigentlicher Meßwert und was als Meß- oder Beobachtungsfehler anzusehen ist, kann nur durch den Bezug auf eine Theorie oder eine Vorhersagetechnik

festgelegt werden: Theoriekonstruktion und ‚Fehlertheorie‘ bzw. Techniken der Fehlerbehandlung gehen Hand in Hand.

Bei den theorieexternen Fehlerkonzepten in der Psychophysik wird der Zusammenhang zwischen den betrachteten theoretischen Größen als deterministisch betrachtet, eine Vorgehensweise, die derjenigen der klassischen Physik gleicht. Fehlerkonzepte werden dann eingeführt, um zwischen den beobachteten und theoretischen Größen zu vermitteln (wobei bereits die Einführung des Begriffes ‚theoretisch‘ gleichsam das notwendige Komplement des Begriffes ‚Fehler‘ ist).

Es ist bemerkenswert, daß die Bemühungen um ein Verständnis des ‚Beobachtungsfehlers‘ in der Physik historisch entscheidende Anstöße zur Entwicklung der Psychophysik geliefert haben. Damit lassen sich umgekehrt Teile der Psychophysik vom Standpunkt der Physik aus betrachtet als Bausteine einer physikexternen Fehlertheorie auffassen, ist es doch die menschliche Wahrnehmung, die lange Zeit in der physikalischen Messung das entscheidende letzte Glied des Meßprozesses bildete. Zwar wußte man seit altersher um die Unzuverlässigkeit der Wahrnehmung als eines Erkenntnisinstruments, doch wurde erst im 18. Jahrhundert die Eigengesetzlichkeit dieses Erkenntnisinstruments zum Objekt systematischer wissenschaftlicher Betrachtung. Der Begriff des Beobachtungsfehlers entstand, und mit den Grundlagen seiner formalen Behandlung wurden zugleich die Grundlagen der Wahrscheinlichkeitstheorie gelegt. Um also zuverlässige physikalische Werte zu gewinnen, mußten Einflußfaktoren eliminiert werden, die sich nicht auf physikalische Faktoren zurückführen lassen, sondern sich aus Eigenschaften der menschlichen Wahrnehmung ergeben. Vor allem in der Astronomie spielte die Reliabilität visueller Wahrnehmungsleistungen eine große Rolle: Bei der Helligkeitsmessung von Sternen mußten vorhandene Lichtstärken unterschieden werden und physikalisch gleiche Lichtintensitäten auch als gleich beurteilt werden. So genau sich das Auge, verglichen mit der physikalischen Bestimmung, auch erwies, so zeigten sich doch in systematischer Weise Variabilitäten des Urteils. In seinen Untersuchungen zur Photometrie von 1760 unterschied Lambert drei Fehlerquellen: die unvermeidliche Unbestimmtheit im Urteil des Auges, die Unachtsamkeit des Beobachters und schließlich die Beschaffenheit des Instruments einschließlich anderer, vom Beobachter unabhängiger Umstände. Zudem findet er, daß für die beiden ersten Fehlerarten die Größe des Fehlers mit seiner Auftretenswahrscheinlichkeit zusammenhängt. Dem arithmetischen Mittel der beobachteten Meßwerte kommt, wie Lambert meint, die „größte Wahrscheinlichkeit“ zu. Bernoulli, Gauß und Laplace entwickelten dann eine formale Theorie der Beobachtungsfehler, die in der Herleitung der Normalverteilung und der Entwicklung der Methode der kleinsten Quadrate gipfelte.

Astronomische Beobachtungen führten auch zu einer weiteren Einsicht in die Systematik von Beobachtungsfehlern: zur „persönlichen Gleichung“. Bei astronomischen Beobachtungen nach einem komplexen Urteilsverfahren, an dem gleichzeitig Auge und Ohr beteiligt waren, hatte 1796 der englische Astronom Maskelyne entdeckt, daß sein Assistent ein bestimmtes Ereignis systematisch eine halbe Sekunde später als er berichtete (woraufhin er ihn entließ). Diese interindividuellen Differenzen machte 1820 der Königsberger Astronom Bessel zum Gegenstand einer eigenständigen Untersuchung, die zu dem Ergebnis führte, daß sich zwischen je zwei Beobachtern eine im Mittel relativ konstante Differenz der Beobachtungszeiten feststellen ließ. Dadurch war es möglich, für jeden Beobachter eine individuelle Fehlergleichung, die „persönliche Gleichung“, zu erstellen.

In der Theorie der Beobachtungsfehler gelangte man schließlich zu folgender Klassifikation: (i) Fehler des Meßinstrumentes selbst, (ii) systematische interindividuelle Differenzen („persönliche Gleichung“), (iii) systematische Fehler (etwa wenn Beobachtungen nicht unter ‚Normbedingungen‘ durchgeführt wurden) und (iv) Zufallsfehler, die durch alle unbekanntes Faktoren der Variabilität hervorgerufen werden.

Fechner erkannte die Bedeutung der formalen Methoden der Fehlerbehandlung, wie sie im Zusammenhang mit der Physik und Astronomie entwickelt worden waren, für die Psychophysik. Der für ihn so wichtige theoretische Begriff der Schwelle kann nämlich nur dann zu experimentellen Beobachtungen in Beziehung gesetzt werden, wenn man ein Verfahren angibt, aus den über Replikationen variierenden Urteilen die Schwelle als eigentlichen Meßwert zu gewinnen. Zu genau diesem Zwecke schlägt Fechner (1860, Bd. I, S. 104ff.) vor, bei der Methode der richtigen und falschen Fälle, dem Vorläufer des Paarvergleichsverfahrens, die Gaußsche Fehlerverteilung zu verwenden, die damit ihre erste Anwendung in der Psychologie erfährt. Diese Anwendung hatte freilich bei Fechner den Charakter eines reinen Hilfsmittels bei der Schwellenbestimmung und stand noch nicht in einem Zusammenhang mit der Konstruktion einer psychologischen Skala. Wundt (1908, S. 536) kritisierte daher Fechners Übertragung physikalischer Fehlermethoden in die Psychologie: „Was für den Physiker ein nicht näher zu untersuchendes störendes Moment ist, das wird also für den Psychologen zur Hauptsache; und was für den Physiker die Hauptsache, der endgültige Mittelwert, das ist für den Psychologen zumeist nur ein nebensächliches Orientierungsmittel. In der psychologischen Verwendung jener physikalischen Methoden der Fehlerelimination von ‚Fehlern‘ zu reden, ist daher streng genommen selbst ein Fehler: der physikalische Beobachtungsfehler ist vielmehr, psychologisch betrachtet, eine Abweichung von irgend einem mittleren Verhalten, deren individuelle psychologische Bedingungen das eigentliche Objekt der Untersuchung bilden.“ Für die Konstruktion psychologischer Skalen impliziert dies die Forderung, Variabilitäten



bereits bei der Entwicklung von Theorien, aus denen sich die Skala ergibt, und nicht erst beim eigentlichen Akt der Messung im Sinne externer Fehlerkonzepte Rechnung zu tragen. Die Variabilität wird damit selbst zum Gegenstand der Modellbildung. Dieser Schritt wurde formal 1927 durch Thurstone geleistet: Er verlagerte die Normalverteilung gleichsam von außen nach innen. Mit der Annahme von internen Diskriminanzprozessen, die als normalverteilt angesehen werden, postulierte Thurstone einen expliziten psychologischen Mechanismus zur Erklärung von Urteilsvariabilität. Dies führte zum oben bereits behandelten Konzept der psychometrischen Funktion; für die Psychophysik stellt dieses Konzept eine der wichtigsten theoretischen Errungenschaften zur Behandlung probabilistischer Urteilsfluktuationen dar; zudem fand es in der Testtheorie sein Gegenstück in der Item-Response-Funktion.

Anders als in der Physik ist man also in der psychologischen Skalierung nicht vorrangig an den von allen variablen und konstanten Fehlern bereinigten Mittelwerten einer Messung interessiert, sondern gerade an den Gesetzmäßigkeiten, die der Variabilität von Urteilsprozessen zugrunde liegt. Demzufolge kommt hier probabilistischen Vorgehensweisen eine besondere Bedeutung zu; diese Haltung wird nachdrücklich von Falmagne (1976, S.66) vertreten: „In the behavioral sciences, the only regularities are statistical ones. This erratic nature of behavioral data makes it almost mandatory that the theories be probabilistic.“

So berechtigt eine solche Haltung im Hinblick auf bestimmte Fragestellungen der Psychologie grundsätzlich ist und so bedeutend die Rolle möglicher probabilistischer Modelle sein kann, so viel Zurückhaltung ist auf der anderen Seite geboten im Hinblick auf ‚Routine-Probabilisierungen‘ deterministischer Theorien ohne ausreichenden theoretischen Rückhalt (wenn sich etwa hinter umfassenden wahrscheinlichkeitstheoretischen Betrachtungen der magere strukturelle Gehalt einer Rating-Skala verbirgt). Eine solche spiegelt sich in besonders extremer Weise in dem gelegentlich geäußerten Gedanken einer allgemeinen (gleichsam theorieinvarianten) formalen Behandlung des Fehlerproblems wider. So wurde in den 70er Jahren vorgeschlagen, die Meßstruktur im Rahmen booleschwertiger oder probabilistischer Modelltheorie zu formulieren, wodurch man jeder Aussage einen Wahrheitsgrad zuordnen kann. Dies führt aber, wie Heyer & Mausfeld (1987) zeigten, zu einer Trivialisierung des Fehlerproblems, da eine Aussage genau dann im booleschwertigen oder probabilistischen Sinne gültig ist, wenn sie von jeder Vp deterministisch erfüllt wird.

Oft stellt sich auch die Forderung der Probabilisierung deterministischer Strukturen als ein Reflex auf ein empiristisches Mißverständnis deterministischer Theorien dar. Auch wenn das Folgende nicht als ein grundsätzliches

Plädoyer für ein deterministisches Vorgehen angesehen werden soll, so scheint uns dieses Problem doch eines Exkurses wert.

Einer naiv-empiristischen Lesart zufolge müßten in aller Regel sowohl physikalische Gesetze wie auch die Axiome einer deterministischen Meßstruktur als falsifiziert angesehen werden. Dem liegt ein Mißverständnis zugrunde, welches in meßtheoretischem Zusammenhang seinen Ausdruck bereits in der Verwendung der Bezeichnung **empirische Struktur** findet. Die Relationen, die eine qualitative Struktur im Sinne der Meßtheorie charakterisieren, erweisen sich nämlich bei näherer Betrachtung nicht als Beobachtungsprädikate im naiven Sinne, sondern als theoretische (Dispositions)-Prädikate. Suppes et al. (1989, S. 300) stellen unmißverständlich fest: „A relational statement, such as  $a \geq b$ , is not considered to be the record of a particular record or experiment but is a theoretical assertion inferred from the data.“ Meßtheoretische Axiome als solche sind also keineswegs direkt empirisch überprüfbar. Gemessen an den entsprechenden empirischen Relationen konstituieren sie theoretische Idealisierungen, in denen sich der jeweils erreichte Stand der theoretischen Perspektive niederschlägt. Auf numerischer Seite findet diese theoretische Idealisierung ihre Entsprechung in der Verwendung der reellen Zahlen, mit denen das Bild einer idealiter beliebig genauen Messungen einhergeht; quantitative Vorhersagen aus der Theorie können also mit beliebiger Exaktheit formuliert werden, so daß von vornherein keine perfekte Übereinstimmung zwischen Theorie und Daten zu erwarten ist. Die Anbindung theoretischer Aussagen an Beobachtungsaussagen kann folglich nur durch theoretische Bindeglieder, Normen oder etablierte Bewertungskriterien erfolgen, die festlegen, wie groß und von welcher Art zulässige Spannungen zwischen theoretischen und empirischen Aussagen sein dürfen. Die Beziehung einer Meßstruktur zu experimentellen Beobachtungen spiegelt also das in der Wissenschaftstheorie umfassend diskutierte Problem von Beobachtung und Theorie wider. Vor diesem Hintergrund erweist sich eine Bewertung der Spannung zwischen den qualitativen Axiomen einer Meßstruktur und den entsprechenden experimentellen Beobachtungen als sehr subtiles Problem (für eine tiefgehende konzeptuelle Analyse sei auf Niederée, 1992a, b, verwiesen).

Deterministischer und probabilistischer Betrachtungsweisen in der Forschung sowie Auffassungen darüber, was man als ein geeignetes Experiment ansieht und was ‚Daten‘ sind, verknüpfen sich zusammen mit Bewertungskriterien und Bindeglieder, durch die erst eine theoretische und eine Beobachtungsaussage miteinander verbunden werden können, zu einem komplexen Wirkungsgefüge. Die Bindeglieder zwischen dem Theoriekern und seinem Gegenstandsbereich sind selbst in einem weiteren Sinne Bestandteil einer inhaltlichen Theorie und in ihren paradigmatischen Kontext eingebunden. Ein Theoriebereich besteht nämlich nicht nur aus dem idealisierten Theoriekern, etwa im Sinne der Miniatur-Theorie einer Meßstruktur, sondern er beinhaltet stets

auch die Formulierung zahlreicher Anfangs- und Randbedingungen sowie Hilfshypothesen, durch die erst die Theorie mit einem Experiment in Beziehung gesetzt werden kann. Im Prozeß der Entstehung einer Theorie bildet sich mit der (zumeist impliziten) Formulierung solcher Hilfshypothesen eine Übereinkunft heraus, was als vernünftige Übereinstimmung zwischen Daten und Theorie anzusehen ist und was nicht. Dieser Konsens, der seinen prototypischen Niederschlag in Lehrbuchtabellen der für bestimmte Gesetze ‚typischen‘ Meßwerte findet, muß als wesentlicher Bestandteil des paradigmatischen Kontextes einer Theorie angesehen werden, d.h. „without the tables, the theory would be essentially incomplete“ (Kuhn, 1961, S.36). Solange es indes keine solchen aus der inhaltlichen Theorie und ihrem paradigmatischen Kontext erwachsenden Kriterien darüber gibt, welche Spannung man zwischen theoretischer Aussage und Beobachtungsaussage zu tolerieren gewillt ist, bleibt eine Entscheidung hierüber willkürlich.

Zweck einer Probabilisierung kann nicht sein, die Theorie gegen Verletzungen zu immunisieren und so ihren empirischen Gehalt zu reduzieren. Mit der Stochastisierung einer psychologischen Theorie muß vielmehr das Ziel verbunden sein, eine inhaltliche Theorie über die internen Mechanismen zu formulieren, welche für die Urteilsfluktuationen verantwortlich sind. Eine Stochastisierung, die nicht an hinreichend viele und hinreichend strenge empirische Restriktionen des betrachteten Gegenstandsbereiches angebunden ist, läuft Gefahr, zum Selbstzweck zu werden und die Anbindung an die substantielle Theoriebildung des jeweiligen Gegenstandsbereiches zu verlieren. Auf die Gefahren probabilistischer Modelle hat insbesondere Krantz (1972c) in seiner Diskussion der meßtheoretischen Perspektive hingewiesen. Ganz anders stellt sich die Behandlung von Beobachtungsfehlern dar, wenn sie nicht lediglich im Kontext einer isolierten Skalenkonstruktion erfolgt, sondern wenn die betrachtete Urteilsaufgabe in eine umfassendere inhaltliche Theorie eingebettet ist. Heyer & Mausfeld (1987) diskutieren unter diesem Aspekt zwei Beispiele aus der Psychophysik.

Damit kommen wir resümierend wieder zu der eingangs gemachten Feststellung zurück, daß Fehlertheorie und substantielle Theorie Hand in Hand gehen müssen. So wichtig es ist, eine Vielfalt formaler Techniken zur ‚Fehler‘-Behandlung verfügbar zu haben - und hierzu haben die Psychophysik und die Abstrakte Meßtheorie entscheidende Anstöße geliefert -, so unsinnig ist es, in aprioristischer Weise die Entwicklung einer allgemeinen Fehlertheorie für psychologische Urteilsprozesse zu verlangen. Es gibt nicht das Fehlerproblem in der Psychophysik oder Psychologie, sondern jeder Theoriebereich wird, ist er nur reichhaltig genug, eigene Arten der Fehlerbehandlung hervorbringen. Sieht man als Ziel einer naturwissenschaftlich orientierten Psychologie die Entwicklung reichhaltiger substantiellen Theorien in spezifischen psychologischen Phänomenbereichen, so stellt die Behandlung von Beobachtungs-

fehlern kein eigenständiges Problem mehr dar, wie beispielsweise ein Blick in die verschiedenen Bereichen einer Psychophysik der perzeptuellen Informationsverarbeitung erkennen läßt.

## **5. Zur Rolle von Invarianzkonzepten bei der Konstruktion psychologischer Skalen**

Die durch die oben beschriebene Verfahren der Skalenkonstruktion erstellten Skalen zielen ursprünglich auf eine Messung von ‚Empfindungen‘. Könnte man diese unabhängig messen, ergäbe sich im Falle eindimensionaler physikalischer Reize die Frage nach der Beziehung einer solchen Empfindungsskala  $u$  zu einer vorgegebenen zugehörigen physikalischen Skala  $f$  (unter der stillschweigenden Annahme, daß mit jedem Reize genau eine Empfindung assoziiert ist). Doch ist eine Skalenkonstruktion auf psychologischer Seite unter direkter Bezugnahme auf ‚Empfindungen‘ nicht möglich. Die Konstruktion einer solchen Skala wird sich daher auf gewisse psychologische Relationen beziehen müssen, welche mit dem Urteilsverhalten von Personen in Zusammenhang stehen, und ‚Empfindungen‘ als theoretische Begriffe (in Anlehnung an alltagssprachliche Vorstellungen) rekonstruieren und zu dem Urteilsverhalten in Beziehung setzen. Wir haben oben mehrere Verfahren, die Skala  $u$  zu konstruieren, kennengelernt, die teilweise auf rein psychologischen Relationen beruhen (z. B. Thurstone-Skalen), teilweise auf einem Zusammenspiel psychologischer und physikalischer Relationen (z. B. im Falle des Webersehen Gesetz sowie die Repräsentation einer Grassmann-Struktur). Die Transformation, die den Zusammenhang von physikalischer Skala  $f$  und psychophysikalischer Skala  $u$  beschreibt, bezeichnet man als psychophysikalische Funktion. (Freilich gibt es nicht zu einer Sinnesmodalität ‚die‘ psychophysikalische Funktion, da diese von  $f$  und  $u$  abhängt.) Häufig spiegeln sich in der Form der psychophysikalischen Funktion qualitativ-strukturelle Beziehungen zwischen den zugrunde liegenden psychologischen und physikalischen Relationen wider. In diese Beziehungen kann sich eine neue Einsicht ausdrücken, die über die bei der Skalenkonstruktion bereits vorausgesetzten Prinzipien hinausgeht, wie es in der Regel der Fall ist, wenn die Skala  $u$  lediglich auf psychologischen Relationen beruht. Im anderen Fall kann es geschehen, daß sich hier nur Prinzipien widerspiegeln, welche bereits der Skalenkonstruktion bereits zugrunde gelegt wurden (wie z.B. bei der mittels des Webersehen Gesetz gewonnenen Fechnerschen logarithmischen Skala).

Als besonders aufschlußreich bei der Analyse derartiger qualitativ-strukturellen Beziehungen erweisen sich Konzepte der Invarianz und Symmetrie, welchen auf numerischer Seite die Konzepte **Skalenniveau**, **zulässige Transforma-**

**tion** und **meaningfulness** entsprechen. Hierdurch lassen sich mitunter Aufschlüsse über die Form der psychophysikalischen Funktion gewinnen. Der folgende Abschnitt beschäftigt sich mit diesen Konzepten; alle Betrachtungen werden nur für den technisch einfachsten Fall durchgeführt, damit die grundsätzlichen Perspektiven, soweit sie hier von Interesse sind, deutlich werden werden.

## 5.1 Skalenniveau, zulässige Transformationen und **meaningfulness-Konzepte**

Während der 30er Jahre fand im Umfeld logisch-empiristischer Wissenschaftstheorie eine lebhafte Diskussion um die Rolle klassifikatorischer, komparativer und quantitativer Begriffe für die empirische Theoriebildung statt. Dabei war man besonders daran interessiert, die mit den Begriffen der Messung und der quantitativen Aussage zusammenhängenden konzeptuellen Fragen zu klären. Hierzu gehörten Fragen nach dem empirischen Gehalt quantitativer Aussagen und die Suche nach Kriterien, wann eine quantitative Aussage als sinnvoll oder als sinnlos anzusehen ist. Cohen & Nagel (1934) unterschieden drei Typen numerischer Skalen: Skalen, deren Zahlen nur zur Identifikation empirischer Objekte dienen, Skalen, die eine empirische Ordnung widerspiegeln, und schließlich Skalen, die eine quantitative empirische Beziehung widerspiegeln. Diese Klassifikation wurde durch Carnap, Hempel und vor allem durch den Mathematiker G.H. Birkhoff präzisiert. Diese Überlegungen wurden innerhalb der Psychologie von S.S. Stevens aufgegriffen und als Theorie der Skalenniveaus bekannt gemacht. Seitdem wurden und werden derartige Fragen vor allem in der Psychologie diskutiert, wo sie als besonders bedrängend empfunden werden.

Die mit der Stevensschen Klassifikation verbundenen Probleme seien an einem einfachen Fall verdeutlicht. Dazu betrachten wir statt quantitativer Aussagen numerische Relationen (Invarianzuntersuchungen von quantitativen Aussagen werden zumeist auf solche der in ihnen enthaltenen numerischen Relationen zurückgeführt, da der Begriff ‚quantitative Aussage‘ sehr viel schwerer präzise formal zu klären ist). Sei mit  $A$  eine Menge von ‚empirischen Objekten‘ bezeichnet und mit  $N$  eine Menge von ‚numerischen Objekten‘ (etwa  $N \subset \mathbb{R}$ ).  $R$  sei eine  $n$ -stellige Relation auf  $N$ . Den Objekten in  $A$  sollen nun Zahlen aus  $N$  zugeordnet werden, welche die Objekte in bestimmten vorgegebenen Aspekten charakterisieren sollen. Eine Abbildung, die dies leistet, heißt eine Skala. Die Menge aller Funktionen, die dies in ‚analoger Weise‘ leisten, heißt Skalenfamilie (bzgl.  $A$  und  $N$ ) und sei mit  $F$  bezeichnet. Nun können wir Kriterien für die Invarianz von Relationen formulieren:

(M1) Sei  $R$  eine  $n$ -stellige Relation auf  $N$  und sei  $T$  eine Menge von Funktionen  $t: N \rightarrow N$ . Dann heißt  $R$  **T-invariant** genau dann, wenn für alle  $t \in T$  und alle  $x_1, \dots, x_n \in N$  gilt:

$$(x_1, \dots, x_n) \in R \Leftrightarrow (t(x_1), \dots, t(x_n)) \in R.$$

(M2)  $R$  heißt **F-invariant** genau dann, wenn für alle  $f, g \in F$  und alle  $a_1, \dots, a_n \in A$  gilt:

$$(f(a_1), \dots, f(a_n)) \in R \Leftrightarrow (g(a_1), \dots, g(a_n)) \in R$$

Das meaningfulness-Kriterium (M2), von Adams, Fagot & Robinson (1965, S. 105f.) als „reference invariance“ bezeichnet, ist zunächst intuitiv verständlicher als (M1), da es gerade die Forderung beinhaltet, daß eine numerische Relation  $R$  invariant unter allen ‚Repräsentationen‘ der empirischen Objekte sein soll, die hinsichtlich der geforderten Charakterisierung ‚gleichermaßen‘ geeignet sind. Das Kriterium (M1) führt dagegen unmittelbar zu der Frage, wie sich eine geeignete Menge  $T$  charakterisieren läßt. Eine einzelne Abbildung  $t$  mit der Eigenschaft  $t \circ f \in F$  für alle  $f$  (wobei mit  $\circ$  die Komposition von Abbildungen bezeichnet wird), heißt **zulässige Transformation** (bzgl.  $F$ ). Wie läßt sich nun die Menge der zulässigen Transformationen bestimmen? Diese Frage ist innerhalb des Stevensschen Ansatzes nicht in befriedigender Weise zu klären. Wie die beiden genannten Kriterien zusammenhängen, wird in dem Fall deutlich, daß für ein beliebiges  $f_0 \in F$  gilt:  $F = (t \circ f_0 \mid t \in T)$ , d.h. mittels  $T$  läßt sich aus jeder beliebigen Skala ganz  $F$  erzeugen. Die  $T$ -Invarianz erweist sich aus dieser Perspektive als verkappte  $F$ -Invarianz; diese ist der grundlegendere Begriff, auch wenn in der Literatur häufig mit der Betonung **Invarianz unter den zulässigen Transformationen** der gegenteilige Eindruck erweckt wird. Erst im Repräsentationsansatz der Messung werden die hier angesprochenen Probleme in befriedigender Weise gelöst. Die Skalenfamilie  $F$  ist dann gerade die Menge der strukturerhaltenden Abbildungen (Homomorphismen bzw. Isomorphismen) zwischen qualitativer und numerischer Struktur, wodurch sich die obige Beschreibung, daß die Abbildungen in  $F$  das Geforderte ‚in analoger Weise‘ leisten, präzise fassen läßt. Das Skalenniveau ergibt sich dabei (bis auf Konjugation von Skalenfamilien, die dem Übergang von einer numerischen Repräsentationsstruktur zu einer anderen entsprechen; vgl. Krantz et al., 1971, S. 99ff., sowie Falmagne & Narens, 1983) aus bestimmten Annahmen über die betrachtete qualitative Struktur. (Hierbei wird auch deutlich, daß die beiden obigen Kriterien nicht notwendig äquivalent sein müssen, da es Situationen geben kann, in denen zwei Homomorphismen  $f$  und  $g$  durch ‚zulässige Transformationen‘ nicht ineinander überführt werden können.) Ausführliche Darstellungen finden sich in Adams, Fagot & Robinson (1965), Pfanzagl (1971), sowie Luce et al. (1990, Kap. 22).

Während damit die Bestimmung der Menge der zulässigen Transformationen, d.h. des Skalenniveaus, für die Repräsentation einer qualitativen Struktur geklärt ist, stellt sich darüber hinaus die Frage nach einer qualitativen Charakterisierung des Sachverhalts, daß eine Struktur mit einem bestimmten Skalenniveau repräsentiert werden kann. Eine wesentliche Rolle in diesem Zusammenhang spielen Symmetriekonzepte, wie das des Automorphismus einer Struktur (als eines strukturinternen Isomorphismus). Ein Automorphismus  $\alpha$  einer Struktur  $A = \langle A, S_1, \dots, S_n \rangle$  ist eine bijektive Abbildung von  $A$  auf sich selbst mit der Eigenschaft, daß für alle  $a_1, \dots, a_{k(j)} \in A$  und für alle  $j = 1, \dots, n$  gilt:

$$(a_1, \dots, a_{k(j)}) \in S_j \Leftrightarrow (\alpha(a_1), \dots, \alpha(a_{k(j)})) \in S_j, \tag{5.1}$$

wobei  $k(j)$  die Stelligkeit der Relation  $S_j$  bezeichne.

Betrachten wir nun das folgende Beispiel: Sei  $A$  eine extensive Struktur, d.h.  $\mathbf{A} = \langle \mathbf{A}, \geq, \circ \rangle$ ,  $\mathfrak{R} = \langle \mathbf{R}e^+, \geq, + \rangle$  und  $f$  ein Isomorphismus zwischen beiden Strukturen. Dann läßt sich zeigen, daß auch  $r \cdot f$  ( $r > 0$ ) ein Isomorphismus ist und daß je zwei Isomorphismen in dieser Weise zusammenhängen. Wir betrachten nun die Abbildung  $\alpha_r = f^{-1}r f$ , d. h.  $\alpha_r(a) = f^{-1}(r f(x))$ ; diese überführt  $A$  in  $A$  und ist ein Automorphismus von  $A$ , dem auf numerischer Seite der durch  $t(x) = r x$  definierte Automorphismus von  $\mathfrak{R}$  entspricht. Umgekehrt gilt: Ist  $\alpha$  ein Automorphismus von  $A$ , dann ist die Abbildung  $f \circ \alpha$  ein Isomorphismus von  $A$  nach  $\mathfrak{R}$ . Derartige Betrachtungen führen, wie erstmals Narens (1981) gezeigt hat, zu einem tieferen qualitativen Verständnis der Skalenniveaus: Jedes Skalenniveau auf numerischer Seite spiegelt interne Symmetrien der qualitativen Struktur wider. Von besonderer Bedeutung in diesem Zusammenhang ist die Narens'sche Klassifikation der Automorphismen einer qualitativen Struktur nach ihrer Reichhaltigkeit und Redundanz. (Die Reichhaltigkeit der Automorphismengruppe wird durch Narens' Konzept der „M-homogeneity“ beschrieben, die angibt, ob es einen Automorphismus gibt, der zwei vorgegebene M-Tupel geordneter Elemente ineinander überführt; die Redundanz wird durch die „N-uniqueness“ erfaßt, die angibt, ob zwei Automorphismen, die an N Punkten übereinstimmen, notwendigerweise identisch sind.) Aufbauend auf dieser Klassifikation weisen Narens und Alpern in mathematisch tiefliegenden Resultaten insbesondere nach, daß (für geordnete Strukturen, die ein Kontinuum bilden) die üblichen Skalentypen fast vollständig die möglichen Symmetrien beschreiben, obgleich auf numerischer Seite a priori beliebige Arten zulässiger Transformationen möglich sein könnten (siehe hierzu Luce et al., 1990, Kap.20).

In entsprechender Weise lassen sich numerischen Invarianzkonzepten solche qualitativer Art gegenüberstellen. Diese beruht auf dem folgenden Begriff der Struktur-Invarianz.

(M3) Sei  $\mathbf{A}$  eine Struktur mit Trägermenge  $A$  und  $S$  eine Relation auf  $A$ .  $S$  heißt **strukturinvariant** (genauer:  $A$ -invariant) genau dann, wenn für alle Automorphismen  $\mathbf{a}$  der Struktur und alle  $a_1, \dots, a_n \in A$  gilt:

$$((a_1), \dots, (a_n)) \in S \Leftrightarrow (\alpha(a_1), \dots, \alpha(a_n)) \in S.$$

Diese Bedingung besagt gerade, daß  $S$  invariant ist unter den Automorphismen der Struktur. Der Zusammenhang zu den zuvor formulierten Invarianz-Konzepten (M1) und (M2) soll wiederum an einem Beispiel verdeutlicht werden. Sei  $\mathbf{A}$  eine qualitative Struktur,  $\mathfrak{R}$  eine isomorphe numerische Struktur und  $f$  ein Isomorphismus zwischen beiden. Betrachten wir nun eine ‚neue‘ qualitative Relation  $S$  auf  $\mathbf{A}$  und die zugeordnete numerische Relation  $R = f(S)$ . Dann ist  $S$  genau dann strukturinvariant, wenn  $R$   $F$ -invariant ist, wobei  $F$  die Menge aller Isomorphismen zwischen den beiden Strukturen ist. (Und dies ist wiederum genau dann der Fall, wenn  $R$   $T$ -invariant ist, wobei  $T$  die Menge aller Automorphismen von  $\mathfrak{R}$  ist.) Für die allgemeinen Beziehungen der aufgeführten Invarianzkriterien zueinander siehe Luce et al. (1990, Kap.22) sowie Narens (1985, S. 156ff.).

Betrachten wir als Beispiel die im Abschnitt 3.3 im Zusammenhang mit dem Webersehen Gesetz eingeführte qualitative Semiordnung  $>_m$ . Sei, wie oben,  $\mathbf{A} = (A, \succsim, 0)$  ein physikalisches Kontinuum und sei  $\mathfrak{R} = \langle \mathbb{R}_{e+}, \succsim, + \rangle$ . Unter Bezugnahme auf die in Gleichung (3) dargestellte Beziehung

$$a >_m b \Leftrightarrow f(a) > f(b) + \Delta_f(b)$$

läßt sich die zugeordnete numerische Relation  $R$  schreiben als

$$R = \{(x, y) \mid x > y + \Delta(y)\},$$

(wobei  $\Delta(f(b)) = \Delta_f(b)$ ).

$f$  ist in diesem Fall eine Verhältnisskala,  $R$  somit  $F$ -invariant, wenn es invariant unter linearen Abbildungen ist. Dies wiederum ist genau dann der Fall, wenn  $A(y) = k \cdot y$ , d.h., wenn das Webersehe Gesetz gilt. Dies nun ist wiederum äquivalent dazu, daß  $>_m$  strukturinvariant ist. Mit anderen Worten: Die Gültigkeit des Webersehen Gesetzes ist dazu äquivalent, daß die Semiordnung  $>_m$  bezüglich der zugrunde liegenden extensiven physikalischen Struktur invariant ist.

Die Grenzen dieser Betrachtungsweise werden deutlich, wenn man Varianten des Webersehen Gesetzes betrachtet, welche der Existenz einer absoluten Schwelle Rechnung tragen: Die Einführung einer absoluten Schwelle ist nämlich mit der Annahme der Strukturinvarianz unverträglich und macht eine geeignete Verfeinerung dieses und verwandter Konzepte erforderlich.



## 5.2 Die ‚möglichen Formen‘ psychophysikalischer Funktionen und Gesetze

In einer klassischen Arbeit beschäftigte sich Luce (1959) mit Bedingungen, unter denen sich die Form der psychophysikalischen Funktion bestimmen läßt. Restriktionen zur Bestimmung ihrer Form gewinnt Luce aus Voraussetzungen eines Skalenniveaus der mit der physikalischen Skala  $f$  und der psychophysikalischen Skala  $u$  verbundenen Skalenfamilien und aus einer metatheoretischen Forderung, die er als „principle of theory construction“ bezeichnet (Luce, 1959, S.85). Jede psychophysikalische Funktion muß danach die Eigenschaft haben, daß eine zulässige Transformation der unabhängigen Variablen (hier also: der physikalischen) zu einer zulässigen Transformation der abhängigen Variablen (hier also: der psychophysikalischen) führt.

Wird die psychophysikalische Funktion (bzgl.  $f$  und  $u$ ) mit  $\Phi$  bezeichnet, d.h. besteht für alle Reize  $a$  die Beziehung  $u(a) = \Phi(f(a))$ , so bedeutet dies gerade, daß zu jeder zulässigen Transformation  $t$  der physikalischen Skala eine zulässige Transformation  $t'$  der psychologischen Skala  $u$  existiert, so daß gilt:

$$(M4) \quad \Phi(t(f(a))) = t'(u(a)).$$

Setzt man beispielsweise voraus, daß die physikalische Skala (genauer Skalenfamilie) ebenso wie die psychophysikalische eine Verhältnisskala ist, so ergibt sich aus (M4) eine Funktionalgleichung, aus welcher (unter bestimmten Stetigkeitsannahmen) folgt, daß  $\Phi$  eine Potenzfunktion sein muß. Setzt man statt dessen voraus, daß  $u$  eine Intervallskala ist, so ist auch eine logarithmische Form von  $\Phi$  möglich. In diesem Sinne behandelt Luce eine Reihe von Kombinationen verschiedener Skalenniveaus. Dies setzt selbstverständlich eine angemessene Festlegung insbesondere des psychophysikalischen Skalenniveaus voraus.

Die Bedingung (M4) läßt sich zu den oben eingeführten qualitativen Invarianzkonzepten in folgender Weise in Beziehung setzen: Sei  $A$  die Menge der physikalischen Reize, und nehmen wir an, daß die betrachtete physikalische Skalenfamilie gerade die Menge aller Isomorphismen zwischen einer qualitativen physikalischen Struktur  $\mathbf{A}$  mit Trägermenge  $A$  und einer numerischen Struktur  $\mathfrak{R}_1$  ist und die psychophysikalische Skalenfamilie die Menge der Isomorphismen zwischen einer psychophysikalischen Struktur  $\mathbf{B}$  mit Trägermenge  $A$  und einer numerischen Struktur  $\mathfrak{R}_2$ . (M4) erweist sich dann gerade als äquivalent zu der Annahme, daß die Relationen der psychophysikalischen Struktur  $A$ -invariant sind.

Vergleichbare und über die Lucuschen hinausgehenden Resultate erzielten Roberts & Rosenbaum (1986), die sich jedoch dem Problem aus einer anderen Perspektive nähern als Luce.

Die psychophysikalische Funktion ist nur ein spezielles Beispiel einer Beziehung zwischen Skalen in der Psychophysik. So wurden von Falmagne & Narens (1983) psychophysikalische Gesetze untersucht, die sich auf zwei numerische Input-Codes und einen numerischen Output-Code beziehen. Für die Beziehungen zwischen den Input- und Output-Codes formulieren sie verschiedene Invarianzprinzipien und untersuchen deren logische Beziehung sowie die Einschränkungen, die sich (unter gewissen strukturellen Zusatzannahmen) aus diesen Invarianzprinzipien für den Zusammenhang von Input- und Output-Codes ergeben.

Invarianz-Argumente dieser Art können also dazu beitragen, von vornherein die Klasse der mit ihnen verträglichen numerischen Gesetze deutlich einzugrenzen. Ein Beispiel hierzu aus der Psychoakustik findet sich in Falmagne & Narens (1983, S.314ff.; s.a. Falmagne, 1985, S.339ff.). Ein bekannter Fall, durch derartige Betrachtungen unzulässige Formen von Gesetzen auszuschließen, ist die Dimensionsanalyse der Physik, zu der die genannten Invarianzbetrachtungen eine Verwandtschaft aufweisen. Die Dimensionsinvarianz als Forderung auf numerischer Seite ist äquivalent mit der Struktur-Invarianz auf Seiten der qualitativen Struktur physikalischer Größen, wie Luce (1978; siehe auch Ramsey, 1976) gezeigt hat.

### 5.3 Wie sinnvoll sind meaningfulness-Betrachtungen?

Die genannten Invarianz-Kriterien basieren auf metatheoretischen Intuitionen, denen in der Physik als Dimensionsanalyse und Symmetriebetrachtungen und in der Mathematik als ein Ordnungsschema auf der Grundlage der Invarianz geometrischer Konzepte unter bestimmten Transformationsgruppen (Kleinsches Programm) eine große Bedeutung zukommt. In der Physik erfüllen fast alle bekannten Gesetze diese Forderungen, und die auf Invarianzkonzepten beruhenden Techniken haben sich als fruchtbar sowohl zum Auffinden möglicher wie auch zur Elimination unbrauchbarer Gesetzesformen erwiesen. Im Kontext der Psychophysik bringen die obigen Invarianzforderungen zudem zum Ausdruck, daß ein System gesetzmäßiger Beziehungen zwischen physikalischen und psychologischen Variablen gewisse Kohärenzeigenschaften und Symmetrien haben soll.

Betrachten wir die Tragweite solcher Invarianzforderungen am Beispiel des Webersehen Gesetzes: Invarianzbetrachtungen der genannten Art führen, wie wir oben gesehen haben, zu einer Auszeichnung des Webersehen Gesetzes unter den möglichen Gesetzen, die das Verhalten der Unterschiedsschwelle in Abhängigkeit vom physikalischen Reiz beschreiben könnten; diese wollen wir kurz als Sensitivitätsgesetze bezeichnen. Dies kann jedoch nicht heißen, daß das Webersehe Gesetz das einzig mögliche Sensitivitätsgesetz ist, wie es eine

aprioristische Lesart von Invarianzbetrachtungen impliziert, die in prototypischer Weise der Begriff **(non-)meaningfulness** und das Lucesche „principle of theory construction“ zum Ausdruck bringen. In der Tat sind nicht-lineare Sensitivitätsgesetze in der Psychophysik häufig zu finden. In entsprechender Weise kritisierte Rozeboom (1962) die Anwendung des Luceschen Prinzips bereits für die Physik unter Verweis auf das exponentielle Gesetz des radioaktiven Zerfalls. Dies veranlaßte Luce (1962), seine Position entsprechend zu modifizieren.

Sensitivitätsgesetze, wie etwa „ $\Delta s$  ist proportional zu  $s^2$ “, die nicht-invariante Relationen  $>_m$  beschreiben, sind nicht in einem absoluten Sinne „non-meaningful“. Anders als im Falle der invarianten Gesetze erfordert die präzise numerische Spezifikation eines eine solche Relation beschreibenden Gesetzes, d.h. der obigen Relation  $R$ , den Bezug auf eine konkrete Skala bzw. Einheit. Unter Zugrundelegung einer bestimmten extensiven Skala  $f$  ergibt sich für das soeben genannte Gesetz eine Spezifikation der Art  $\Delta_f(a) = c_f f(a)^2$ . Die Nicht-Invarianz drückt sich nun darin aus, daß der Wechsel zu einer anderen Skala mit einer Änderung der Konstanten  $c_f$  einhergeht. Derartige Skalenabhängigkeiten numerischer Spezifikationen von Gesetzen werden durch Pfanzagl (1971, S. 50) Konzept der **meaningful parametrization** erfaßt.

(M5) Eine Familie von Relationen  $(R)_{f \in F}$  heißt **meaningfully parametrized** genau dann, wenn für alle  $f, g \in F$  und alle  $a_1, \dots, a_n \in A$  gilt:

$$(f(a_1), \dots, f(a_n)) \in R_f \Leftrightarrow (g(a_1), \dots, g(a_n)) \in R_g.$$

(Siehe hierzu auch Luce et al., 1990, S. 278). Im obigen Beispiel ist  $R_f = \{(x, y) \mid x = y + c_f y^2\}$ . Die Möglichkeit, Skalenabhängigkeiten in der Art von (M5) zuzulassen, von der auch die Physik Gebrauch macht, zeigt die Grenzen eines aprioristischen meaningfulness-Verständnisses.

Dies bedeutet jedoch nicht, daß den genannten Invarianzkonzepten und verwandten Intuitionen in den Wissenschaften nicht dennoch eine erhebliche Rolle zukommt, wie es bereits das Beispiel der Physik lehrt (vgl. v.Fraasen, 1989). Sie bilden ein wesentliches Element theoretischen Denkens jenseits eines reinen *curve fittings*, in dem sich die Haltung spiegelt, ‚intrinsisch Gesetzhaftes‘ von ‚kontingenten empirischen Generalisierungen‘ scheidet zu wollen. Diese Haltung ist oftmals mit Auffassungen einer ‚Theorie-Ästhetik‘ verbunden, wie sie seit Pythagoras die Naturwissenschaften insbesondere in der Form von Symmetriebetrachtungen prägt (vgl. Guthrie, 1962, S. 236ff.; v. Fritz, 1981, S. 47ff.).

Invarianzkonzepte weisen einen engen Zusammenhang mit Definierbarkeitskonzepten auf. So sind insbesondere alle in einer Struktur  $A$  (etwa im Sinne Prädikatenlogik erster Stufe) definierbaren Relationen  $A$ -invariant. (Die Um-

kehrung gilt jedoch nur in einem eingeschränkten Sinne.) Läßt sich also eine neue Relation unter Verwendung der Grundrelationen einer Struktur definieren, so ist sie strukturinvariant. Dies eröffnet eine interessante neue Perspektive für meaningfulness-Betrachtungen: Man geht von einer vorgegebenen Struktur aus und betrachtet deren Grundrelationen als **meaningful**; sodann definiert man relativ zu dieser gegebenen Struktur solche Relationen als **meaningful**, die sich (in einem zu präzisierenden Sinne) aus den Grundrelationen definieren lassen. Diese Art des Zugangs wurde von Narens (1988) vorgeschlagen. Auch hier ist wieder vor einem aprioristischen Mißverständnis zu warnen, da es natürlich keine absoluten Kriterien dafür geben kann, welche Struktur als ‚Fundament‘ derartigen Betrachtungen zugrunde zu legen ist. Narens & Mausfeld (1992) ziehen Definierbarkeitskonzepte heran, um im Rahmen eines erweiterten Invarianz-Klassifikationskriteriums für die Psychophysik Kriterien dafür anzugeben, wann zwei physikalische Situationen als äquivalent angesehen werden sollen. Hat man ein solches Kriterium gewonnen, so trägt das folgende psychophysikalische Äquivalenzprinzip der asymmetrischen Beziehung zwischen physikalischen und psychologischen Relationen Rechnung: Eine notwendige Voraussetzung dafür, daß eine psychophysikalische Aussage einen psychologischen Gehalt hat, ist, daß ihr Wahrheitswert unabhängig von der Wahl äquivalenter physikalischer Systeme ist. Dem von Narens & Mausfeld (1992) entwickelten Kriterium zufolge sind zwei physikalischer Strukturen mit gleicher Trägermenge genau dann äquivalent, wenn sie isomorph sind und sich jede Relation der einen Struktur durch eine Relation der anderen Struktur (in einer hinreichend reichhaltigen logischen Sprache) definieren läßt. Im Falle zweier stetiger extensiver Strukturen mit Repräsentationen  $f_1$  bzw.  $f_2$  ist dies äquivalent zu der Bedingung, daß es ein  $r \in \mathbf{Re}^+$  gibt, so daß für die beiden Repräsentationen gilt:

$$f_1^{-1}[f_1(x) + f_1(y)] = f_2^{-1}[f_2(x)^r + f_2(y)^r]^{1/r}.$$

Ersetzt man also in einer extensiven Struktur mit Operation  $0$ , und Repräsentation  $f_1$  die extensive Operation durch eine Operation  $O_2$ , für die gilt:

$$x O_2 y = f_1^{-1}[f_1(x)^r + f_1(y)^r]^{1/r},$$

so erhält man eine physikalisch äquivalente Struktur. Man kann leicht zeigen, daß physikalisch äquivalente Strukturen dieselbe Automorphismengruppe haben (da eine Transformation der Art  $f \rightarrow \alpha f$  die beiden obigen Ausdrücke nicht ändert).

Gegenüber der klassischen **meaningfulness-Konzeption**, wie sie insbesondere durch die obigen Bedingungen (M1)-(M3) beschreiben wurde, bezieht das Äquivalenzprinzip nicht nur die numerische Repräsentation, sondern die qualitativen physikalischen Operationen selbst mit in die Invarianzforderung ein und verschärft damit die Bedingungen, unter denen quantitative Aussagen und

numerische Relationen in einer psychophysikalischen Struktur als psychologisch relevant anzusehen sind.

Im Lichte des Äquivalenzprinzips erweisen sich eine Reihe quantitativer Aussagen der Psychophysik als nicht psychologisch relevant: Wie leicht erkennbar ist, ergibt sich beispielsweise aus dem Äquivalenzprinzip, daß die durch Größenschätzung erhaltenen Exponenten nicht die untersuchte Sinnesmodalität charakterisieren, sondern ebenfalls von der verwendeten physikalischen Operation abhängen (im Beispiel der Lautstärke verdoppelt sich der Exponent, wenn man den physikalischen Reiz in Einheiten des Schalldrucks statt der Schallenergie mißt). Sinnesmodalitäten, deren Exponent größer als 1 ist als „expanders“ und solche, deren Exponent kleiner als 1 ist als „compressors“ zu klassifizieren, wie Stevens (e. g. 1974, S.370) dies unternimmt, ist dem Äquivalenzprinzip zufolge nicht sinnvoll, da einer solchen Klassifikation kein empirisch-psychologischer Gehalt zukommt.

Die obigen Ausführungen legen nahe, Invarianz-Forderungen in Psychophysik und Psychologie nicht als strikte aprioristische ‚Sinnkriterien‘ zu betrachten, sondern als ein heuristisches Forschungsmittel. Beide Arten von meaningfulness-Betrachtungen, die auf Symmetrieintuitionen basierenden Invarianzbetrachtungen und die durch eine logisch-reduktionistische Perspektive motivierten Definierbarkeitsbetrachtungen sind aprioristische Metaprinzipien, durch die Willkürliches bei der Theoriebildung ausgeschlossen werden soll und die angeben, welche Konzepte im weitesten Sinne in der theoretischen Sprache formulierbar bzw. ausdrückbar sind. Der Wert eines *meaningfulness*-Kriteriums hängt natürlich in entscheidender Weise von der gewählten theoretischen Sprache bzw. von der Ausgangsstruktur ab (für deren Wahl sich andererseits kaum aprioristische Kriterien finden lassen). So können derartige Betrachtungen nur eine Art Faustregel sein, denn zu diesen Prinzipien in Widerspruch stehende Fälle können, wie Rozebooms (1962) Beispiel des exponentiellen Zerfallgesetzes zeigt, stets durch Zusatzannahmen eingeführt werden; doch weisen meaningfulness-Betrachtungen darauf hin, daß solche Zusatzannahmen explizit gemacht werden müssen, und erhöhen dadurch gleichsam die Hürde für Willkürelemente.

Daß eine zu strikte Handhabung von meaningfulness-Forderungen einer angemessenen Theorienentwicklung gar entgegenstehen kann, läßt sich an einem Beispiel aus der Farbwahrnehmung verdeutlichen: Die empirisch gefundene Nicht-Linearität des sog. Blau-Gelb-Kanals der Opponenten Theorie der Farbcodierung bedeutet, daß die Einführung sog. Urfarben in die Grassmann-Struktur *non-meaningful* ist, da sie deren Automorphismen zerstört. Diese *non-meaningfulness* auf der Ebene der Grassmann-Struktur bedeutet aber einen echten Erkenntnisgewinn.

Dieses Beispiel macht deutlich, daß unter dem Aspekt von Forschungsheuristiken *meaningfulness*-Betrachtungen notwendigerweise in einem gewissen Sinne konservativ sind. Jeder echte Zuwachs einer Theorie an empirischem Gehalt kann dazu führen, daß die neue Theorie weniger Automorphismen aufweist als die alte (wurde man wie in einem Prosastück von J.L. Borges eine Theorie zulassen, die eine 1-1-Darstellung der Welt wäre, so hätte diese bis auf die Identität keine Automorphismen mehr). Zwischen einem Erkenntnisfortschritt und *meaningfulness*-Betrachtungen besteht also eine gewisse Spannung, und nur bezogen auf die jeweilige Stufe der Theoriebildung können diese ein Instrument zur Klärung methodologischer Probleme sein. In jedem Fall ist die Bezeichnung ***meaningfulness***, die sich historisch aus den semantischen Sinnkriterien des logischen Empirismus herleitet, eine sehr unglückliche Wortwahl (bestenfalls könnte man von einer konditionalen ***meaningfulness*** sprechen); das Konzept hat mit Attributen wie ‚sinnvoll‘ und ‚bedeutsam‘ wenig gemein, kann ihnen in einem Prozeß der Theorieentwicklung gar entgegenstehen.

Ihrem Ursprunge nach spiegeln sich in Invarianzkriterien physikalische Sichtweisen wider. Somit stellt sich die Frage nach der Angemessenheit der Übertragung einer solchen Perspektive in die Psychophysik. Diese Frage muß als offen angesehen werden, steht doch, so scheint es, die Psychophysik in Aspekten der Theoriebildung der Biologie näher als der Physik. Aus neurophysiologischer Perspektive ist eine strenge und globale Erfüllung von Invarianzkriterien nicht zu erwarten, was indes approximative Gültigkeit nicht ausschließt. Nähert man sich der Psychophysik aus funktionalistischer Sicht (***computational approach***), so können geeignet verfeinerte und inhaltlich eingebundene Invarianzkonzepten wiederum von entscheidender Bedeutung sein. Hierauf verweisen schon, wenn auch in subtiler Weise, die Intuitionen, welche an den Begriff der Invarianten gebundenen sind (vgl. Mausfeld, in diesem Band, Kap. 4), die wiederum mit den Konstanzleistungen des Wahrnehmungssystems zusammenhängen.

## **Literatur**

- Adams, E. W., Fagot, R. F. & Robinson, R. (1965). A theory of appropriate statistics. ***Psychometrika***, 30, 99-127.
- Anderson, N. H. (1982). ***Methods of information integration theory***. New York: Academic Press.
- Arbuckle, J. & Larimer, J. (1976). The number of two-way tables satisfying certain additivity axioms. ***Journal of Mathematical Psychology***, 13, 89-100.
- Baird, J. C. & Noma, E. (1978). ***Fundamentals of scaling and psychophysics***. New York: Wiley.

- Bock, R. D. & Jones, L. V. (1968). **The measurement and prediction of judgement and choice**. San Francisco: Holden-Day.
- Böhme, G., van den Daele, W. & Krohn, W. (1974). Die Finalisierung der Wissenschaft. In W. Dietrich (Hrsg.) **Theorien der Wissenschaftsgeschichte** (S. 276-311). Frankfurt: Suhrkamp.
- Bourbaki, N. (1966). **General Topology. Vol. ZZ**. Paris: Hermann.
- Brentano, E (1974/1924). **Psychologie vom empirischen Standpunkt**. Bd. I. Leipzig.
- Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. **Psychological Science, 3**, 186-190.
- Cohen, M. R. & Nagel, E. (1934). **An introduction to logic and scientific method**. London: Routledge & Kegan Paul.
- Colonius, H. (1984). **Stochastische Theorien individuellen Wahlverhaltens**. Berlin: Springer.
- Debreu, G. (1959). Topological methods in cardinal utility theory. In K. J. Arrow, S. Karlin & P. Suppes (Eds.), **Mathematical models in the social sciences. (S. 16-26)**. Stanford: Stanford University Press.
- Drösler, J. (1989). **Quantitative psychology**. Göttingen: Hogrefe.
- Falmagne, J.-C. (1976). Random conjoint measurement and loudness summation. **Psychological Review, 83**, 65-79.
- Falmagne, J.C. (1979). On a class of probabilistic conjoint measurement models: Some diagnostic properties. **Journal of Mathematical Psychology, 19**, 73-88.
- Falmagne, J.-C. (1985). **Elements of psychophysical theory**. Oxford: Clarendon.
- Falmagne, J.C. (1986). Psychophysical measurement and theory. In K.R. Boff, L. Kaufman & J. P. Thomas (Eds.), **Handbook of perception and human performance. Vol. I. Sensory processes and perception**. New York: Wiley.
- Falmagne, J.-C. & Narens, L. (1983). Scales and meaningfulness of quantitative laws. **Synthese, 55**, 287-325.
- Fechner, G. Th. (1860). **Elemente der Psychophysik. Bd. I & II**. Leipzig: Breitkopf & Härtel.
- Fechner, G. Th. (1877). **Zn Sachen der Psychophysik**. Leipzig.
- Fechner, G.Th. (1888). über die psychischen Maßprinzipien und das Webersche Gesetz. **Philosophische Studien, 4**, 161-230.
- Fishburn, P. C. (1992). Induced binary probabilities and the linear ordering polytope: A status report. **Mathematical Social Sciences, 23**, 67-80.
- v. Fritz, K. (1971). **Grundprobleme der Geschichte der antiken Wissenschaft**. Berlin: de Gruyter.
- Gigerenzer, G. & Strube, G. (1983). Are there limits to binaural additivity of loudness? **Journal of Experimental Psychology: Human Perception and Performance, 9**, 126-136.
- Gould, S. J. (1981). **The mismeasure of man**. New York: Norton.
- Grassmann, H. (1853). Zur Theorie der Farbmischung. **Poggendorffs Annalen der Physik und Chemie, 89**, 69-84.

- Guthrie, W. K. C. (1962). **A history of greek philosophy. The earlier Presocratics and the Pythagoreans.** Cambridge: Cambridge University Press.
- Helmholtz, H. v. (1887). Zählen und Messen erkenntnistheoretisch betrachtet. In **Philosophische Aufsätze Eduard Zeller gewidmet.** Leipzig: Fues.
- Heyer, D. & Mausfeld, R. (1987). On errors, probabilistic measurement and boolean valued logic. **Methodika, 1**, 113-138.
- Heyer, D. & Niederée, R. (1992). Generalizing the concept of binary choice systems induced by rankings: One way of probabilizing deterministic measurement structures. **Mathematical Social Sciences, 23**, 31-44.
- Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Mass. **Berichte und Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Classe, 53**, 1-46.
- Holman, E. W. & Marley, A. A. J. (1974). Stimulus and response measurement. In E. C. Carterette & M.P. Friedman (Eds.), **Handbook of perception. Vol. ZZ. (S. 173-213).** New York: Academic Press.
- Iverson, J. (1991). Probabilistic measurement theory. In J.-P. Doignon & J.-C. Falma-gne (Eds.), **Mathematical Psychology. Current developments (S. 135-155).** Berlin: Springer.
- Kuhn, T. S. (1961). The function of measurement in modern science. In H. Wolf (Ed.), **Quantification. A history of the meaning of measurement in the natural and social sciences.** (S. 31-63). Indianapolis: Bobbs-Merrill.
- Krantz, D. H. (1972a). A theory of magnitude estimation and cross-modality matching. **Journal of Mathematical Psychology, 9**, 168-199.
- Krantz, D.H. (1972b). Visual scaling. In L.M. Hurvich & D. Jameson (Eds.), **Handbook of sensory physiology, VII/4, Visual psychophysics.** (S. 660-689). Berlin: Springer.
- Krantz, D.H. (1972c). Measurement structures and psychological laws. **Science, 175**, 1427-1435.
- Krantz, D. H., Luce, R. D., Suppes, P. & Tversky, A. (1971). **Foundations of measurement. Vol. I. Additive and polynomial representations.** New York: Academic Press.
- von Kries, J. (1882). über die Messung intensiver Größen und über das sogenannte psychophysische Gesetz. **Vierteljahresschrift für wissenschaftliche Philosophie, 6**, 257-294.
- Krull, W. (1960). über die Endomorphismen von total geordneten Archimedischen Abelschen Halbgruppen. **Mathematische Zeitschrift, 74**, 81-90.
- Laming, D.R. J. (1986). **Sensory analysis.** New York: Academic Press.
- Levitt, W. J. M., Riemersma, J. B. & Bunt, A. A. (1972). Binaural additivity in loudness. **British Journal of Mathematical and Statistical Psychology, 25**, 51-68.
- Levine, M.V. (1970). Transformations that render curves parallel. **Journal of Mathematical Psychology, 7**, 410-443.
- Luce, R.D. (1959). On the possible psychophysical laws. **Psychological Review, 66**, 81-95.
- Luce, R.D. (1962). Comments on Rozeboom's criticisms of 'On the possible psychophysical laws'. **Psychological Review, 69**, 548-551.



- Luce, R. D. (1977). Thurstone's discriminial processes fifty years later. *Psychometrika*, 42, 461-489.
- Luce, R. D. (1978). Dimensionally invariant numerical laws correspond to meaningful qualitative relations. *Philosophy of Science*, 41, 1-16.
- Luce, R.D. (1990). „On the possible psychophysical laws" revisited: Remarks on cross-modal matching. *Journal of Mathematical Psychology*, 97, 66-77.
- Luce, R.D. & Galanter, E. (1963a). Discrimination. In R.D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology, Vol. I (S. 191-243)*. New York: Wiley.
- Luce, R. D. & Galanter, E. (1963b). Psychophysical scaling. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology, Vol. Z (S. 245-307)*. New York: Wiley.
- Luce, R.D. & Green, D.M. (1974). Detection, discrimination, and recognition. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception. Psychophysical judgment and measurement. Vol. ZZ. (S. 300-342)*. New York: Academic Press.
- Luce, R. D., Krantz, D.H., Suppes, P. & Tversky, A. (1990). *Foundations of measurement, Vol. III. Representation, axiomatization, and invariance*. New York: Academic Press.
- Luce, R.D. & Krumhansl, C.L. (1988). Measurement, scaling, and psychophysics. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology. Vol. I. Perception and motivation. (S. 3-73)*. New York: Wiley.
- Luce, R. D. & Narens, L. (1993). Fifteen problems concerning the representational theory of measurement. In P. Humphreys (Ed.), *Essays in honor of Patrick Suppes*. Amsterdam: Kluwer.
- Luce, R. D. & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology. Vol. III. (S. 249-410)*. New York: Wiley.
- Martin, G. (1956). *Klassische Ontologie der Zahl*. Köln: Kölner Universitätsverlag.
- McClelland, G. (1977). A note on Arbuckle and Larimer „The number of two-way tables satisfying certain additivity axioms". *Journal of Mathematical Psychology*, 15, 292-295.
- McClelland, G. H. & Coombs, C. H. (1975). Ordmet: A general algorithm for constructing all numerical solutions to ordered metric data. *Psychometrika*, 50, 269-290.
- Merkel, J. (1888). Die Abhängigkeit zwischen Reiz und Empfindung. *Philosophische Studien*, 4, 541-594.
- Mittelstraß, J. & Schroeder-Heister, P. (1986). Zeichen, Kalkül, Wahrscheinlichkeit. Elemente einer Mathesis universalis bei Leibniz. In H. Stachowiak (Hrsg.), *Handbuch pragmatischen Denkens I*. Hamburg: Meiner.
- Narens, L. (1981). On the scales of measurement. *Journal of Mathematical Psychology*, 24, 249-275.
- Narens, L. (1985). *Abstract measurement theory*. Cambridge: MIT Press.

- Narens, L. (1988). Meaningfulness and the Erlanger program of Felix Klein. *Mathématiques Informatique et Sciences Humaines*, 101, 61-72.
- Narens, L. (1993). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology* (to appear).
- Narens, L. & Luce, R. D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin*, 99, 166-180.
- Narens, L. & Mausfeld, R. (1992). On the relation of the psychological and the physical in psychophysics. *Psychological Review*, 99, 467-479.
- von Neumann, J. & Morgenstern, O. (1947). *Theory of games and economic behavior*. 2nd Ed. Princeton: Princeton University Press.
- Niederée, R. (1987). On the reference to real numbers in fundamental measurement: A model-theoretic approach. In E. Roskam & R. Suck (Eds.) *Progress in mathematical psychology* (S. 3-24). Amsterdam: North-Holland.
- Niederée, R. (1992a). *Maß und Zahl. Logisch-modelltheoretische Untersuchungen zur Theorie fundamentaler Messung*. Frankfurt: Lang.
- Niederée, R. (1992b). What do numbers measure? A new approach to fundamental measurement. *Mathematical Social Sciences*, 24, 237-276.
- Pfanzagl, J. (1971). *Theory of measurement*. 2nd ed. Wien: Physica.
- Plateau, M. J. (1872). Sur la Mesure des Sensations physiques, et sur la Loi qui Lie l'Intensité des ces Sensations a l'Intensité de la Cause Excitante. *Bulletin de l'Académie Royale des Sciences, des Lettres et des Beaux Arts de Belgique*, 33, 376-388.
- Ramsey, F.P. (1931). *The foundations of mathematics and other logical essays*. New York: Harcourt Brate.
- Ramsey, J.O. (1976). Algebraic representations in physical and behavioral sciences. *Synthese*, 32, 419-453.
- Restle, E. & Greeno, J.G. (1970). *Introduction to mathematical psychology*. Reading, Mass.: Addison-Wesley.
- Roberts, F. S. (1979). *Measurement theory with applications to decision making, Utility, and the social sciences*. Reading, Mass.: Addison-Wesley.
- Roberts, F.S. & Rosenbaum, Z. (1986). Scale type, meaningfulness, and the possible psychophysical laws. *Mathematical Social Sciences*, 12, 77-95.
- Robertson, T., Wright, F.T. & Dykstra, R. L. (1988). *Order restrictions in statistical inference*. Chichester: Wiley.
- Roskam, E. (1983). Allgemeine Datentheorie. In H. Feger & J. Bredenkamp (Hrsg.), *Enzyklopädie der Psychologie. Messen und Testen*. (S. 1-135). Göttingen: Hogrefe.
- Rozeboom, W. W. (1962). The untenability of Luce's principle. *Psychological Review*, 69, 542-547.
- Schneider, B., Parker, S. & Stein, D. (1974). The measurement of loudness using direct comparisons of sensory intervalls. *Journal of Mathematical Psychology*, 11, 259-273.
- Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64, 153-181.

- Stevens, S.S. (1974). Perceptual magnitude and its measurement. In E.C. Carterette & M.P. Friedman (Eds.), **Handbook of perception. Vol. II. Psychophysical judgment and measurement.** (S.361-389). New York: Academic Press.
- Suppes, P. (1962). Models of data. In E. Nagel, P. Suppes & A. Tarski (Eds.), **Logic, methodology and philosophy of science** (S.252-261). Stanford: Stanford University Press.
- Suppes, P., Krantz, D.H., Luce, R.D. & Tversky, A. (1989). **Foundations of measurement. Vol. II. Geometrical, threshold, and probabilistic representations.** New York: Academic Press.
- Tack, W. H. (1983). Psychophysische Methoden. In H. Feger & J. Bredenkamp (Hrsg.), **Enzyklopädie der Psychologie. Messen und Testen.** (S. 346-426). Göttingen: Hogrefe.
- Wandell, B. & Luce, R.D. (1978). Pooling peripheral information: Averages versus extrem values. **Journal of Mathematical Psychology, 17, 220-235.**
- Wundt, W. (1908). **Grundzüge der physiologischen Psychologie. Bd. I.** (6. Aufl.). Leipzig: Engelmann.

## 14. Kapitel

# Hypothesenprüfung

**Edgar Erdfelder und Jürgen Bredenkamp**

Hypothesenprüfungen stehen im Dienste des übergeordneten Zieles wissenschaftlichen Arbeitens, beobachtbare Ereignisse und Gesetzmäßigkeiten stichhaltig **erklären** zu können. Wissenschaftliche Ereignis- oder Gesetzeserklärungen setzen in der Psychologie wie in anderen Erfahrungswissenschaften Hypothesen oder Theorien voraus, die u. a. dem empirischen Adäquatheitskriterium genügen müssen und sich somit - wie man sagt - „bewährt“ haben (vgl. Westmeyer, 1973; Groeben & Westmeyer, 1981<sup>2</sup>; Gadenne, Kapitel 7 und 9 dieses Bandes). Die methodologischen Probleme, mit denen sich das vorliegende Kapitel auseinandersetzt, resultieren aus der Tatsache, daß durch empirische Untersuchungen im Regelfall nicht ohne weiteres festgestellt werden kann, ob Hypothesen und Theorien wahr oder falsch sind. Dies hat zum Teil **logische**, zum Teil aber auch **forschungspraktische** Gründe. Zu den logischen Gründen zählt etwa die Unmöglichkeit der Verifikation allgemeiner Hypothesen (Popper, 1982<sup>7</sup>), die als Bestandteile des Explanans wissenschaftlicher Erklärungsargumente von zentraler Bedeutung sind, aber auch die Tatsache, daß statistische Hypothesen, die in der Psychologie eine wichtige Rolle spielen, mit allen denkbaren empirischen Ereignissen prinzipiell kompatibel sind. In forschungspraktischer Hinsicht ergibt sich nicht selten das Problem, daß auch über deterministische und singuläre Hypothesen ohne Allgemeinheitsgrad empirisch nicht eindeutig entschieden werden kann, weil die Untersuchungsdaten fehlerbelastet sind oder eine direkte Hypothesenprüfung an einzelnen Individuen Verletzungen von Annahmen zur Folge hätte, die Bestandteile der zu prüfenden Hypothesen sind. Alle genannten Probleme und die daraus ableitbaren Konsequenzen für die Forschungspraxis werden in den folgenden Abschnitten anhand von Beispielen verdeutlicht und präzisiert.

Obwohl der modalen psychologischen Hypothese (PH) weder Verifizierbarkeit noch Falsifizierbarkeit im Sinne Poppers (1982<sup>7</sup>) zugesprochen werden kann, liegt kein Grund zur Resignation vor. Die Methodologie der Prüfung von PHn muß so ausgearbeitet sein, daß sie Regeln in die Hand gibt, deren Anwendung psychologische Untersuchungen so informativ macht, daß sich

Forscher bei einer bestimmten Befundlage **entschließen** können, die Theorie fallen zu lassen bzw. durch eine bessere Theorie zu ersetzen. Diese Position entspricht nach Lakatos (1970) nicht einem „dogmatischen“, sondern einem „methodologischen Falsifikationismus“. Der dogmatische Falsifikationismus geht von einer unfehlbaren empirischen Basis aus: Wenn eine Theorie falsifiziert wird, ist ihre Falschheit bewiesen. Im Rahmen des methodologischen Falsifikationismus kann eine empirisch nicht bewährte Theorie dagegen u.U. wahr sein. Demzufolge ist nicht nur das Risiko eines fälschlichen Bewährungsurteils (mangelnde „Strenge“ einer Untersuchung), sondern auch das Risiko eines fälschlichen Nichtbewährungsurteils (mangelnde „Fairneß“ einer Untersuchung) bei der methodologischen Analyse hypothesenprüfender empirischer Untersuchungen im Blick zu behalten (vgl. Gadenne, 1976, S.70; Hager, 1987, 1992; Westermann, 1987).

In diesem Kapitel wird eine Methodologie der Überprüfung von PHn vorgestellt, die den genannten Forderungen Rechnung trägt. Sie kann „deduktivistisch“ in dem Sinne genannt werden, daß sie auf induktive Elemente, wie sie etwa in der Methodologie Campbell und Stanleys (1963) auszumachen sind (Gadenne, 1976, 1984; vgl. auch Erdfelder, Kapitel 2 dieses Bandes), soweit wie möglich verzichtet (zum Begriff der Induktion vgl. Westermann & Gerjets, Kapitel 10 dieses Bandes). Sie kann weiterhin als Methodologie der **statistischen** Prüfung von PHn bezeichnet werden, da sie davon ausgeht, daß - wie in der psychologischen Forschungspraxis üblich - statistische Hypothesen (SHn) als Prüfinstanzen für PHn fungieren. Mit dieser Schwerpunktsetzung wollen wir keineswegs suggerieren, daß die gegenwärtig dominierende Forschungspraxis die allein mögliche oder allein sinnvolle Form empirischer Psychologie darstellt. Empirische Prüfungen von PHn müssen nicht notwendig auf statistische Hypothesentests hinauslaufen. Andere, deskriptivstatistische oder auch nichtstatistische Wege der Überprüfung sind denkbar und u.U. auch sinnvoll. Dennoch scheint uns der Versuch einer rationalen Rekonstruktion der statistischen Prüfung von PHn allein deshalb von besonderer Bedeutung zu sein, weil dieses Vorgehen offenbar von der großen Mehrzahl empirisch arbeitender Psychologen präferiert wird. Dies allein rechtfertigt die in diesem Kapitel gewählte Perspektive.

Im ersten Abschnitt werden die wichtigsten Aspekte einer deduktivistischen Methodologie empirischer Untersuchungen in der Psychologie kurz skizziert, wobei insbesondere die Funktion der Inferenzstatistik im Überprüfungsprozeß in den Mittelpunkt gerückt wird. Dieser Abschnitt stützt sich wesentlich auf Arbeiten von Bredenkamp (1969, 1972, 1980), Gadenne (1976, 1984) und Hager und Westermann (1983b, 1983c). Der zweite Abschnitt ist der Explikation der Begriffe „Strenge“, „Fairneß“ und „Validität“ einer Hypothesenprüfung sowie den methodologischen Folgerungen gewidmet, die sich aus der Forderung nach möglichst validen Untersuchungen ergeben. Zentrale Be-

standteile der deduktivistischen Methodologie werden in diesem Abschnitt formal analysiert und begründet. Der dritte Abschnitt gibt einige Empfehlungen zur statistischen Entscheidungsstrategie, die möglichst strenge und faire Prüfungen gewährleisten soll. Der vierte Abschnitt schließlich hat speziell das Theorie-Empirie-Überbrückungsproblem zum Gegenstand, d.h. die Frage, wie in geeigneter Form eine Beziehung zwischen theoretischen Begriffen in PHn und beobachtbaren Tatbeständen hergestellt werden kann.

## **1. Grundzüge einer deduktivistischen Theorie hypohesentestender Untersuchungen**

Empirische Untersuchungen in der Psychologie zeichnen sich fast ausnahmslos dadurch aus, daß substanzwissenschaftliche Fragestellungen mittels statistischer Hypothesentests beantwortet werden. Bemerkenswert ist dieses Vorgehen vor allem deshalb, weil die interessierende PH und die letztlich getestete SH im Regelfall nicht identisch sind. PHn beziehen sich typischerweise auf psychische Vorgänge, die im einzelnen Individuum ablaufen. Beispiele hierfür sind die all-gemeinpsychologischen Hypothesen „Imaginale Verarbeitungsprozesse führen im Vergleich zu sprachlichen Verarbeitungsprozessen zu überlegenen Gedächtnisleistungen“, „Erwachsene Personen bewältigen konzeptuelle Probleme durch das Testen von Hypothesen“, „Die Empfindungsstärke ist eine logarithmische Funktion der Reizstärke“, um nur einige Beispiele zu nennen. Auf einzelne Individuen bezogene Hypothesen finden sich nicht nur in der Allgemeinen Psychologie, sondern z.B. auch in der Sozialpsychologie: „Wer öffentlich unter Druck eine Meinung äußert, die seiner eigenen Meinung widerspricht, tendiert dazu, seine Meinung in Richtung der Äußerung abzuändern.“ Statistisch getestet werden aber in der Allgemeinen Psychologie wie in der Sozialpsychologie häufig Populationsaussagen der Art, daß die Mittelwerte aus zwei oder mehr Populationen, in denen die interessierende abhängige Variable (AV) jeweils normal und mit gleichen Varianzen verteilt ist, identisch bzw. nicht identisch sind. Wenn dieses Vorgehen nicht schlicht am eigentlichen Zweck der Untersuchung, nämlich eine vorgegebene PH testen zu wollen, vorbei gehen soll, muß es eine Verknüpfung zwischen PH und SH geben, die so gestaltet ist, daß der statistische Test aussagekräftig in bezug auf die PH ist.

### 1.1 Warum werden psychologische Hypothesen statistisch geprüft?

Anhand zweier Beispiele soll zunächst untersucht werden, warum PHn auf dem - wie es scheint - „Umweg“ über die Statistik geprüft werden. Das erste Beispiel bezieht sich auf die Invarianzhypothese (**total time hypothesis**, vgl.

Baddeley, 1990) des verbalen Lernens, nach der die Gesamtlernzeit bis zur Beherrschung eines Stoffes eine individuelle Konstante ist, unabhängig davon, in wieviele Lernversuche diese Zeit aufgeteilt wird. Es muß demnach für alle Individuen  $T_i = Y_{ij} \cdot t_j$  gelten, wobei  $T_i$  die Gesamtlernzeit des Individuums  $i$ ,  $t_j$  die pro Lernversuch gewährte Bearbeitungszeit und  $Y_{ij}$  die benötigte Anzahl der Lernversuche von Individuum  $i$  bei Bearbeitungszeit  $t_j$  pro Lernversuch ist. Verglichen werden üblicherweise die mittleren Gesamtlernzeiten von Versuchspersonen (Vpn), die den verschiedenen Versuchsbedingungen (unterschiedliche Werte für  $t_j$ ) zufällig zugewiesen wurden (Bugelski, 1962). Unter diesen Umständen folgt aus der Gültigkeit der Invarianzhypothese für alle Individuen die statistische Hypothese  $E(T) = E(Y_j) \cdot t_j$ , wobei die Gesamtlernzeit  $T$  und die Anzahl der Lernversuche  $Y_j$  bei Bearbeitungszeit  $t_j$  nunmehr als Zufallsvariablen (ZVn) aufzufassen sind. Von zentraler Bedeutung für die Implikationsbeziehung ist, daß mit den verschiedenen Werten für  $t_j$  nicht andere Variablen wie z.B. die Lernfähigkeit der Vpn korreliert sind (vgl. Abschnitt 2.4). Derartige Konfundierungen sollen durch die Randomisierung verhindert werden.

Die Anwendung der Statistik ist hier offenbar sinnvoll. Die o. g. SH entspricht der Nullhypothese ( $H_0$ ) einer Varianzanalyse, wenn vorausgesetzt wird, daß die Gesamtlernzeiten innerhalb aller Bedingungen normal und mit gleichen Varianzen verteilt sind. Wird durch den statistischen Test die  $H_0$  nicht zurückgewiesen, so liegt ein Ergebnis vor, das der Invarianzhypothese nicht widerspricht. Muß dagegen  $H_0$  zurückgewiesen werden, widerspricht der Ausgang des Tests der PH.

Dennoch ist zu fragen, warum die Invarianzhypothese **auf diese** Weise geprüft wird. Denkbar wäre auch ein Vorgehen, bei dem jede Vp unter allen Versuchsbedingungen lernt und sodann geprüft wird, ob die Invarianzhypothese für jede Vp bestätigt werden kann. Ginge man so vor, müßte dennoch auch in diesem Fall Statistik angewendet werden. Wegen der Fehlerbehaftetheit der Messungen sind nicht alle Gesamtlernzeitwerte eines Individuums unter konstanten Lernbedingungen identisch, und der statistische Test hätte zu prüfen, ob trotz dieser Abweichungen ein Ergebnis vorliegt, das  $H_0$  entspricht. Allerdings müßte in einem derartigen Experiment mit wiederholten Messungen aus ersichtlichen Gründen jede Vp je Versuchsbedingung unterschiedliche Lernstoffe bewältigen. Um Aufwärmefekte zu vermeiden, müßte die Untersuchung jeder Vp auf mehrere Tage verteilt werden, was nicht ausschließt, daß sie Strategien der Auseinandersetzung mit der Art des Lernstoffes erwerben, die das spätere Lernen schneller ablaufen lassen. Derartige Aufgaben- und Strategievariablen wären für jedes Individuum mit den interessierenden Versuchsbedingungen vermischt, so daß die Invarianzhypothese je Individuum nicht ohne Ausschaltung dieser Störvariablen geprüft werden könnte. Erreicht werden könnte in einer Untersuchung, in der jede Vp unter allen Bedingungen

$t_j$ , untersucht wird, daß **im Durchschnitt** Aufgaben- und Strategievariablen nicht mit  $t_j$  korrelieren. In diesem Fall ließe sich wieder eine auf interindividuell aggregierte Daten bezogene SH prüfen. Auch die intraindividuelle Variation der Bedingungen schließt also in diesem Beispiel - und in vielen anderen auch - die Überprüfung der PH am einzelnen Individuum aus. Zu viele Störvariablen beeinflussen den Ausgang dieser Prüfung. Das gilt nicht für die Prüfung einer Populationsaussage, die, wie gesagt, sinnvoll ist, wenn sie mit der PH deduktiv Verknüpfbar ist.

Betrachten wir noch ein zweites Beispiel. Die PH sei: „Wer unter Druck eine Meinung äußert, die seiner eigenen Meinung widerspricht, tendiert dazu, seine Meinung in Richtung der Äußerung zu verändern.“ Diese Hypothese gehört zur Theorie der kognitiven Dissonanz. In ihr treten theoretische Terme auf, deren Operationalisierung in einem Experiment von Festinger und Carlsmith (1959) versucht wurde. Darauf ist hier nicht einzugehen (vgl. dazu Abschnitt 4). Geprüft werden könnte diese PH an jedem Individuum, wenn folgendermaßen vorgegangen wird: Zuerst ist ohne Druckanwendung die Meinung zu einem Sachverhalt zu erkunden. Danach muß die Vp unter Druckanwendung dazu gebracht werden, eine ihrer eigenen Auffassung widersprechende Meinung zu äußern, bevor geprüft wird, ob sich die tatsächliche Auffassung geändert hat. Die Gefahr, daß ein Versuchsteilnehmer die Hypothese durchschaut und willentlich Ergebnisse produziert, die ihr entsprechen oder widersprechen, ist groß. Deshalb verbot sich für Festinger und Carlsmith (1959) die wiederholte Erfassung von Meinungen jeder Vp. Sie wiesen die Vpn zufällig den Versuchsbedingungen „Druck“ und „kein Druck“ zu und erhoben zu einem Zeitpunkt, als die Vpn glaubten, das Experiment sei schon beendet, nur einmal die Meinung zu dem Experiment. Verglichen wurden die durchschnittlichen Meinungen der Versuchsgruppen. Das ist sinnvoll, wenn aus der PH folgt, daß im Durchschnitt die unter Druck stehenden Vpn eine als ausgesprochen monoton empfundene Tätigkeit nicht als so langweilig beurteilen wie die Vpn der Kontrollgruppe. Auch hier wurde eine auf Individuen bezogene PH auf dem Umweg über die Prüfung einer Populationsaussage getestet, um den Einfluß von Störvariablen zu kontrollieren.

Derartige Beispiele ließen sich beliebig fortsetzen (vgl. Bredenkamp 1972, 1980). Das Testen von statistischen Populationsaussagen, denen auf Individuen bezogene PHn vorgeordnet sind, ist typisch für weite Bereiche der Psychologie. Sinnvoll ist dieses Vorgehen aber nur dann, wenn eine Verknüpfung zwischen beiden Aussagen konstruiert werden kann. Im Falle der Invarianzhypothese war dies die implikative Verknüpfung von PH und  $H_0$ :  $\mu_1 = \mu_2 = \dots = \mu_m$ . Im Fall der kognitiven Dissonanzhypothese war dies die Implikationsbeziehung  $PH \Rightarrow H$ :  $\mu_1 < \mu_2$ . Selbstverständlich muß diese implikative Beziehung im Einzelfall bewiesen werden (vgl. Abschnitt 2.4), und außerdem müssen bestimmte statistische Annahmen gelten, wenn man als sta-



tistisches Prüfverfahren die Varianzanalyse bzw. den t-Test verwendet. Wenn wir von diesen Annahmen zunächst einmal absehen, läßt sich sagen, daß bei Gültigkeit der implikativen Verknüpfung ein Ergebnis, das für die Verneinung der implizierten SH spricht, notwendig, aber nicht hinreichend für die Falsifikation der PH ist<sup>1</sup>. Eine notwendige, aber nicht hinreichende Bedingung für die Falsifikation liegt vor, weil im Rahmen eines sophistizierten methodologischen Falsifikationismus eine Hypothese erst falsifiziert werden sollte, wenn eine bessere zur Verfügung steht (Lakatos, 1970).

## 1.2 Die Notwendigkeit der simultanen Kontrolle von $\alpha$ und $\beta$

Wenn eine implikative Beziehung zwischen PH und SH vorliegt, ist weiterhin zu gewährleisten, daß neben der statistischen Fehlerwahrscheinlichkeit  $\alpha$  auch die Fehlerwahrscheinlichkeit  $\beta$  kontrolliert wird. Bekanntlich ist  $\alpha$  die Wahrscheinlichkeit für eine falsche Verwerfung von  $H_0$ ,  $\beta$  die für eine falsche Akzeptanz von  $H_0$ . Die Forderung auch der Kontrolle von  $\beta$ , die sich - wie in Abschnitt 2 gezeigt wird - aus der Verknüpfung von PH und SH ergibt, zeigt die Relevanz der Neyman-Pearson-Testtheorie für die psychologische Forschung auf (vgl. Ostmann & Wutke, Kapitel 16 dieses Bandes). Sie erlaubt, vor der Durchführung einer Untersuchung unter Annahme bestimmter Verteilungsvoraussetzungen den Stichprobenumfang so zu bestimmen, daß eine Abweichung bestimmter Größe von  $H_0$  mit der Wahrscheinlichkeit  $1-\beta$  entdeckt werden kann<sup>2</sup>. Auf bestimmte Probleme und Lösungsvorschläge im Zusammenhang mit der Verfolgung von Entscheidungsstrategien im Rahmen der Neyman-Pearson-Testtheorie gehen wir in Abschnitt 3 ein (ausführlich dazu Bredenkamp 1972, 1980; Hager & Westermann, 1983c; Hager, 1987; Ostmann & Wutke, Kapitel 16 dieses Bandes). In Abschnitt 2 wird der Bezug der Wahrscheinlichkeiten  $\alpha$  und  $\beta$  zur Strenge und Fairneß einer Untersuchung dargestellt.

Die implikative PH-SH-Verknüpfung und die Durchführung eines Signifikanztests, der neben  $\alpha$  auch  $\beta$  kontrolliert, können die Falsifikation einer PH selbstverständlich niemals logisch erzwingen. Bei kleinen Fehlerwahrscheinlichkeiten ist aber die Untersuchung so aussagekräftig, daß man sich u. U. zu einer Falsifikation **entschließen** kann. Die Strategie ist also dem methodologischen Falsifikationismus zuzuordnen. Lakatos (1970, S. 109) betont ausdrücklich, daß „the Neyman-Pearson approach rests completely on methodological falsificationism“. Daß in der Psychologie die nach Neyman-Pear-

1 Die Alternative zur implizierten SH hat hier eine ähnliche Funktion wie die falsifizierenden Basissätze in der Falsifikationstheorie von Popper (1982').

2 Die Sequentialstatistik nach Wald (z.B. Wald, 1948) ist in manchen Fällen eine ökonomische Alternative zu Entscheidungsstrategien, die von einer Festlegung des Stichprobenumfangs a priori ausgehen.

sonstrategien testbaren Populationsaussagen häufig nicht die interessierenden PHn sind, liegt nicht an der fehlenden Falsifizierbarkeit im Sinne des dogmatischen Falsifikationismus, sondern ausschließlich daran, daß PHn im Gegensatz zu Populationsaussagen auf individuelle Prozesse abzielen. Die auf Individuen bezogenen PHn wurden von Bredenkamp (1972) als deterministische oder statistische Allaussagen interpretiert, die sich auf eine offene Population von Vpn beziehen. Derartige Aussagen sind notwendig, um dem Erklärungsanspruch genügen zu können (vgl. Groeben & Westmeyer, 1981<sup>2</sup>). Zugleich wirft diese Interpretation aber auch das Problem der Rechtfertigung von Verteilungsannahmen statistischer Tests auf. Hierüber sagen individuenbezogene PHn gewöhnlich nichts aus. Der nächste Abschnitt zeigt, wie dieses Problem im Rahmen der experimentalpsychologischen Forschung gelöst werden kann.

### 1.3 Die Bedeutung von Randomisierungstests

Verteilungsannahmen werden hinfällig, wenn statt der üblichen parametrischen Signifikanztests sog. Randomisierungstests Verwendung finden. Wir erläutern das Prinzip des Randomisierungstests, das von R. A. Fisher stammt und für die experimentalpsychologische Forschung von Edgington (1969) propagiert wurde, anhand einer varianzanalytischen Fragestellung, wie sie z.B. der Prüfung der Invarianzhypothese (vgl. Abschnitt 1.1) zugrunde liegt. Sei  $n_j$  ( $j = 1, \dots, 3$ ) die Anzahl der unter der  $j$ -ten Versuchsbedingung erhobenen Meßwerte. Es sind dann  $(n_1 + n_2 + n_3)! / (n_1! n_2! n_3!)$  Aufteilungen dieser Werte auf die drei Versuchsbedingungen möglich, die unter  $H_0$  alle gleich wahrscheinlich sind. Für jede Aufteilung kann die bekannte varianzanalytische  $F$ -Statistik berechnet werden. Fällt die  $F$ -Statistik für die empirisch tatsächlich eingetretene Aufteilung der Meßwerte unter die  $100 \cdot \alpha$  %-Aufteilungen mit dem größten  $F$ -Wert, wird  $H_0$  zurückgewiesen; andernfalls wird  $H_0$  beibehalten.

Das Verfahren macht keine Verteilungsvoraussetzungen. Es prüft, ob für die **Vpn des Experiments**  $H_0$  zurückzuweisen ist; Verallgemeinerungen auf andere Personen, die nicht am Experiment teilgenommen haben, sind nicht möglich und werden auch nicht angestrebt. In der deduktivistischen Methodologie sind alle Personen, die zum Geltungsbereich der PH gehören, gleichermaßen repräsentativ. Wenn mit mehr als einer Vp experimentiert wird, dann nicht, um die Repräsentativität einer Untersuchung zu erhöhen, sondern um PHn strenger überprüfen zu können.

Der Randomisierungstest paßt ideal zur deduktivistischen Methodologie. Allerdings ist eine Kontrolle des  $\beta$ -Fehlers vor der Durchführung eines Experiments nur unter problematischen Zusatzannahmen und nur mit sehr großem Rechenaufwand möglich (vgl. Gabriel & Hsu, 1983). Da jedoch häufig die

Stichprobenverteilung der Statistik eines Randomisierungstests mit der eines parametrischen Tests approximativ übereinstimmt, hat Bredenkamp (1980) letzteren als rechnerische Vereinfachung eines Randomisierungstests interpretiert. Diese Interpretation, die auf empirischen Robustheitsstudien beruht (Bredenkamp, 1980; Willmes, 1987), verbindet die Vorteile eines Randomisierungstests mit denen der Neyman-Pearson-Theorie für parametrische Tests. Will man sich darauf nicht verlassen, kann man auch tatsächlich Randomisierungstests durchführen und den Stichprobenumfang nach Erwägungen festlegen, die in Abschnitt 3.4 behandelt werden.

## 1.4 Das Theorie-Empirie-Überbrückungsproblem

In PHn werden universelle theoretische Begriffe wie „imaginale Verarbeitung“, „Gedächtnisleistung“, „Empfindungsstärke“ usw. verwendet. Dies ist notwendig, wenn dem Erklärungsanspruch Rechnung getragen werden soll, da Gesetzhypothesen in ihrem Geltungsanspruch über raumzeitliche Beschränkungen hinausgehen müssen. Obwohl die Grenze zwischen Beobachtungs- und Theoriensprache nicht scharf ist, ist die Unterscheidung theoretischer und empirischer Begriffe nützlich in dem Sinne, daß bestimmte Probleme der Operationalisierung angesprochen werden können (vgl. auch Gadenne, Kapitel 7 dieses Bandes). Wir gehen hier nur auf eines dieser Probleme ein, nämlich inwieweit die empirischen Variablen die in der psychologischen Hypothese benannten Vorgänge repräsentieren. U. E. ist das die schwierigste Frage überhaupt, die im hypothetisch-deduktiven Forschungsprozeß auftritt. Die Operationalisierung setzt bestimmte Hilfshypothesen voraus, damit aus Theorie, Anfangsbedingungen und Hilfsannahmen eine Prognose abgeleitet werden kann, die empirisch zu prüfen ist. Oftmals dürfte unklar sein, welche Hilfshypothesen zu präferieren sind. Ob wir z.B. über Instruktionen oder über die Auswahl bestimmter Lernmaterialien imaginale Verarbeitungsprozesse besser induzieren, ist ungewiß. Für diesen Fall wurde von einigen Autoren die **konzeptuelle Replikation** vorgeschlagen. Eine theoretische Variable konzeptuell zu replizieren heißt, sie unterschiedlich zu operationalisieren. Dies kann in ein- und demselben Versuch (intraexperimentelle Replikation) - es liegen dann multifaktorielle Versuche vor, wenn sich die Replikation auf die **W** bezieht, oder multivariate Versuche, wenn sie sich auf die AV bezieht - oder in verschiedenen Versuchen geschehen (interexperimentelle Replikation). Wenn in einem Versuch mehrere Indikatoren für die interessierenden Variablen vorliegen, kann man sie als fehlerbelastete Indikatorvariablen von latenten Variablen ansehen und entsprechende statistische Analysen durchführen (vgl. Abschnitt 4). Allerdings sind dann auch im experimentellen Kontext Randomisierungstests nicht mehr durchführbar, und der Einsatz von statistischen

Verfahren der Modellgeltungsprüfung ist neu zu rechtfertigen (vgl. dazu Abschnitt 4.4).

Bei interexperimenteller Replikation der interessierenden theoretischen Größe(n) kann der Bewährungsgrad der zu testenden PH unter Zuhilfenahme des Bayes'schen Theorems quantifiziert werden. Wenn im einfachsten Fall über alle Untersuchungen a und b konstant sind und vor dem Forschungsprogramm  $H_0$  und  $H_1$  für gleichermaßen wahrscheinlich gehalten wurden, gilt nach t Untersuchungen, daß  $p_t(H_0; \text{Daten}) = Q/(Q+1)$ , wobei  $Q = \frac{a^s A(1-a)^{t-s} \bar{A}^{bt-s}}{s}$  und s die Anzahl signifikanter Resultate ist. Ist  $Q > 1$  und damit  $p_t(H_0; \text{Daten}) > .5$ , bekräftigt die Evidenz eher  $H_0$  als  $H_1$ ; bei  $Q < 1$  ist es umgekehrt. Was hieraus für die inhaltliche Hypothese folgt, hängt davon ab, ob  $H_0$  oder  $H_1$  aus ihr abgeleitet wurde. Das Verfahren funktioniert allerdings nur unter der Annahme, daß die Apriori-Wahrscheinlichkeit für alle Parameter in der Indifferenzzone ( $H'_1$ ) Null ist. Ein Wert  $Q > 1$  muß deshalb nicht für  $H_0$  sprechen, sondern kann auch für  $H'_1$  sprechen, die von einer geringeren Effektstärke als  $H_1$  ausgeht. Dies kann ohne weitere Datenerhebung untersucht werden, indem für die neu festgelegte Effektstärke das  $\beta$  in jeder Untersuchung berechnet und Q neu gebildet wird. Ist Q nunmehr kleiner als 1, wird  $H'_1$  bekräftigt, was darauf hinweist, daß vor dem Forschungsprogramm die Effektstärke zu groß angesetzt wurde.

## 1.5 Die Bedeutung der Situationsvalidität

Der Geltungsanspruch für PHn wird nie auf bestimmte Versuchsleiter, Versuchsräume usw. relativiert. Hat man Vermutungen darüber, daß derartige Gültigkeitseinschränkungen bestehen oder daß eine Hypothese nur für Personen mit bestimmten Merkmalen gilt, muß dies über den Test sog. disordinaler Wechselwirkungen geprüft werden. Eine Wechselwirkung zweier Wn A und B heißt disordinal, wenn a) die Rangordnung der Erwartungswerte über den B-Faktor sich zwischen den Stufen des A-Faktors unterscheidet, b) sich die Rangordnung der Erwartungswerte über den A-Faktor zwischen den Stufen des B-Faktors unterscheidet oder c) beides der Fall ist (vgl. z.B. Bredenkamp, 1980)<sup>3</sup>. Der Nachweis, daß eine Wechselwirkung disordinal ist, ist nicht einfach zu führen. Verfahren zur Bestimmung der Disordinalität sind Bestandteile einer deduktivistischen Methodologie (vgl. dazu Bredenkamp, 1982; Hager & Westermann, 1983c; Hager, 1987, 1992).

3 Bredenkamp (1980) differenziert darüber hinaus zwischen semidisordinalen Wechselwirkungen (a, b) und vollständig disordinalen Wechselwirkungen (c).

## 2. Eine formale Analyse der statistischen Prüfung psychologischer Hypothesen

Die in Abschnitt 1.1 geschilderten Probleme einer „direkten“ empirischen Beurteilung von PHn anhand von Beobachtungen am einzelnen Individuum sind u. E. ein wesentlicher Grund dafür, daß die zuerst in den Agrarwissenschaften populär gewordene Idee statistischer Hypothesentests in der Psychologie begeistert aufgegriffen und angewandt wurde. Kann jedoch auf diese Weise die in der Einleitung dieses Kapitels geforderte „strenge und zugleich faire“ Prüfung von PHn erreicht werden? Um diese Frage beantworten zu können, muß zunächst geklärt werden, wie eine „möglichst strenge und zugleich faire“ **statistische** Prüfung von PHn zu explizieren ist. Im Anschluß daran ist zu untersuchen, welche methodologischen Konsequenzen eine geeignete Erklärung nach sich zieht.

### 2.1 Was ist eine strenge und faire statistische Prüfung psychologischer Hypothesen?

Westermann (1987, S. 37f.) und Hager (1987, 1992) haben im Anschluß an Arbeiten von Hager und Westermann (1983a, 1983b) sowie Westermann und Hager (1986) ein auf Gadenne (1976) zurückgehendes Strengeskriterium dadurch konkretisiert, daß sie zunächst zwei Fehlerwahrscheinlichkeiten  $e_U$ , und  $f_U$  eingeführt haben. Dabei ist  $e_U$  als die Wahrscheinlichkeit eines **fälschlichen Bewährungsurteils** und  $f_U$  als die Wahrscheinlichkeit eines **fälschlichen Nichtbewährungsurteils bzgl. einer zu prüfenden PH** in einer konkreten empirischen Untersuchung U definiert:

$$e_U := p(\text{PH gilt in U als bewährt; PH trifft in U nicht zu}) \quad (2.1)$$

$$f_U := p(\text{PH gilt in U als nichtbewährt; PH trifft in U zu}). \quad (2.2)$$

Das Zeichen „;“ ist hier als „unter der Annahme der Gültigkeit von“ zu lesen; bedingte Wahrscheinlichkeiten werden demgegenüber durch das Zeichen „|“ ausgedrückt. Man kann nun den **Grad der Strenge der Prüfung von PH in U** gleichsetzen mit der Wahrscheinlichkeit  $1-e_U$  für korrekte Nichtbewährungsurteile und analog die **Fairneß der Prüfung von PH in U** mit der Wahrscheinlichkeit  $1-f_U$  für korrekte Bewährungsurteile. Es ist leicht zu erkennen, daß eine isolierte Forderung nach „maximaler Strenge“ methodologisch nicht sinnvoll ist. Ihr könnte ja ganz einfach dadurch entsprochen werden, daß man unabhängig von der Befundlage ausschließlich Nichtbewährungsurteile bzgl. PH abgibt. Dies aber hieße, daß keinerlei Information über den Wahrheitswert von PH vermittelt wird. Westermann (1987) und Hager (1987) ersetzen deshalb die Forderung nach **Strenge** durch die Forderung nach **Validität**. Validität

ist für sie der **Durchschnitt von Streng und Fairneß**. Diese Definition hat den Nachteil, daß mangelnde Streng durch erhöhte Fairneß und umgekehrt mangelnde Fairneß durch erhöhte Streng kompensierbar ist. Will man sicherstellen, daß eine Untersuchung nur dann valide genannt wird, wenn sie sowohl streng als auch fair ist, empfiehlt sich ein komparativer Validitätsbegriff der folgenden Art (vgl. auch Hager & Westermann, 1983b, S.69/70; eine graphische Veranschaulichung liefert Abbildung 1):

**Definition 1.** Seien U und V zwei empirische Untersuchungen zur Überprüfung von PH; dann erfolgt in U genau dann eine validere Prüfung von PH als in V (kurz: U ist **valider als** V), wenn  $e_U < e_V$  oder  $f_U < f_V$ , aber weder  $e_U > e_V$  noch  $f_U > f_V$ .

In Worten ausgedrückt heißt dies, daß mindestens eine der beiden Fehlerwahrscheinlichkeiten in U kleiner und keine größer als in V ist.

**Definition 2.** U und V können ferner **gleich valide** (zur Überprüfung von PH) genannt werden, wenn  $e_U = e_V$ , sowie  $f_U = f_V$  gilt.

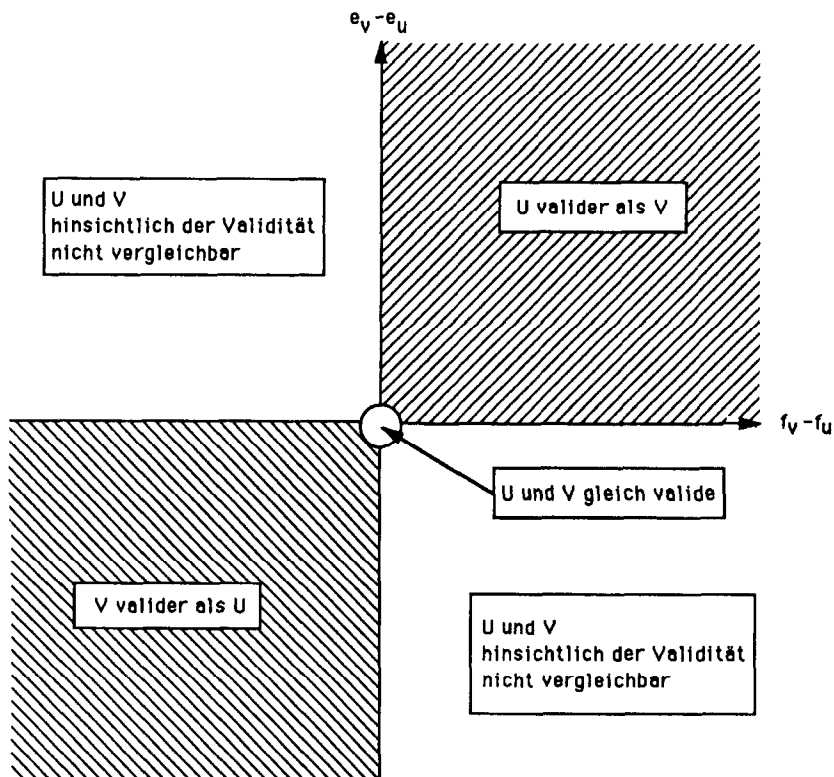


Abb. 1: Graphische Veranschaulichung verschiedener komparativer Validitätsbegriffe

Der Nachteil dieser Definitionsvorschläge, daß zwei Untersuchungen **hinsichtlich der Validität nicht** vergleichbar sein können (wenn U nicht valider als V, V nicht valider als U sowie U und V nicht gleich valide sind), fällt gegenüber dem o.g. Vorteil u.E. nicht ins Gewicht.

Wir werden im folgenden die Forderung nach möglichst strengen und zugleich fairen Hypothesenprüfungen mit der Forderung nach Validität im soeben definierten Sinne gleichsetzen (vgl. auch Hager & Westermann, 1983a, 1983b; Erdfelder, 1993a, 1993b). Im Kontext **statistischer** Prüfungen von PHn ist nun allerdings die empirische Untersuchung U noch genauer zu spezifizieren. Entscheidend ist hier vor allem das Konzept der **statistischen Hypothese**, die in noch näher zu spezifizierender Weise aus der PH gewonnen werden muß. Hiermit ist nicht nur eine einzelne Parameterhypothese (z. B. Mittelwertshypothese) gemeint, sondern eine Parameterhypothese oder eine logische Verknüpfung verschiedener Parameterhypothesen in einem bestimmten **empirisch interpretierten** statistischen Modell. Wenn eine bestimmte statistische Hypothese  $SH_U$  spezifiziert wird, ist also immer auch eine konkrete Untersuchungssituation einschließlich der unabhängigen und abhängigen (Zufalls-)Variablen, des Versuchsplans und des Versuchsablaufs damit verbunden. Darüber hinaus ist mit der  $SH_U$  ein **Modell** festgelegt (oder eine statistische Oberhypothese sensu Stegmüller, 1973), auf dessen Hintergrund die Daten statistisch analysiert werden. Hierbei wird es sich häufig um einen Spezialfall des Allgemeinen Linearen Modells oder des Log-Linearen Modells handeln; prinzipiell sind aber auch andere statistische Rahmenmodelle als Oberhypothesen denkbar.

Zusätzlich zur  $SH_U$  ist eine **Entscheidungsstrategie**  $ES_U$  festzulegen, die angibt, welche Klasse denkbarer empirischer Daten in U zu einer Annahme und welche zu einer Ablehnung der  $SH_U$  führt<sup>4</sup>. Das Paar  $(SH_U, ES_U)$  bildet dann die empirisch-statistische Untersuchung, mittels der die PH beurteilt werden soll. Die Beurteilung geschieht in der Weise, daß eine Annahme der  $SH_U$  zu einem Bewährungsurteil und eine Ablehnung der  $SH_U$  zu einem Nichtbewährungsurteil bzgl. PH führt. Die statistische Entscheidung über  $SH_U$  wird also gewissermaßen auf die PH übertragen.

Die Forderung nach validen Untersuchungen kann nun wie folgt konkretisiert werden:

**Regel  $R_1$ :** Zu einer gegebenen PH wähle  $SH_U$  und  $ES_U$  derart, daß  $(SH_U, ES_U)$  möglichst valide ist.

4 Die Entscheidungsstrategie entspricht dem, was man gemeinhin unter einem „statistischen Test“ versteht. Sie umfaßt die Festlegung einer Teststatistik T, einer Entscheidungsregel  $d(T = t)$  und einer Menge von Entscheidungsalternativen, die hier als  $A = \{H_0, H_1\}$  vorgegeben ist (vgl. dazu Ostmann & Wutke, Kapitel 16 dieses Bandes, Abschnitt 2).

## 2.2 Determinanten von Strenge und Fairneß

Die Umsetzung der methodologischen Regel  $R_1$  verlangt eine Wahl von  $SH_U$  und  $ES_U$  in der Weise, daß die Wahrscheinlichkeiten fälschlicher Bewährungs- und fälschlicher Nichtbewährungsurteile bzgl. PH möglichst klein sind. Doch wie lassen sich diese Fehlbeurteilungswahrscheinlichkeiten reduzieren oder u. U. sogar kontrollieren, so daß sie bestimmte (kleine) Werte nicht überschreiten? Erdfelder (1993a) hat gezeigt, daß  $e_U$  und  $f_U$  von vier Parametern abhängen, die wir in dieser Arbeit wie folgt einführen wollen:

$$\varepsilon := p(\text{Annahme von } SH_U \mid SH_U \text{ trifft nicht zu}) \quad (2.3)$$

$$\phi := p(\text{Ablehnung von } SH_U \mid SH_U \text{ trifft zu}) \quad (2.4)$$

$$g := p(SH_U \text{ trifft zu; PH trifft in } U \text{ nicht zu}) \quad (2.5)$$

$$h := p(SH_U \text{ trifft nicht zu; PH trifft in } U \text{ zu}) \quad (2.6)$$

Die beiden zuerst aufgeführten Fehlerwahrscheinlichkeiten, die wir dem Vorschlag Hagers (1987, 1992; siehe auch Hager & Westermann, 1983a, 1983b) folgend mit  $\varepsilon$  und  $\phi$  notieren, sind **statistische Irrtumswahrscheinlichkeiten**, d. h. **Wahrscheinlichkeiten für fälschliche Annahmen und Ablehnungen von  $SH_U$**  bei Anwendung der festgelegten Entscheidungsstrategie  $ES_U$ . Ist z.B.  $SH_U$  eine statistische Alternativhypothese ( $H_1$ ), so gilt  $\varepsilon = \alpha$  und  $\phi = \beta$ , wobei  $\alpha$  und  $\beta$  die bekannten statistischen Irrtumswahrscheinlichkeiten erster bzw. zweiter Art sind. Umgekehrt gilt  $\varepsilon = \beta$  und  $\phi = \alpha$ , wenn  $S_U$  eine  $H_0$  ist. Ist  $SH_U$ : allgemein eine bestimmte Konjunktion, Disjunktion oder eine andere Verknüpfung verschiedener Null- und/oder Alternativhypothesen, so sind  $\varepsilon$  und  $\phi$  mehr oder minder komplizierte Funktionen der Irrtumswahrscheinlichkeiten der durchgeführten Einzeltests. Die Kontrolle von  $\varepsilon$  und  $\phi$  bereitet - wie sich zeigen läßt - vielleicht praktische, aber jedenfalls keine prinzipiellen Probleme (vgl. bereits Hager & Westermann, 1983a, Westermann & Hager, 1986). Wir werden auf diesen Punkt in Abschnitt 3.3 zurückkommen.

In diesem Abschnitt werden wir uns vor allem den in den Gleichungen (2.5) und (2.6) eingeführten Wahrscheinlichkeiten  $g$  und  $h$  widmen. Beide Wahrscheinlichkeiten betreffen das Verhältnis von PH und daraus gewonnener  $SH_U$ . Sie können am einfachsten als Wahrscheinlichkeiten für Wahrheitswertdiskrepanzen zwischen PH und  $SH_U$  charakterisiert werden:  $g$  ist die Wahrscheinlichkeit einer wahren  $SH_U$  im Falle einer ungültigen PH,  $h$  umgekehrt die Wahrscheinlichkeit einer falschen  $SH_U$  im Falle einer zutreffenden PH (Erdfelder, 1993a).

Wie hängen nun  $e_U$  und  $f_U$  von den o.g. Parametern ab? Es muß lediglich untersucht werden, auf welchen Wegen man zu Fehlbeurteilungen der zu testenden PH gelangen kann (vgl. im Detail Erdfelder, 1993a). Betrachten wir



zunächst die Fehlerwahrscheinlichkeit  $e_U$ . Wenn PH in U nicht zutrifft, kommt es zu einer Fehlbeurteilung genau dann, wenn  $SH_U$  angenommen wird. Dies ist auf zwei Wegen möglich<sup>5</sup>: (a)  $SH_U$  trifft zu und wird korrekt angenommen und (b)  $SH_U$  trifft nicht zu, wird aber dennoch fälschlich angenommen. Die Wahrscheinlichkeit für Weg (a) ist gleich dem Produkt der Wahrscheinlichkeiten  $g$  und  $(1-\phi)$ , die für Weg (b) gleich dem Produkt der entsprechenden Wahrscheinlichkeiten  $(1-g)$  und  $\epsilon$ . Da die Wege disjunkt sind, addieren sich die Wahrscheinlichkeiten beider Wege zu  $e_U$  auf:

$$\begin{aligned} e_U &= p(\text{Annahme von } SH_U; \text{ PH trifft in U nicht zu}) \\ &= g \cdot (1-\phi) + (1-g) \cdot \epsilon \\ &= \epsilon + (1-\epsilon-\phi) \cdot g. \end{aligned} \quad (2.7)$$

Wir sehen, daß  $e_U$  für gegebene statistische Irrtumswahrscheinlichkeiten  $\epsilon$  und  $\phi$  eine lineare Funktion von  $g$  mit Ordinatenabschnitt  $\epsilon$  und Steigung  $1-\epsilon-\phi$  ist. Da diese Steigung niemals negativ werden kann (vgl. Erdfelder, 1993b), ist  $\epsilon$  zugleich der kleinste Wert, den  $e_U$  annehmen kann. Mit Wachsen  $g$  wächst auch  $e_U$ . Falls  $g$  und  $\phi$  gegeben sind und  $g$  kleiner als 1 ist, können wir zusätzlich sagen, daß  $e_U$  mit  $\epsilon$  wächst.

Eine analoge Analyse läßt sich für  $f_U$  durchführen. Auch hier können zwei Wege zu einer falschen Beurteilung der PH führen, die in diesem Fall aus der Ablehnung von  $SH_U$  resultiert. Weg (a) führt über eine gültige  $SH_U$  zu einer fälschlichen Verwerfung von  $SH_U$  und hat die Wahrscheinlichkeit  $(1-h) \cdot \phi$ ; Weg (b) mit Wahrscheinlichkeit  $h \cdot (1-\epsilon)$  führt über eine ungültige  $SH_U$  zu einer korrekten Zurückweisung. Analog zu (2.7) ergibt sich dann:

$$\begin{aligned} f_U &= p(\text{Verwerfung von } SH_U; \text{ PH trifft in U zu}) \\ &= (1-h) \cdot \phi + h \cdot (1-\epsilon) \\ &= \phi + (1-\phi-\epsilon) \cdot h. \end{aligned} \quad (2.8)$$

Für gegebene statistische Irrtumswahrscheinlichkeiten ist (2.8) eine linear steigende Funktion von  $h$  mit dem kleinstmöglichen Wert  $\phi$ . Falls  $h < 1$  ist, wächst  $f_U$  unter sonst gleichen Bedingungen mit  $\phi$  und fällt mit  $\epsilon$ .

<sup>5</sup> Um die Darstellung zu vereinfachen, wird hier angenommen, daß  $H_0$  und  $H_1$  den Parameterraum ausschöpfen. Die folgenden Formeln (2.7) und (2.8) sind zu erweitern, wenn diese Annahme fallengelassen wird. Die methodologischen Folgerungen in Abschnitt 2.3 werden hierdurch nicht wesentlich tangiert.

### 2.3 Methodologische Folgerungen aus der Forderung nach Strenge und Fairneß

Hager und Westermann (erstmalig 1983a, 1983b) haben behauptet, daß  $e_U$  und  $f_U$  durch die statistischen Irrtumswahrscheinlichkeiten  $\epsilon$  und  $\phi$  beeinflußt werden können. Diese Behauptung ist richtig, wie die Gleichungen (2.7) und (2.8) zeigen. Zugleich wird aber deutlich, daß die Art der Abhängigkeit zwischen statistischen und PH-Fehlbeurteilungswahrscheinlichkeiten keineswegs eine einfache ist. Sehen wir einmal von den Grenzfällen  $g = 1$  und  $h = 1$  ab, trifft es zwar zu, daß - ceteris paribus - eine Herabsetzung von  $\epsilon$  eine Herabsetzung von  $e_U$  und eine Reduzierung von  $\phi$  eine Reduzierung von  $f_U$  bewirkt (vgl. z.B. Hager, 1987, 1992 und Hussy & Möller, Kapitel 11 dieses Bandes). Dar- aus kann aber nicht ohne weiteres geschlossen werden, daß die Reduzierung der statistischen Irrtumswahrscheinlichkeiten eine Erhöhung der Validität zur Folge hat, denn eine Herabsetzung von  $\epsilon$  führt leider i.a. auch zu einer Erhöhung von  $f_U$  ebenso wie eine Reduzierung von  $\phi$  eine Erhöhung von  $e_U$  zur Folge hat. Wir können vorerst festhalten:

**Beobachtung 1.** Seien U und V zwei Untersuchungen mit  $g > 0$  und  $h > 0$ , die sich nur darin unterscheiden, daß genau eine der beiden statistischen Irrtumswahrscheinlichkeiten in U kleiner gewählt wurde. Dann sind U und V hinsichtlich der Validität nicht vergleichbar. Insbesondere ist ausgeschlossen, daß U valider als V ist.

Angenommen, wir streben eine Erhöhung der Validität durch Reduzierung beider Irrtumswahrscheinlichkeiten  $\epsilon$  und  $\phi$  an. Unter welchen Bedingungen ist dies möglich? Wenn wir  $\epsilon$  um den Betrag  $k_\epsilon$  ( $0 < k_\epsilon < \epsilon$ ) verkleinern, führt dies einerseits zu einer Reduzierung von  $e_U$ , andererseits aber auch zu einem Anwachsen von  $f_U$ . Dieser unerwünschte Anstieg muß durch Reduzierung der zweiten statistischen Irrtumswahrscheinlichkeit  $\phi$  kompensiert werden. Man kann zeigen (Erdfelder, 1993b), daß die Verkleinerung von  $\phi$  um den Betrag  $k_\phi = h / (1-h) \cdot k_\epsilon$  die gerade notwendige Kompensation leistet, d.h. ein Anwachsen von  $f_U$  verhindert. Natürlich führt diese  $\phi$ -Änderung wiederum zu einem unerwünschten Anstieg von  $e_U$ , der nur dann kleiner als die durch  $\epsilon$ -Änderung erzielte Verbesserung ist, falls  $g + h < 1$  (Erdfelder, 1993b). Das gleiche Resultat ergibt sich, wenn man eine Validitätsverbesserung zunächst durch Reduzierung von  $f_U$  über  $\phi$  versucht. Wir können demnach konstatieren:

**Beobachtung 2.** Seien U und V zwei Untersuchungen, die sich lediglich darin unterscheiden, daß  $\epsilon$  und  $\phi$  in U kleiner als in V gewählt wurden. Dann kann U nur dann valider als V sein, falls  $g + h < 1$ .

Die Forderung nach Reduktion der statistischen Irrtumswahrscheinlichkeiten ist also lediglich bei Gültigkeit einer Randbedingung aus der Forderung nach Validität ableitbar. Ehe wir kleinere  $\varepsilon$ - und  $\phi$ -Werte fordern, muß sichergestellt sein, daß die Wahrscheinlichkeiten  $g$  und  $h$  für Wahrheitswertdiskrepanzen zwischen  $PH$  und  $SH_U$  zusammen nicht größer als 1 sind.

Ein naheliegender Gedanke wäre es, mit folgender Forderung zu beginnen:

**Regel  $R_2$ :** Zu einer gegebenen  $PH$  wähle  $SH_U$  so, daß die Wahrscheinlichkeiten  $g$  und  $h$  den Wert 0 annehmen.

Diese Forderung folgt aus Regel  $R_1$ . Die oben erwähnten Eigenschaften der Gleichungen (2.7) und (2.8) garantieren, daß von zwei Untersuchungen, die sich lediglich hinsichtlich der Wahrscheinlichkeiten  $g$  oder  $h$  unterscheiden, grundsätzlich diejenige valider sein wird, die Regel  $R_2$  genügt. Aber wie kann gezeigt werden, daß eine Untersuchung Regel  $R_2$  genügt? Dies kann nur über einen Äquivalenzbeweis geschehen, der Wahrheitswertdiskrepanzen zwischen  $PH$  und  $SH_U$  prinzipiell ausschließt. Praktisch bedeutet die Umsetzung der Regel  $R_2$  daher einerseits eine stochastische Formulierung der  $PH$  und andererseits eine Spezifikation von  $SH_U$  derart, daß  $SH_U$  zugleich notwendig und hinreichend für die Gültigkeit von  $PH$  in  $U$  ist. Der Äquivalenzbeweis ist im Kalkül der Stochastik zu erbringen.

Diese Forderungen erscheinen auf den ersten Blick ungewöhnlich streng, so daß es angezeigt ist, sich mit ihrer Realisierbarkeit zu beschäftigen. Zunächst muß festgehalten werden, daß es zur stochastischen Beweisführung keine Alternative gibt, wenn man sich einmal auf die methodologische Regel  $R_1$  in Verbindung mit den Definitionen 1 und 2 festgelegt hat. Die Definition von Strenge, Fairneß und Validität in termini von Wahrscheinlichkeiten legt die formale Sprache fest, in der die Argumentation zu führen ist. Ist die formale Sprache festgelegt, so muß die  $PH$  in eben diese Sprache „übersetzt“<sup>6</sup> werden, sofern sie nicht von vornherein schon in dieser Sprache formuliert wurde. Andernfalls sind Behauptungen über Implikations- oder Äquivalenzbeziehungen zwischen  $PH$  und  $SH_U$  weder beweisbar noch widerlegbar; wie Regel  $R_2$  zeigt, können wir aber hierauf nicht verzichten.

Glücklicherweise ist mit der Festlegung auf die Stochastik ein äußerst flexibler Rahmen gegeben. In dieser Sprache lassen sich deterministische und nichtdeterministische Aussagen über qualitative (diskrete) und quantitative ZVn formulieren, die beobachtbar (manifest) oder nicht beobachtbar (latent) sein können (einführend z.B. Steyer, 1988, Kapitel 15 dieses Bandes). Tatsächlich sind

<sup>6</sup> Es ist klar, daß ein solcher „Übersetzungsvorgang“ im Regelfall auch eine Explikation der zumeist umgangssprachlich formulierten  $PH$  beinhalten muß. Dies ist auf die zumeist gegebene Vagheit und Mehrdeutigkeit der ursprünglichen Formulierung in einer nicht-formalen Sprache zurückzuführen und insofern nicht zu vermeiden.

PHn, die in diesem Rahmen nicht angemessen formulierbar sind, nur schwer vorstellbar. So überrascht es nicht, daß PHn, zu denen empirische Überprüfungen vorgeschlagen wurden, die Regel  $R_2$  genügen, in der psychologischen Literatur einen immer breiteren Raum einnehmen. Ein prominentes und besonders typisches Beispiel bilden verschiedene probabilistische Modelle additiv-verbundener Messung (vgl. Falmagne, 1979), die u.a. zur Überprüfung von PHn über binaurale Lautheitssummation in der Psychoakustik verwendet wurden (z.B. Falmagne, Iverson & Markovici, 1979).

Allerdings ist es nicht leicht, eine  $SH_U$  zu finden, die Regel  $R_2$  genügt. Vor allem der Nachweis der Implikation  $SH_U \text{ ti } PH$  ist normalerweise nicht trivial, so daß man u.U. behelfsweise an schwächere Formen der Überprüfung denken muß, als sie Regel  $R_2$  verlangt. Auch andere Gründe können ausschlaggebend dafür sein, die Forderung nach Äquivalenz von PH und  $SH_U$  fallenzulassen: Wie wir in Abschnitt 1.1 gesehen haben, ist oftmals eine Überprüfung von PHn auf Individuumsebene prinzipiell ausgeschlossen, so daß zwangsläufig auf die Aggregatebene übergegangen werden muß. Aggregataussagen können aber bestenfalls notwendige, niemals aber zugleich auch hinreichende Bedingungen für die Gültigkeit (allgemeinpsychologischer) PHn sein, die für die Individuumsebene Gültigkeit beanspruchen (vgl. Bredenkamp, 1972, 1980). Will man den in der Psychologie sehr häufig anzutreffenden und oftmals gar nicht vermeidbaren Übergang auf die Aggregatebene nicht von vornherein als sinnlos abtun, wird man sich der folgenden Abschwächung von Regel  $R_2$  zuwenden müssen:

**Regel  $R_2$ :** Zu einer gegebenen PH wähle  $SH_U$  so, daß die Wahrscheinlichkeit  $h$  den Wert 0 annimmt.

Praktisch wird mit  $R_2$  die Forderung nach Äquivalenz von PH und  $SH_U$  auf die Forderung nach einer **Implikationsbeziehung**  $PH \Rightarrow SH_U$  reduziert. Diese Regel dürfte im Falle einer präzise formulierten, empirisch gehaltvollen PH wesentlich leichter als  $R_2$  zu erfüllen sein.

Welche Konsequenzen hat die abgeschwächte Regel  $R_2$ ? Wenn die Wahrscheinlichkeit  $g$  keiner Normierung unterworfen wird, so heißt das nach Gleichung (2.7) zunächst, daß die Wahrscheinlichkeit eines fälschlichen Bewährungsurteils für die PH nicht kontrolliert werden kann. **Nur Nichtbewährungsurteile können demnach bei kontrollierter Fehlerwahrscheinlichkeit  $f_U = \phi$  erfolgen (vgl. Gleichung 2.8), nicht aber Bewährungsurteile.** Dennoch ist Regel  $R_2$  streng genug, um eine Erhöhung der Validität über die Reduktion von statistischen Irrtumswahrscheinlichkeiten zu ermöglichen. Die Validität kann problemlos durch Verkleinerung von  $\epsilon$  erhöht werden, da diese Maßnahme  $e_U$  senkt (falls  $g < 1$ ) und  $f_U$  nicht tangiert. Prinzipiell kann auch an eine Validitätserhöhung über Reduzierung von  $\phi$  um den Betrag  $k_\phi$  in Verbindung mit einem mindestens um den Betrag  $g/(1-g) \cdot k_\phi$  verringerten  $\epsilon$

gedacht werden. Diese Maßnahme würde  $f_U$  senken und sicherstellen, daß  $e_U$  zumindest nicht erhöht wird. Da  $g$  jedoch im hier interessierenden Fall unbekannt ist, nutzt die angegebene E-Adjustierungsformel wenig. Es empfiehlt sich deshalb,  $\phi$  nicht zu klein zu wählen, da man ansonsten Gefahr läuft, zu viele fälschliche Bewährungsurteile zu erhalten.

Fassen wir zusammen. Aus dem Validitätsmaximierungspostulat  $R_1$ , von dem wir ausgegangen sind, folgt, daß wann **immer möglich** Regel  $R_2$  angewendet werden sollte. In Verbindung mit  $R_2$  folgt aus  $R_1$  auch die methodologische Regel:

**Regel  $R_3$ :** Zu einer PH wähle eine Untersuchung  $U$  mit einer Entscheidungsstrategie  $ES_U$  so, daß die statistischen Irrtumswahrscheinlichkeiten  $\epsilon$  und  $\phi$  kontrollierbar und möglichst klein sind.

Die Kombination von  $R_2$  und  $R_3$  stellt sicher, daß PH-Bewährungsurteile bei kontrollierter (kleiner) Fehlerwahrscheinlichkeit  $e_U = \epsilon$  und Nichtbewährungsurteile bei kontrollierter (kleiner) Fehlerwahrscheinlichkeit  $f_U = \phi$  erfolgen.

Läßt sich dieses ideale Vorgehen nicht realisieren, so empfiehlt sich die Realisierung der abgeschwächten Regel  $R_2$  in Verbindung mit der methodologischen Regel:

**Regel  $R_3$ :** Zu einer PH wähle eine Untersuchung  $U$  mit einer Entscheidungsstrategie  $ES_U$  so, daß  $\phi$  und  $\epsilon$  kontrollierbar sind. Ferner sollte  $\epsilon$  möglichst klein gewählt werden.

Die Kombination von  $R_2$  und  $R_3$  stellt die Kontrolle der Wahrscheinlichkeit  $f_U = \phi$  für fälschliche Nichtbewährungsurteile sicher. Die Wahrscheinlichkeit  $e_U$  für fälschliche Bewährungsurteile kann mangels Normierung von  $g$  zwar nicht kontrolliert, aber u.U. (in unbekanntem Ausmaß) reduziert werden, wenn man die folgende Forderung ergänzt:

**Regel  $R_4$ :** Zu einer PH wähle  $SH_U$  so, daß  $g$  möglichst klein (und  $U$  somit möglichst streng) ist.

Aus Regel  $R_4$  kann abgeleitet werden, welche von mehreren aus einer PH deduzierbaren SHn zur Überprüfung herangezogen werden sollte. Es gilt (Erdfelder, 1993b):

**Beobachtung 3.** Seien  $(SH_U, ES_U)$  und  $(SH_V, ES_V)$  zwei Untersuchungen zur Prüfung einer PH, die beide der Regel  $R_2$  genügen. Es gelte ferner  $SH_U \Rightarrow SH_V$ , nicht aber  $SH_V \Rightarrow SH_U$ . Dann ist  $g$  in  $U$  kleiner als in  $V$  und - sofern  $\epsilon$  und  $\phi$  sich zwischen beiden Untersuchungen nicht unterscheiden-  $U$  somit auch valider als  $V$ .

Man sollte also immer die „bestimmtere“ (empirisch gehaltvollere) SH als Überprüfungsinstanz heranziehen, sofern Wahlmöglichkeiten bestehen (vgl. auch Bredenkamp, 1984). Diese Maßnahme erhöht die Strenge einer Untersuchung, ohne die Fairneß zu senken.

Die Kombination der Regeln  $R_2$ ,  $R_3$  und  $R_4$  ist die zweitbeste Umsetzung der Regel  $R_1$ , die denkbar ist. Andere Umsetzungen von  $R_1$ , die nicht auf der Forderung basieren, daß  $g$  oder  $h$  Null sind, sind zwar formal konzipierbar, dürften aber mangels zwingender Begründungen für bestimmte Wahrscheinlichkeitswerte von  $g$  und  $h$ , die dann größer Null und kleiner Eins sein müßten, nicht realisierbar sein. Wie wir anhand von Beobachtung 2 sehen, ist aber nicht einmal die Forderung nach kleinen statistischen Irrtumswahrscheinlichkeiten  $\epsilon$  und  $\phi$  gerechtfertigt, wenn die Möglichkeit besteht, daß  $g$  und  $h$  zusammen größer als 1 sind.

## 2.4 Randomisierte oder nichtrandomisierte Untersuchungen?

Die im letzten Abschnitt besprochenen Folgerungen aus der Forderung nach möglichst validen Untersuchungen beinhalten Antworten auf konkrete Probleme der psychologischen Versuchsplanung, die den Regeln nicht auf den ersten Blick anzusehen sind. Ein Beispiel hierfür ist die Frage, ob von der Kontrolltechnik der Randomisierung Gebrauch zu machen ist (vgl. Strack & Rehm, Kapitel 12 dieses Bandes). Als Begründung für die Notwendigkeit der Randomisierung wird häufig genannt, daß allein randomisierte Untersuchungen eine kausale Interpretation eines eventuellen Effekts der UV(n) auf die AV(n) ermöglichen. Diese Begründung ist natürlich genauso problematisch wie der Kausalitätsbegriff, auf den sie Bezug nimmt. Da der Kausalitätsbegriff wissenschaftstheoretisch nach wie vor als nicht geklärt betrachtet werden kann und insbesondere unklar ist, in welchem Sinne experimentellen Befunden kausale Interpretierbarkeit zugesprochen werden kann, hat Steyer (1992 und Kapitel 15 dieses Bandes) einige Kausalitätsbegriffe vorgeschlagen, die mit dem (randomisierten) Experiment sinnvoll in Verbindung zu bringen sind. Einer der Steyerschen Kausalitätsbegriffe - der Begriff der schwachen kausal-regressiven Abhängigkeit - ist erfüllt, wenn die interessierende UV und potentielle Störvariablen stochastisch unabhängig sind. Bekanntlich stellt aber das Randomisierungsprinzip genau diese Unabhängigkeit von UV(n) und Störvariablen sicher. Insofern kann man mit Steyer (1992) feststellen, daß Randomisierung immer dann angestrebt werden sollte, wenn eine kausale Interpretierbarkeit der Versuchsergebnisse erwünscht ist.

Man kann nun weiterfragen, **warum** eine kausale Interpretierbarkeit der Versuchsergebnisse erwünscht sein sollte. Auf dem in den vorstehenden Abschnitten dargelegten methodologischen Hintergrund läßt sich diese Frage wie folgt

beantworten: **Randomisierung ist im Regelfall deshalb anzustreben, weil für viele PHn nur auf diese Weise sichergestellt werden kann, daß die Implikation  $PH \Rightarrow SH_U$  und damit  $h = 0$  erfüllt ist. Hieraus folgt: Relativ zu einer nicht-randomisierten Untersuchung erhöht Randomisierung die Fairneß ( $1-e_U$ ) und damit die Validität einer Untersuchung.**

Zur Begründung wollen wir den einfachen, aber dennoch typischen Fall einer PH betrachten, die einen Effekt einer dichotomen UV X auf eine AV Y behauptet. Als Beispiel, das auch von Hager (1987) zur Illustration herangezogen wird, sei die PH „Konkrete Wörter ( $X = 1$ ) werden besser behalten als abstrakte Wörter ( $X = 0$ )“ herangezogen. Die AV Y entspricht in diesem Fall dem Gedächtnismaß (z. B. einem Reproduktions- oder Rekognitionsmaß), auf das sich die PH bezieht. Bei Licht besehen enthalten derartige PHn oft noch zwei Zusatzannahmen, die zumeist nicht explizit genannt, aber durchaus mitgedacht werden: (a) Neben dem X-Effekt gibt es möglicherweise einen additiven Effekt unbekannter Fehlervariablen mit Erwartungswert 0, so daß nicht Y, sondern der bedingte Erwartungswert von Y unter X linear von X abhängt; (b) die lineare Beziehung zwischen bedingtem Erwartungswert und X gilt *ceteris paribus*, d.h. bei Konstanzhaltung anderer Störeinflüsse auf Y. Bei den Störgrößen kann es sich z.B. um Persönlichkeitseigenschaften der Vpn oder auch um situative Kontextbedingungen handeln, die neben der interessierenden **W** Einfluß auf die interessierende AV der Untersuchung nehmen.

Bezeichnen wir die (u.U. mehrdimensionale) unbekannte Störgröße mit W und ihre möglichen Realisationen mit w, so lautet eine angemessene Formulierung der skizzierten PH:

Für alle Werte **W = w**:

$$E(Y|X, W = w) = a_w + b_w \cdot X, \text{ wobei } b_w > 0. \quad (2.9)$$

Für nicht konstantgehaltene Werte der Störvariablen gilt somit

$$E(Y|X, W) = f(W) + g(W) \cdot X, \quad (2.10)$$

wobei  $g(W)$  positiv reellwertig sein muß. Letztlich geprüft - z.B. via t-Test oder Mann-Whitney-Test - wird aber die  $SH_U$

$$E(Y|X) = a + b \cdot X, \text{ wobei } b > 0. \quad (2.11)$$

Ist dieses Vorgehen gerechtfertigt? Natürlich kann die PH gemäß Gleichung (2.9) nicht aus der  $SH_U$  gemäß Gleichung (2.11) deduziert werden und deshalb können PH und  $SH_U$  auch nicht äquivalent sein. Der bestmögliche Weg der Überprüfung der PH (mit  $g = h = 0$ ) ist somit nicht gegeben. Ist aber wenigstens die Implikation  $PH \Rightarrow SH_U$  erfüllt, so daß  $h = 0$  gilt? Ohne weitere Zusatzannahmen folgt Gleichung (2.11) nicht aus Gleichung (2.9). Wenn die

Störvariable  $W$  nicht konstantgehalten wird und  $X$  und  $W$  abhängig sind, kann bekanntlich trotz Gültigkeit von (2.9) ein negativer Effekt  $b < 0$  in Gleichung (2.11) auftreten, der fälschlich die Ungültigkeit der PH nahelegt (vgl. Steyer, 1992, Kapitel 15 dieses Bandes). Dies muß man ausschließen, wenn man von der PH auf die  $SH_U$  übergehen will. Ist  $W$  bekannt und vollständig kontrollierbar, so kann dies durch die experimentellen Kontrolltechniken der **Eliminierung**, **Konstanthaltung** und der **systematischen Variation** der Störeinflüsse oder auch über **Parallelisierung** erfolgen (vgl. Strack & Rehm, Kapitel 12 dieses Bandes). Ist  $W$  nicht (vollständig) bekannt oder nicht (vollständig) kontrollierbar, so bleibt als **einzig möglicher Weg** die **Randomisierung**, welche die stochastische Unabhängigkeit (stU) von  $X$  und  $W$  garantiert. Man kann dann aus Gleichung (2.10) herleiten:

$$\begin{aligned}
 E(Y|X) &= E((E(Y|X, W) + F)X) \quad (\text{da } F := Y - E(Y|X, W)) \\
 &= E(E(Y|X, W)|X) + E(FX) \\
 &= E(E(Y|X, W)|X) + 0 \\
 &= E((f(W) + g(W) \cdot X)|X) \quad (\text{nach Gleichung 2.10}) \\
 &= E(f(W)|X) + E(g(W)|X) \cdot X \\
 &= a + b \cdot X, \quad b > 0 \quad (\text{aufgrund stU von } X \text{ und } W).
 \end{aligned}
 \tag{2.12}$$

Ähnliche Argumentationen lassen sich auch für andere PHn führen, die ceteris-paribus-Klauseln beinhalten. Aus Platzgründen wollen wir dies jedoch nicht weiter ausführen.

Das Fazit dieses Abschnitts lautet, daß eine Verwendung der Randomisierung bei Überprüfungen von PHn geboten ist, die ceteris-paribus-Klauseln bezüglich unbekannter oder nicht vollständig kontrollierbarer Störeinflüsse beinhalten. In diesem Fall ermöglicht die Randomisierung die Deduktion prüfbarer statistischer Hypothesen, welche die Störgrößen nicht mehr beinhalten. Dies **ist der eigentliche Zweck der Randomisierung**.

## 2.5 Sind statistische Aggregathypothesen zulässig?

Sehr viele PHn - z.B. die im letzten Abschnitt diskutierte Hypothese über die Abhängigkeit der Gedächtnisleistung von der Konkrettheit des Wortmaterials - sind allgemeinspsychologischer Natur und behaupten demzufolge etwas über (beliebige) **einzelne** Individuen eines offenen Individuenbereichs (vgl. Abschnitt 1). Überprüft werden diese Hypothesen aber sehr häufig anhand von SHn, die sich auf **über Individuen aggregierte Daten** beziehen. Die Frage, ob diese oftmals gar nicht vermeidbare Vorgehensweise methodologisch vertretbar ist, läßt sich in ähnlicher Weise wie die Frage nach dem Zweck der Randomisierung beantworten. Wie schon erwähnt, kann beim Übergang **von**



Individuen auf Aggregate bestenfalls eine notwendige, nicht aber auch eine hinreichende Bedingung für die Gültigkeit der vorgeordneten PH resultieren. Da wir andere Wege der statistischen Prüfung einer PH oben ausgeschlossen haben, bleibt nur eine Möglichkeit, die Überprüfung allgemeinpsychologischer Hypothesen über statistische Aggregathypothesen zu rechtfertigen: **Die statistische Aggregathypothese muß aus der allgemeinpsychologischen Hypothese logisch folgen** (vgl. bereits Estes, 1956).

Im Regelfall ist die Frage, ob eine Aggregathypothese aus einer allgemeinpsychologischen Hypothese folgt, nicht so leicht zu beantworten, wie es auf den ersten Blick vielleicht den Anschein hat. Im letzten Abschnitt haben wir beispielsweise gesehen, daß aus der Hypothese „Für alle Individuen gilt, daß konkrete Wörter *ceteris paribus* besser als abstrakte Wörter behalten werden“ keineswegs ohne weiteres folgt „Im Mittel (über Individuen) werden konkrete Wörter besser als abstrakte Wörter behalten“. Die Implikation gilt nur, wenn randomisiert wurde. Ebenso wie in diesem Beispiel muß jede PH daraufhin analysiert werden, welche Aggregathypothesen unter welchen Zusatzannahmen daraus ableitbar sind und welche nicht.

Bei PHn, die quantitative (durch eine mathematische Funktion beschreibbare) Gesetzmäßigkeiten zwischen einer  $W$   $X$  und einer  $AV$   $Y$  behaupten, ist besondere Vorsicht geboten. Es gilt im allgemeinen nicht, daß die über Individuen gemittelten Daten demselben Funktionstyp wie die individuellen Funktionen folgen. Z. B. ist das Potenzgesetz für interindividuell gemittelte Daten nicht erfüllt, wenn es für einzelne Individuen mit interindividuell variablem Exponenten gilt. Andere Beispiele dafür, daß ein bestimmter Funktionstyp für individuumsbezogene Daten nicht auch für die interindividuell gemittelten Daten gelten muß, präsentieren u. a. Bakan (1954) und Estes (1956). Die gleichen Autoren haben auch hinreichende Bedingungen dafür angegeben, daß ein bestimmter Funktionstyp invariant gegenüber arithmetischer Mittelwertbildung ist. Bei der Beantwortung der Frage, ob eine konkrete Aggregathypothese aus einer allgemeinpsychologischen Gesetzhypothese ableitbar ist, können diese Kriterien wertvolle Dienste leisten.

## 2.6 Skalenniveau und Statistik

Seit Ende der vierziger Jahre wird in der einschlägigen psychologischen Methodenliteratur - insbesondere in der Zeitschrift **Psychological Bulletin** - eine Debatte darüber ausgetragen, ob das Skalenniveau psychologischer Variablen die Wahl der (Prüf-) Statistik mitbestimmen sollte oder nicht. Von bedingungsloser Zustimmung bis hin zu bedingungsloser Ablehnung wurden nahezu alle denkbaren Standpunkte zu dieser Frage vertreten. Eine Zusammenfassung der frühen Debatte findet man bei Bredenkamp (1972); neuere Ar-

beiten sind in Luce, Krantz, Suppes und Tversky (1990) zitiert und knapp zusammengefaßt.

Unsere Auffassung, die weitgehend derjenigen von Luce et al. (1990, S.299) entspricht, kann wie folgt zusammengefaßt werden: **Bei der Wahl der Prüf-Statistik zu einer vorgegebenen  $SH_U$  spielt das Skalenniveau keine Rolle, bei der Wahl einer  $SH_U$  zu einer vorgegebenen PH aber eine ganz entscheidende.** Diese Position ergibt sich zwingend aus den Konsequenzen des Validitätsmaximierungspostulats, die wir am Ende von Abschnitt 2.3 zusammenfassend dargestellt haben. Nehmen wir zunächst an, eine aus der PH ableitbare  $SH_U$  sei bereits gegeben. Dann kommt es lediglich darauf an, ein Prüfverfahren festzulegen, das eine Kontrolle von  $\epsilon$  und  $\phi$  ermöglicht und dabei möglichst effizient ist, d.h. möglichst kleine Irrtumswahrscheinlichkeiten bei möglichst kleinem Stichprobenumfang sicherstellt. Hieraus folgt in keiner Weise, daß die Prüfstatistik bzw. der Wahrheitswert von Aussagen über numerische Werte der Prüfstatistik invariant gegenüber zulässigen Transformationen von  $UV_n$  und  $AV_n$  sein muß. Also kann zu einer vorgegebenen  $SH_U$  eine Prüfstatistik völlig unabhängig vom Skalenniveau der Variablen ausgewählt werden.

Ganz anders sieht es bei der Wahl einer  $SH_U$  zu einer vorgegebenen PH aus. Wenn wir mit Regel  $R_2$  oder  $R_2'$  fordern, daß die  $SH_U$  aus der PH logisch folgt, **so** setzt dies die **Bedeutsamkeit** der  $SH_U$  in dem Sinne voraus, daß der **Wahrheitswert der  $SH_U$  invariant gegenüber zulässigen Transformationen der involvierten  $ZV_n$**  ist. Andernfalls wäre die Antwort auf die Frage, ob  $PH \Rightarrow SH_U$  gilt, von der willkürlichen Auswahl einer numerischen Zuordnung abhängig.

Auch bei der Untersuchung von Invarianzeigenschaften statistischer Hypothesen ist Sorgfalt geboten. Ein gängiges Vorurteil ist beispielsweise, daß monotone Transformationen von  $ZV_n$  die Rangordnung von Erwartungswerten invariant lassen, so daß z.B. die gerichtete Mittelwertshypothese  $E(Y | X = 1) > E(Y | X = 0)$  auch bei Rangskalenniveau von  $Y$  bedeutsam ist. Dies stimmt nicht. Nehmen wir beispielsweise an, daß  $Y$  eine Reaktionszeitvariable ist, die in jeder von zwei Gruppen  $X = 1$  bzw.  $X = 0$  lognormalverteilt ist mit  $E(Y | X = 1) > E(Y | X = 0)$ . Wenn  $Y$  Rangskalenniveau aufweist, dürfen wir z.B. die Transformation  $Y^* := \ln(Y)$  anwenden. Unter den genannten Bedingungen muß  $Y^*$  innerhalb der Gruppen normalverteilt sein. Man kann zeigen, daß die gruppenspezifischen Erwartungswerte von  $Y$  und  $Y^*$  dann in folgender Beziehung zueinander stehen (vgl. Johnson & Kotz, 1970, S. 115, Gleichung 6.1):

$$E(Y | X) = \exp(E(Y^* | X) + 0.5 \cdot \text{Var}(Y^* | X)) \quad (2.13)$$

Es hängt also von der Varianz von  $Y^*$  innerhalb der Gruppen ab, ob auch  $E(Y^* | X = 1) > E(Y^* | X = 0)$  gilt oder nicht. Wenn man sich im Rahmen eines

statistischen Modells bewegt, das die Varianzen nicht näher normiert, ist demzufolge die gerichtete Mittelwertshypothese bei Rangskalenniveau nicht bedeutsam. Man kann nun darauf verweisen, daß man das ALM zugrunde legt, welches  $\text{Var}(Y^* | X = 1) = \text{Var}(Y^* | X = 0)$  vorschreibt. In diesem Fall wäre zwar die gerichtete Mittelwertshypothese bei Rangskalenniveau bedeutsam. Allerdings wäre dann das zugrundeliegende Modell selbst nicht bedeutsam, denn man verläßt es, wenn man eine beliebige nichtlineare (aber monotone) Transformation von  $Y$  vornimmt. Wir können uns also nicht damit begnügen, Invarianzeigenschaften für die Parameterhypothese zu fordern; die Gültigkeit des statistischen Modells selbst muß invariant gegenüber zulässigen Transformationen der ZVn sein. Beides ist gemeint, wenn wir fordern, daß nicht Statistiken, sondern SHn bedeutsam (*meaningful*) sein müssen.

### 3. Empfehlungen zur statistischen Entscheidungsstrategie

Die in den Regeln  $R_2$  und  $R_3$  geforderte Kontrolle der statistischen Fehlerwahrscheinlichkeiten ist möglich, wenn vor der Durchführung einer Untersuchung neben der Wahrscheinlichkeit  $\alpha$  für die fälschliche Zurückweisung der  $H_0$  eine Effektgröße spezifiziert wird, die mit der Wahrscheinlichkeit  $1-\beta$  zu einer korrekten Zurückweisung von  $H_0$  führen soll. Bei Gültigkeit bestimmter statistischer Annahmen läßt sich dann der Stichprobenumfang bestimmen, der benötigt wird, um  $\alpha$  und  $\beta$  nicht größer als die Werte werden zu lassen, die festgelegt wurden. Die Bestimmung des Stichprobenumfangs für gegebene Effektstärke sowie gewünschtes  $\alpha$  und  $\beta$  kann z.B. mit Hilfe der Tabellen Cohens (1977 oder 1988) oder Hagers (1987) vorgenommen werden. Inzwischen liegen aber auch komfortable und leicht bedienbare Programme für Personalcomputer vor, die die Planung von Stichprobenumfängen sowie die Durchführung anderer Teststärkeanalysen erheblich erleichtern (z. B. Faul, Erdfelder & Buchner, 1993).

#### 3.1 Zur Festlegung von $\alpha$ und $\beta$

Wie in Abschnitt 2.3 ausgeführt, lassen sich Strenge und Fairneß einer Untersuchung nur dann vollständig über  $\alpha$  und  $\beta$  kontrollieren, wenn  $g$  und  $h$  Null sind. Wird dagegen die  $SH_U$  durch die PH nur impliziert, ist also  $h = 0$  und  $g > 0$ , so ist die Kontrolle von Strenge und Fairneß nicht mehr so einfach. Wir schlagen vor, in diesem Fall von der Fairneß auszugehen, die  $1-\phi$  beträgt (vgl. Gleichung 2.8, Abschnitt 2.2). Ist festgelegt worden, wie groß die Fairneß mindestens sein soll, so läßt sich sagen, daß die Strenge der Prüfung um so

größer wird, je kleiner  $\epsilon$  gewählt wird. Für verschiedene Werte von  $\epsilon$  und  $g$  läßt sich nach der Formel

$$1 - e_U = (1 - \epsilon) \cdot (1 - g) + \phi \cdot g \quad (3.1)$$

die Strenge bestimmen. Diese beträgt z.B. .82 für  $g = .1$ ,  $\phi = .1$  und  $\epsilon = .1$ . Bei  $g = .1$ ,  $\phi = .1$  und  $\epsilon = .2$  beträgt die Strenge .73. Wie klein  $\epsilon$  gewählt werden kann, hängt natürlich auch von den zur Verfügung stehenden Ressourcen ab. Wenn z.B. eine PH die SH<sub>U</sub> H<sub>1</sub>:  $\mu_1 > \mu_2$  in einem Zwei-Gruppen-Experiment impliziert, sind nach Formel (3.2)

$$n = \frac{(z_{(1-\alpha)} + z_{(1-\beta)})^2}{2f^2} \quad (3.2)$$

$N = 2 \cdot n = 2 \cdot 33$  Vpn notwendig, um einen Effekt der Größe  $f^2 = .1$  (vgl. Cohen, 1977) bei einseitigem Test zum Niveau  $\epsilon = \alpha = .1$  mit der Wahrscheinlichkeit  $1 - \phi = 1 - \beta = .9$  entdecken zu können. In dieser Formel sind  $z_{1-\alpha}$  bzw.  $z_{1-\beta}$  die z-Werte zum (1- $\alpha$ ),100- bzw. (1- $\beta$ )100-Perzentil der Standardnormalverteilung; sie gilt, wenn die AV in den Gruppen normal und mit gleichen Varianzen verteilt ist (vgl. Bredenkamp, 1969). Wird  $\alpha$  auf .2 erhöht, reduziert sich der Stichprobenumfang auf  $2 \cdot 23$  Vpn. Dies ändert nichts an der Fairneß, reduziert aber die Strenge der Prüfung.

### 3.2 Zur Festlegung der Effektstärke

Erfahrungsgemäß bereitet die Festlegung der Effektstärke die meisten Probleme. Möglicherweise ist das scheinbare Fehlen rationaler Kriterien der Effektgrößenfixierung dafür entscheidend, daß Signifikanztests so häufig ohne Kontrolle von  $\beta$  durchgeführt werden. Allerdings löst dieses Vorgehen das Problem keineswegs. Nach Formel (3.2) hat man sich im Falle einer unreflektierten Wahl des Stichprobenumfangs für jedes  $\alpha$  und  $\beta$  implizit auf eine unbekannt kritische Effektstärke festgelegt (vgl. Bredenkamp, 1969). Durch die Planung des Stichprobenumfangs nach vorgegebenen Werten für  $\alpha$ ,  $\beta$  und die Effektgröße macht man die Prämissen der Entscheidungsstrategie dagegen explizit und durchschaubar.

Nach welchen Kriterien können Effektgrößen festgelegt werden? Ohne Anspruch auf Vollständigkeit seien hier drei Möglichkeiten aufgeführt. Die erste Möglichkeit besteht darin, aus vorangegangenen Untersuchungen die Effektstärke zu schätzen (vgl. dazu Bredenkamp, 1980). Weiterhin kann man Überlegungen zur Reliabilität der AV einbeziehen. Wenn man annimmt, daß die experimentelle Fehlervarianz sich im Sinne der klassischen Testtheorie additiv aus wahrer Varianz und Meßfehlervarianz zusammensetzt, und festlegt, daß die durch die experimentellen Effekte bedingte Varianz nicht kleiner als die

Meßfehlervarianz sein sollte, impliziert dies die Spezifikation  $f^2 = 1 - r_{yy}$ , wobei  $r_{yy}$  die Reliabilität der AV ist (vgl. Bredenkamp, 1972). Bei reliablen Variablen lohnt demnach die Aufdeckung kleiner Effekte, die relativ viele Vpn erforderlich macht'. Schließlich bleibt die dritte Möglichkeit, sich an den von Cohen (1977) geschaffenen Konventionen zu orientieren, der „kleine“, „mittlere“ und „große“ Effekte ( $f^2 = .01$ ,  $f^2 = .0625$  und  $f^2 = .16$ ) unterscheidet. Wenn  $SH_U$  eine  $H_0$  ist, mag man bereits an der Aufdeckung „kleiner“ Abweichungen von  $H_0$  interessiert sein. Ist  $SH_U$  dagegen eine  $H_0$ , wird man vielleicht „mittlere“ oder „große“ Effekte fordern, ehe man die zugrundeliegende PH als bewährt betrachtet.

Allerdings können Cohens (1977) Konventionen auch Probleme aufwerfen. Diese sollen am Beispiel einer monotonen Trendhypothese erläutert werden. Grundlegend ist hier die Unterscheidung zwischen der Effektgröße zu Lasten des Kontrastes zweier Mittelwerte  $f_c^2$  und der gesamten Effektstärke  $f_t^2$  über alle Bedingungen hinweg. Werden zwei Mittelwerte  $\mu_j$  und  $\mu_k$  verglichen, gilt

$$f_c^2 = (\mu_j - \mu_k)^2 / (2m\sigma^2), \quad (3.3)$$

wobei  $m$  die Gesamtzahl der Versuchsgruppen und  $\sigma^2$  die Varianz innerhalb der Bedingungen ist (vgl. Bredenkamp, 1984). Formel (3.3) ist ein Spezialfall von

$$f_c^2 = \frac{\sum_{j=1}^m u_j \mu_j^2 / (m \sum_{j=1}^m u_j^2)}{\sigma^2}, \quad (3.4)$$

wobei  $u_j$  das Gewicht ist, das dem  $j$ -ten Mittelwert im Kontrast zugewiesen wird. Der in Gleichung (3.4) über dem großen Bruchstrich stehende Ausdruck ist die durch die lineare Regression von  $\mu_j$  auf  $u_j$  ( $j = 1, \dots, m$ ) aufgeklärte Varianz, so daß  $f_c^2$  nichts anderes als das Verhältnis dieser Varianz zur Binnenvarianz  $\sigma^2$  ist. Bei orthogonalen Kontrasten gilt (vgl. Bredenkamp, 1984):

$$\sum_{c=1}^{m-1} f_c^2 = f_t^2. \quad (3.5)$$

Für nonorthogonale Kontraste jeweils zweier Mittelwerte machen wir uns folgende Beziehung zunutze:

7 Auch wenn die Reliabilität nicht bekannt ist, kann diese Formel hilfreich sein. Dann ist für  $r_{yy}$  eine erwartete Reliabilität einzusetzen und  $f^2$  entsprechend darauf abzustimmen.

$$\frac{\sum_{j \neq j'} (\mu_j - \mu_{j'})^2}{m(m-1)} = \frac{\sum_{j=1}^m (\mu_j - \mu)^2}{m-1} = \frac{m}{m-1} \sigma^2 f_t^2. \quad (3.6)$$

Ersetzen von  $(\mu_j - \mu_{j'})^2$  durch  $2m\sigma^2 f_c^2$  (vgl. Gleichung 3.3) liefert:

$$\frac{2}{m} \sum_{c=1}^{\frac{m(m-1)}{2}} f_c^2 = f_t^2. \quad (3.7)$$

Die Summe auf der linken Seite der Gleichung läuft hierbei über alle  $m(m-1) / 2$  möglichen paarweisen Mittelwertsvergleiche eines  $m$ -Gruppen-Designs. Bei  $m = 6$  besagt z. B. die Hypothese eines streng monoton steigenden Trends  $H_1: \mu_1 < \mu_2 < \mu_3 < \mu_4 < \mu_5 < \mu_6$ . Angenommen, es wird für jeden der fünf Tests, die sich auf die  $H_0$  der Gleichheit zweier benachbarter Mittelwerte beziehen,  $f_{c1}^2 = .01$  festgelegt. Bis auf die Konstante  $2m\sigma^2$  beträgt dann der Unterschied zwischen benachbarten Mittelwerten mindestens  $f_{c1} = .1$ , zwischen Mittelwerten, deren Indextdifferenz 2 beträgt,  $f_{c2} = .2$  usw.; diese  $f_c$ -Werte sind zu quadrieren und aufzuaddieren (s. Gleichung 3.7). So entstehen fünf Werte  $f_{c1}^2 = .01$ , vier Werte  $f_{c2}^2 = .04$ , drei Werte  $f_{c3}^2 = .09$ , zwei Werte  $f_{c4}^2 = .16$  und ein Wert  $f_{c5}^2 = .25$ . Ihre Summe beträgt 1.05. Multiplikation mit dem Faktor  $2/m = 1/3$  ergibt  $f_t^2 = .35$ . Das heißt: Soll für jeden Test ein laut Cohen (1977) „kleiner“ Effekt entdeckt werden, so ist impliziert, daß der globale Effekt  $f_t^2 = .35$  beträgt und damit größer ist als ein „großer“ Effekt. Bei  $f_{c1}^2 = .16$  ergibt sich  $f_t^2 = 5.6$ . Um einen „großen“ Effekt  $f_{c1}^2 = .16$  bei einseitigem Test,  $\alpha = .05$  und  $\beta = .05$  entdecken zu können, sind aber bereits unter jeder Bedingung  $n = 12$ , insgesamt also 72 Vpn erforderlich. Dieser Wert ergibt sich aufgrund von Formel (3.8) (Bredenkamp, 1984), die eine Verallgemeinerung von (3.2) ist<sup>8</sup>:

$$n = \frac{(z_{(1-\alpha)} + z_{(1-\beta)})^2}{m f_{c1}^2}. \quad (3.8)$$

Für die Festlegung der Effektgröße resultiert nach den vorangegangenen Ausführungen folgendes: Da die  $H_1$  einer Varianzanalyse wegen ihrer Unbestimmtheit wohl kaum jemals die durch die PH implizierte oder mit ihr äquivalente  $SH_U$  ist, sollte sich die Effektgrößenbestimmung auf die interessierenden Kontraste beziehen. Hierbei ist immer zu berücksichtigen, was aus diesen Festlegungen für  $f_t^2$  resultiert. Da  $f_t^2 = \rho^2 / (1-\rho^2)$  bzw.  $\rho^2 = f_t^2 / (1+f_t^2)$  -

<sup>8</sup> Formel (3.8) führt approximativ zu denselben Ergebnissen wie eine Stichprobenumfangsplanung nach Cohen (1977, Kap. 9), wenn man berücksichtigt, daß Cohen von **zweiseitigen** Tests ausgeht. Die Differenz der berechneten  $n$ -Werte ist maximal 1.

wobei  $\rho^2$  der in der zugrundeliegenden Population durch die experimentellen Bedingungen aufgeklärte Anteil an der Gesamtvarianz ist -, bedeutet z.B. ein  $f_t^2 = 5.6$ , daß 85 % der Gesamtvarianz durch die Bedingungsvariation aufgeklärt wird. Das ist ein unrealistisch großer Wert, auf den man sich in der Regel kaum festlegen würde. Reduziert man  $f_c^2$ , so resultiert ein realistisches  $f_t^2$ , aber die erforderlichen Stichprobenumfänge werden so groß, daß das Experiment kaum zu realisieren ist. Realisierbare Werte ergeben sich, wenn man sich entschließt, die Anzahl der Bedingungen zu reduzieren und  $\alpha$  oder  $\beta$  zu erhöhen.

Die obigen Erörterungen gelten in analoger Form auch für andere spezielle Mittelwertskontraste. Wenn dagegen die  $H_0$  einer Varianzanalyse die durch die PH implizierte  $SH_U$  ist, muß bei der Effektgrößenbestimmung von  $f_t^2$  ausgegangen werden.

### 3.3 Mehrfache Signifikanztests

Manchmal besteht die abgeleitete  $SH_U$  aus der Konjunktion verschiedener SHn. Bei der eben besprochenen  $SH_U$  eines monotonen Trends und  $m = 3$  gilt z.B.  $SH_U \Leftrightarrow ((H_1^{(1)} : \mu_1 < \mu_2) \wedge (H_1^{(2)}: \mu_2 < \mu_3))$ .

Tabelle 1: Statistische Fehlerwahrscheinlichkeiten bei der Entscheidung über eine Konjunktion zweier Alternativhypothesen

	Zutreffend ist			
	$SH_U$	$\sim SH_U$		
Entscheidung für $H_1^{(1)} \wedge H_1^{(2)}$	$H_0^{(1)} \wedge H_0^{(2)}$	$H_0^{(1)} \wedge H_1^{(2)}$	$H_1^{(1)} \wedge H_0^{(2)}$	
$H_1^{(1)} \wedge H_1^{(2)}$	$1 - \phi = (1 - \beta)^2$	$\epsilon_1 = \alpha^2$	$\epsilon_2 = \alpha(1 - \beta)$	$\epsilon_3 = \alpha(1 - \beta)$
$\sim(H_1^{(1)} \wedge H_1^{(2)})$	$\phi = 1 - (1 - \beta)^2$	$1 - \epsilon_1 = 1 - \alpha^2$	$1 - \epsilon_2 = 1 - \alpha(1 - \beta)$	$1 - \epsilon_3 = 1 - \alpha(1 - \beta)$
$H_1^{(1)} \vee H_1^{(2)}$	$1 - \phi = 1 - \beta^2$	$\epsilon_1 = 1 - (1 - \alpha)^2$	$\epsilon_2 = 1 - \beta(1 - \alpha)$	$\epsilon_3 = 1 - \beta(1 - \alpha)$
$\sim(H_1^{(1)} \vee H_1^{(2)})$	$\phi = \beta^2$	$1 - \epsilon_1 = (1 - \alpha)^2$	$1 - \epsilon_2 = \beta(1 - \alpha)$	$1 - \epsilon_3 = \beta(1 - \alpha)$

Tabelle 1 verdeutlicht die Wahrscheinlichkeiten richtiger und falscher Entscheidungen für  $SH_U$  unter der Annahme, daß bei den beiden erforderlichen Signifikanztests  $\alpha_1 = CZ, = \alpha$  und  $\beta_1 = \beta_2 = \beta$  ist. Wenn die PH die  $SH_U$  eines streng monotonen Trends impliziert, gilt weiterhin, daß die Wahrscheinlichkeit  $h$  für das Nichtzutreffen dieser Hypothese bei Gültigkeit der PH Null ist. Die Fairneß ist dann unter Zugrundelegung der Entscheidungsstrategie, nur dann für  $SH_U$  zu entscheiden, wenn  $H_1^{(1)}$  und  $H_1^{(2)}$  angenommen werden, genau  $1 - \phi = (1 - \beta)^2$ . Wenn die Fairneß der Untersuchung nicht kleiner als ein

bestimmter Wert werden soll, ist in diesem Fall für die einzelnen Tests  $\beta$  zu adjustieren, während eine Adjustierung des  $\alpha$ -Fehlers keine Vorteile bringt, da die Wahrscheinlichkeit für die fälschliche Annahme von  $H_1^{(1)}$  und  $H_1^{(2)}$  ohnehin kleiner als  $\alpha$  ist (vgl. Tabelle 1, obere Hälfte). Eine Adjustierung nach der Formel  $\beta \cdot = \beta / (\text{Anzahl Tests})$  stellt sicher, daß die Fairneß nicht unter den gewünschten Wert  $1-\beta$  fällt.

Westermann und Hager haben verschiedentlich (z.B. 1986) auf die Notwendigkeit der Adjustierung von  $\beta$  hingewiesen, wenn eine konjunktive Verknüpfung von Alternativhypothesen aus der PH abgeleitet wird. Bei einer konjunktiven Verknüpfung von Nullhypothesen ist dagegen die Adjustierung von  $\alpha$  erforderlich. Zum Fall mehrerer  $H_0$ - und  $H_1$ -Hypothesen vgl. Westermann und Hager (1986).

Zurück zum Beispiel. Wie groß ist die Strenge der Prüfung?  $SH_U$  kann auf dreierlei Weise nicht zutreffen (vgl. Tabelle 1), und die Wahrscheinlichkeiten hierfür seien  $g_1$ ,  $g_2$  und  $g_3$  genannt. In diesem Fall gilt:

$$e_U = g_1\varepsilon_1 + g_2\varepsilon_2 + g_3\varepsilon_3 + (1 - g_1 - g_2 - g_3)(1 - \phi). \quad (3.9)$$

Ohne Kenntnis der  $g$ -Wahrscheinlichkeiten läßt sich über die Strenge der Entscheidungsstrategie nichts sagen.

Die bisher betrachtete Entscheidungsstrategie (obere Hälfte der Tabelle 1) sei mit Westermann und Hager (1986) streng genannt, während die Strategie, auf Richtigkeit der Hypothese  $H_1^{(1)} \wedge H_1^{(2)}$  zu entscheiden, wenn wenigstens einer der Tests  $H_1$  bekräftigt, schwach genannt sei. Die Wahrscheinlichkeit für richtige und falsche Entscheidungen beim Verfolgen dieser Strategie gibt die untere Hälfte von Tabelle 1 wieder. Wie man erkennt, ist die Fairneß der „strengen“ Strategie kleiner als die der „schwachen“, während die Strenge größer ist, da jeder einzelne Summand in Gleichung (3.9) für die strenge Strategie kleiner ist. Durch eine Adjustierung des  $\beta$ -Fehlers kann man diesen Nachteil der „schwachen“ Strategie nicht ausgleichen, während der Fairneß-Nachteil der „strengen“ Strategie durch eine Reduktion des  $\beta$ -Fehlers kompensierbar ist.

### 3.4 Nonparametrische Verfahren

Die bislang besprochenen Stichprobenumfangsplanungen sind an Annahmen über die Verteilung der Teststatistik gebunden. Annahmeärmere Hypothesenprüfungen lassen sich in experimentellen Untersuchungen mit Hilfe von Randomisierungstests durchführen. Da die asymptotische relative Effizienz (A.R.E.) des Randomisierungstests relativ zum F-Test bzw. t-Test im Normalverteilungsfall gleich Eins ist, kann man die Tabellen Cohens (1977, 1988)



oder das Programm von Faul et al. (1993) benutzen, um den Stichprobenumfang für einen Randomisierungstest zu bestimmen (Willmes, 1987), sofern ein solcher für die Testsituation existiert. Seien  $N_p$  und  $N_v$  die Stichprobenumfänge eines auf der Annahme der Normalverteilung basierenden parametrischen bzw. eines verteilungsfreien Tests. Bei gleichem  $\beta$  wird mit Wachsen  $N_p$  die entdeckbare Effektgröße immer kleiner werden.  $N_v$  ist der Stichprobenumfang des verteilungsfreien Tests, der diese Effektgröße mit gleicher Teststärke entdeckt. Die A.R.E. ist der Wert, dem der Quotient  $N_p / N_v$  mit Wachsen  $N_p$  zustrebt. Wenn A.R.E. = 1, kann man erwarten, daß auch im finiten Fall bei nicht zu kleinem Stichprobenumfang die Teststärkeunterschiede vernachlässigbar sind. Sind die Verteilungsvoraussetzungen des parametrischen Tests nicht erfüllt, dürfte der Randomisierungstest das teststärkere Verfahren sein.

Bei großem  $N$  ist allerdings die Anzahl möglicher Permutationen so groß, daß die Durchführung eines exakten Randomisierungstests unmöglich wird. Durchführbar sind aber Tests, die auf einer Zufallsstichprobe aus allen möglichen Permutationen beruhen. Für derartige Monte-Carlo-Lösungen kann entsprechend den Empfehlungen von Willmes (1987) der Stichprobenumfang durch das  $N$  im Normalverteilungsfall, dividiert durch die sog. Dwass-Effizienz, bestimmt werden. Wenn  $T(\alpha)$  und  $T_s(\alpha)$  die Teststärken des exakten Tests bzw. der Monte-Carlo-Lösung mit dem Simulationsumfang  $S$  sind, gilt (vgl. Willmes 1987, S.400):

$$T(\alpha) - T_s(\alpha) \leq T(\alpha) \cdot (1 - e^D_{S,\alpha}), \quad (3.10)$$

wobei  $e^D_{S,\alpha}$  die Dwass-Effizienz zum Simulationsumfang  $S$  ist. Willmes (1987, S. 400f.) präsentiert Tabellen und Approximationsformeln, welche die Bestimmung des erforderlichen Simulationsumfangs  $S$  für vorgegebene Dwass-Effizienzen und  $\alpha$ -Niveaus ermöglichen. Umgekehrt ist natürlich auch die Bestimmung der Dwass-Effizienz für gegebenes  $\alpha$  und  $S$  möglich. Bei  $\alpha = .05$  und  $S = 1219$  ist  $e^D$  z.B. .95 (Willmes, 1987, S.401, Tab. 8.2).

Viele der auf Rängen beruhenden nonparametrischen Verfahren wie der Kruskal-Wallis-H-Test, der U-Test nach Mann-Whitney usw. sind **de facto** Randomisierungstests. Das erforderliche  $N$  läßt sich approximativ bestimmen, wenn der für den Normalverteilungsfall bestimmte Stichprobenumfang durch die A.R.E. des nonparametrischen Tests dividiert wird, die für verschiedene Verfahren z. B. bei Lienert (1973) angegeben ist (Bredenkamp, 1980; Willmes, 1987). Ob derartige Verfahren eingesetzt werden sollten, hängt von der aus der PH abgeleiteten  $SH_U$  ab. Die auf Rängen beruhenden Randomisierungstests prüfen die  $H_0$  gleicher Rang-Erwartungswerte für verschiedene experimentelle Bedingungen. Nur wenn sich die  $SH_U$  auf Erwartungswerte von Rängen bezieht, sind diese Tests indiziert. Dies wird z.B. dann häufig der Fall

sein, wenn für die AV Ordinalskalenniveau unterstellt werden muß. Mittelwertshypothesen sind in diesem Fall nicht bedeutsam (vgl. Abschnitt 2.6), wohl aber Hypothesen über Erwartungswerte von Rängen.

#### **4. Stochastische Modelle mit latenten Variablen als Bestandteile einer deduktivistischen Methodologie**

Wenn PHn theoretische Größen beinhalten, deren Ausprägungen der direkten Beobachtung nicht zugänglich sind (wie z. B. „Intelligenz“, „Gedächtnisspur“, „Empfindungsintensität“, „Erfolgsmotiv“ usw.), stellt sich die grundlegende Frage, wie eine logische Beziehung zwischen derartigen PHn und beobachtbaren Tatbeständen hergestellt werden kann, die dann eventuell in Form von SHn formulierbar sind (vgl. Abschnitt 1.4). Macht man sich klar, daß PHn, die - wie die Invarianzhypothese des verbalen Lernens (vgl. Abschnitt 1.1) - ausschließlich beobachtbare Größen enthalten, eher die Ausnahme als die Regel sind, wird deutlich, welcher Stellenwert dem Theorie-Empirie-Überbrückungsproblem in der psychologischen Forschung zukommt.

##### 4.1 Probleme der Operationalisierung

Die in der psychologischen Forschung übliche Methode des Umgangs mit dem Theorie-Empirie-Überbrückungsproblem ist die sog. Operationalisierung theoretischer Größen. „Operationalisierung“ darf dabei nicht im Sinne von „operationaler Definition“ verstanden werden. Operationale Definitionen ziehen bekanntlich Implikationen auch für nicht beobachtete Untersuchungseinheiten nach sich und sind schon aus diesem Grunde abzulehnen (vgl. Herrmann, 1973). Eine „Operationalisierung“ entspricht eher dem, was in der analytischen Wissenschaftstheorie als „bilateraler Reduktionssatz“ bezeichnet wird. Es handelt sich hierbei um eine bedingte Definition, die eine Äquivalenz von theoretischer und empirischer Größe unter einer **empirischen (situativen) Randbedingung** behauptet (vgl. Westmeyer, 1972; Herrmann, 1973). Während allerdings in bilateralen Reduktionssätzen von perfekten, deterministischen Beziehungen zwischen theoretischen und empirischen Größen ausgegangen wird, beruhen Operationalisierungen eher auf der vagen Idee einer nicht-perfekten, „ungefahren“ Korrespondenz zwischen theoretischer und empirischer Größe.

Eine theoretische Größe operationalisieren bedeutet in praktischer Hinsicht, ihr (mindestens) eine empirische Indikatorvariable zuzuordnen, die mehr oder minder fehlerbelastet ist, d.h. in nicht perfekt eindeutiger Beziehung zur theoretischen Größe steht. Wenn diese Zuordnung erfolgt ist, werden alle

theoretischen Größen der PH durch ihre empirischen Indikatoren (z.B. HAWIE-IQ, Reproduktions- oder Rekognitionsleistung, Größenschätzung, Kategorialurteil, TAT-Score usw.) ersetzt. Aus der so resultierenden modifizierten Hypothese, die wir in Anlehnung an Hager (1987) als **empirische Hypothese** (EH) bezeichnen wollen, wird dann eine  $SH_U$  abgeleitet und statistisch geprüft (vgl. Hussy & Möller, Kapitel 11 dieses Bandes). Die statistische Entscheidung wird anschließend der Beurteilung der PH zugrunde gelegt.

Ist dieses Vorgehen gerechtfertigt? Dies ist selbst dann nicht notwendigerweise der Fall, wenn eine Implikation  $EH \Rightarrow SH_U$  oder gar eine Äquivalenz  $EH \Leftrightarrow SH_U$  nachweislich vorliegt. Zu fordern ist nämlich aufgrund von Regel  $R_2$  bzw.  $R_2'$  eine Äquivalenz- oder Implikationsbeziehung nicht zwischen EH und  $SH_U$ , sondern zwischen PH und  $SH_U$ . Über die Beziehung zwischen PH und  $SH_U$  ist jedoch zunächst nichts bekannt, da EH und PH in keiner logisch determinierten Beziehung zueinander stehen.

Wie kann gezeigt werden, daß die aus einer EH gewonnene  $SH_U$  nicht aus der eigentlich zu prüfenden PH folgt? Hierzu muß man jeweils im Einzelfall ein stochastisches Modell formulieren, das (a) eine geeignete Darstellung der zentralen theoretischen Aussagen der PH beinhaltet (sog. „Strukturmodell“) und (b) Beziehungen zwischen den als latente ZVn aufzufassenden theoretischen Größen und den empirischen ZVn der  $SH_U$  herstellt (sog. „Meßmodell“). Kann man nun zeigen, daß zu einem gegebenen Strukturmodell (der „eigentlichen“ PH also) mindestens ein plausibles Meßmodell denkbar ist, derart, daß die Negation von  $SH_U$  aus der Konjunktion von Struktur- und Meßmodell ableitbar ist, dann ist damit der Nachweis des Nichtbestehens einer Implikationsbeziehung  $PH \Rightarrow SH_U$  erbracht. Beispiele findet man z.B. bei Batchelder und Riefer (1986) sowie Erdfelder (1991).

## 4.2 Die stochastische Formulierung psychologischer Hypothesen mit theoretischen Größen

In der Psychologie gibt es eine lange Tradition psychophysikalischer Skalierungsmodelle, deren Bedeutung für die Lösung des Theorie-Empirie-Überbrückungsproblems außerhalb der Psychophysik lange Zeit übersehen wurde. Der Grundgedanke dieser Modelle besteht darin, theoretische psychologische Größen - wie z.B. die durch einen Reiz ausgelöste Empfindungsintensität - durch latente ZVn oder Parameter der Verteilung latenter ZVn (kurz: latente Parameter) zu repräsentieren. „Latent“ soll in diesem Zusammenhang einfach besagen, daß die Ausprägungen der entsprechenden ZVn nicht beobachtbar sind oder zumindest nicht beobachtet wurden, so daß die Schätzung der latenten Parameter nicht anhand einer Stichprobe aus der Verteilung der latenten

ZVn erfolgen kann. Lösbar wird das Problem der Schätzung latenter Parameter dadurch, daß über ein geeignetes Meßmodell eine Beziehung zwischen den Verteilungen latenter und beobachteter ZVn hergestellt wird, derart, daß Schätzer für die latenten Parameter aus Schätzern für die Parameter der Verteilung beobachtbarer ZVn ableitbar sind.

Wann ist ein Meßmodell „geeignet“? Unseres Erachtens sind einerseits formale, andererseits aber auch sog. inhaltliche Kriterien ausschlaggebend dafür, ob ein bestimmtes stochastisches Modell als Werkzeug zur Überprüfung einer bestimmten PH eingesetzt werden sollte. Die formalen Kriterien lassen sich am besten anhand der durch das Meßmodell (bzw. die entsprechenden Modellgleichungen) definierten Abbildung  $f: \Lambda \rightarrow \Theta$  erläutern, wobei  $\Lambda$  die Menge möglicher latenter Parameterwerte und  $\Theta$  die Menge der möglichen Parameterwerte der beobachtbaren Verteilung ist. Durch die Abbildung  $f$  wird jedem  $\lambda \in \Lambda$  (jedem u.U. mehrdimensionalen latenten Parameter) genau ein  $\theta \in \Theta$  (ein u. U. mehrdimensionaler Parameter der beobachtbaren Verteilung) zugeordnet.

Es lassen sich nun drei Kriterien formulieren: (1) Die Abbildung  $f$  sollte möglichst nicht surjektiv sein, d.h. die Bildmenge  $f(\Lambda)$  sollte eine echte Teilmenge von  $\Theta$  sein. Konkret bedeutet dies, daß Parameterkonstellationen der beobachtbaren Verteilung existieren (nämlich alle  $\theta \in \Theta \setminus f(\Lambda)$ ), die bei Gültigkeit der Modellgleichungen nicht vorkommen dürften. Ist diese Situation gegeben, so ist es prinzipiell möglich, einen Modellgeltungstest zu konstruieren. Da dieser Test nur eine von mehreren Möglichkeiten der empirischen Adäquatheitsprüfung darstellt (vgl. Punkt 3 weiter unten), wollen wir die Nichtsurjektivität von  $f$  nicht zwingend fordern. Prinzipiell sollen also auch Modelle zugelassen werden, die keine beobachtbaren Verteilungen ausschließen und für die sich demzufolge kein Modellanpassungstest konstruieren läßt<sup>9</sup>. Wenn allerdings die Abbildung  $f$  nicht surjektiv ist, so wollen wir fordern, daß der Modellanpassungstest positiv ausfällt. Auf statistische Probleme, die sich im Zusammenhang mit der Modellgeltungsprüfung stellen, gehen wir in Abschnitt 4.4 ein.

(2) Die Abbildung  $f$  sollte injektiv sein, d. h. unterschiedlichen latenten Parametern sollten auch unterschiedliche Parameter der beobachtbaren Verteilung zugeordnet sein. Ist diese Bedingung erfüllt, so heißt das Modell auch „global identifizierbar“. Die globale Identifizierbarkeit stellt sicher, daß eine Funktion  $g: f(\Lambda) \rightarrow \Lambda$ , nämlich  $g \Leftrightarrow f^{-1}$ , definiert werden kann, die es erlaubt, die interessierenden latenten Parameter aus den Parametern der beobachtbaren Verteilung rückzurechnen. Dies ist eine Voraussetzung für die Schätzbarkeit der

<sup>9</sup> Die Two-High Threshold-Theorie (vgl. Snodgrass & Corwin, 1988) ist ein Beispiel für ein derartiges saturiertes Modell, sofern sie auf lediglich eine Versuchsbedingung (eine Treffer- und eine falsche Alarmrate) angewendet wird.

latentem Parameter. Dennoch ist die globale Identifizierbarkeit keine generell unverzichtbare Forderung. Impliziert nämlich die zu prüfende PH keine bestimmte  $SH_U$  über die latenten Parameter, sondern lediglich die Gültigkeit des entsprechenden Modells **unabhängig von konkreten Parameterwerten, so** ist die Forderung nach Injektivität von  $f$  verzichtbar. Grundsätzlich müssen nur die Parameter identifizierbar sein, für die sich aus der zu prüfenden PH konkrete Hypothesen ableiten lassen. Probleme, die sich bei der Modellgeltungsprüfung nichtidentifizierbarer Modelle stellen, werden ebenfalls in Abschnitt 4.4 angesprochen.

(3) Ist ein Modell global identifizierbar oder sind zumindest einige latente Parameter des Modells identifizierbar, so muß das Modell bezüglich der identifizierbaren Parameter **konstruktvalid** sein. Der Begriff der Konstruktvalidität wird hier in Anlehnung an Cronbach und Meehl (1956) gebraucht. Gemeint ist damit, daß experimentelle UVn, die gemäß der zu prüfenden PH bestimmte theoretische Größen beeinflussen, einen hypothesenkonden Effekt auf genau die latenten Parameter ausüben müssen, welche die entsprechenden theoretischen Größen im stochastischen Modell repräsentieren.

Alle drei genannten Kriterien lassen sich sehr schön am Beispiel der Signaldeckungstheorie (Green & Swets, 1974) erläutern. Zunächst muß die Abbildung  $f$  untersucht werden, die den latenten Parametern  $d'$  (Sensitivität) und  $\beta$  (Antworttendenz) Parameter der beobachtbaren Verteilung - nämlich Wahrscheinlichkeiten eines Treffers und eines falschen Alarms - zuordnet. Eine nähere Analyse zeigt, daß diese Abbildung surjektiv und injektiv ist, wenn nur eine Versuchsbedingung realisiert wurde, d.h. nur eine Treffer- und nur eine falsche Alarmrate vorliegt. Das Modell ist unter diesen Umständen somit nicht testbar, wohl aber global identifizierbar, d.h. es lassen sich Schätzer für  $d'$  und  $\beta$  ableiten. Eine Möglichkeit, die Konstruktvalidität des Modells zu überprüfen, besteht darin, Treffer- und falsche Alarmraten unter  $k$  verschiedenen Versuchsbedingungen zu erheben, die sich z.B. hinsichtlich der Auszahlungen für falsche und korrekte Ja-Antworten unterscheiden. Es läßt sich nun die PH prüfen, daß die Auszahlungen die Antworttendenzen  $\beta_j$ , nicht aber die Sensitivitäten  $d'_j$  ( $j = 1, \dots, k$ ) beeinflussen. Hierzu ist lediglich ein simultanes signaldeckungstheoretisches Modell für alle  $k$  Versuchsbedingungen mit der Parameterrestriktion  $d'_1 = d'_2 = \dots = d'_k$  zu formulieren. Dieses Modell entspricht der zu testenden  $SH_U$ . Es umfaßt nicht mehr 2, sondern  $2k$  Modellgleichungen (für  $k$  Paare von Treffer- und falschen Alarmraten). Während  $\Omega$  somit nun  $2k$ -dimensional ist, ist  $\Lambda$  lediglich  $k+1$ -dimensional, da neben  $k$  Antworttendenzparametern  $\beta_j$  lediglich ein Sensitivitätsparameter  $d'$  vorgesehen ist. Eine Analyse dieser Abbildung zeigt, daß sie nicht surjektiv und injektiv ist, so daß (a) ein Modellgeltungstest konstruierbar und (b) das Modell global identifizierbar ist. Der Modellgeltungstest hätte in diesem Fall die  $H_0$  zu überprüfen, daß alle  $k$  Paare von Treffer- und falschen Alarmwahr-

scheinlichkeiten auf einer durch die Modellgleichungen definierten ROC-Kurve liegen. Geht der Test insignifikant aus und beeinflusst die UV „Auszahlung“ die  $\beta$ -Parameter in der erwarteten Richtung, so liegt eine Bestätigung der zu prüfenden PH vor. Das Modell kann als **für diese empirische Anwendungssituation** konstruktvalide gelten. Weitere Möglichkeiten der Konstruktvalidierung bestehen darin, andere UVn zu variieren, für die PHn einen Effekt auf die Sensitivität  $d'$  oder die Antworttendenz  $\beta$  behaupten.

### 4.3 Eine Auswahl wichtiger stochastischer Rahmenmodelle

Die Möglichkeit, das Theorie-Empirie-Überbrückungsproblem mittels stochastischer Skalierungsmodelle anzugehen, wird in der Psychophysik schon seit längerer Zeit genutzt. Erst in jüngerer Zeit zeichnet sich jedoch ab, daß das Prinzip der stochastischen Modellbildung auch in anderen Bereichen der Psychologie nutzbar gemacht werden kann. Dies gilt vor allem für die Bereiche, in denen die psychophysikalischen Skalierungsmodelle ohne methodische Modifikationen zur Beantwortung neuer psychologischer Fragestellungen einsetzbar sind, z.B. bei der Analyse von Rekognitionsdaten in der Gedächtnispsychologie (vgl. Snodgrass & Corwin, 1988; Bredenkamp & Erdfelder, 1993). Die Verfügbarkeit schneller Computer und damit die praktische Realisierbarkeit aufwendiger numerischer Verfahren der Parameterschätzung und Modellprüfung läßt darüber hinaus inzwischen die Anwendung weitaus komplizierterer stochastischer Modelle mit latenten Variablen zu. Die entsprechenden stochastischen Rahmenmodelle sind durchweg nicht neu, schieden jedoch in der Vergangenheit häufig einfach deshalb als Werkzeuge der Hypothesenprüfung aus, weil die zugehörige statistische Analyse nicht praktikabel war.

Ohne Anspruch auf Vollständigkeit sollen einige wichtige Rahmenmodelle angesprochen werden, die als Instrumente der Überprüfung von PHn in **Zukunft** noch an Bedeutung gewinnen könnten. Zunächst wären hier **Strukturgleichungsmodelle** zu nennen. Sie gehören zu den derzeit populärsten Modellen mit latenten Variablen, nicht zuletzt wohl aufgrund der allgemeinen Verfügbarkeit benutzerfreundlicher Computerprogramme (z. B. Jöreskog & Sörbom, 1988). PHn, die sich als lineare Regressionshypothesen formulieren lassen, d.h. einfache und multiple Regressions- sowie Pfadmodelle, können im Rahmen von Strukturgleichungsmodellen angemessen formuliert und geprüft werden; dies gilt insbesondere dann, wenn die involvierten ZVn „latent“ sind und lediglich meßfehlerbelastete Indikatoren vorliegen. Strukturgleichungsmodelle decken damit den Bereich komplett ab, der sich mit den Schlagworten „Faktorenanalyse“ und „Pfadanalyse mit latenten und manifesten Variablen“ grob charakterisieren läßt. Die Modellgleichungen linearer Strukturgleichungsmodelle drücken die Kovarianzen beobachteter ZVn als

Funktion latenter Parameter (latenter Pfadkoeffizienten und Faktorladungen sowie Varianzen und ggf. auch Kovarianzen latenter ZVn) aus. Zugrunde gelegt wird dabei immer ein lineares Meßmodell, d.h. es wird angenommen, daß beobachtete und latente ZVn in linearer Beziehung zueinander stehen. Eine Darstellung der Grundlagen linearer Strukturgleichmodelle unter besonderer Berücksichtigung nichtrekursiver Modelle und der Identifizierbarkeitsproblematik gibt Andres (1990). Die Anwendungen linearer Strukturgleichmodelle dürften in erster Linie im Bereich der Differentiellen Psychologie und im Bereich der Entwicklungspsychologie liegen. Für die experimentelle Psychologie dürften primär simultane Strukturgleichungsmodelle über mehrere Gruppen von Bedeutung sein. Mit diesem Spezialfall, der z.B. via LISREL (Jöreskog & Sörbom, 1988) analysierbar ist, lassen sich Varianzanalysen für latente AVn durchführen, soweit mehrere Indikatorvariablen dieser AV verfügbar sind (intraexperimentelle Replikation der AV) und ein lineares Meßmodell unterstellt werden kann (LISREL-MANOVA).

Eine weitere wichtige Modellklasse, die in der Psychologie erst in jüngster Zeit Beachtung findet, bilden **sof. finite Mischverteilungen** (Titterington, Smith & Makov, 1985; Erdfelder, 1990a; Rost & Langeheine, 1991). Finite Mischverteilungen stellen einen geeigneten formalen Rahmen zur Prüfung von PHn zur Verfügung, die die Untersuchungseinheiten (z.B. Vpn) in mehrere disjunkte und exhaustive „latente Klassen“ unterteilen, für die jeweils unterschiedliche Verteilungen beobachteter ZVn gelten. „Latent“ bedeutet hier wiederum, daß eine exakte Indikatorvariable für die Klassenzugehörigkeit einer Untersuchungseinheit nicht vorliegt. Es können daher nicht die Verteilungen der beobachteten ZVn innerhalb der latenten Klassen, sondern nur ihre „Mischung“ über alle Klassen hinweg beobachtet werden.

Ein wichtiger Spezialfall finiter Mischverteilungsmodelle ist das auf P. Lazarsfeld zurückgehende Modell mehrerer latenter Klassen (**latent class model**, vgl. Lazarsfeld & Henry, 1968). Es ergibt sich aus dem allgemeinen finiten Mischverteilungsmodell durch die Annahme, daß die beobachteten ZVn zum einen qualitativ und zum anderen lokal (d.h. innerhalb der latenten Klassen) stochastisch unabhängig sind. Obwohl dieses Modell bislang vorwiegend in der Soziologie und in der Differentiellen und Diagnostischen Psychologie eingesetzt wurde, ist es für alle Bereiche der Psychologie von außerordentlicher Bedeutung. So läßt sich etwa zeigen, daß multinomiale probabilistische Ereignisbäume - von Riefer und Batchelder (1988) als Instrumente zur Prüfung kognitionspsychologischer Hypothesen und zur Messung kognitiver Prozesse propagiert - nichts anderes sind als spezielle Modelle mehrerer latenter Klassen (Erdfelder, 1990b).

Von großer Bedeutung sind ferner **Markoff-Modelle** (Wickens, 1982), deren Einsatz bislang auf stochastische Lerntheorien (vgl. Tack, 1976) weitgehend

beschränkt blieb, die sich aber zur Formulierung von PHn, welche sich auf zeitlich erstreckte Prozesse beziehen, auch in anderen Bereichen der Psychologie prinzipiell eignen. **Reaktionszeitmodelle** (Townsend & Ashby, 1983; Luce, 1986) bilden eine in formaler Hinsicht heterogene Modellklasse, welche eine Vielzahl kognitionspsychologischer PHn zu formulieren und zu prüfen gestattet.

#### 4.4 Probleme der Modellgeltungsprüfung

In Abschnitt 4.2 wurde festgestellt, daß die Modellgleichungen stochastischer Modelle möglichst eine Abbildung  $f$  definieren sollten, die nicht surjektiv ist. In diesem Fall ist eine Teilklasse denkbarer beobachtbarer Verteilungen mit der Gültigkeit des Modells unvereinbar. Prinzipiell ist es somit möglich, einen Modellgeltungstest zu konstruieren, der prüft, ob der (u.U. mehrdimensionale) Parameter  $\theta_e$ , welcher die empirische Verteilung charakterisiert, im „erlaubten Bereich“, d.h. in der Bildmenge  $f(\Lambda)$  des Raumes latenter Parameter, oder außerhalb des erlaubten Bereichs liegt. Ein solcher Modellgeltungstest kann natürlich nur ein statistischer Hypothesentest sein, da die Modellgeltungshypothese sich nicht auf Statistiken (Kennwerte der beobachteten Stichprobe), sondern auf Parameter (Kennwerte der zugrundeliegenden Verteilung beobachtbarer ZVn) bezieht. Das Problem besteht somit zunächst darin, eine Teststatistik und eine Stichprobenverteilung dieser Teststatistik unter der Nullhypothese

$$H_0: \theta_e \in f(\Lambda) \quad (4.1)$$

abzuleiten. Diese  $H_0$  entspricht der  $SH_U$ , die aus der vorgeordneten PH abgeleitet wurde. Eine Entscheidung für  $H_0$  ist demnach als Bewährungsurteil bzgl. PH, eine Entscheidung gegen  $H_0$  als Nichtbewährungsurteil bzgl. PH zu interpretieren.

Wilks (1938) hat vorgeschlagen, die transformierte Likelihood-Quotienten-Statistik

$$L^2 := -2 \cdot \ln(L(M_0) / L(M_1)) \quad (4.2)$$

zum Vergleich der durch verschiedene stochastische Modelle erzielten Datenanpassungen heranzuziehen. In Gleichung (4.2) bezeichnet  $L(M_0)$  die maximale Likelihood<sup>10</sup> für eine gegebene Datenstichprobe unter einem stochastischen Modell  $M_0$  und  $L(M_1)$  analog die maximale Likelihood für den gleichen Datensatz unter einem „allgemeineren“ stochastischen Modell  $M_1$ . Die Be-

<sup>10</sup> Gemeint ist die Likelihood, die resultiert, wenn die freien Parameter des Modells durch ihre Maximum-Likelihood (ML) - Schätzer ersetzt werden. Die Existenz von ML-Schätzern wird vorausgesetzt.



deutung von „allgemeiner“ wird klarer, wenn man sich die Modellgleichungen der beiden Modelle anschaut, welche zwei Abbildungen  $f_0: \Lambda_0 \rightarrow 0$  und  $f_1: \Lambda_1 \rightarrow 0$  in die Menge möglicher empirischer Verteilungen definieren. Gilt für die beiden Bildmengen die Teilmengenrelation  $f_0(A_0) \subset f_1(A_1)$ , ist also jeder  $M_0$ -konforme Parameter  $o \in E$  zugleich  $M_1$ -konform, so heißt  $M_1$  „allgemeiner“ als  $M_0$ . Geht es - wie im hier zu diskutierenden Fall der Modellgeltungshypothese gemäß Gleichung (4.1) - in erster Linie nicht um den Vergleich zweier stochastischer Modelle, sondern um die statistische Evaluation eines bestimmten Modells  $M_0$ , wählt man zweckmäßigerweise  $M_1$  als saturiertes Modell, so daß  $f_1(A_1) = 0$ .

Wilks (1938) hat gezeigt, daß die Statistik  $L^2$  unter bestimmten Regularitätsbedingungen bei Gültigkeit von  $H_0$  asymptotisch zentral  $\chi^2$ -verteilt ist, wobei sich die Freiheitsgrade  $df$  aus der Anzahl **unabhängiger** Parameterrestriktionen ergeben, die notwendig sind, um das Modell  $M_1$  in den Spezialfall  $M_0$  zu überführen. Im Regelfall entspricht  $df$  gerade der Differenz zwischen der Anzahl freier Parameter in  $M_1$  und in  $M_0$ . Darüber hinaus läßt sich nachweisen, daß die Wilksschen Bedingungen ebenfalls hinreichend sind, um für  $N \rightarrow \infty$  eine nonzentrale  $X^2_{(df)}$ -Verteilung von  $L^2$  bei Ungültigkeit des Modells  $M_0$  (d.h.  $o \in O \setminus f_0(\Lambda)$ ) zu garantieren. Dies eröffnet die Möglichkeit der Kontrolle von  $\beta$ , z.B. mit Hilfe der  $\chi^2$ -Tabellen von Cohen (1977 oder 1988) oder mit Hilfe des Programms von Faul et al. (1993).

Die simultane Kontrolle von  $\alpha$  und  $\beta$  scheint also bei Verwendung der Wilksschen Statistik auf den ersten Blick keine Probleme zu bereiten. Leider täuscht dieser Eindruck, da gerade bei Modellgeltungstests für die im letzten Abschnitt erwähnten Modelle die Regularitätsbedingungen, die zur Herleitung der asymptotischen Verteilung von  $L^2$  benötigt werden, oftmals nicht erfüllt sind. Grundsätzlich gibt es die Möglichkeit, die genannten Probleme in jedem Einzelfall durch Adjustierungen der  $L^2$ -Statistik anzugehen, derart, daß nach einer „korrigierten“  $L^2$ -Statistik gesucht wird, die unter  $H_0$  annähernd einer **Standard- $\chi^2$ -Verteilung** folgt. Dieses Vorgehen war in einigen Fällen erfolgreich (vgl. Titterton et al., 1985), verlangt allerdings als Rechtfertigung immer eine Simulationsstudie, die die Adäquatheit der verwendeten Korrekturformel belegt. Wenn ein solcher Aufwand schon getrieben werden muß, gibt es u. E. bessere, generell anwendbare Alternativen, wie z.B. die von Aitkin, Anderson und Hinde (1981) zur Prüfung von Latent-Class-Modellen vorgeschlagene Vorgehensweise: Sie approximierten die Stichprobenverteilung von  $L^2$  über eine Monte-Carlo-Studie, in der wiederholt Stichproben des (empirisch realisierten) Umfangs  $N$  nach einem Modell generiert wurden, das aus der Ersetzung aller freien Modellparameter durch ihre ML-Schätzer resultiert.

Aus der Sicht einer deduktivistischen Methodologie ergibt sich ein weiteres Problem, das nicht ausgeklammert werden soll: das Problem der Rechtferti-

gung der Verteilungsannahmen, die den Inferenzverfahren (und auch den ML-Schätzern) zugrunde liegen. Beziehen sich die aus PHn abgeleiteten SHn auf Mittelwerte verschiedener experimenteller Bedingungen, so gibt es - wie in Abschnitt 1.3 ausgeführt - prinzipiell die Möglichkeit, die üblichen parametrischen Tests als approximative Randomisierungstests ohne **jede** Verteilungs**annahme** zu rechtfertigen. Gibt es eine ähnliche Möglichkeit auch für Modellgeltungstests, die auf  $L^2$  basieren? Eine endgültige Antwort auf diese Frage kann derzeit noch nicht gegeben werden. Wir vermuten jedoch, daß die von Efron (1979) vorgeschlagene Bootstrap-Methode hier u. U. weiterhilft. Die Bootstrap-Methode ist im Kern eine Konkretisierung der Idee, daß die empirisch vorliegende Stichprobe die bestmögliche Schätzung für die zugrundeliegende Population ist. Ist über die Population nichts bekannt, was über die vorliegende Stichprobe hinausgeht, so liegt es nahe, die Stichprobenverteilung einer Statistik in der Weise zu approximieren, daß man aus der empirisch vorliegenden Stichprobe des Umfangs  $N$  weitere Stichproben des Umfangs  $N$  **mit Zurücklegen** zieht (sog. Bootstrap-Stichproben). Für jede aus der empirischen Stichprobe gezogene Bootstrap-Stichprobe wird die zur Diskussion stehende Statistik berechnet, so daß eine Bootstrap-Stichprobenverteilung dieser Statistik resultiert.

Efron hat diese Methode vor allen Dingen zur verteilungsfreien Bestimmung von Konfidenzintervallen und Standardschätzfehlern herangezogen. Hierzu liegt inzwischen auch ein beachtliches Arsenal an mathematisch-statistischer Literatur vor, das die positiven Eigenschaften der Bootstrap-Methode analytisch zu begründen erlaubt (vgl. das Literaturverzeichnis von Sievers, 1990). Simulationsstudien von Sievers (1990) deuten darüber hinaus an, daß die Leistungsfähigkeit der Bootstrap-Methode möglicherweise weit über das hinausreicht, was bislang in der mathematischen Statistik untersucht wurde. Seine Ergebnisse zeigen in beeindruckender Weise, daß einiges dafür spricht, auch statistische Inferenzverfahren im Lichte der Bootstrap-Methode zu betrachten: Die zentralen Verteilungen einiger üblicher F-Statistiken (ANOVA und Hotellings  $T^2$ ) werden bei Gültigkeit der parametrischen Verteilungsannahmen nahezu perfekt durch die entsprechenden Bootstrap-Verteilungen approximiert. Sind die Standardannahmen dagegen verletzt, so halten die parametrischen Tests das  $\alpha$ -Niveau z.T. nicht ein, während das Bootstrap-Verfahren unter allen Bedingungen die nominellen Fehlerrisiken bewahrt. Damit wird der Versuch nahegelegt, die Bootstrap-Methode auch zur Absicherung von Modellgeltungstests heranzuziehen. Ob dieser Versuch in eine generelle Empfehlung einmünden kann, muß von weiteren Resultaten zur Bootstrap-Methode abhängig gemacht werden.

## 4.5 Einwände gegen stochastische Modellbildung zwecks Überprüfung psychologischer Hypothesen

Gegen die Verwendung formaler Modelle in der Psychologie sind verschiedentlich Einwände erhoben worden, auf die abschließend kurz eingegangen werden soll. Eine recht ausführliche Liste dieser Argumente hat Deppe (1977, Kapitel 8) zusammengestellt und zugleich Gegenargumente geliefert, die aufzeigen, daß die Kritikpunkte eigentlich nicht formale Modelle per se, sondern lediglich bestimmte Formen ihrer Verwendung betreffen. Sehr viele Kritikpunkte werden hinfällig, wenn deutlich gemacht wird, daß Modellbildung nicht auf eine kaum zu leistende isomorphe Abbildung des Gegenstandsgebietes abzielt. Die Art der Verwendung, von der in dieser Arbeit die Rede ist, hat z.B. wesentlich bescheidenere Ziele. Stochastische Modelle interessieren lediglich insoweit, als sie mathematische Formulierungen (notwendiger Bedingungen) von PHn für bestimmte Untersuchungssituationen darstellen. Sie sind somit lediglich **Werkzeuge der Hypothesenprüfung** und sonst nichts.

Sind sie geeignete Werkzeuge in dem Sinne, daß sie eine optimale Überprüfung von PHn erlauben? Die Antwort auf diese Frage hängt davon ab, was man genau unter einer PH versteht. Betrachten wir den Prototyp der Frustrations-Aggressions-Hypothese, die für eine bestimmte Untersuchungssituation formuliert wird: „Für alle Vpn x der Untersuchung gilt: Wenn x frustriert wird, reagiert x aggressiv.“ Diese Hypothese läßt sich als latente **2·2-Kontingenztafel** mit einer leeren Zelle im Rahmen eines finiten Mischverteilungsmodells formulieren: Die latente Klasse der frustrierten und nicht aggressiven Personen muß hypothesengemäß leer sein, alle anderen latenten Klassen (frustriert und aggressiv, nicht frustriert und aggressiv, nicht frustriert und nicht aggressiv) können dagegen beliebig frequentiert sein. Damit hat man allerdings noch kein testbares Modell. Das **Strukturmodell** bedarf der Ergänzung durch ein **Meßmodell**, welches die Beziehung zu beobachtbaren ZVn herstellt. Erst die Konjunktion von Struktur- und Meßmodell kann über einen Modellgeltungstest geprüft werden.

Versteht man unter der PH eine Konjunktion von Struktur- und Meßmodell, so kann offenbar ohne Einschränkung davon gesprochen werden, daß stochastische Modelle eine geeignete Methode der Überprüfung von PHn sind. Versteht man dagegen unter einer PH ausschließlich das, was im Rahmen des stochastischen Modells „Strukturmodell“ genannt wird, so kann der Modellgeltungstest nicht ohne weiteres als Test der PH interpretiert werden. Denkbar ist ja, daß das stochastische Modell aufgrund eines ungeeigneten Meßmodells verworfen werden muß, obwohl der „psychologische Kern“ - das Strukturmodell - durchaus zutrifft. Dieses Problem ist als Duhem-Quine-Problem bekannt (z.B. Gadenne, 1984, Kapitel 9 dieses Bandes) und keineswegs spe-

zifisch für Hypothesenprüfungen mittels stochastischer Modelle. Grundsätzlich gelingt es fast nie, eine direkte Implikationsbeziehung zwischen einer primär interessierenden wissenschaftlichen Hypothese und bestimmten Datenklassen herzustellen. Fast immer ist eine **Konjunktion** von wissenschaftlicher Hypothese und Hilfsannahmen erforderlich, um empirisch prüfbare Folgerungen ableiten zu können. Treten diese Folgerungen empirisch nicht ein, ist es logisch immer vertretbar, das Scheitern auf die Hilfsannahmen und nicht auf die wissenschaftliche Kernhypothese zu attribuieren.

Als Ausweg aus dem Duhem-Quine-Problem bleibt nur die nüchterne Einsicht, daß isolierten psychologischen Kernhypothesen ohne Meßmodell für die involvierten theoretischen Größen nicht sinnvoll empirische Prüfbarkeit attestiert werden kann. Die o. g. Frustrations-Aggressions-Hypothese ist also keine prüfbare PH, solange man sich nicht auf ein bestimmtes Meßmodell für „Frustration“ und „Aggression“ festgelegt hat. Ohne ein solches Meßmodell ist die Frustrations-Aggressions-Hypothese lediglich eine Heuristik, die in Verbindung mit Intuitionen über adäquate Operationalisierungen prüfbare PHn unregen kann. Bezüglich einer Heuristik fragt man aber sinnvollerweise nicht nach Wahrheit und Falschheit, sondern nach Fruchtbarkeit oder Unfruchtbarkeit.

### Literatur

- Aitkin, M., Anderson, D. & Hinde, J. (1981). Statistical modeling of data on teaching styles (with discussion). *Journal of the Royal Statistical Society, A*, 144, 419-461.
- Andres, J. (1990). **Grundlagen linearer Strukturgleichungsmodelle**. Frankfurt: Lang.
- Baddeley, A. (1990). **Human memory**. London and Hove: Erlbaum.
- Bakan, D. (1954). A generalization of Sidman's results on group and individual functions, and a criterion. *Psychological Bulletin*, 51, 63-64.
- Batchelder, W. H. & Riefer, D.M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical and Statistical Psychology*, 39, 129-149.
- Bredenkamp, J. (1969). Über die Anwendung von Signifikanztests bei theoriestendenden Experimenten. *Psychologische Beiträge*, 11, 275-285.
- Bredenkamp, J. (1972). **Der Signifikanztest in der psychologischen Forschung**. Frankfurt: Akademische Verlagsgesellschaft.
- Bredenkamp, J. (1980). **Theorie und Planung psychologischer Experimente**. Darmstadt: Steinkopff.
- Bredenkamp, J. (1982). Verfahren zur Ermittlung des Typs der statistischen Wechselwirkung. *Psychologische Beiträge*, 24, 56-75 und 309.
- Bredenkamp, J. (1984). Anmerkungen und Korrekturen zu Hager & Westermann: Entscheidung über statistische und wissenschaftliche Hypothesen: Probleme bei

mehrfachen Signifikanztests zur Prüfung **einer** wissenschaftlichen Hypothese. *Zeitschrift für Sozialpsychologie*, 15, 224-229.

- Bredenkamp, J. & Erdfelder, E. (1993). Methoden der Gedächtnispsychologie. In D. Albert & K.-H. Stapf (Hrsg.), **Gedächtnis** (= Enzyklopädie der Psychologie, Themenbereich C, Serie II, Band 4). Göttingen: Hogrefe (im Druck).
- Bugelski, B. R. (1962). Presentation time, total time, and mediation in paired-associate learning. *Journal of Experimental Psychology*, 63, 409-412.
- Campbell, D.T. & Stanley, J. C (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), **Handbook of research on teaching**. Chicago: Randy McNally.
- Cohen, J. (1977). **Statistical power analysis for the behavioral sciences** (revised edition). New York: Academic Press.
- Cohen, J. (1988). **Statistical power analysis for the behavioral sciences** (2nd edition). Hillsdale: Erlbaum.
- Cronbach, L.J. & Meehl, P. E. (1956). Construct validity in psychological tests. In H. Feigl & M. Scriven (Eds.), **Minnesota studies in the philosophy of science. Volume Z: The foundations of science and the concepts of psychology and psychoanalysis**. Minneapolis: University of Minnesota Press.
- Deppe, W. (1977). **Formale Modelle in der Psychologie**. Stuttgart: Kohlhammer.
- Edgington, E. S. (1969). **Statistical inference: the distribution-free approach**. New York: Mc Graw-Hill.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.
- Erdfelder, E. (1990a). Deterministic developmental hypotheses, probabilistic rules of manifestation, and the analysis of finite mixture distributions. In A. von Eye (Ed.), **Statistical methods in longitudinal research. Volume II: Time series and categorical longitudinal data** (pp. 471-509). Boston: Academic Press.
- Erdfelder, E. (1990b). Probabilistische Ereignisbäume als restringierte univariate Latent-Class-Modelle. In D. Frey (Hrsg.), **Bericht über den 37. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1990. Band 1: Kurzfassungen** (S.590-591). Göttingen: Hogrefe.
- Erdfelder, E. (1991). Prädiktionsanalyse und die Analyse latenter Klassen: Welches Verfahren für welche Fragestellung? In A. von Eye (Hrsg.), **Prädiktionsanalyse. Vorhersagen mit kategorialen Variablen** (S. 157-191). Weinheim: Psychologie Verlags Union.
- Erdfelder, E. (1993a). **Entscheidungen über statistische und wissenschaftliche Hypothesen: Welche Fehlerwahrscheinlichkeiten sind zu reduzieren?** (Manuskript, zur Veröffentlichung eingereicht). Bonn: psychologisches Institut der Universität.
- Erdfelder, E. (1993b). **Entscheidungen über statistische und wissenschaftliche Hypothesen: Konsequenzen des Validitätspostulats**. (Manuskript, zur Veröffentlichung eingereicht). Bonn: Psychologisches Institut der Universität.
- Estes, W. K. (1956). The problem of inference from curves based on groups data. *Psychological Bulletin*, 53, 134-140.
- Falmagne, J. C. (1979). On a class of probabilistic conjoint measurement models: Some diagnostic properties. *Journal of Mathematical Psychology*, 19, 73-88.

- Falmagne, J. C., Iverson, G. J. & Marcovici, S. (1979). Binaural loudness summation: Probabilistic theory and data. *Psychological Review*, 86, 25-43.
- Faul, F., Erdfelder, E. & Buchner, A. (1993). **GPOWER: A general power analysis program**. (Manuskript, zur Veröffentlichung eingereicht). Bonn: Psychologisches Institut der Universität.
- Festinger, L. & Carlsmith, J.M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58, S.203-210.
- Gabriel, K. R. & Hsu, C.-F. (1983). Evaluation of the power of revandomization tests, with application to weather modification experiments. *Journal of the American Statistical Association*, 78, 766-775.
- Gadenne, V. (1976). **Die Gültigkeit psychologischer** Untersuchungen. Stuttgart: Kohlhammer.
- Gadenne, V. (1984). **Theorie und Erfahrung in der psychologischen Forschung**. Tübingen: Mohr.
- Green, D.M. & Swets, J.A. (1974). **Signal detection theory and psychophysics**. New York: Krieger.
- Groeben, N. & Westmeyer, H. (1981<sup>2</sup>). **Kriterien psychologischer Forschung**. München: Juventa.
- Hager, W. (1987). Grundlagen einer Versuchsplanung zur Prüfung empirischer Hypothesen der Psychologie. In G. Lüer (Hrsg.), **Allgemeine Experimentelle Psychologie** (S. 43-264). Stuttgart: Fischer.
- Hager, W. (1992). **Jenseits von Experiment und Quasi-Experiment. Zur Struktur psychologischer Versuche und zur Ableitung von Vorhersagen**. Göttingen: Hogrefe.
- Hager, W. & Westermann, R. (1983a). Entscheidung über statistische und wissenschaftliche Hypothesen: Probleme bei mehrfachen Signifikantests zur Prüfung einer wissenschaftlichen Hypothese. *Zeitschrift für Sozialpsychologie*, 14, 106-117.
- Hager, W. & Westermann, R. (1983b). Zur Wahl und Prüfung statistischer Hypothesen in psychologischen Untersuchungen. *Zeitschrift für experimentelle und angewandte Psychologie*, 30, 67-94.
- Hager, W. & Westermann, R. (1983c). Planung und Auswertung von Experimenten. In J. Bredenkamp und H. Feger (Hrsg.), **Hypothesenprüfung** (= Enzyklopädie der Psychologie, Serie Forschungsmethoden der Psychologie, Band 5, S. 24-238). Göttingen: Hogrefe.
- Herrmann, T. (1973). **Persönlichkeitsmerkmale. Bestimmung und Verwendung in der psychologischen Wissenschaft**. Stuttgart: Kohlhammer.
- Jöreskog, K.G. & Sörbom, D. (1988). **LZSREL 7. A guide to the program and its applications**. Chicago: SPSS Inc.
- Johnson, N. L. & Kotz, S. (1970). **Distributions in statistics. Continuous univariate distributions** - 1. New York: Wiley.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programs. In I. Lakatos & A. Musgrave (eds.), **Criticism and the growth of knowledge** (pp. 91-195). Cambridge: University Press.
- Lazarsfeld, P. F. & Henry, N. W. (1968). **Latent structure analysis**. Boston: Houghton Mifflin.

- Lienert, G. A. (1973). **Verteilungsfreie Methoden der Biostatistik, Band 1**. Meisenheim am Glan: Hain.
- Luce, R. D. (1986). **Response times. Their role in inferring elementary mental organization**. New York: Oxford University Press.
- Luce, R.D., Krantz, D.H., Suppes, P. & Tversky, A.P. (1990). **Foundations of measurement. Volume III: Representation, axiomatization, and invariance**. San Diego: Academic Press.
- Popper, K. R. (1982<sup>7</sup>). **Logik der Forschung**. Tübingen: Mohr.
- Riefer, D. M. & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. **Psychological Review**, *91*, 318-339.
- Rost, J. & Langeheine, J. (1991). Mischverteilungsmodelle: die Methodologie der kommenden Jahre. In D. Frey (Hrsg.), **Bericht über den 37. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1990, Band 2** (S. 622-626). Göttingen: Hogrefe.
- Sievers, W. (1990). Bootstrap-Konfidenzintervalle und Bootstrap-Akzeptanz-Bereiche hypothesenprüfender Verfahren. **Zeitschrift für experimentelle und angewandte Psychologie**, *37*, 85-123.
- Snodgrass, J. G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. **Journal of Experimental Psychology: General**, *117*, 34-50.
- Stegmüller, W. (1973). **Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band 4: Personelle und statistische Wahrscheinlichkeit. 2. Halbband: Statistisches Schließen**. Berlin: Springer.
- Steyer, R. (1988). Conditional expectations: An introduction to the concept and its applications in empirical sciences. **Methodika**, *2*, 53-78.
- Steyer, R. (1992). **Theorie kausaler Regressionsmodelle**. Stuttgart: Fischer.
- Tack, W. H. (1976). **Stochastische Lernmodelle**. Stuttgart: Kohlhammer.
- Titterton, D.M., Smith, A. F. M. & Makov, U.E. (1985). **Statistical analysis of finite mixture distributions**. New York: Wiley.
- Townsend, J. T. & Ashby, F. G. (1983). **Stochastic modeling of elementary psychological processes**. Cambridge: Cambridge University Press.
- Wald, A. (1948). **Sequential analysis**. New York: Wiley.
- Westermann, R. (1987). Wissenschaftstheoretische Grundlagen der experimentellen Psychologie. In G. Lüer (Hrsg.), **Allgemeine Experimentelle Psychologie (S. 5-42)**. Stuttgart: Fischer.
- Westermann, R. & Hager, W. (1986). Error probabilities in educational and psychological research. **Journal of Educational Statistics**, *11*, S. 117-146.
- Westmeyer, H. (1972). **Logik der Diagnostik**. Stuttgart: Kohlhammer.
- Westmeyer, H. (1973). **Kritik der psychologischen Unvernunft**. Stuttgart: Kohlhammer.
- Wickens, T.D. (1982). **Models for behavior. Stochastic processes in psychology**. San Francisco: Freeman.
- Wilks, S.S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. **Annals of Mathematical Statistics**, *9*, 60-62.

Willmes, K. (1987). **Beiträge zu Theorie und Anwendung von Permutationstests in der uni- und multivariaten Datenanalyse.** (Unveröffentlichte Dissertation). Trier: Fachbereich 1 - Psychologie der Universität.

### **Autorenhinweis**

Die Verfasser danken den Mitautoren dieses Bandes sowie den Herren Dr. J. Andres und Prof. Dr. A. Iseler für kritische Kommentare zur Erstfassung des vorliegenden Kapitels. Lisa Irmen hat das Manuskript korrekturgelesen und Formulierungsänderungen vorgeschlagen, die wir größtenteils übernommen haben. Auch ihr sei herzlich gedankt.



## Stochastische Modelle

**Rolf Steyer**

“ ... the true logic for this world is the calculus of probabilities ...”

J. Clerk Maxwell

In den Humanwissenschaften wie der Psychologie, Soziologie und Ökonomie verstehen sich viele Forscher als empirische Wissenschaftler. Neben dem Kriterium der **logischen Widerspruchsfreiheit** ist für sie die **Erfahrung** oder **Empirie** das wesentliche Korrektiv für Theorien. Die Theorien einer **empirischen** Wissenschaft müssen etwas über unsere **Erfahrung** aussagen. Nur dann sind sie an ihr überprüfbar. Die **deduktivistisch** oder **falsifikationistisch** orientierten Wissenschaftler verlangen von einer empirischen Theorie, daß aus ihr Aussagen über die Empirie **logisch** abgeleitet werden können (vgl. hierzu das Abgrenzungsproblem bei Papper, 1984). Aus der Sicht einer deduktivistischen Methodologie sind die **logische Widerspruchsfreiheit** und die **logische Ableitbarkeit** von Aussagen über beobachtbare Sachverhalte die beiden wichtigsten Kriterien, denen eine empirische Theorie genügen sollte. Alle weiteren Kriterien dienen nur noch dazu, verschiedene empirische Theorien untereinander zu bewerten.

Sowohl um die logische Widerspruchsfreiheit überprüfen zu können, als auch um Aussagen über die Empirie ableiten zu können, muß eine Theorie in einer formalen Sprache formuliert sein oder in diese übersetzt werden können, denn nur dort sind die Regeln des logischen Schließens anwendbar (s. dazu auch Erdfelder & Bredenkamp, Kap. 14, Abschnitt 4.5, in diesem Band). Dabei ist allerdings anzumerken, daß bisher nur wenige empirische Theorien diese beiden Kriterien erfüllen, die m.E. aber dennoch als anzustrebende Ideale unverzichtbar sind, jedenfalls dann, wenn es auf Präzision ankommt. Für viele Alltagszwecke reichen natürlich auch umgangssprachlich formulierte Theorien aus.

**Aufgaben stochastischer Modelle:** Stochastische Modelle, **um die es in diesem** Artikel geht, genügen den o.g. beiden Kriterien. Sie sind in einer formalen

Sprache - der Sprache der Wahrscheinlichkeitstheorie - formuliert und erlauben daher die Überprüfung ihrer logischen Widerspruchsfreiheit und die Ableitung von Aussagen über die Empirie. Stochastische Modelle sind insbesondere für Anwendungen in der Psychologie und den benachbarten Sozialwissenschaften wegen des **Meßfehlerproblems** und allgemein des **Problems der multiplen Determiniertheit** geeignet: Beobachtungen und Messungen sind fehlerbehaftet und die zu erklärenden Phänomene haben viele verschiedene Ursachen, die man nur selten alle kennt, so daß deterministische Erklärungen selten möglich sind. In einer empirischen Theorie haben stochastische Modelle im wesentlichen **zwei Hauptaufgaben** (zur Beziehung zwischen Theorie und Modell siehe Baizer-, Moulines & Sneed, 1987); sie explizieren die Verknüpfung zwischen:

- (a) empirischen und theoretischen Begriffen und damit das **Meßmodell**,
- (b) den theoretischen (bzw. empirischen) Begriffen und damit die **Abhängigkeitsbegriffe**.

Betrachten wir als Beispiel die Hypothese „Frustration führt zu Aggression“! Hier kommen die beiden theoretischen Begriffe „Frustration“ und „Aggression“ vor, die mit dem Abhängigkeitsbegriff „führt zu“ verknüpft sind. Alle drei Begriffe haben zunächst nur umgangssprachliche Bedeutungen, die für den Alltag auch hinreichend präzise sein mögen. Dem Präzisionsanspruch einer empirischen Wissenschaft genügt die umgangssprachliche Formulierung der Frustrations-Aggressions-Hypothese jedoch nicht, da sie allzu viele Fragen offen läßt. Ist mit ihr die deterministische Aussage gemeint:

- (a) Für alle Menschen gilt: wenn sie frustriert sind, reagieren sie aggressiv?

Oder ist lediglich die folgende probabilistische Aussage gemeint:

- (b) Für alle Menschen gilt: wenn sie frustriert sind, ist die Wahrscheinlichkeit, daß sie aggressiv reagieren, erhöht?

Sind „Frustration“ und „Aggression“ qualitative, komparative oder gar metrische Begriffe? Wären es nur qualitative Begriffe, die nur die Ausprägungen „vorhanden\* und „nicht vorhanden“ haben, dann wären nur die obige deterministische (a) und die probabilistische Präzisierung (b) der Frustrations-Aggressions-Hypothese möglich. Handelt es sich aber um komparative oder gar metrische Begriffe, wären auch weitere Präzisierungen möglich wie z.B.:

- (c) Für alle Menschen gilt: je mehr sie frustriert werden, desto stärker reagieren sie aggressiv.
- (d) Für alle Menschen gilt: je mehr sie frustriert werden, desto höher ist die Wahrscheinlichkeit, daß sie aggressiv reagieren.
- (e) Für alle Menschen gilt: je mehr sie frustriert werden, desto höher ist der Erwartungswert ihrer Aggressivität.

Dieses Beispiel ließe sich leicht über mehrere Seiten fortsetzen. Aber auch mit den obigen Ausführungen dürfte folgendes schon hinreichend klar geworden sein: Damit

aus der Frustrations-Aggressions-Hypothese eine Hypothese einer **deduktiven** empirischen Wissenschaft werden kann, müssen die theoretischen Begriffe „Frustration“ und „Aggression“, aber auch der Abhängigkeitsbegriff „führt zu“ präzisiert werden, und zwar so, daß eine Verknüpfung zwischen theoretischen Begriffen und beobachtbaren Sachverhalten hergestellt wird, die logisch-mathematische Deduktionen ermöglicht. Umgangssprachlich formulierte Hypothesen wie die Frustrations-Aggressions-Hypothese als „wissenschaftliche Hypothesen“ zu bezeichnen (s. z.B. Hager, 1987, **1992**), halte ich für bedenklich. Zwar sind es Hypothesen, die Wissenschaftler zu einem bestimmten Zeitpunkt im Prozeß der Entwicklung ihrer Theorien haben, aber sie sind m.E. nicht das, wohin Theorienentwicklung zielen sollte. Für solche umgangssprachlich formulierte Hypothesen, die im Wissenschaftsprozeß durchaus nützlich sind, gibt es in der Regel nicht nur eine einzige, sondern viele verschiedene Möglichkeiten der Präzisierung. Daher können umgangssprachlich formulierte Hypothesen m. E. nicht das Endziel, sondern nur das Rohmaterial darstellen, aus dem nach entsprechender Bearbeitung Hypothesen konstruiert werden können, aus denen sich empirische Aussagen logisch deduzieren lassen. Sicherlich kommen Wissenschaftler bei ihrer Theorienentwicklung nicht nur mit Deduktion aus, sondern müssen auch an vielen Stellen induktive Schritte tun (s. dazu Westermann & Gerjets, Kap. 10 in diesem Band), aber dennoch sollte man nicht verwischen, wo man induktiv und wo man deduktiv arbeitet. Im letzteren Fall weiß man nämlich, wo man absolute Sicherheit über die Gültigkeit der Schlußfolgerungen hat, im ersteren Fall dagegen, fehlt diese Sicherheit. Die Unterscheidung zwischen deduktiven und induktiven Schlüssen ist also insbesondere bei der Theorienkritik und -revision von Bedeutung, da sie unsere Aufmerksamkeit auf diejenigen Stellen zu richten erlaubt, die möglicherweise falsch sein können, wohingegen andere Teile der Theorie schon aus logischen Gründen nicht falsch sein können. Im Abschnitt 3 werden wir ein Beispiel für den letzten Fall kennenlernen, nämlich die Unkorreliertheit von Fehler- und True-Score-Variablen.

Auch andere Beispiele zeigen, daß unsere Umgangssprache zwar viele Begriffe enthält, mit denen wir nichtdeterministische Abhängigkeiten beschreiben können, aber für viele Zwecke ist sie zu unpräzise. Was bedeutet z.B. die Aussage „Rauchen fördert Lungenkrebs“. Ist damit eine korrelative Aussage über eine Population gemeint? Wenn ja, handelt es sich um eine lineare Abhängigkeit oder gibt es bestimmte Schwellen, an denen die Wahrscheinlichkeit an Lungenkrebs zu erkranken, stärker ansteigt? Ist die Abhängigkeit in allen Teilpopulationen gleich, oder gibt es Populationen, die trotz Rauchen weniger gefährdet sind? Ist gar eine kausale Abhängigkeit gemeint und wenn ja in welchem Sinn? (Eine deterministische Abhängigkeit ist wohl auszuschließen.) Handelt es sich vielleicht gar nicht um eine Populationsaussage, sondern bezieht sie sich auf jedes Individuum? Auch hier stellt sich die Frage nach der Art der Abhängigkeit und nach eventuellen interindividuellen Differenzen.

Die Mathematik hat uns in den letzten Jahrhunderten und insbesondere seit dem Erscheinen von Kolmogoroffs (1970) Buch „Grundbegriffe der Wahrscheinlichkeitsrechnung“ (im Jahre 1933) eine exakte Sprache, die Wahrscheinlichkeitstheorie oder Stochastik, zur Verfügung gestellt, mit der wir die Phänomene der Sozialwissenschaften und ihre Abhängigkeiten präzise beschreiben können. Der Preis dafür ist allerdings, daß wir zunächst diese formale

Sprache lernen müssen, um die Bedeutung verschiedener Arten stochastischer Abhängigkeiten verstehen zu können.

**überblick:** Der diesem Artikel vorgegebene Umfang verbietet es, exakte Definitionen und mathematische Sätze darzustellen. Stattdessen kann es nur darum gehen, den Leser zum Studium der Wahrscheinlichkeitstheorie und ihrer Anwendungen in der Psychologie zu motivieren. Dies hoffe ich dadurch zu erreichen, daß ich die Relevanz stochastischer Modelle zur Konstruktion empirischer wissenschaftlicher Theorien herauszuarbeiten versuche. Dies ist natürlich nicht möglich, ohne vorher die **Bestandteile eines stochastischen Modells** einzuführen und ihre Bedeutung bei ihrer Anwendung in der Psychologie zu erläutern. Dies umreißt den Gegenstand des **Abschnitts 1**.

Im **Abschnitt 2** geht es um verschiedene **Arten stochastischer Abhängigkeiten**. In **Abschnitt 2.1** wird die **stochastische Abhängigkeit** von Ereignissen und Variablen behandelt und im **Abschnitt 2.2** sind es verschiedene Arten **regressiver** und **korrelativer Abhängigkeit**. Dabei wird darauf hingewiesen, daß es nicht nur Abhängigkeiten zwischen zwei Variablen oder Ereignissen gibt, sondern daß oft erst die gleichzeitige Betrachtung vieler Variablen oder Ereignisse ein angemessenes Bild der Realität ergibt. Selbst bei einer bivariaten, aber noch mehr bei einer multivariaten Betrachtung, müssen wir zwischen verschiedenen Arten stochastischer Abhängigkeit unterscheiden, die nicht nur von methodischem, sondern auch von inhaltlichem Interesse sind, da sie sozusagen einen abstrahierten Inhalt darstellen, der vielen Anwendungen gemeinsam ist. Jede Art dieser stochastischen Abhängigkeiten ist auch inhaltlich anders zu interpretieren. Im **Abschnitt 2.3** geht es um kausale **regressive Abhängigkeiten**. Das Kausalitätsproblem stellt sich z.B. dann, wenn zu vermuten ist, daß eine betrachtete stochastische Abhängigkeit einer Variablen Y von einer weiteren Variablen X nicht durch eine Ursache-Wirkungs-Beziehung zwischen diesen beiden Variablen zustande kommt, sondern dadurch, daß beide von einer oder mehreren „Drittvariablen“ beeinflußt werden. Eng verknüpft mit dem Kausalitätsproblem sind die verschiedenen Techniken der experimentellen Versuchsplanung, wie z.B. die randomisierte Aufteilung der Beobachtungseinheiten auf die Versuchsbedingungen. Die in vielen umgangssprachlich formulierten Theorien vorkommende **Ceteris-paribus-Klausel** („unter sonst gleichen Bedingungen gilt: ...“) ist als Versuch anzusehen, Hypothesen über kausale Abhängigkeiten zu formulieren. Als problematisch ist eine derartige Formulierung m.E. deswegen zu bewerten, weil damit meist Unmögliches - die Konstanthaltung **aller** Störvariablen - gefordert wird. Dies wird aber selbst im randomisierten Experiment nicht erreicht.

Die Thematik des **Abschnitts 3** knüpft an der ersten der o.g. Aufgaben stochastischer Modelle an, der Verknüpfung zwischen theoretischen und empirischen Begriffen. Es wird exemplarisch gezeigt, wie man in **stochastischen**

**Meßmodellen** theoretische Begriffe konstruieren kann. Dabei wird deutlich, daß man über die übliche Praxis hinausgehen kann, nach der Konstrukte nur aus Namen (wie z.B. „Intelligenz“, „Bildhaftigkeit“) bestehen, mit denen wir bestimmte Assoziationen und relativ vage Hypothesen verknüpfen. Stattdessen kann man Konstrukte mathematisch konstruieren, indem man bestimmte Annahmen über die Eigenschaften bekannter Größen einführt und daraus die Existenz des Konstrukts ableitet, dessen Skalenniveau deduziert sowie die explizite mathematische Beziehung zwischen Konstrukt und beobachtbaren Variablen angibt, woraus sich auch die genaue inhaltliche Bedeutung des Konstrukts erschließt. Es liegt auf der Hand, daß es erst mit derartigen Konstrukten möglich wird, eine deduktivistische empirische Wissenschaft zu betreiben, die auf logisch-mathematischer Argumentation und nicht nur auf den weithin üblichen Plausibilitätsüberlegungen basiert.

Damit dürfte klar geworden sein, daß es in diesem Artikel weder um einen Überblick über die verschiedenen stochastischen Modelle gehen soll, noch um einen Überblick über deren Anwendungen in den Sozialwissenschaften. Dazu sei auf einige Arbeiten verwiesen, die wenigstens für bestimmte Teilbereiche einen solchen Überblick versuchen, nämlich Austin und Wolfle (1991), Bentler (1986), Lewis (1986), Mulaik (1986), Rost und Strauß (1992) sowie Torgerson (1986).

## 1. *Bestandteile und Gegenstand stochastischer Modelle*

Was sind stochastische Modelle? Aus welchen Bestandteilen bestehen sie? Was ist ihr Gegenstand? Allgemein gesprochen handelt es sich dabei um Gebilde, die in der Sprache der Wahrscheinlichkeitstheorie formuliert sind, aus bestimmten Bestandteilen bestehen und bestimmte Phänomene beschreiben und erklären sollen, die in **Zufallsexperimenten** auftreten können. Dieser Gegenstand klingt zunächst sehr bescheiden, beinhaltet aber weit mehr, als es auf den ersten Blick erscheinen mag.

### 1.1 Zufallsexperiment

Alle Aussagen, die im Rahmen eines stochastischen Modells formuliert werden, wie z.B. über die Korrelation zweier Variablen oder irgendeine andere Art des Zusammenhangs zwischen Variablen oder Ereignissen, beziehen sich auf das betrachtete Zufallsexperiment und können daher nicht ohne weiteres auf andere Zufallsexperimente übertragen werden. Bei stochastischen Modellen geht es also nicht um generelle Allaussagen, sondern lediglich um die Gesetzmäßigkeiten in dem jeweils betrachteten Zufallsexperiment. Das zu-

grundlegende Zufallsexperiment ist der „Geltungsbereich“ oder das „empirische Phänomen“, von dem in stochastischen Aussagen die Rede ist. Dies schließt natürlich nicht aus, daß in komplexeren Theorien generelle Aussagen über Klassen von stochastischen Modellen gemacht werden, was zu einem entsprechend größeren Geltungsbereich der Theorie führt.

Dies ist in den Sozialwissenschaften nicht anders als in den klassischen Naturwissenschaften, allerdings sind dort die Randbedingungen, die ein Zufallsexperiment definieren, oft überdauernder als in den Sozialwissenschaften. Die Masseverteilung in einem Holzwürfel ist bspw. über viele Jahrzehnte konstant und kann sich eigentlich nur durch die leicht zu diagnostizierenden Aktivitäten eines Holzwurms verändern. Das Zufallsexperiment „Werfen des Würfels“ kann daher nahezu beliebig oft wiederholt werden. In den Sozialwissenschaften dagegen können manche Zufallsexperimente überhaupt nicht wiederholt werden. Das Zufallsexperiment „Lösen einer bestimmten Intelligenztestaufgabe durch Fritz Müller“ kann wegen des zu erwartenden Lerneffekts unter gleichen Bedingungen nicht wiederholt werden. Dennoch ist es auch hier sinnvoll, ein Zufallsexperiment zu unterstellen und Modelle darüber zu formulieren, wie bspw. die Lösungswahrscheinlichkeit von der Fähigkeit der Person und der Schwierigkeit der Aufgabe abhängt (s. dazu das Rasch-Modell z.B. bei de Gruijter & van der Kamp, **1984**; Fischer, **1974**; **1981**, **1983**, **1988**; Kubinger, **1988**; Rasch, **1980**; Rost, **1988**; Steyer & Eid, **1993**; Tutz, **1989**).

Stochastische Modelle haben es zwar oft, aber nicht immer mit Massenphänomenen zu tun, die in großen Stichproben auftreten. Wenn wir bspw. eine Vorhersage der Wahrscheinlichkeit benötigen, mit der Fritz Müller eine bestimmte Aufgabe löst, so können wir dies auch dann tun, wenn diese Aufgabe nur ein einziges Mal zu lösen ist. Jede Entscheidung (z.B. im beruflichen Kontext), eine bestimmte Person für eine bestimmte Aufgabe auszuwählen, beruht letztlich auf der Abschätzung einer solchen Wahrscheinlichkeit. Natürlich benötigen wir dazu Informationen über die Fähigkeit der betreffenden Person und die Aufgabenschwierigkeit, die z.B. aus einem ähnlichen Zufallsexperiment stammen können. Mit stochastischen Modellen erhebt man also keinen Allgemeinheitsanspruch, sondern bezieht sich prinzipiell auf ein bestimmtes, noch durchzuführendes Zufallsexperiment, das auch nicht unbedingt wiederholbar sein muß. Wie bereits erwähnt, spricht jedoch nichts dagegen, stochastische Modelle in umfassendere Theorien einzubetten, für die dann ein allgemeinerer Geltungsanspruch erhoben werden kann.

**Beispiel 1:** Ein Zufallsexperiment liegt bspw. vor, wenn wir eine Person aus einer Menge von Personen (einer Grundgesamtheit oder **Population**) ziehen, einen psychologischen Test bearbeiten lassen und ihr Antwortmuster registrieren. Diese Art von Zufallsexperimenten wird bei vielen stochastischen Meßmodellen zugrundegelegt (s. ausführlicher Steyer & Eid, 1993). Das Ziehen einer Stichprobe vom Umfang  $N$  besteht aus dem  $N$ -maligen Wiederholen dieses Zufallsexperiments, wobei allerdings die Grundgesamtheit mit jedem Ziehen um die bereits gezogene Person kleiner wird.

**Beispiel 2:** Bei anderen Arten von Zufallsexperimenten wird nur das Antwortmuster **einer** gegebenen **Person** als Ergebnis des Zufallsexperiments betrachtet (s. z.B. Rost, 1988). Bei den einfachsten Meßmodellen wird das Zustandekommen eines Testwerts durch das Zusammenwirken der zu messenden Eigenschaft und eines Meßfehlers erklärt. Bei komplexeren Meßmodellen können die Schwierigkeit des Meßinstruments (des Tests oder eines einzelnen Items; s. z.B. Rasch, 1980), situationale Effekte und zusätzliche Personeigenschaften als weitere Variablen hinzugezogen werden (s. z.B. Steyer, Ferring & Schmitt, 1992), um das Zustandekommen eines Testwerts zu erklären.

**Beispiel 3:** Ein anderes, in der Psychologie typisches Zufallsexperiment, besteht aus dem Ziehen einer Person aus einer Grundgesamtheit, ihrer Zuweisung zu einer von mehreren experimentellen Bedingungen und der Registrierung ihres Werts auf einer oder mehreren Kriteriumsvariablen. Neben Meßfehlern und anderen nicht bekannten Störvariablen erklären in solchen Modellen die experimentellen Bedingungen das Zustandekommen der Werte der Kriteriumsvariablen (s. z.B. Kap. 10 von Steyer, 1992).

## 1.2 Wahrscheinlichkeitsraum

Notwendiger Bestandteil jedes stochastischen Modells ist ein Wahrscheinlichkeitsraum, der die (formal)sprachliche Repräsentation des betrachteten Zufallsexperiments und damit des betrachteten empirischen Phänomens darstellt. Ein Wahrscheinlichkeitsraum besteht aus drei Komponenten:

- (a) der Menge der möglichen Ergebnisse (des betrachteten Zufallsexperiments),
- (b) der Menge der möglichen Ereignisse,
- (c) dem Wahrscheinlichkeitsmaß.

Die **Menge der möglichen Ergebnisse** beschreibt die formale Struktur des empirischen Phänomens. Die **Menge der möglichen Ereignisse** gibt an, von welchen Ereignissen man sprechen können will, und das **Wahrscheinlichkeitsmaß** ist eine Funktion, die jedem möglichen Ereignis eine Wahrscheinlichkeit zuordnet. In Anwendungen geschieht diese Zuordnung der Wahrscheinlichkeiten meist nicht explizit, da diese Wahrscheinlichkeiten gar nicht bekannt sind. Wenn man von **der** Wahrscheinlichkeit eines Ereignisses spricht, wird aber bereits vorausgesetzt, daß diese Zuordnung existiert.

Ich verwende hier die etwas umständlichen (und auch unüblichen) Bezeichnungen „**mögliche** Ergebnisse“ und „mögliche Ereignisse“, um damit die Unterschiede zwischen einem in einem bereits durchgeführten Zufallsexperiment **aufgetretenen** Ergebnis bzw. Ereignis einerseits und einem **möglichen** Ergebnis bzw. Ereignis in einem (betrachteten, noch durchzuführenden) Zufallsexperiment andererseits hervorzuheben. Ziel stochastischer Modelle ist nämlich nicht in erster Linie die Beschreibung der Sy-

stematik bereits beobachteter Ereignisse, sondern die Angabe der Gesetzmäßigkeiten, die das Zufallsexperiment und die dabei **möglichen** Ereignisse charakterisieren. Indirekt werden damit natürlich auch die in einem durchgeführten Experiment **tatsächlich** aufgetretenen Ereignisse erklärt.

**Menge der (möglichen) Ergebnisse:** Zur Illustration betrachten wir das oben bereits angeführte Zufallsexperiment des Ziehens einer Person  $u$  aus einer Menge  $U$  (Grundgesamtheit) und des Registrierens ihres Testergebnisses, das z.B. aus den möglichen Kombinationen des Lösens (+) oder Nichtlöstens (-) von zehn Aufgaben bestehen möge. Ein mögliches Ergebnis des betrachteten Zufallsexperiment wäre dann z.B.: (Fritz Meier, +, +, -, +, -, -, -, -, +, +). Demnach wurde also Fritz Meier gezogen und dieser löste die Aufgaben 1, 2, 4, 9, 10 und die anderen fünf Aufgaben nicht. Die Menge  $\Omega$  der möglichen Ergebnisse dieses Zufallsexperiments ist dann ein Kreuzprodukt der Form  $\Omega = U \times \mathbf{M}$ , wobei  $\mathbf{M} := \{+, -\}^{10} = \{+, -\} \times \dots \times \{+, -\}$  (10 Faktormengen).

Ziel eines stochastischen Modells in einem derartigen Zufallsexperiment könnte z.B. sein zu beschreiben, wie die **Lösungswahrscheinlichkeit** einer Aufgabe von ihrer **Schwierigkeit** und der **Fähigkeit** der gezogenen Person abhängt. Darüber hinaus erlaubt ein solches Modell erst die **Konstruktion der Schwierigkeits- und Fähigkeitsbegriffe**, die für die Lösung der betrachteten Aufgaben relevant sind. Beide Begriffe beziehen sich zunächst nur auf das betrachtete Zufallsexperiment.

**Menge der (möglichen) Ereignisse:** Ereignisse in einem solchen Zufallsexperiment sind Teilmengen von  $\Omega$ . Beim oben beschriebenen Zufallsexperiment ist das Ereignis, Fritz Meier zu ziehen, (Fritz Meier)  $\times \mathbf{M}$ . Das Ereignis, Fritz Meier **oder** Franz Müller zu ziehen, ist  $\{\text{Fritz Meier}\} \times \mathbf{M} \cup \{\text{Franz Müller}\} \times \mathbf{M}$ , und das Ereignis, daß die gezogene Person (wer auch immer das ist) die erste Aufgabe löst, ist  $U \times \{+\} \times \{+, -\}^9$ . In jedem Fall ist das aufgeführte Ereignis eine Teilmenge von  $\Omega$ .

Die möglichen Ereignisse kann man wieder zu einer Menge zusammenfassen, z.B. der Menge aller Teilmengen von  $\Omega$ , der **Potenzmenge** von  $\Omega$ . Man muß jedoch nicht immer alle Teilmengen von  $\Omega$  als mögliche Ereignisse betrachten. Wahrscheinlichkeiten können auch dann schon sinnvoll definiert werden, wenn man nur eine Teilmenge der Potenzmenge von  $\Omega$  betrachtet, die man als  $\sigma$ -Algebra bezeichnet und mit Zeichen wie z.B.  $\mathcal{A}$  notiert (zur Definition einer  $\sigma$ -Algebra s. Bauer, 1978, S. 16). Eine  $\sigma$ -Algebra - und daher jede Ereignismenge - ist abgeschlossen gegenüber abzählbaren Vereinigung und Schnittmengenbildungen, d.h. die Vereinigungs- und die Schnittmenge von Elementen aus  $\mathcal{A}$  sind ebenfalls Elemente aus  $\mathcal{A}$ .

**Wahrscheinlichkeitsmaß:** Jedem (möglichen) Ereignis  $A$  ist durch das Wahrscheinlichkeitsmaß (W-Maß)  $P$  eine Wahrscheinlichkeit  $P(A)$  zugeordnet. In den meisten Anwendungen sind diese Wahrscheinlichkeiten unbekannt. Den-



noch muß natürlich vorausgesetzt werden, daß diese Wahrscheinlichkeiten existieren. Empirische Untersuchungen dienen in der Regel dazu, diese Wahrscheinlichkeiten zu schätzen oder allgemeiner formuliert, einige Aussagen über diese Wahrscheinlichkeiten machen zu können.

Die Wahrscheinlichkeit eines Ereignisses soll eine Zahl zwischen 0 und 1 (einschließlich) sein. Eine weitere wichtige definierende Eigenschaft ist die Additivität eines Wahrscheinlichkeitsmaßes, d.h. die Eigenschaft

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots, \quad (1.1)$$

falls diese Ereignisse **paarweise disjunkt** sind, falls also für jedes Paar dieser Ereignisse gilt:  $A_i \cap A_j = \emptyset$ , falls  $i \neq j$ . Wenn die Ereignisse paarweise disjunkt sind, dann addieren sich also ihre Einzelwahrscheinlichkeiten zur Wahrscheinlichkeit dafür, daß mindestens eines dieser Ereignisse eintritt. Man beachte auch, daß es erst mit der Einführung eines Wahrscheinlichkeitsmaßes sinnvoll wird, von Ereignissen etc. zu sprechen. Vorher handelt es sich nur um Teilmengen der zugrundegelegten Menge  $\Omega$ .

Durch die beiden Mengen  $\Omega$ ,  $\mathcal{A}$  und durch das W-Maß  $\mathbf{P}$ , d.h. durch den Wahrscheinlichkeitsraum (W-Raum)  $\langle \Omega, \mathcal{A}, \mathbf{P} \rangle$ , ist ein Zufallsexperiment beschreibbar. Damit stellt der W-Raum die formalsprachliche Repräsentation des in einem stochastischen Modell betrachteten empirischen Phänomens dar. In einem solchen W-Raum stecken prinzipiell alle Informationen und alle Aussagen, die man über ein betrachtetes Zufallsexperiment formulieren kann. In  $\Omega$  sind alle möglichen Ergebnisse aufgeführt, die bei diesem Zufallsexperiment auftreten können, in  $\mathcal{A}$  sind alle Ereignisse angegeben, von denen man in diesem Kontext sprechen kann und mit dem W-Maß  $\mathbf{P}$  liegen die Wahrscheinlichkeiten aller (möglichen) Ereignisse fest, auch wenn sie in der Regel unbekannt sind. Damit liegt auch fest, wie diese Ereignisse voneinander abhängen, da die Schnittmengen von Ereignissen auch Ereignisse sind (s. dazu unten die Definition der stochastischen Abhängigkeit von Ereignissen).

### 1.3 Zufallsvariablen

**Zufallsvariablen** (Zv) oder, synonym, **stochastische Variablen**, ordnen jedem Ergebnis der Ergebnismenge  $\Omega$  einen Wert zu (z.B. dem Ergebnis „Augenzahl eins“ beim Werfen eines Würfels den Wert 1). Zufallsvariablen sind also Abbildungen mit dem Definitionsbereich  $\Omega$ . Die zugeordneten Werte können Zahlen, aber auch Elemente beliebiger anderer Mengen sein (s. dazu Bauer, 1978, S. 136). Sind die Werte der Zufallsvariablen reelle Zahlen, dann sind unter bestimmten Voraussetzungen **Erwartungswerte** (theoretische Mittelwerte), **Varianzen**, **Kovarianzen** und **Korrelationen** definiert. Diese Größen kennzeichnen bestimmte Eigenschaften der **Verteilung** bzw. der gemeinsamen Ver-

teilung von Zufallsvariablen. (Zur exakten Definition dieser Begriffe sei auf Bauer, 1978, oder Steyer & Eid, 1993, verwiesen.) Mit solchen numerischen Zufallsvariablen kann man quantitative Gesetze formulieren. Dadurch lassen sich empirische Phänomene sehr viel einfacher beschreiben, als dies durch ausschließliche Verwendung von Ereignissen möglich wäre.

Zufallsvariablen bilden die Ergebnisse  $\omega \in \Omega$  eines Zufallsexperimentes nach einer **festen** Zuordnungsvorschrift ab. Die Ergebnisse  $\omega \in \Omega$  sind zufällig und damit indirekt auch die Werte einer Zufallsvariablen. Zufallsvariablen können sehr systematisch voneinander abhängen, auch wenn diese Systematik in der Regel nicht deterministisch ist. So kann man bspw. die Körpergröße (X) und das Geschlecht (Y) als Zufallsvariablen in einem Zufallsexperiment einführen, das aus dem Ziehen einer Person aus einer Population und dem Registrieren des X- und Y-Wertes besteht. Die beiden Variablen X und Y werden sicherlich nicht unabhängig sein, sondern in einer bestimmten Stärke miteinander in Zusammenhang stehen.

Bei Ereignissen haben wir zwischen einem **möglichen** Ereignis und einem **tatsächlich aufgetretenen** Ereignis unterschieden. Die analoge Unterscheidung bei Zufallsvariablen ist die zwischen der Zufallsvariablen selbst und ihrem tatsächlich aufgetretenen Wert bzw. ihrer Realisierung.

## 1.4 Zusammenfassende Bemerkungen

Natürlich konnten hier nur einige grundlegende stochastische Begriffe eingeführt werden. Vollständigere Einführungen, die auch für empirische Wissenschaftler geeignet sind, findet man z.B. bei Bosch (1986), Härtter (1987), Hinderer (1980), Krickeberg und Ziezold (1979), Rényi (1977) und Viertl (1990). An den mathematisch vorgebildeten Leser wenden sich Ash (1972), Bauer (1978), Breiman (1968), Gänsler und Stute (1977), Loeve (1977, 1978) und Plachky (1981).

Mit den oben behandelten drei Bestandteilen des W-Raums und den Zufallsvariablen sind die elementaren Bausteine stochastischer Modelle angegeben. Zugleich ist damit der Rahmen abgesteckt, in dem sich stochastische Modelle bewegen. Allgemein geht es also bei einem stochastischen Modell um die Beschreibung der Gesetzmäßigkeiten des Auftretens von Ereignissen und der Realisierungen von Zufallsvariablen sowie um die Beschreibung der Abhängigkeiten zwischen Ereignissen und zwischen Zufallsvariablen in einem Zufallsexperiment. Aus bereits vorliegenden Daten, die aus einem bereits durchgeführten Zufallsexperiment stammen, versucht man, bestimmte Parameter zu schätzen und Hypothesen über deren Größe zu testen. Diese Parameter sollen aber nicht primär die vorliegenden **Daten** oder die tatsächlich aufgetretenen Ereignisse beschreiben, sondern die Gesetzmäßigkeiten des betrachteten Zufallsexperiments charakterisieren. **Das empirische Phänomen, um das es in erster Linie bei stochastischen Modellen geht, sind also nicht die erhobenen Da-**

**ten und deren Systematisierung, sondern das (durch den  $W$ -Raum repräsentierte) betrachtete Zufallsexperiment und die darin herrschenden Gesetzmäßigkeiten.** Die Daten sind - bildlich gesprochen - lediglich die Spuren, aus denen wir auf die Gesetze schließen wollen, die das Zufallsexperiment steuern. Hat man diese Gesetze erkannt, werden damit natürlich auch die Daten und deren Abhängigkeitsmuster erklärt. Stochastische Modelle unterscheiden sich daher in ihrer Zielsetzung von anderen Modellen und Verfahren (z.B. Multidimensionale Skalierung, Clusteranalyse), die oft nur der Beschreibung und Strukturierung vorliegender Daten dienen.

Zur Beschreibung der Gesetzmäßigkeiten eines Zufallsexperiments werden die (einfachen und bedingten) Wahrscheinlichkeiten von Ereignissen bzw. die Verteilungen von Zufallsvariablen mittels Aussagen über ihre Erwartungswerte, Varianzen u.ä. charakterisiert. Bei der Beschreibung der **Abhängigkeiten** zwischen Zufallsvariablen macht man Aussagen über ihre gemeinsamen Verteilungen, z.B. durch Aussagen über ihre Korrelation, bedingten Erwartungswerte, bedingten Varianzen etc. Verschiedene Arten von in stochastischen Modellen verwendeten Abhängigkeiten und Gesetzen sollen im folgenden Abschnitt behandelt werden.

## 2. Stochastische Abhängigkeitsbegriffe

Abhängigkeiten zwischen Ereignissen oder Variablen sind von zentralem Interesse in jeder empirischen Wissenschaft. Dabei ist oft nicht die Frage, **ob** eine Abhängigkeit zwischen zwei oder mehr Variablen besteht, sondern **wie** diese Abhängigkeit aussieht. Selbst auf der Ebene von Ereignissen können dabei recht komplexe Phänomene auftreten. So kann es z.B. sein, daß drei Ereignisse paarweise unabhängig sind, aber dennoch aus jeweils zwei Ereignissen eine perfekte Vorhersage des dritten möglich ist. In komplexen Systemen, die in der Psychologie und den anderen Sozialwissenschaften betrachtet werden, haben wir es mit dem **Problem der multiplen Determiniertheit** zu tun. Damit ist klar, daß man sich nicht auf die Betrachtung bivariater Abhängigkeiten beschränken, sondern multivariate Zusammenhänge betrachten sollte. Die Behaltensleistung in einer Gedächtnisaufgabe bspw. hängt nicht nur von der Gedächtnisstärke der betrachteten Person ab, sondern auch von der Bildhaftigkeit des Lernmaterials, seiner Bedeutungshaltigkeit, dem Grad der Aktiviertheit der Person, der Stärke ihrer Motiviertheit, bei dieser Aufgabe eine gute Leistung zu erbringen, dem Grad ihrer Müdigkeit, ihrem Trainingsstand, der Art der vorherigen Tätigkeit (Transfereffekte) etc. Die verschiedenen Arten stochastischer Abhängigkeiten zwischen Ereignissen und zwischen Zufallsvariablen, um die es in diesem Abschnitt geht, sind daher von zentraler

Bedeutung für stochastische Modelle, die empirische Phänomene in den Sozialwissenschaften beschreiben und erklären sollen.

## 2.1 Stochastische Abhängigkeit

Der Begriff der stochastischen Abhängigkeit ist sowohl für Ereignisse als auch für Zufallsvariablen definiert. Wir betrachten zunächst die stochastische Abhängigkeit zwischen Ereignissen.

### 2.1.1 Stochastische Abhängigkeit zwischen Ereignissen

Bei der Einführung der stochastischen Abhängigkeit bzw. Unabhängigkeit zwischen Ereignissen kann man vom Begriff der bedingten Wahrscheinlichkeit  $P(A|B)$  ausgehen. Ist diese verschieden von  $P(A)$ , dann sind die beiden Ereignisse  $A$  und  $B$  **abhängig**.

**Bedingte Wahrscheinlichkeit:** Zunächst soll daher der Begriff der bedingten Wahrscheinlichkeit eingeführt werden. Sei  $P(B) > 0$ . Dann heißt die Zahl  $P(A|B)$  **bedingte Wahrscheinlichkeit des Ereignisses A gegeben das Ereignis B** (kurz: B-bedingte Wahrscheinlichkeit von A) genau dann, wenn

$$P(A|B) := \frac{P(A \cap B)}{P(B)}. \quad (2.1)$$

Die linke Seite der obigen Gleichung ist **nicht** etwa so zu lesen, daß **A|B** eine Wahrscheinlichkeit  $P$  zugeordnet wird. Da  $A$  und  $B$  Mengen sind, ist der Ausdruck **A|B** gar nicht definiert. Stattdessen ist der Ausdruck  $P(A|B)$  **so zu verstehen**, daß dem Ereignis  $A$  eine bedingte Wahrscheinlichkeit gegeben  $B$  zugewiesen wird. Um dies auch in der Schreibweise auszudrücken, kann man statt  $P(A|B)$  auch  $P_B(A)$  schreiben. Genau wie  $P$  ist auch  $P_B$  ein  $W$ -Maß auf  $\mathcal{A}$  und verfügt daher über alle Eigenschaften eines  $W$ -Maßes (s. insbesondere G1.1.1). Für verschiedene Ereignisse  $B$  und  $C$  ist im allgemeinen natürlich auch die B-bedingte Wahrscheinlichkeit  $P(A|B) = P_B(A)$  von der C-bedingten Wahrscheinlichkeit  $P(A|C) = P_C(A)$  verschieden.

**Stochastische Unabhängigkeit zweier Ereignisse:** Im Falle der Unabhängigkeit sollten die bedingten und unbedingten Wahrscheinlichkeiten gleich sein, d. h.  $P(A|B) = P(A)$  und  $P(B|A) = P(B)$ . In diesem Fall gelten also:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A) \quad \text{und} \quad P(B|A) = \frac{P(A \cap B)}{P(A)} = P(B). \quad (2.2)$$

Um auch den Fall  $P(A) = 0$  und/oder  $P(B) = 0$  zuzulassen, multipliziert man beide Seiten dieser Gleichungen mit  $P(B)$  bzw.  $P(A)$  und definiert Unabhängigkeit über die resultierende Gleichung. Zwei Ereignisse  $A$  und  $B$  heißen **unabhängig** (bezüglich des Wahrscheinlichkeitsmaßes  $P$ , kurz:  **$P$ -unabhängig**) genau dann, wenn

$$P(A \cap B) = P(A) \cdot P(B). \quad (2.3)$$

Der oben definierte Begriff der Unabhängigkeit zweier Ereignisse ist also ein symmetrischer Begriff. Sprechweisen, die eine Asymmetrie nahelegen, wie „das Ereignis  $A$  ist unabhängig vom Ereignis  $B$ “, sollte man daher möglichst vermeiden. Man beachte auch, daß hier ein spezieller Begriff der Unabhängigkeit definiert wurde, der sich auf eine bestimmte Eigenschaft der **Wahrscheinlichkeiten** der beiden Ereignisse und ihrer Schnittmenge bezieht. Daher spricht man auch von **stochastischer** Unabhängigkeit oder Unabhängigkeit **bzgl. des  $W$ -Maßes  $P$** . Die **stochastische Abhängigkeit** zweier Ereignisse ist als Negation ihrer stochastischen Unabhängigkeit definiert.

**Stochastische Unabhängigkeit mehrerer Ereignisse:** Wenn wir z.B. drei Ereignisse  $A$ ,  $B$  und  $C$  betrachten, dann kann es durchaus sein, daß für alle drei möglichen Paare stochastische Unabhängigkeit besteht, man aber dennoch aus zwei von ihnen das dritte perfekt vorhersagen kann (**Meehlsches Paradoxon**; vgl. Krauth & Lienert, 1973). Würde man sich also nur auf die Betrachtung **paarweiser** Abhängigkeit oder Unabhängigkeit beschränken, würde man in einem solchen Fall eine perfekte Abhängigkeit nicht entdecken. Bei der Definition der stochastischen Unabhängigkeit dreier Ereignisse postuliert man also nicht nur für jedes Paar die Gleichung 2.3, sondern auch

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C). \quad (2.4)$$

Gilt diese Gleichung nicht, dann heißen die drei Ereignisse **stochastisch abhängig**. Aus Gleichung 2.4 folgt 2.3 für jedes Paar dieser drei Ereignisse.

Komplexe stochastische Abhängigkeitsstrukturen zwischen Ereignissen und zwischen kategoriellen Variablen können auf verschiedene Arten beschrieben werden (s. z.B. Goodman & Kruskal, 1979). Man kann solche Abhängigkeitsstrukturen sehr gut mit **log-linearen Modellen** beschreiben, deren Darstellung jedoch hier nicht möglich ist. Stattdessen sei auf Andersen (1990), Bishop, Fienberg und Holland (1977), Langeheine (1986), Hartung (1989) und Knoke und Burke (1980) verwiesen.

### 2.1.2 Stochastische Abhängigkeit zwischen Zufallsvariablen

Die Definition der stochastischen Unabhängigkeit von Zufallsvariablen ist nicht ganz so elementar wie die zwischen Ereignissen. Verbal und notwendigerweise etwas unpräzise kann man formulieren, daß zwei stochastische Variablen  $X$  und  $Y$  stochastisch unabhängig sind, genau dann, wenn alle mit ihnen verknüpften Ereignisse im oben definierten Sinn unabhängig sind; genauer, wenn für alle Ereignisse  $A$ , daß  $X$  einen Wert in einer bestimmten

Teilmenge ihrer Wertemenge annimmt und **B**, daß Y einen Wert in einer bestimmten Teilmenge ihrer Wertemenge annimmt, die Gleichung 2.3 gilt. Eine präzise Definition der Unabhängigkeit von Zufallsvariablen findet man z.B. bei Bauer (1978) oder bei Steyer und Eid (1993).

Eine andere Umschreibung der stochastischen Unabhängigkeit zwischen zwei Zufallsvariablen ist, daß kein Aspekt der Verteilung der einen Variablen von der Ausprägung (dem Wert) der anderen Variablen abhängt. Insbesondere hängen weder der Erwartungswert noch die Varianz oder irgendein anderer Kennwert der Verteilung der einen Variablen ab vom Wert, den die andere Variable annimmt. Die stochastische Abhängigkeit ist natürlich wieder durch die Negation der Unabhängigkeit definiert. Da die Definition der stochastischen Unabhängigkeit von Zufallsvariablen auf die Unabhängigkeit von Ereignissen zurückgeführt wird, kann man die Definition der Unabhängigkeit mehrerer Zufallsvariablen ebenfalls auf die entsprechende Definition für Ereignisse zurückführen.

Die stochastische Unabhängigkeit von Zufallsvariablen spielt z.B. bei randomisierten Experimenten eine Rolle. Bei einem randomisierten Experiment wird eine Beobachtungseinheit (z. B. eine Versuchsperson) definitionsgemäß nach Zufall einer der experimentellen Bedingungen zugeordnet. Damit wird die stochastische Unabhängigkeit zwischen der Treatmentvariablen und allen Variablen hergestellt, die eine Eigenschaft der Beobachtungseinheit **vor** der Zuweisung repräsentieren. Dies hat wichtige Konsequenzen für die kausale Interpretierbarkeit der Abhängigkeit der Kriteriums- von der Treatmentvariablen (näheres dazu in Abschnitt 2.3).

## 2.2 Regressive und korrelative Abhängigkeiten

Während der Begriff der stochastischen **Unabhängigkeit** bei vielen stochastischen und statistischen Modellen eine wichtige Rolle spielt, sind Aussagen über eine stochastische **Abhängigkeit** zu unspezifisch. Es gibt nämlich sehr viele verschiedene Arten stochastischer Abhängigkeit. Die wichtigsten Arten sind die regressiven und korrelativen Abhängigkeiten.

Aussagen über regressive und korrelative Abhängigkeiten können dann getroffen werden, wenn neben einem W-Raum mindestens zwei Zufallsvariablen X und Y vorliegen. Bei **korrelativen** Abhängigkeiten müssen beide Variablen numerisch sein (d.h. Zahlen als Werte annehmen). Aussagen über **regressive** Abhängigkeiten dagegen sind auch dann möglich, wenn nur der Regressand Y numerisch ist. Der Regressor X dagegen kann qualitativ sein, also z.B. die Werte „Experimentalgruppe“ oder „Kontrollgruppe“ annehmen.

Bei einer Aussage über eine regressive Abhängigkeit einer numerischen Zufallsvariablen Y von einer Zufallsvariablen X macht man eine Aussage über die **Regression** oder **bedingte Erwartung**  $E(Y|X)$ . Bei der Regression  $E(Y|X)$

des **Regressanden**  $Y$  auf den **Regressor**  $X$  handelt es sich um eine weitere Zufallsvariable, deren Werte die **bedingten Erwartungswerte**  $E(Y|X = x)$  sind (zur genaueren Definition s. z.B. Bauer, 1978, Steyer, 1988, 1992 oder Steyer & Eid, 1993, Anhang G). Aussagen über eine solche Regression sind nicht nur der Kern der einfachen oder multiplen Regressionsanalyse, sondern auch der Varianzanalyse, der Faktorenanalyse und der Strukturgleichungsmodelle.

Aussagen über Regressionen können in verschiedener Weise formuliert werden, z.B.

- (a) als Regressionskurve in einem Kartesischen Koordinatensystem,
- (b) als Säulendiagramm, mit dem man bedingte Wahrscheinlichkeiten oder Erwartungswerte angibt,
- (c) als Tabelle, in der man bedingte Erwartungswerte in Gruppen angibt,
- (d) als Pfaddiagramm oder auch
- (e) als Gleichung.

Die Darstellungsform hat jedoch nichts mit der logischen Struktur zu tun, die gemeint ist, wenn von Regressionen die Rede ist. In allen genannten Fällen geht es um Aussagen darum, wie die bedingten Erwartungswerte  $E(Y|X = x)$  einer Variablen  $Y$  von den Werten einer (bzw. mehrerer) Variablen  $X$  (bzw.  $X_1, \dots, X_m$ ) abhängen, oder um globale Aussagen darüber, wie stark diese regressive Abhängigkeit ist, z.B. durch Angabe des **Determinationskoeffizienten**

$$R_{Y|X}^2 := \text{Var}[E(Y|X)] / \text{Var}(Y). \quad (2.5)$$

Diese Kenngröße gibt den Anteil der Variation der bedingten Erwartungswerte  $E(Y|X = x)$  um den theoretischen Erwartungswert  $E(Y)$  an der Gesamtvarianz der betrachteten  $Y$ -Variablen an, die sich ihrerseits additiv aus der Varianz  $\text{Var}[E(Y|X)]$  der Regression  $E(Y|X)$  und der Varianz  $\text{Var}(\epsilon)$  der Fehlervariablen  $\epsilon := Y - E(Y|X)$  zusammensetzt.

Der Begriff der Regression eignet sich dazu, verschiedene Arten von Abhängigkeiten und Unabhängigkeiten des Regressanden  $Y$  vom Regressor  $X$  zu beschreiben. **Regressive Unabhängigkeit** der Variablen  $Y$  von  $X$  liegt vor genau dann, wenn gilt:

$$E(Y|X) = E(Y). \quad (2.6)$$

Andernfalls heißt  $Y$  von  $X$  **regressiv abhängig**. Bei regressiver Unabhängigkeit nimmt also die Regression  $E(Y|X)$  für jede Ausprägung  $x$  von  $X$  den gleichen Wert, nämlich  $E(Y)$  an. Ist  $X$  eine kontinuierliche numerische Variable, kann man den Fall der regressiven Unabhängigkeit auch durch eine Regressionsgerade darstellen, die parallel zur  $X$ -Achse verläuft.

### 2.2.1 Einfache regressive Abhängigkeit

Besteht eine regressive Abhängigkeit eines Regressanden  $Y$  von einem Regressor  $X$ , so kann sie verschiedener Art sein. Die einfachste Art der regressiven Abhängigkeit ist die **lineare**, die dann vorliegen kann, wenn es sich bei  $X$  um einen **numerischen** Regressor handelt. **Einfache lineare regressive Abhängigkeit** ist dadurch definiert, daß die Regression  $E(Y|X)$  eine lineare Funktion von  $X$  ist:

$$E(Y|X) = \alpha_0 + \alpha_1 \cdot X, \quad \text{wobei } \alpha_0, \alpha_1 \in \mathbb{R}. \quad (2.7)$$

Kann  $X$  nur zwei verschiedene Werte annehmen, dann muß sich die Regression  $E(Y|X)$  sogar als eine solche lineare Funktion von  $X$  darstellen lassen. Um dies zu erkennen, trage man die beiden Punkte  $[x_1, E(Y|X = x_1)]$  und  $[x_2, E(Y|X = x_2)]$  in ein 2-dimensionales Koordinatensystem ein. Dabei sieht man, daß zwei Punkte **immer** auf einer Geraden liegen, die sich durch den Achsenabschnitt  $c_0$  und die Steigung  $a_1$  angeben läßt. Nimmt  $X$  mehr als zwei verschiedene Werte an, muß die regressive Abhängigkeit möglicherweise auch durch eine andere Funktion dargestellt **werden** (z.B. als Polynom).

**Interpretation der Regressionskoeffizienten:** In Gleichung 2.7 lassen sich  $a_0$  als Achsenabschnitt und  $a_1$  als Steigung einer Geraden in einem 2-dimensionalen Koordinatensystem interpretieren. Demnach gibt  $CC_1$  die Differenz  $E(Y|X = x_1) - E(Y|X = x_2)$  der bedingten Erwartungswerte an, wenn die Differenz  $x_1 - x_2$  zweier Werte von  $X$  gleich eins ist.

In vielen Büchern findet man eine falsche Interpretation des Regressionskoeffizienten  $CC_1$ , derzufolge dieser Koeffizient angäbe, wie sich der bedingte Erwartungswert von  $Y$  verändert, wenn man  $X$  um eine Einheit verändert. Bei einer solchen Interpretation wird jedoch übersehen, daß mit einer **Veränderung** einer  $X$ -Variablen von einem ganz anderen Zufallsexperiment gesprochen wird, als wenn man  $X$  und  $Y$  lediglich **beobachtet**. Bei einer Intervention zur Veränderung von  $X$  können bspw. auch indirekte Effekte auf  $Y$  auftreten, die den direkten Effekt einer Veränderung von  $X$  auf  $Y$  überlagern. In den Sozialwissenschaften müssen wir immer damit rechnen, daß wir es mit einem ganzen System vieler interdependenter Variablen zu tun haben, bei dem wir zwar die Kovariation zweier Variablen beobachten und mit einer Regression beschreiben können; aber aus diesen Beobachtungsergebnissen können wir nicht **logisch** - allenfalls heuristisch - schließen, was passiert, wenn wir in das System eingreifen, da damit ein neues System geschaffen wird. In Termini der Wahrscheinlichkeitstheorie formuliert, liegt mit einer Intervention ein anderes Zufallsexperiment vor, das durch einen anderen  $W$ -Raum repräsentiert wird.

### 2.2.2 Einfache lineare quasi-regressive Abhängigkeit

Ist die Regression **nicht** linear, kann man dennoch betrachten, wie gut die Abhängigkeit einer numerischen Zufallsvariablen  $Y$  von einer zweiten numerischen Zufallsvariablen  $X$  durch eine **lineare** Funktion beschrieben werden



kann, wobei diese Funktion dann aber nicht als Regression im oben behandelten Sinn interpretiert werden kann. Eine solche lineare Funktion, die die nach dem Kleinst-Quadrat-Kriterium optimale lineare Beziehung zwischen  $X$  und  $Y$  angibt, bezeichne ich als **lineare Quasi-Regression**. (Müller, 1975, spricht stattdessen von der „Regression 2. Art“.) Diese darf nicht mit dem oben behandelten Begriff der einfachen linearen Regression verwechselt werden. Unter der linearen Quasi-Regression ist diejenige lineare Funktion  $f(X) = \alpha + \beta \cdot X$  von  $X$  zu verstehen, die die folgende Funktion von  $a$  und  $\mathbf{b}$  - das sogenannte Kleinst-Quadrat-Kriterium - minimiert:

$$LS(a, \mathbf{b}) := E [(Y - (a + \mathbf{b} \cdot X))^2]. \quad (2.8)$$

Weist  $LS(a, \mathbf{b})$  für die Zahlen  $\alpha$  und  $\beta$  ihr Minimum auf, bezeichnet man die Funktion  $f(X) = \alpha + \beta \cdot X$  auch mit  $\hat{Y}$ . **Die so definierte lineare Quasi-Regression  $\hat{Y}$  ist nur dann mit der (echten) Regression  $E(Y|X)$  identisch, wenn  $E(Y|X)$  tatsächlich eine lineare Funktion von  $X$  ist.** Läßt sich  $E(Y|X)$  dagegen z.B. nur durch eine Parabel beschreiben, dann unterscheiden sich die Regression und die lineare Quasi-Regression voneinander. Im allgemeinen gibt also die lineare Quasi-Regression  $\hat{Y}$  **nicht** an, wie die bedingten Erwartungswerte  $E(Y|X = x)$  von den Werten  $x$  von  $X$  abhängen. Stattdessen gibt  $\hat{Y}$  **nur einen Wert an, der dann im Sinn des Kleinst-Quadrat-Kriteriums (s. Gl.2.8)** eine optimale Vorhersage für  $Y$  ist, falls man sich für die Vorhersage auf die Klasse der **linearen** Funktionen von  $X$  beschränkt. Der einzige Grund, lineare Quasi-Regressionen und nicht die echte Regression zu verwenden, könnte in der größeren Einfachheit liegen. Dies wird aber mit einem Verzicht an Information bezahlt. Natürlich kann man auch die Begriffe einer quadratischen, kubischen Quasi-Regression etc. einführen. Dabei ist lediglich  $a + \mathbf{b} \cdot X$  durch die entsprechende quadratische, kubische Funktion etc. von  $X$  zu ersetzen.

### 2.2.3 Korrelative Abhängigkeit

Die wohl am häufigsten verwendete Kenngröße zur Beschreibung einer Abhängigkeit zwischen zwei numerischen Zufallsvariablen ist der Korrelationskoeffizient

$$Kor(X, Y) := Cov(X, Y) / [Std(X) \cdot Std(Y)], \quad (2.9)$$

der als Kovarianz geteilt durch das Produkt der beiden Standardabweichungen definiert ist, falls die beiden Standardabweichungen größer 0 sind; andernfalls ist  $Kor(X, Y) := 0$ . (Zur Definition der Kovarianz siehe z.B. Bauer, 1978, oder Steyer & Eid, 1993.) Mit dieser Kenngröße wird die Stärke der durch Gleichung 2.7 beschriebenen **linearen** regressiven Abhängigkeit des Regressanden  $Y$  vom Regressor  $X$  angegeben. Liegt eine **lineare** regressive Abhängigkeit vor,

dann ist der quadrierte Korrelationskoeffizient mit dem in Gleichung 2.5 definierten Determinationskoeffizienten identisch. Ist die Regression dagegen **nicht** linear, dann gibt der Korrelationskoeffizient an, wie stark der durch die **lineare Quasi-Regression** beschreibbare Zusammenhang zwischen X und Y ist.

### 2.2.4 Partielle lineare regressive Abhängigkeit

Betrachtet man **zwei** numerische Regressoren X und Z, ergeben sich wiederum verschiedene Abhängigkeitsarten des Regressanden Y von den beiden Regressoren X und Z. Die einfachste Art ist dabei beschreibbar durch die multiple Regressionsgleichung

$$E(Y|X, Z) = \beta_0 + \beta_1 X + \beta_2 Z, \quad \beta_0, \beta_1, \beta_2 \in \mathbb{R}. \quad (2.10)$$

Dabei kann es durchaus sein, daß der **partielle Regressionskoeffizient**  $\beta_1$  nicht mit dem entsprechenden einfachen Regressionskoeffizienten  $\mathbf{a}_1$  aus der einfachen linearen Regression (s. Gl. 2.7) identisch ist. In diesem Fall heißen X und **Z konfundiert**. Setzt man die Gültigkeit der Gleichungen 2.7 und 2.10 voraus, dann tritt der Fall  $\alpha_1 = \beta_1$  dann und nur dann ein, wenn (a) Z von X regressiv unabhängig ist (s. Gl.2.6) oder (b) wenn  $\beta_2$  gleich 0 ist. Im allgemeinen kann es durchaus sein, daß  $\mathbf{a}_1$  positiv und  $\beta_1$  negativ sind und umgekehrt, oder, daß einer der beiden Parameter gleich null ist und der andere stark von null verschieden. (Ein Beispiel findet man in Steyer, 1992, Kap.3). Man beachte, daß hier von den wahren Parametern die Rede ist und nicht von ihren Schätzungen anhand einer Stichprobe.

Spätestens hier wird klar, daß es verschiedene Arten regressiver Abhängigkeit gibt, die in Anwendungen auch inhaltlich völlig verschiedenes besagen und die auf den ersten Blick sogar widersprüchlich erscheinen können, bspw. dann, wenn  $\mathbf{a}_1 < 0$  und  $\beta_1 > 0$ . Es gibt also nicht **die** Abhängigkeit einer Variablen Y von einer Variablen X, nicht einmal **die regressive** Abhängigkeit der Variablen Y von X und selbst nicht **die partielle regressive** Abhängigkeit der Variablen Y von X, da auch die partiellen regressiven Abhängigkeiten der Variablen Y von X, die in multiplen Regressionen mit zwei, drei, vier Regressoren beschrieben werden, völlig verschieden aussehen können.

### 2.2.5 Bedingte regressive Abhängigkeit

Zum Verständnis der Bedeutung der regressiven Abhängigkeit eines Regressanden Y bei der Berücksichtigung von zwei (oder mehr) Regressoren X und Z ist die Betrachtung der **bedingten Regression**  $E_Z$  von Nutzen, d.h.

der Regression von  $Y$  auf  $X$  bzgl. des bedingten Wahrscheinlichkeitsmaßes  $P_{Z=z}$ . Bezeichnet  $Z$  bspw. die Geschlechtsvariable mit den beiden Werten 1 (für Frauen) und 0 (für Männer), dann ist  $E_{Z=0}(Y|X)$  die Regression von  $Y$  auf  $X$  in der Subpopulation der Männer.

Die wichtigste Eigenschaft der durch die Gleichung 2.10 beschriebenen Art regressiver Abhängigkeit ist, daß die bedingte Regression  $E_{Z=z}(Y|X)$  bei festem Wert  $z$  von  $Z$  wiederum eine **lineare** Funktion von  $X$  ist, und zwar mit dem **gleichen Steigungskoeffizienten**  $\beta_1$  für alle möglichen Werte  $z$  von  $Z$ . Die Achsenabschnitte dieser bedingten Regressionen für verschiedene Werte  $z$  von  $Z$  unterscheiden sich dagegen, falls  $\beta_2$  ungleich null ist:

$$E_{Z=z}(Y|X) = (\beta_0 + \beta_2 z) + \beta_1 X, \quad \beta_0, \beta_1, \beta_2 \in \mathbb{R}. \quad (2.11)$$

Die durch  $\beta_1$  beschriebene Abhängigkeit des Regressanden  $Y$  von  $X$  ist damit über alle Werte von  $Z$  generalisierbar. Dies ist keineswegs selbstverständlich, wie in den nachfolgenden Abschnitten deutlich wird. Darüber hinaus ist damit gezeigt, daß man den **partiellen Regressionskoeffizienten**  $\beta_1$  als **einfachen Regressionskoeffizienten in der bedingten Regression**  $E_{Z=z}(Y|X)$  verstehen kann. (Durch Vertauschung von  $X$  und  $Z$  erhält man die analoge Aussage für  $\beta_2$ .)

Eine weitere Art regressiver Abhängigkeit resultiert, wenn man zwar bei jedem gegebenen Wert  $z$  von  $Z$  eine lineare Regression von  $Y$  auf  $X$  hat, dabei jedoch der Steigungskoeffizient vom jeweiligen Wert von  $Z$  abhängt. Etwas präziser läßt sich diese Art der Abhängigkeit durch die Gleichung

$$E(Y|X, Z) = g_0(Z) + g_1(Z) \cdot X \quad (2.12)$$

charakterisieren, wobei  $g_0(Z)$  und  $g_1(Z)$  beliebige reellwertige Funktionen von  $Z$  bezeichnen. Für feste Werte  $z$  von  $Z$  resultiert hier die **bedingte lineare Regression**

$$E_{Z=z}(Y|X) = g_0(z) + g_1(z) \cdot X, \quad (2.13)$$

deren Steigungskoeffizienten vom Wert  $z$  von  $Z$  abhängen und durch die Funktion  $g_1(Z)$  berechnet werden können, falls diese Funktion bekannt ist.

Ist nun  $g_1(Z)$  keine konstante Funktion, die für alle Werte  $z$  von  $Z$  den gleichen Wert  $\beta_1$  annimmt, sagt man auch, daß  $Z$  den Effekt von  $X$  **moderiert** (besser wäre eigentlich: **modifiziert**). Man spricht dann daher auch von einem **Moderatormodell** und nennt  $Z$  einen **Moderator**. Anwendungen findet man z.B. bei Bredenkamp (1984a, b) oder Erdfelder und Steyer (1984).

Ein recht häufig vorkommendes spezielles Moderatormodell liegt vor, wenn es sich bei den Funktionen  $g_0$  und  $g_1$  um **lineare** Funktionen von  $Z$  handelt:

$$\begin{aligned} E(Y|X, Z) &= (\beta_0 + \beta_2 Z) + (\beta_1 + \beta_3 \cdot Z) X \\ &= \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 X \cdot Z. \end{aligned} \quad (2.14)$$

Das Moderatormodell wird gelegentlich mit dem oben bereits erwähnten Phänomen der Konfundierung verwechselt (s. z.B. Wermuth, 1989), was von Schmitt (1990) sowie von Schmitt und Götz-Baltes (1992) kritisiert wurde. Die wesentliche Eigenschaft eines **Moderatormodells** ist, daß sich die bedingten Regressionskoeffizienten  $g_1(z)$  für verschiedene Werte  $z$  von  $Z$  voneinander unterscheiden. Bei der **Konfundierung** dagegen wird vorausgesetzt, daß die bedingten Regressionskoeffizienten für alle Werte  $z$  von  $Z$  gleich  $\beta_1$  sind. Dieser Koeffizient  $\beta_1$  ist nur nicht mit dem einfachen Regressionskoeffizienten  $\alpha_1$  identisch.

Bei allen bisherigen Betrachtungen waren die bedingten Regressionen **lineare** Funktionen von  $X$ . Dieser Fall wird am allgemeinsten durch die Gleichung 2.12 beschrieben, die sich leicht verallgemeinern läßt, z.B. zu

$$E(Y|X, Z) = g_0(Z) + g_1(Z) \cdot X + g_2(Z) \cdot X^2, \quad (2.15)$$

falls  $X$  mindestens drei verschiedene Werte annehmen kann.

Das **allgemeine Prinzip der bedingten Regression** ist die Beschreibung der regressiven Abhängigkeit des Regressanden  $Y$  vom Regressor  $X$  bei konstanten Ausprägungen einer oder mehrerer anderer Variablen  $Z$ . Dabei beachte man, daß  $Z$  in den Gleichungen 2.12 und 2.15 durchaus auch für eine vektorielle Variable stehen kann. Ist z.B.  $Z = (Z_1, Z_2)$  ein 2-dimensionaler Vektor numerischer Zufallsvariablen und handelt es sich bei  $g_0(Z)$  und  $g_1(Z)$  jeweils um lineare Funktionen von  $Z_1$  und  $Z_2$ , dann ist ein Spezialfall von Gleichung 2.12

$$E(Y|X, Z) = (\beta_0 + \beta_1 Z_1 + \beta_2 Z_2) + (\gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2) \cdot X \quad (2.16)$$

Auch wenn diese Formeln recht kompliziert erscheinen, sollte man dennoch bedenken, daß damit eine große Vereinfachung erzielt werden kann, wenn man beschreiben will, wie die regressive Abhängigkeit des Regressanden  $Y$  vom Regressor  $X$  von den Ausprägungskombinationen der Variablen  $Z_1$  und  $Z_2$  modifiziert wird.

## 2.2.6 Zusammenfassende Bemerkungen

Bei Regressions- und Korrelationsmodellen äußert sich das Problem der multiplen Determiniertheit darin, daß im allgemeinen **mit jedem** neu **hinzugenommenen Regressor eine andere regressive (bzw. korrelative) Abhängigkeit** betrachtet wird. Auf der Modellebene schlägt sich dies darin nieder, daß sich der einfache bzw. partielle Regressionskoeffizient und damit die jeweils betrachtete einfache bzw. partielle regressive Abhängigkeit mit jedem neu hinzugenommenen Regressor ändern kann, und zwar **nicht** nur graduell und **nicht** nur wegen stichprobenbedingter Schätzfehlerprobleme. So können aus **positi-**

ven Abhängigkeiten **negative** oder auch Unabhängigkeiten werden und umgekehrt (s. auch die Erläuterungen zu G1.2.10).

Demnach gibt es nicht **den Effekt** eines Regressors X auf einen Regressanden Y, sondern prinzipiell mindestens sovieler Effekte von X auf Y wie es weitere Variablen gibt, die man zu dem gerade betrachteten Regressor X als zusätzliche Regressoren hinzuziehen kann. Man siehe hierzu auch die Literatur zu „Suppressorvariablen“ und „Moderatorvariablen“ (z. B. Bartussek, 1970; Cheek, 1982; Darlington, 1968; Ghiselli, 1963; Kenny & Judd, 1984; Moosbrugger, 1981; Penner & Wymer, 1983; Saunders, 1956; Schmitt, 1990).

### 2.3 Kausale regressive Abhängigkeiten

Nach den obigen Ausführungen ist leicht einzusehen, daß Regressionsmodelle (und folglich auch Spezialfälle wie Varianz- und faktoranalytische Modelle) nicht zur Theorienformulierung oder -prüfung beitragen können, wenn die mit ihnen definierten Abhängigkeitsbegriffe so vielfältig und zunächst auch beliebig sind. Regressive Abhängigkeiten sind damit **für sich genommen** zur Theorienformulierung ungeeignet. Tatsächlich kommt in psychologischen Hypothesen oft eine **kausale** Terminologie wie „Effekt“, „Wirkung“, „führt zu“ u.ä. vor. Oft ist die Hypothese über einen kausalen Effekt auch mit Hilfe der **Ceteris-paribus-Klausel** formuliert, derzufolge bestimmte Effekte nur bei Konstanzhaltung aller potentiellen Störvariablen postuliert werden. Hinter dieser Klausel verbirgt sich der auf John Stuart Mill (1862) zurückgehende Kausalitätsbegriff, daß, wenn alle anderen Variablen konstant sind und nur X variiert, die Kovariation zwischen X und der später auftretenden Variablen Y auf X als Ursache zurückzuführen ist.

Wie wir sehen werden, ist damit ein sehr strenger Kausalitätsbegriff involviert. Mit der Ceteris-paribus-Klausel bezieht man sich außerdem auf eine in der Regel nicht herstellbare Realität. In den meisten Fällen ist die Konstanzhaltung aller potentiellen Störvariablen schlechterdings nicht möglich. Muß man außerdem mit Interaktionen rechnen, stellt sich die Frage, auf welchen Werten man die jeweiligen Störvariablen konstanthalten soll, da Interaktion ja bedeutet, daß die Effekte nicht auf jeder Stufe der konstantgehaltenen Variablen gleich sind. Diesen Problemen kann man entgehen, wenn man die durch die Ceteris-paribus-Klausel formulierte Aussage über eine kausale Abhängigkeit **indirekt** überprüft. Erdfelder und Bredenkamp (in diesem Band, Kap. 14, Abschnitt 2.4) zeigen nämlich, daß aus einer derartigen Aussage unter bestimmten Voraussetzungen noch Konsequenzen ableitbar sind, die sich in einem randomisierten Experiment überprüfen lassen.

Eine Alternative dazu ist, die *Ceteris-paribus*-Klausel auch schon in der Theorie oder Hypothese durch die ***Ceteris-distributionibus-paribus-Klausel*** zu ersetzen, derzufolge nur noch die ***Verteilungen*** aller potentiellen Störvariablen konstantgehalten sind. Dies erreicht man durch die Kontrolltechnik der Randomisierung. Durch die dabei vorgenommene zufällige Aufteilung der Vpn auf die Versuchsbedingungen stellt man die stochastische Unabhängigkeit der Treatmentvariablen mit allen potentiellen Störvariablen her, deren Werte bereits vor der experimentellen Behandlung festliegen. (Dabei beachte man, daß hier **nicht** von Abhängigkeiten in einer Stichprobe, sondern von den wahren Abhängigkeiten die Rede ist.) Die durch die Randomisierung hergestellte stochastische Unabhängigkeit impliziert aber, daß die Verteilungen der potentiellen Störvariablen in allen experimentellen Bedingungen gleich sind. Dadurch werden die Effekte der anderen potentiellen Störvariablen gleichmäßig auf die Versuchsbedingungen verteilt. Durch die Randomisierung werden also die Störvariablen nicht konstantgehalten, auch werden deren Effekte nicht eliminiert. Stattdessen werden die ***Verteilungen*** der potentiellen Störvariablen innerhalb der Treatment-Bedingungen gleichgehalten. Der Nachteil ist allerdings, daß man im randomisierten Experiment nur **durchschnittliche** Effekte - gemittelt über die nichtbeachteten potentiellen Störvariablen - analysiert. Diese Gedanken sollen in den folgenden Abschnitten näher erläutert werden.

### 2.3.1 Vorbereitende Definitionen

Um die obigen Vorstellungen zu präzisieren, ist es zunächst angebracht, den **allgemeinen** Begriff der **Konfundierung** zu definieren, wie er im Kontext von Regressionsmodellen verwendet werden kann. (Im Abschnitt 2.2 hatten wir uns auf den Fall einer **linearen** Regression beschränkt.) Dabei spielt zunächst neben  $X$  nur ein einziger weiterer Regressor  $W$  eine Rolle.

Unkonfundiertheit: Seien  $Y$  eine numerische sowie  $X$  und  $W$  beliebige Zufallsvariablen auf einem gemeinsamen  $W$ -Raum. Dann heißt  $X$  mit  $W$  bzgl. der Regression  **$E(Y|X)$  unkonfundiert** genau dann, wenn gilt:

(u<sub>1</sub>)  $X$  und  $W$  sind stochastisch unabhängig,  
oder

(u<sub>2</sub>) es existiert eine numerische Funktion  $f$  derart, daß für  $f(W)$  gilt:  
$$E(Y|X, W) = E(Y|X) + f(W).$$

$X$  heißt mit  $W$  bzgl.  $E(Y|X)$  **konfundiert**, wenn weder (u<sub>1</sub>) noch (u<sub>2</sub>) gilt.

Im Fall (u<sub>1</sub>) sind die bedingten Wahrscheinlichkeiten  $P(W=w|X=x)$  mit den unbedingten Wahrscheinlichkeiten  $P(W=w)$  identisch und im Fall (u<sub>2</sub>) stimmt die einfache Regression  **$E(Y|X)$**  bis auf eine additive Konstante mit den bedingten Regressionen  **$E_{W=w}(Y|X) = E(Y|X) + f(w)$**  überein.

Bei gewöhnlichen Regressionsmodellen ist die Unkonfundiertheit im allgemeinen **nicht** gewährleistet, es sei denn, die Bedingung ( $u_1$ ) oder die Bedingung ( $u_2$ ) gilt aus **empirischen** Gründen. Im Rahmen eines randomisierten Experiments dagegen ist die Unkonfundiertheit **immer** erfüllt, falls es sich bei  $W$  um eine Variable handelt, deren Werte bereits **vor** der Applikation des Treatments festliegen, da durch die Randomisierung die Unabhängigkeit von  $X$  und  $W$  **bergestellt** wird. Dabei beachte man wieder, daß hier nicht von der Unabhängigkeit in einer Stichprobe die Rede ist.

**Konsequenzen für die Interpretation:** Die Unkonfundiertheit hat wichtige Konsequenzen für die Interpretation der durch die Regression  $E(Y|X)$  beschriebenen Abhängigkeit. Es läßt sich nämlich zeigen, daß die bedingten Erwartungswerte  $E(Y|X=x)$  dann als **Mittelung** der bedingten Erwartungswerte  $E(Y|X=x, W=w)$  - gemittelt über die Verteilung von  $W$  - interpretiert werden können, d.h. im Falle der Unkonfundiertheit von  $X$  und  $W$  gilt für jeden Wert  $E(Y|X=x)$  der Regression  $E(Y|X)$ :

$$E(Y|X=x) = \sum_w E(Y|X=x, W=w) \cdot P(W=w), \quad (2.17)$$

wobei die Summierung über alle Werte  $w$  von  $W$  vorzunehmen ist. Daß die Unkonfundiertheit die Gültigkeit der Gleichung 2.17 impliziert, erkennt man folgendermaßen: Da die Gleichung

$$E(Y|X=x) = \sum_w E(Y|X=x, W=w) \cdot P(W=w|X=x) \quad (2.18)$$

allgemeingültig ist, sieht man sofort, daß im Fall der stochastischen Unabhängigkeit von  $X$  und  $W$  die Gleichung 2.17 erfüllt ist, denn dann gilt  $P(W=w|X=x) = P(W=w)$  [s. Bedingung ( $u_1$ )]. Gleichung 2.17 ist aber auch dann erfüllt, wenn für alle Werte von  $X$  und  $W$  gilt:  $E(Y|X=x, W=w) = E(Y|X=x) + f(w)$  [s. Bedingung ( $u_2$ )]. Dies aber sind die beiden Bedingungen, von denen zumindest eine erfüllt ist, wenn die Unkonfundiertheit gilt.

Ist die Unkonfundiertheit erfüllt, resultieren die bedingten Erwartungswerte  $E(Y|X=x)$  demzufolge aus der Mittelung der bedingten Erwartungswerte  $E(Y|X=x, W=w)$  über die Verteilung von  $W$ . Dies ist weitaus mehr, als bei einem gewöhnlichen Regressionsmodell gilt. Dort kann es nämlich z.B. durchaus sein, daß für **jeden** festen Wert einer potentiellen Störvariablen  $W$  eine **negative** lineare regressive Abhängigkeit besteht, aber dennoch eine positive lineare regressive Abhängigkeit beobachtet wird, wenn man nur die Regression  $E(Y|X)$  betrachtet (s. dazu ein ausführliches Beispiel bei Steyer, 1992, Kap.3). Dies ist genau der bereits früher behandelte Fall  $\alpha_1 > 0$  und  $\beta_1 < 0$  (s. die Gleichungen 2.7 und 2.10).

Warum ist es so wichtig, daß man diese Mittelungsinterpretation vornehmen kann? Angenommen, die psychologische Hypothese postuliert bei konstantgehaltenen Störvariablen - repräsentiert durch die mehrdimensionale Zufallsvariable  $\mathbf{W} = (\mathbf{W}_1, \dots, \mathbf{W}_m)$  - einen **positiven regressiv linearen Effekt** von X auf Y, dann läßt sich diese Hypothese wie folgt formulieren (s. Erdfelder & Bredenkamp, Kap. 14, Abschnitt 2.4, in diesem Band):

$$E(Y|X, \mathbf{W}) = g_0(\mathbf{W}) + g_1(\mathbf{W}) \cdot X,$$

wobei  $g_1(\mathbf{w}) > 0$  für alle Werte  $\mathbf{w}$  von  $\mathbf{W}$ . (2.19)

Kann man nun  $\mathbf{W}$  nicht beobachten - und man wird nur selten alle potentiellen Störvariablen beobachten können -, dann läßt sich diese Hypothese dennoch **indirekt** überprüfen, indem man durch Randomisierung dafür sorgt, daß X und  $\mathbf{W}$  stochastisch unabhängig sind. Dann gilt nämlich auch, daß die Funktion  $g_1(\mathbf{W})$  von X regressiv unabhängig ist, d.h.

$$E[g_1(\mathbf{W})|X] = E[g_1(\mathbf{W})] = \alpha_1, \tag{2.20}$$

wobei  $\alpha_1$  der Steigungskoeffizient der linearen Regression  $E(Y|X) = \alpha_0 + \alpha_1 \cdot X$  ist. Der Erwartungswert einer Variablen, die nur positive Werte annehmen kann [hier: die Werte der Funktion  $g_1(\mathbf{W})$ ], ist aber notwendigerweise ebenfalls positiv. Aus der obigen psychologischen Hypothese kann man unter der Voraussetzung eines randomisierten Experiments (Steyer, 1992, zeigt, daß dafür auch die schwächere Annahme der Gleichung 2.17 reicht) also ableiten, daß auch der Koeffizient  $\alpha_1$  der einfachen linearen Regression  $E(Y|X) = \alpha_0 + \alpha_1 \cdot X$  positiv ist. Entscheidet man aufgrund eines randomisierten Experiments dann, daß  $\alpha_1$  **nicht** positiv ist, kann damit die psychologische Hypothese als falsifiziert gelten (s. Erdfelder & Bredenkamp, Kap. 14, Abschnitt 2.4 in diesem Band). Die Gültigkeit der Gleichung 2.17 ist demnach nicht nur aus ästhetischen Gründen wünschenswert, sondern auch **notwendige** Voraussetzung, wenn man bestimmte psychologische Hypothesen im Sinne der Falsifikation empirisch überprüfen will.

**Additive Dekomponierbarkeit:** Wem die oben erläuterte „Mittelungsinterpretation“ zu wenig erscheint, der kann zusätzlich noch die Bedingung der „Additiven Dekomponierbarkeit“ fordern, die eine Interaktion zwischen X und einer potentiellen Störvariablen  $\mathbf{W}$  verbietet. Diese Bedingung ist z. B. erfüllt, wenn Gleichung 2.10 gilt. In diesem Fall setzt sich die Regression also **additiv** aus den Effekten der beiden Regressoren zusammen. Der interpretative Vorteil wurde ausführlich im Abschnitt 2.2.5 behandelt.

Eine allgemeine Definition der Additiven Dekomponierbarkeit, die nicht auf den Fall der multiplen **linearen** Regression beschränkt ist, kann man wie folgt formulieren: Seien Y eine numerische und X und W beliebige Zufallsvariablen



auf einem gemeinsamen  $W$ -Raum. Dann heißt die Regression  $E(Y|X, W)$  **additiv dekomponierbar** genau dann, wenn gilt:

- (a) es existieren zwei numerische Funktionen  $g(X)$  und  $h(W)$  von  $X$  bzw. von  $W$ , für die gilt:

$$E(Y|X, W) = g(X) + h(W). \quad (2.21)$$

**Konsequenzen für die Interpretation:** Man kann nun zeigen, daß, wenn  $X$  und  $W$  bzgl.  $E(Y|X)$  unkonfundiert sind und die Additive Dekomponierbarkeit erfüllt ist, sich die Regression  $E(Y|X, W)$  durch die Gleichung

$$E(Y|X, W) = E(Y|X) + f(W) \quad (2.22)$$

darstellen läßt, wobei  $f(W)$  wiederum eine Funktion von  $W$  ist. Gilt aber die Gleichung 2.22, so hat dies wichtige Konsequenzen für die Interpretation der durch die Regression  $E(Y|X)$  beschriebenen Abhängigkeit. Für jeden festen Wert  $w$  von  $W$  gilt dann nämlich:

$$E_{W=w}(Y|X) = E(Y|X) + f(w), \quad (2.23)$$

d.h. die durch  $E(Y|X)$  beschriebene Abhängigkeit gilt für jeden Wert  $w$  von  $W$ , wenn man von der additiven Konstanten  $f(w)$  absieht. Das Ignorieren von  $W$  ist in diesem Fall unbedenklich.

Repräsentiert  $W$  bspw. die Zugehörigkeit zu einer Subpopulation (z. B. Männer oder Frauen), so ist die regressive Abhängigkeit des Regressanden  $Y$  vom Regressor  $X$  in der Regel in jeder dieser Subpopulationen gleich und auch gleich wie in der Gesamtpopulation, abgesehen von der additiven Konstanten  $f(w)$ . Anschaulich bedeutet das, daß die  $(W=w)$ -bedingten Regressionen  $E_{W=w}(Y|X)$  parallel verlaufen müssen, und auch parallel zur (unbedingten) Regression  $E(Y|X)$ .

Die Additive Dekomponierbarkeit läßt nicht zu, daß zwischen  $X$  und der Störvariablen  $W$  eine Interaktion im varianzanalytischen Sinn besteht. Dies ist eine Bedingung, die nicht einmal im randomisierten Experiment hergestellt werden kann, denn trotz zufälliger Zuweisung der Vpn auf die Versuchsbedingungen ist es möglich, daß die experimentelle Behandlung (repräsentiert durch die Zufallsvariable  $X$ ) auf  $Y$  bei Männern anders wirkt als bei Frauen ( $W$  repräsentiere die Geschlechtsvariable). Lediglich die Unkonfundiertheit mit den oben beschriebenen interpretativen Konsequenzen (die „Mittelungsinterpretation“) läßt sich durch die Randomisierung herstellen.

### 2.3.2 Schwache kausale regressive Abhängigkeit

Im randomisierten Experiment gilt die Unkonfundiertheit nicht nur für eine einzige, sondern für alle potentiellen Störvariablen, jedenfalls dann, wenn man die Menge der potentiellen Störvariablen mit der Menge derjenigen Variablen gleichsetzt, deren Werte bereits vor der Applikation des Treatments festliegen. Da im randomisierten Experiment die Unkonfundiertheit erfüllt ist, gilt in diesem Fall auch die **Schwache Kausalität**, die in diesem Kontext nichts anderes meint, als daß X mit **jeder** potentiellen Störvariablen bzgl. Y unkonfundiert ist, und der Regressor X dem Regressanden **Y vorgeordnet** ist.

Anstelle der Unkonfundiertheit kann man zur Definition der Schwachen **Kausalität** auch die etwas schwächere Bedingung verwenden, daß Gleichung 2.17 für **alle** potentiellen Störvariablen gilt. Beschränkt man sich der Einfachheit halber auf den Fall diskreter potentieller Störvariablen (der allgemeine Fall wird bei Steyer, 1992, behandelt), so kann man die **schwache kausale regressive Abhängigkeit** also wie folgt definieren: Seien Y eine numerische und X eine beliebige Zufallsvariable auf einem gemeinsamen W-Raum. Dann heißt Y von **X schwach kausal regressiv abhängig** genau dann, wenn gelten:

(w<sub>1</sub>) Für alle potentiellen Störvariablen gilt Gleichung 2.17.

(w<sub>2</sub>) X ist Y **vorgeordnet**.

Die Bedingung (w<sub>2</sub>) ist im randomisierten Experiment natürlich ebenfalls automatisch erfüllt, da man den Regressanden Y erst nach der Setzung der durch X repräsentierten experimentellen Bedingungen beobachtet (zur Formalisierung des Begriffs der Vorgeordnetheit s. Steyer, 1992, Kap. 6). Man beachte, daß Unabhängigkeit hier als Spezialfall der Abhängigkeit betrachtet wird.

### 2.3.3 Starke kausale regressive Abhängigkeit

Die interpretativen Probleme, die durch mögliche Interaktionen entstehen können, wurden oben ausführlich diskutiert. Will man nun Abhängigkeiten betrachten, bei denen angenommen wird, daß keine Interaktionen auftreten, muß man neben der Unkonfundiertheit auch die Additive Dekomponierbarkeit postulieren und zwar für **alle** potentiellen Störvariablen. Dies führt aber zur Bedingung der **Starken Kausalität**, die besagt, daß außer der Vorgeordnetheit auch für **jede** potentielle Störvariable die Additive Dekomponierbarkeit gilt sowie, daß X und **W** bzgl.  $E(Y|X)$  unkonfundiert sind.

Eine etwas schwächere Definition der Starken Kausalität läßt sich wie folgt formulieren: Seien Y eine numerische und X eine beliebige Zufallsvariable auf

einem gemeinsamen W-Raum. Dann heißt  $Y$  von  $X$  **stark** kausal **regressiv abhängig** genau dann, wenn gelten:

( $s_1$ ) Für alle potentiellen Störvariablen existiert eine numerische Funktion  $f$  derart, daß für  $f(W)$  die Gleichung 2.22 gilt.

( $s_2$ )  $X$  ist  $Y$  **vorgeordnet**.

Wie oben bereits bemerkt, läßt sich die Starke Kausalität nicht mehr durch Techniken der Versuchsplanung herstellen. Man bedenke aber, daß diese Bedingung immer noch schwächer ist, als die weithin verbreitete Vorstellung, alle wichtigen Variablen in die Regression einzubeziehen.

### 2.3.4 Zusammenfassende Bemerkungen

Wegen dem Problem der multiplen Determination und dem Meßfehlerproblem werden in den Sozialwissenschaften häufig Regressionsmodelle verwendet. Damit ist man jedoch mit dem Grundproblem von Regressionsmodellen konfrontiert: Mit jedem zusätzlichen Regressor können eine Vielzahl neuer Abhängigkeiten eines gegebenen Regressanden  $Y$  von einem gegebenen Regressor  $X$  betrachtet werden. Regressionsmodelle und stochastische Modelle im allgemeinen sind daher für die Entwicklung, Formulierung und Prüfung von Theorien nur dann von Interesse, wenn sie **zu kausalen** Regressionsmodellen vervollständigt werden.

Ein kausales unterscheidet sich von einem nichtkausalen Regressionsmodell im wesentlichen durch zweierlei:

- (a) Durch eine prozeßhafte Betrachtung des zugrundeliegenden Zufallsexperiments, wodurch eine Vorgeordnetheit zwischen den Variablen und Ereignissen angenommen werden kann.
- (b) Eine Kausalitätsbedingung wie z.B. die Schwache Kausalität, in der eine bestimmte Beziehung zwischen der betrachteten Regression und den potentiellen Störvariablen postuliert wird.

Im randomisierten Experiment ist die Schwache Kausalität **immer** erfüllt. In nichtrandomisierten Studien dagegen sind sowohl die Schwache als auch die Starke Kausalität falsifizierbar, wenn man statistische Entscheidungen zwischenschaltet. In randomisierten Experimenten dagegen kann nur die Starke Kausalität falsch sein, da diese Interaktionen zwischen der betrachteten Treatmentvariablen  $X$  und potentiellen Störvariablen ausschließt. Bei **schwachen** kausalen regressiven Abhängigkeiten dagegen, die in jedem randomisierten Experiment vorliegen, sind solche Interaktionen zugelassen.

Zu einer vollständigeren und mathematisch exakteren Darstellung der Theorie kausaler Regressionsmodelle sei auf Steyer (1992) verwiesen. Dort findet man

auch kommentierte Hinweise auf die umfangreiche Literatur zu anderen (auch anderen **stochastischen**) Theorien kausaler Abhängigkeit. Am interessantesten sind dabei m. E. die Weiterentwicklungen der Theorie von Suppes (1970, 1981) durch Stegmüller (1983) und vor allem durch Spohn (1980, 1983, 1990, 1991, im Druck). Weiter sei auf die Darstellung **bedingter** kausaler Regressionsmodelle bei Steyer (1992) hingewiesen, die wegen der multivariaten Natur empirischer Phänomene in vielen Fällen nützlicher sein können, als die hier dargestellten einfachen kausalen Regressionsmodelle.

### 3. Stochastische Meßmodelle

Stochastische Meßmodelle sind spätestens seit Thurstone (1931) Kern der psychometrischen Literatur (Fischer, 1974; Formann, 1984; Gulliksen, 1950; Hambleton & Swaminathan, 1985; Kubinger, 1988; Langeheine & Rost, 1988; Lewis, 1986; Lord & Novick, 1968; Rost, 1988; Rost & Strauß, 1992; Steyer, Wender & Widaman, 1993). Allerdings stehen in diesen Arbeiten statistische Probleme des Schätzens von Parametern und Testens von Hypothesen im Vordergrund. Meßtheoretische Fragen, wie die Untersuchung des Skalenniveaus u.ä., wurden bisher weitgehend vernachlässigt (s. jedoch Fischer, 1974, 1988; Steyer, 1989; Steyer & Eid, 1993). Die meßtheoretischen Fragen sind dagegen Kern der Arbeiten von Falmagne (1985), Krantz, Luce, Suppes und Tversky (1971), Luce, Krantz, Suppes und Tversky (1990), Roberts (1979) und Suppes, Krantz, Luce und Tversky (1989).

In diesem Abschnitt soll nun exemplarisch gezeigt werden, wie man in stochastischen Meßmodellen **Konstrukte konstruieren** kann. In einen Meßmodell werden die Verknüpfungen zwischen theoretischen und empirischen Begriffen expliziert. Ohne ein Meßmodell wären aus Aussagen über theoretische Begriffe keine logischen Implikationen für die Empirie möglich, d.h. erst durch ein Meßmodell kann eine Theorie zu einer **empirischen** Theorie werden, in dem Sinn, daß aus ihr Aussagen über die Empirie logisch ableitbar sind. Aus der Sicht einer deduktivistischen Methodologie (s. Erdfelder & Bredenkamp, Kap. 14 in diesem Band) sind also Meßmodelle von zentraler wissenschaftstheoretischer Bedeutung.

Zur Illustrierung eines solchen Meßmodells wähle ich das mathematisch einfachste und bekannteste stochastische Meßmodell, nämlich das **Modell essentiell T-äquivalenter Variablen**. Die analoge Argumentation kann man aber z.B. mit dem Rasch-Modell (s. dazu Steyer, 1989, oder ausführlicher, Steyer & Eid, 1993) und mit jedem anderen stochastischen Meßmodell führen. Die Auswahl des Modells essentiell  **$\tau$ -äquivalenter** Variablen zur exemplarischen Darstellung eines stochastischen Meßmodell ist nicht nur wegen seiner Einfachheit und Verbreitung, sondern auch dadurch begründet, daß seine Struktur als Meßmodell lange Zeit verkannt wurde. So wurden z.B. die Modelle der klassischen Testtheorie von Suppes und Zinnes (1963) als atheoretisches „pointer

measurement“ (p.21) abgetan. Dieser Vorwurf trifft m.E. jedoch nur auf viele unbeachtete Anwendungen, vielleicht auch auf atheoretische Darstellungen (s. z.B. Lienert, **1961**), nicht jedoch auf die Modelle selbst zu, wie in diesem Abschnitt gezeigt werden soll.

### 3.1 Grundbegriffe der Klassischen Theorie psychometrischer Tests

Bevor das eigentliche Meßmodell eingeführt werden kann, müssen zuvor einige Grundbegriffe bereitgestellt werden, die als Antwort auf das **Meßfehlerproblem** entwickelt wurden.

Angenommen wir wollen aufgrund eines Testergebnisses Aussagen über eine bestimmte Persönlichkeitseigenschaft der getesteten Person machen. Würden wir nun den Test mehrmals vorlegen, dann würden wir feststellen, daß die Person bei der ersten Vorgabe andere Antworten auf die Fragen (Items) eines Persönlichkeitstests gibt, als bei der zweiten oder dritten Meßgelegenheit. Auch bei der Verrechnung der Antworten zu einem Gesamtwert würden wir in der Regel feststellen, daß der resultierende Testwert bei jeder Meßgelegenheit ein anderer wäre. Demnach müßten wir dann auch verschiedene Aussagen über die betreffende Persönlichkeitseigenschaft der Person machen, je nachdem, ob wir die erste, zweite oder dritte Meßgelegenheit betrachten.

Zur Erklärung des Sachverhalts, daß die Testwerte der Person bei jeder der drei Meßgelegenheiten anders ausfallen, sind prinzipiell drei Möglichkeiten denkbar:

- (a) Die zu messende Eigenschaft verändert sich zwischen den Messungen.
- (b) Die Unterschiede kommen durch Meßfehler zustande.
- (c) Sowohl Meßfehler als auch Veränderungen der Eigenschaft sind für die Unterschiede verantwortlich.

Erklärung (a) scheint nur sinnvoll, wenn die beobachteten Veränderungen der Testwerte systematisch sind, wenn die betreffende Person z.B. auf **allen** Items des Tests höhere Werte als bei einer früheren Vorgabe erreicht. Veränderungen der Antworten zwischen zwei Meßgelegenheiten fallen jedoch oft unsystematisch und widersprüchlich aus. Bei einigen Items wird die Person gleiche Antworten geben, bei anderen erzielt sie höhere und bei wieder anderen niedrigere Werte. In derartigen Fällen, die sehr oft auftreten, scheint die Erklärung (a) unbefriedigend.

Bei den beiden Erklärungsmöglichkeiten (b) und (c) wird vorausgesetzt, daß sich der beobachtete Testwert aus einem Meßfehler und einem Wert zusammensetzt, der die tatsächliche Eigenschaft der Person repräsentiert. Für jede Person wird also ein Wert gedacht, der **nicht** mit einem Meßfehler behaftet ist, wofür sich die Bezeichnung wahrer **Wert** eingebürgert hat. Prinzipiell läßt sich dies auf verschiedene Weise präzisieren.

**Wahrer Wert und Fehler:** Die in der Psychologie am häufigsten verwendete Präzisierung der Vorstellung eines **wahren Werts** einer Person ist seine Definition als Erwartungswert einer Testwertvariablen  $Y_i$  bzgl. der intraindividuellen (d.h. der Person-bedingten) Verteilung von  $Y_i$  (Gulliksen, 1950; Lord & Novick, 1968). Betrachtet man ein Zufallsexperiment, bei dem eine Person zufällig aus einer Menge  $U$  von Personen (der Population) gezogen wird und

ihre Werte in  $m$  Tests oder Testteilen festgestellt werden, so kann man neben den  $m$  Testwertvariablen  $Y_i$ , die durch die entsprechenden Auswertungsvorschriften der Tests definiert sind, die zugehörigen True-Score-Variablen  $\tau_i$  betrachten, deren Werte die eben definierten wahren Werte der Person sind. Entsprechend werden die **Meßfehlervariablen**  $\varepsilon_i$  als Differenz  $Y_i - \tau_i$  definiert. Bei dieser Präzisierung der Idee eines wahren Wertes wird also jede Testwertvariable  $Y_i$  additiv in ihre True-Score- und ihre Fehlervariable dekomponiert, d.h. für  $i = 1, \dots, m$  gilt:

$$Y_i = \tau_i + \varepsilon_i. \quad (3.1)$$

Aus den oben beschriebenen Definitionen der True-Score- und der Fehlervariablen lassen sich u. a. die folgenden Eigenschaften ableiten (s. z.B. Knoche, 1990; Novick, 1966; Steyer, 1989; Tack, 1980; Zimmerman, 1975):

$$E(\varepsilon_i) = 0, \quad (3.2)$$

$$\text{Cov}(\varepsilon_i, \tau_j) = 0, \quad (3.3)$$

$$\text{Var}(Y_i) = \text{Var}(\tau_i) + \text{Var}(\varepsilon_i), \quad (3.4)$$

$$E(\varepsilon_i | \tau_j) = 0. \quad (3.5)$$

Diese Gleichungen sind Spezialfälle der Eigenschaften des Residuums bzgl. einer bedingten Erwartung oder Regression (s. z.B. Steyer, 1992 oder Steyer & Eid, 1993). Keine dieser Gleichungen kann in einer empirischen Anwendung falsch sein, es sei denn, man ginge von einer anderen Definition der True-Score- und der Fehlervariablen aus. Dies ist ein Beispiel für den in der Einleitung angesprochenen Sachverhalt, daß ein Teil einer Theorie schon aus logischen Gründen wahr ist. An dieser Stelle kann also eine Theorienrevision nicht ansetzen, wohl dagegen bei den zugrundegelegten Definitionen und Annahmen. Letztere sind allerdings rein formaler Natur und stellen für Anwendungen keine wesentliche Restriktionen dar, da sie lediglich das betrachtete empirische Phänomen strukturieren und so einer mathematischen Behandlung zugänglich machen. Die einzige Voraussetzung, die zu den Gleichungen 3.1 bis 3.5 führt, ist die Endlichkeit der Varianz der Y-Variablen.

Man beachte, daß die in den älteren Darstellungen der Klassischen Testtheorie (Gulliksen, 1950; Lehmann, 1983; Lord & Novick, 1968; Wottawa, 1980) als Axiom betrachtete Annahme der Unkorreliertheit der Fehlervariablen  $\varepsilon_i$  untereinander keine Folgerung aus den oben vorgenommenen Definitionen der  $\tau_i$  bzw.  $\varepsilon_i$  ist, d.h. es gilt nicht unbedingt  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ ,  $i \neq j$ . Die Unkorreliertheit der Fehlervariablen **untereinander** kann in empirischen Anwendungen also durchaus falsch sein (s. dazu Tack, 1980; Zimmerman & Williams, 1977). Tatsächlich ist die Unkorreliertheit der Fehlervariablen zwar für viele mathematische Ableitungen bequem, aber nicht unbedingt für alle Zwecke notwendig.

Reliabilität: Eine notwendige Voraussetzung für die Brauchbarkeit einer Meßwertvariablen  $Y_i$  und des ihr zugrundeliegenden Meßinstruments ist, daß ihre Varianz nicht ausschließlich aus Fehlervarianz besteht. Der andere Extremfall,

in dem die Fehlervarianz  $\text{Var}(\epsilon_i)$  gleich 0 ist, ist zwar erstrebenswert, aber in den Sozialwissenschaften eher die Ausnahme als die Regel. Eine einfache Kenngröße, die das Ausmaß der Fehlerbehaftetheit bzw. **Unreliabilität** einer Variablen  $Y_i$  angibt, ist der Varianzanteil  $\text{Var}(\tau_i) / \text{Var}(Y_i)$ .

Dividiert man beide Seiten der Gleichung 3.4 durch die Varianz von  $Y_i$ , so erhält man:  $1 = \text{Var}(Y_i) / \text{Var}(Y_i) = [\text{Var}(\tau_i) / \text{Var}(Y_i)] + [\text{Var}(\epsilon_i) / \text{Var}(Y_i)]$ . Den Kennwert

$$\text{Rel}(Y_i) := 1 - \text{Var}(\epsilon_i) / \text{Var}(Y_i) = \text{Var}(\tau_i) / \text{Var}(Y_i) \quad (3.6)$$

bezeichnet man als **Reliabilität**. Man kann zeigen, daß es sich dabei um einen speziellen Determinationskoeffizienten handelt (s. Gl.2.5). Es gilt nämlich außer Gleichung 3.6 auch:  $\text{Rel}(Y_i) = \text{Var}[E(Y_i | r_i)] / \text{Var}(Y_i)$ . Damit sind nun die Grundbegriffe bereitgestellt, mit denen sich das Modell essentiell  $\tau$ -äquivalenter Variablen formulieren läßt.

## 3.2 Das Modell essentiell $\tau$ -äquivalenter Variablen

Ausgangspunkt sind  $m$  Testwertvariablen  $Y_i$  sowie deren True-Score- und Fehlervariablen  $\tau_i$  bzw.  $\epsilon_i$ . Die **Annahme der essentiellen  $\tau$ -Äquivalenz** läßt sich nun wie folgt formulieren: Zu jedem Paar  $i, j = 1, \dots, m$  existiert eine reelle Konstante  $h_{ij}$ , für die gilt:

$$\tau_i = \tau_j + \lambda_{ij}. \quad (3.7)$$

Diese Annahme besagt, daß die True-Score-Variablen aller  $m$  betrachteten  $Y$ -Variablen identisch sind bis auf eine Verschiebung um eine additive Konstante (d.h. eine Translation), nämlich  $h_{ij}$ .

### 3.2.1 Existenz

Betrachten wir nun den Fall  $j = 1$  und bezeichnen  $\tau_1$  mit  $\eta$  und  $-\lambda_{i1}$  mit  $\lambda_i$ , dann resultiert aus der obigen Gleichung für alle  $i = 1, \dots, m$ :

$$\tau_i = \eta - \lambda_i. \quad (3.8)$$

Diese Gleichung bezeichne ich als das **Fundamentalgesetz der subtraktiven Parametrisierung des Modells essentiell äquivalenter Variablen**. Ich ziehe hier die subtraktive Parametrisierung gegenüber der (logisch äquivalenten) additiven Parametrisierung  $\tau_i = \eta + \kappa_i$  vor, weil sich  $h_i$  in Gleichung 3.8 in vielen Anwendungen als **Schwierigkeit** des Tests interpretieren läßt. Aus der Annahme der essentiellen  $\tau$ -Äquivalenz folgt also die **Existenz** einer latenten Varia-

blen  $\eta$ , die für alle Y-Variablen gemeinsam ist, und eines für die Variable  $Y_i$  spezifischen Kennwerts  $L_i$ . Setzt man nämlich die Gleichungen 3.8 und 3.1 zusammen, dann erhält man für alle  $i = 1, \dots, m$ :

$$Y_i = \eta - \lambda_i + \varepsilon_i. \quad (3.9)$$

Dabei ist die latente Variable  $\eta$  eine theoretische Größe, deren Werte in den üblichen Anwendungen die Fähigkeit, Einstellung oder Eigenschaft der gezo-genen Person charakterisiert, wohingegen  $\lambda_i$  ein Kennwert für den durch die Variable  $Y_i$  repräsentierten Test(teil) ist.

### 3.2.2 Eindeutigkeit und Bedeutsamkeit

Offenbar ist aber  $\eta$  durch die Annahme der essentiellen  $\tau$ -Äquivalenz nicht völlig **eindeutig** definiert, denn wir haben  $\eta$  oben willkürlich mit  $z_1$  gleichge-setzt. Wir hätten  $\eta$  ebensogut mit  $z_2$  oder mit  $z_1 + \alpha$  gleichsetzen können, wobei  $\alpha$  eine beliebige reelle Zahl sein kann. Auch dann kommt man zu den Gleichungen 3.8 und 3.9. Demnach ist  $\eta$  nur eindeutig bis auf Translationen, d.h. Verschiebungen um eine additive Konstante definiert. Mit anderen Worten,  $\eta$  ist **differenzskaliert**. Dieses Skalenniveau reicht aus, damit bspw. Aus-sagen über die Varianz von  $\eta$  und über die Reliabilität der  $Y_i$  im meßtheore-tischen Sinn **bedeutsam** sind. Der Wahrheitswert dieser Aussagen ist also **in-variant unter den zulässigen Transformationen**. Auf dem gleichen Weg kann man sich überlegen, daß auch die Größen  $\lambda_i$  differenzskaliert sind.

Alle obigen Überlegungen basieren auf einer einzigen Annahme, der essen-tiellen  $\tau$ -Äquivalenz. Mit der Gültigkeit dieser Annahme stehen und fallen z.B. alle Aussagen über die Existenz und Eindeutigkeit der theoretischen Va-riablen  $\eta$ . Daher ist es angebracht zu überlegen, ob und wann diese Annahme in empirischen Anwendungen erfüllt sein kann und wie man sie gegebenenfalls überprüfen kann.

In der Regel werden zwei Parallelförmigen eines Tests gerade so konstruiert, daß die durch die betreffenden Auswertungsvorschriften definierten beiden Testwertvariablen  $Y_1$  und  $Y_2$  essentiell  $\tau$ -äquivalent sind. Dabei ist allerdings zu bedenken, ob nicht jede lineare Transformation  $Y'_i = a + b Y_i$ ,  $i = 1, 2$ , die gleiche Information wie die ur-sprünglichen beiden Variablen  $Y_i$  enthält. In diesem Fall wäre zu überlegen, ob man nicht ein Meßmodell konstruiert, in dem nicht die beiden **Variablen  $Y_i$** , sondern die beiden **Familien** aller linearen Transformation der beiden  $Y_i$  den Ausgangspunkt dar-stellen. Dies hätte den Effekt, daß die einzuführende theoretische Größe  $\eta$  nur inter-vallskaliert wäre.



### 3.2.3 Testbarkeit

Die erste empirisch prüfbare Konsequenz aus dem Modell essentiell  $\tau$ -äquivalenter Variablen ist die Gleichheit der Erwartungswerte der Differenzvariablen  $Y_i - Y_j$  in Subpopulationen  $U^{(1)}, U^{(2)} \subset U$ . Dabei wird keine weitere Annahme als die der essentiellen t-Äquivalenz vorausgesetzt (zum Beweis s. Steyer, 1988, 1989, oder auch Steyer & Eid, 1993):

$$E^{(1)}(Y_i - Y_j) = E^{(2)}(Y_i - Y_j), \quad i, j = 1, \dots, m. \quad (3.10)$$

In Anwendungen kann man also die Annahme der essentiellen  $\tau$ -Äquivalenz prüfen, indem man untersucht, ob die Erwartungswerte der Differenzvariablen  $Y_i - Y_j$ ,  $i, j = 1, \dots, m$ , in zwei **verschiedenen** Subpopulationen identisch sind. Die obige Gleichung ist eine Hypothese über die Gleichheit der Erwartungswerte von  $m(m-1)/2$  (Differenz-)Variablen. Fügt man die Annahmen eines Stichprobenmodells, insbesondere die Annahmen der Normalverteilung und der Varianzhomogenität, hinzu, dann ist im Fall  $m = 2$  das statistische Standardverfahren zur Prüfung der o.g. Hypothese der t-Test und im Fall  $m > 2$  der multivariate t-Test, der sich z.B. mit der SPSS-Prozedur MANOVA leicht durchführen läßt.

Der zweite Typ empirisch überprüfbarer Folgerungen aus der Annahme der essentiellen  $\tau$ -Äquivalenz betrifft die Kovarianzen der betrachteten Y-Variablen. Dabei wird allerdings die Bedingung der essentiellen  $\tau$ -Äquivalenz in Konjunktion mit einer zusätzlichen Annahme geprüft, nämlich der Unkorreliertheit der Fehlervariablen  $\epsilon_i$  untereinander:

$$\text{Cov}(\epsilon_i, \epsilon_j) = 0, \quad i \neq j. \quad (3.11)$$

Da  $\eta$  eine Translation (Verschiebung um eine additive Konstante) von einer True-Score-Variablen  $\tau_i$  ist, von der wir schon nach Gleichung 3.3 wissen, daß sie unkorreliert mit den Fehlervariablen ist, gilt außerdem für  $i = 1, \dots, m$ :

$$\text{Cov}(\epsilon_i, \eta) = 0. \quad (3.12)$$

Diese beiden Gleichungen führen nun zu der in der folgenden Gleichung angegebenen Struktur der Kovarianzen der Y-Variablen, wobei  $i, j = 1, \dots, m$ :

$$\text{Cov}(Y_i, Y_j) = \begin{cases} \text{Var}(\eta), & i \neq j, \\ \text{Var}(\eta) + \text{Var}(\epsilon_i), & i = j. \end{cases} \quad (3.13)$$

Gemäß dieser Gleichung haben verschiedene essentiell  $\tau$ -äquivalente Variablen  $Y_i$  und  $Y_j$ ,  $i \neq j$ , jeweils die gleiche **Kovarianz**, falls die Fehler unkorreliert sind. Diese Kovarianz ist zugleich die Varianz von  $\eta$ . Dabei beachte man, daß hier nicht von den Stichprobenkovarianzen die Rede ist, sondern von den wahren (Populations-)Kovarianzen.

Zur Illustration sei der Spezialfall  $m = 3$  betrachtet. Die Kovarianzmatrix  $\Sigma$  der Y-Variablen kann dann gemäß Gleichung 3.13 wie folgt notiert werden:

$$\Sigma = \begin{bmatrix} \text{Var}(\eta) + \text{Var}(\varepsilon_1) & \text{Var}(\eta) & \text{Var}(\eta) \\ \text{Var}(\eta) & \text{Var}(\eta) + \text{Var}(\varepsilon_2) & \text{Var}(\eta) \\ \text{Var}(\eta) & \text{Var}(\eta) & \text{Var}(\eta) + \text{Var}(\varepsilon_3) \end{bmatrix} .. \quad (3.14)$$

Als theoretische Parameter kommen hier also die Varianzen von  $\eta$  und die Varianzen der Fehlervariablen vor. Gemäß Gleichung 3.13 impliziert das Modell essentiell  $\tau$ -äquivalenter Variablen mit unkorrelierten Fehlern die Gleichheit der Kovarianzen  $\text{Cov}(Y_1, Y_2) = \text{Cov}(Y_1, Y_3) = \text{Cov}(Y_2, Y_3)$ . Im Fall  $m = 3$  hat das Modell also bereits empirisch testbare Konsequenzen für die Kovarianten der Y-Variablen, die ja, falls kein Modell essentiell  $\tau$ -äquivalenter Variablen vorläge, beliebige Werte annehmen könnten.

Zur praktischen Durchführung eines solchen Tests kann man ein Programm zur Analyse von Kovarianzstrukturen (zur Einführung in Kovarianzstrukturmodelle s. z.B. Bollen, 1989, Hayduk, 1987, oder Saris & Stronkhorst, 1984) wie z.B. LISREL 7 (Jöreskog & Sörbom, 1989) verwenden. Dies wird ausführlich von Steyer und Eid (1993) demonstriert, die auch eine weitere testbare Konsequenz des Modells essentiell  $\tau$ -äquivalenter Variablen behandeln.

### 3.2.4 Identifizierbarkeit

Die Gleichung 3.13 zeigt nicht nur, daß die Kovarianzmatrix der Y-Variablen eine bestimmte Struktur haben muß, wenn die beiden Modellannahmen gelten, sondern auch, wie man die theoretischen Parameter, die Varianzen von  $\eta$  und  $\%_i$  aus den Varianzen und Kovarianzen der Y-Variablen bestimmen (identifizieren) kann.

Die nächsten beiden Gleichungen geben an, wie die Varianzen  $\text{Var}(\eta)$  und  $\text{Var}(\varepsilon_i)$  der latenten Variablen aus den Varianzen und Kovarianzen der beobachtbaren Variablen  $Y_i$  bestimmt werden können, wenn die Annahmen der essentiellen  $\tau$ -Äquivalenz und unkorrelierter Fehler gültig sind. Unter diesen beiden Annahmen gelten für  $i, j = 1, \dots, m$ :

$$\text{Var}(\eta) = \text{Cov}(Y_i, Y_j), \quad i \neq j, \quad (3.15)$$

$$\text{Var}(\varepsilon_i) = \text{Var}(Y_i) - \text{Cov}(Y_i, Y_j), \quad i \neq j. \quad (3.16)$$

Gemäß diesen Gleichungen sind die Varianzen  $\text{Var}(\eta)$  und  $\text{Var}(\varepsilon_i)$  im Modell essentiell  $\tau$ -äquivalenter Variablen bereits bei  $m = 2$  Y-Variablen identifiziert, da nur von zwei verschiedenen Indizes  $i$  und  $j$  die Rede ist.

Die **Varianzanteile** der  $Y_i$ , die durch die latente Variable  $\eta$  determiniert sind, die Reliabilitäten also, können nach der folgenden Formel bestimmt werden, wenn die Annahmen der essentiellen  $\tau$ -Äquivalenz und unkorrelierter Fehler gelten. Dann folgt nämlich für  $i, j = 1, \dots, m$ :

$$Rel(Y_i) = \frac{Cov(Y_i, Y_i)}{Var(Y_i)}, \quad i \neq j. \quad (3.17)$$

Sind außerdem auch noch die Fehlervarianzen gleich, d.h. gilt  $Var(E_i) = Var(\epsilon_i)$ ,  $i \neq j$ , dann folgt auch

$$Kor(Y_i, Y_j) = Rel(Y_i). \quad (3.18)$$

Dies ist die Gleichung, die den Verfahren zur Bestimmung der Reliabilität über die **Paralleltest-** und über die **Retestkorrelation** zugrundeliegt.

### 3.3 Zusammenfassende Bemerkungen

Stochastische Meßmodelle dienen zwei wichtigen Zwecken: Zum einen wird in einem solchen Modell die theoretische, vom Meßfehler bereinigte Größe konstruiert und zum anderen erlauben solche Modelle abzuschätzen, wie groß der Meßfehler ist, mit dem man bei einer Messung rechnen muß. Die einzelnen Schritte bei der Konstruktion eines Konstrukts im Rahmen eines stochastischen Meßmodells kann man wie folgt zusammenfassen: Ausgangspunkt sind Zufallsvariablen  $Y_i$ ,  $i = 1, \dots, m$ , deren Werte beobachtbares Verhalten der betrachteten Beobachtungseinheit (hier: einer Person) repräsentieren. Die Modellannahmen bestehen aus Aussagen über bestimmte Eigenschaften der Verteilungen dieser Zufallsvariablen (hier: ihrer person-bedingten Erwartungswerte - der True scores - und deren Residuen). Daraus läßt sich die **Existenz** einer oder mehrerer theoretischer Größen (hier  $\eta$  und die Parameter  $\lambda_i$ ) deduzieren, und man kann feststellen, wie **eindeutig** diese theoretischen Größen durch die Modellannahmen definiert sind. Dies ist gleichbedeutend mit der Feststellung ihres **Skalenniveaus**. Daraus läßt sich ableiten, welche Aussagen, die die theoretischen Größen involvieren, **bedeutsam** sind, in dem Sinn, daß der Wahrheitswert dieser Aussagen invariant **unter den zulässigen Transformationen** ist. In einem weiteren Schritt kann man untersuchen, wie sich die theoretischen Größen aus den Verteilungskennwerten der Ausgangsvariablen  $Y_i$  bestimmen oder **identifizieren** lassen. Um die Überprüfung des Modells zu ermöglichen, läßt sich schließlich feststellen, welche **empirisch testbaren Konsequenzen** sich aus den Modellannahmen ableiten lassen.

Die statistischen Tests und Schätzungen der Parameter basieren dann auf weiteren Annahmen wie z.B. der unabhängigen identischen Verteilung in der Stichprobe, die durch das **N-malige** Durchführen des oben behandelten Zufallsexperiments entsteht. Oft (aber nicht immer, s. z.B. Browne, 1984) sind auch Annahmen über bestimmte Verteilungsklassen (wie z.B. Normalverteilung) nötig. Diese statistischen Annahmen und Modelle sind zwar für Anwendungen unerlässlich, aber für die Logik des eigentlichen stochastischen Modells, das oben skizziert wurde, das sich auf das zufällige Ziehen **einer einzigen** Beobachtungseinheit bezieht, irrelevant.

#### 4. Zusammenfassung und Ausblick

In diesem Beitrag wurde dargelegt, wie in stochastischen Modellen ihre beiden Aufgaben, die Explikation von **Meßmodellen** und von **Abhängigkeitsbegriffen** gelöst werden können. Beide Aufgaben dienen dem für eine deduktivistische Methodologie zentralen Ziel, die **logische Ableitung von Aussagen über die Empirie aus der Theorie** zu **ermöglichen**. Nur dann, wenn eine solche logische Ableitbarkeit möglich ist, sind theoretische Aussagen auch falsifizierbar. Um die Deduzierbarkeit empirischer Aussagen zu gewährleisten, sind mindestens zwei Dinge vonnöten: **Erstens** müssen die in der Theorie vorkommenden theoretischen Begriffe (wie z. B. „Frustration“ und „Aggression“) mit Beobachtbarem verknüpft werden, was durch Meßmodelle geschieht (s. dazu Abschnitt 3). **Zweitens** müssen aber auch die postulierten Abhängigkeiten zwischen den theoretischen Begriffen in einer Sprache formuliert werden, die logische Ableitungen für beobachtbare Sachverhalte erlaubt. Die umgangssprachliche Formulierung „Frustration führt zu Aggression“ gestattet vielleicht Folgerungen nach Plausibilitätsüberlegungen, aber keinerlei logische Ableitungen, die m.E. in einer entwickelten wissenschaftlichen Disziplin zu fordern sind. Genau an dieser Stelle können wir die in Abschnitt 2 behandelten stochastischen Abhängigkeitsbegriffe verwenden. Sie stellen die Sprache zur Formulierung der Aussagen über die **Abhängigkeiten** zwischen den theoretischen Begriffen bereit. Dies ist nicht nur für die Wissenschaft, sondern auch für die Gesellschaft von großer Bedeutung, wenn man die möglichen Anwendungsfelder in der Epidemiologie („die radioaktiven Emissionen **führen** zu Leukämie“), Ökologie („die CO<sub>2</sub>-Abgase **führen** zu Waldsterben“), Sozialpolitik („Arbeitslosigkeit **führt** zu Alkoholismus“), Verkehrspolitik („Verbreiterung der Straße **reduziert** die Unfallgefahr“) etc. bedenkt.

In der Psychologie sind meist nur **stochastische** Modelle realistisch, und zwar sowohl als Meßmodelle als auch als Abhängigkeitsmodelle. Die stochastische Natur dieser Modelle kompliziert zwar den Prozeß der Falsifikation durch die Notwendigkeit, eine **Entscheidung** über die Gültigkeit einer Hypothese aufgrund statistischer Untersuchungen und Überlegungen zu treffen (s. Ostmann & Wutke, Kap. 16 in diesem Band), aber dennoch bleibt die Falsifizierbarkeit das entscheidende Kriterium für eine Abgrenzung wissenschaftlicher empirischer Theorien von vorwissenschaftlichen Theorien. Letztere sind zwar notwendige Schritte im Wissenschaftsprozess, sind aber m.E. nur als Zwischenlösungen anzusehen.

## 4.1 Stochastische Meßmodelle

Stochastische Meßmodelle explizieren die Verknüpfung zwischen theoretischen und empirischen Begriffen. Ohne eine klare Verknüpfung ihrer theoretischen Begriffe mit Beobachtbarem wären Disziplinen wie z.B. die Psychologie, Soziologie oder Wirtschaftswissenschaften keine empirischen Wissenschaften, sondern Philosophie. Neben der Verknüpfung zwischen Theorie und Empirie können Meßmodelle einige allgemeine Probleme der Sozialwissenschaften lösen, die im folgenden kurz diskutiert werden.

**Das Meßfehlerproblem:** Sozialwissenschaftliche Beobachtungen und Messungen sind in der Regel meßfehlerbehaftet. Stochastische Meßmodelle können diese Meßfehler berücksichtigen und darüber Auskunft geben, wie **stark** diese Meßfehlerbehaftetheit ist. (Siehe hierzu die Meßfehlervarianz und die Reliabilität in Abschnitt 3.)

**Das Problem situationaler Spezifität:** Neben dem Meßfehlerproblem erschwert auch die Veränderung und Variabilität der zu messenden Eigenschaften und Zustände von Personen das Messen in der Psychologie. Das entsprechende gilt auch für die Messung von Eigenschaften anderer Objekte wie z. B. Gruppen oder sozialer Institutionen. Wie können wir z.B. abschätzen, welcher Anteil eines Testwerts auf Meßfehler, welcher Anteil auf die Besonderheit der Situation, in der der Test vorgelegt wird, und welcher Anteil auf die zu messende überdauernde Eigenschaft des betrachteten Objekts zurückgeht? Während für das Meßfehlerproblem schon seit einigen Jahrzehnten Lösungen existieren, gibt es erst in jüngster Zeit befriedigende Lösungsmöglichkeiten für das Problem der situationalen Spezifität von Messungen, die auch für nichtexperimentelle Beobachtungsstudien brauchbar sind (s. z. B. Schmitt & Steyer, 1990; Steyer, 1987, 1988; Steyer & Schmitt, 1990a, b; Steyer, Ferring & Schmitt, 1992).

Das Problem der situationalen Spezifität ist grundsätzlich nur in solchen Modellen zu lösen, in denen wiederholte Messungen berücksichtigt werden. Im Rahmen von Strukturgleichungsmodellen sind dazu die Bücher von Möbus und Schneider (1986) sowie die Artikel von Möbus und Nagl (1983) und von Jöreskog und Sörbom (1977) zu nennen. Auch im Rahmen der probabilistischen Testtheorie gibt es dazu Ansätze (s. z.B. Andersen, 1985, 1988; Fischer, 1989; Langeheine & van de Pol, 1990a, b; Rost & Spada, 1983).

**Das Problem der Methodenspezifität:** In vielen Anwendungen der obigen Modelle hat sich herausgestellt, daß die Berücksichtigung von Meßfehlern und situativen Effekten noch immer nicht differenziert genug ist, um der tatsächlichen Komplexität psychologischer Messungen gerecht zu werden. Es stellte sich nämlich heraus, daß es auch noch **Meßmethoden-spezifische Effekte** gibt.

Selbst wenn Tests konstruiert werden, um die gleiche Eigenschaft zu messen, gelingt dies oft nicht perfekt. Neben der zu messenden Eigenschaft hängen die Testwerte auch noch von systematischen Effekten ab, die für den Test (das Meßinstrument) spezifisch sind. Dies kommt dadurch zum Ausdruck, daß die Testwertvariablen  $Y_{ik}$  und  $Y_{il}$ , die durch den gleichen Test  $i$  zu zwei verschiedenen Meßgelegenheiten  $k$  und  $l$  erhoben werden, höher miteinander korrelieren, als es durch Modelle erklärbar wäre, die nur Meßfehler, und situative Effekte berücksichtigen. Jöreskog (1979), Saris und van Meurs (1990) sowie Steyer, Ferring und Schmitt (1992) behandeln daher Modelle, in denen neben den o.g. Problemen auch das Problem der **Methoden-Spezifität** berücksichtigt wird.

**Das Problem heterogener Populationen:** Ein weiteres Problem, das man inzwischen im Rahmen stochastischer Meßmodelle lösen kann, ist das **Problem heterogener Populationen**. Oft zeigt sich, daß sich Personen in ihren **Antwortstilen** bei der Beantwortung von Items eines Fragebogens voneinander unterscheiden. Dies kann man dadurch in einem stochastischen Meßmodell berücksichtigen, daß man davon ausgeht, daß eine heterogene Population vorliegt, bei der in jeder Subpopulation eine andere Beziehung zwischen den Antworten auf die Items und den latenten Variablen besteht. So bedeutet „volle Zustimmung“ bei einer Person mit einer Tendenz zu extremen Antworten in bezug auf die zu messende Einstellung nicht das gleiche, wie die gleiche Antwort einer Person, die keine solche Tendenz zu Extremantworten aufweist (s. dazu Rost, 1991; Rost & Langeheine, 1991).

## 4.2 Stochastische Abhängigkeitsbegriffe

Neben den oben diskutierten Problemen, die im Rahmen stochastischer Meßmodelle gelöst werden können, sind stochastische Modelle aber auch zur Formulierung von Hypothesen über **Abhängigkeiten** zwischen theoretischen Variablen, aber auch zwischen empirischen Variablen unerläßlich. Die Relevanz stochastischer Abhängigkeitsbegriffe für eine deduktivistische Methodologie ergibt sich aus folgender Überlegung: Kommen an nur einer Stelle stochastische Begriffe wie Erwartungswert, Wahrscheinlichkeit, Korrelation o.ä. vor, und dies trifft auf den größten Teil empirischer Untersuchungen zu, so sind nur dann Deduktionen möglich, wenn die theoretischen Aussagen selbst bereits in Termini der Wahrscheinlichkeitstheorie formuliert sind oder wenigstens direkt in diese Sprache übersetzt werden können.

Beispiele für verschiedene in psychologischen Hypothesen vorkommende Abhängigkeitsbegriffe wurden ausführlich im Abschnitt 2 behandelt. Es waren dies die Begriffe der einfachen regressiven Abhängigkeit, der korrelativen Abhängigkeit, der partiellen regressiven Abhängigkeit, der bedingten regressiven

Abhängigkeit, sowie der starken und der schwachen kausalen regressiven Abhängigkeit. Auch die mit den verschiedenen Arten stochastischer Abhängigkeit behandelten Präzisierungen von Abhängigkeitsbegriffen dienen letztlich dazu, dem Hauptanliegen der deduktivistischen Methodologie, der logischen Ableitbarkeit empirisch falsifizierbarer Aussagen, Rechnung zu tragen.

### 4.3 Ausblick

Stochastische Modelle dienen nicht nur der Lösung spezieller Probleme innerhalb der Mathematischen Psychologie, obwohl sie auch dort zweifelsohne eine wichtige Rolle spielen (s. z.B. Colonius, 1984; DeGreef & van Buggenhaut, 1984; Fischer, **1993**; Holling, **1989**; Roskam & Suck, 1987). In diesem Beitrag sollte stattdessen gezeigt werden, daß sie eine zentrale und unverzichtbare Funktion für alle Bereiche der „gewöhnlichen“ empirischen Sozialwissenschaften einnehmen, jedenfalls dann, wenn man sich zu einer deduktivistischen Methodologie bekennt, deren Kernpostulat die Falsifizierbarkeit empirischer wissenschaftlicher Theorien ist.

Dazu sei angemerkt, daß man auch den Strukturalismus (Balzer, Moulines & Sneed, **1987**; Gähde, 1983; Gähde & Stegmüller, 1986; Gähde, Jagodzinski & Steyer, 1993; Stegmüller, 1979, 1986; Stephan, **1990**; Westermann, 1987; Westmeyer, 1989, 1992) als einen wichtigen Ansatz im Rahmen einer deduktivistischen Methodologie empirischer Wissenschaften ansehen kann, ist doch sein Hauptanliegen, die **logische Struktur wissenschaftlicher Theorien** (s. Sneed, 1971) zu explizieren. Auch wenn sich in den Untersuchungen der strukturalistischen Wissenschaftstheorie herausgestellt hat, daß bestimmte Theorieteile in der Regel für sich genommen nicht falsifizierbar sind, ist damit das Falsifizierbarkeitspostulat keineswegs ad acta gelegt. In Termini des Strukturalismus bedeutet es nichts anderes, als daß der **empirische Gehalt** (s. Diederich, 1981) einer empirischen wissenschaftlichen Theorie nicht leer sein darf.

### Literatur

- Andersen, E. B. (**1985**). Estimating latent correlations between repeated testings. *Psychometrika*, 50, 3-16.
- Andersen, E.B. (1989). Comparison of latent structure models. In R. Langeheine & J. Kost (Eds.), **Latent trait and latent class models** (pp. 207-229). New York: Plenum.
- Andersen, E. B. (1990). **The statistical analysis of categorical data**. Berlin: Springer.
- Ash, R. A. (1972). **Real analysis and probability**. New York: Academic Press.

- Austin J.T. & Wolfle, L.M. (1991). Annotated bibliography of structural equation modelling: Technical work. *British Journal of Mathematical and Statistical Psychology*, **44**, 93-152.
- Balzer, W., Moulines, C.U. & Sneed, J.D. (1987). *An architectonic for science*. Dordrecht: Reidel.
- Bartussek, D. (1970). Eine Methode zur Bestimmung von Moderatoreffekten. *Diagnostica*, **16**, 57-76.
- Bauer, H. (1978). *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie* (3. Auflage). Berlin: de Gruyter.
- Bentler, P. M. (1986). Structural modeling and Psychometrika: An historical perspective on growth and achievements. *Psychometrika*, **51**, 35-51.
- Bishop, Y.M.M., Fienberg, S.E. & Holland, P. W. (1977). *Discrete multivariate analysis*. Boston: MIT Press.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bosch, K. (1986). *Elementare Einführung in die Wahrscheinlichkeitsrechnung*. Braunschweig: Vieweg.
- Bredenkamp, J. (1984a). Theoretische und experimentelle Analysen dreier Wahrnehmungstäuschungen. *Zeitschrift für Psychologie*, **192**, 47-61.
- Bredenkamp, J. (1984b). Theoretische und experimentelle Analysen einiger Wahrnehmungstäuschungen. *Archiv für Psychologie*, **136**, 281-291.
- Breiman, L. (1968). *Probability*. Reading, MA: Addison-Wesley.
- Browne, M.W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62-83.
- Cheek, J.M. (1982). Aggregation, moderator variables, and the validity of personality tests: A peer rating study. *Journal of Personality and Social Psychology*, **43**, 1254-1269.
- Colonus, H. (1984). *Stochastische Theorien individuellen Wahlverhaltens*. Berlin: Springer.
- Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, **69**, 161-182.
- DeGreef, E. & van Buggenhaut, J. (1984) (Eds.). *Trends in Mathematical Psychology*. Amsterdam: North-Holland.
- De Gruijter, D. N. M. & van der Kamp, L.J. T. (1984). *Statistical models in psychological and educational testing*. Lisse: Swets & Zeitlinger.
- Diederich, W. (1981). *Strukturalistische Rekonstruktionen: Untersuchungen zur Bedeutung, Weiterentwicklung und interdisziplinären Anwendung des strukturalistischen Konzepts wissenschaftlicher Theorien*. Braunschweig: Vieweg.
- Erdfelder, E. & Steyer, R. (1984). Zur Psychophysik einiger Größentäuschungen. *Psychologische Beiträge*, **26**, 639-646.
- Falmagne, C. (1985). *Elements of psychophysical theory*. New York: Oxford University Press.
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Huber.



- Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika*, **46**, 59-77.
- Fischer, G. H. (1983). Neuere Testtheorie. In H. Feger & J. Bredenkamp (Hrsg.), *Messen und Testen. Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie 1: Forschungsmethoden der Psychologie, Band 3* (S. 604-692). Göttingen: Hogrefe.
- Fischer, G. H. (1988). Spezifische Objektivität: Eine wissenschaftstheoretische Grundlage des Rasch-Modells. In K. Kubinger (Hrsg.), *Moderne Testtheorie* (S. 87-11). Weinheim: Psychologie Verlags Union.
- Fischer, G. H. (1989). An IRT-based model for dichotomous longitudinal data. *Psychometrika*, **54**, 599-624.
- Fischer, G. H. (1993). Proceedings of the 22nd Meeting of the European Mathematical Psychology Group in Vienna. New York: Springer.
- Formann, A. (1984). *Die Latent-Class-Analyse*. Weinheim: Beltz.
- Gähde, U. (1983). *T-Theoretizität und Holismus*. Frankfurt am Main: Lang.
- Gähde, U. & Stegmüller, W. (1986). An argument in favor of the Duhem-Quine-thesis: From the structuralist point of view. In: L.E. Hahn & P.H. Schilpp (Eds.), *The Philosophy of W. v. Quine*. The library of living philosophers, Vol. 15, La Salle, IL.: Open Court Publishing Company.
- Gähde, U., Jagodzinski, W. & Steyer, R. (1992). On a structuralist reconstruction of latent state-trait theory. In H. Westmeyer (Ed.), *The structuralist program in psychology: Foundations and applications* (pp. 105-119). Toronto: Hogrefe-Huber.
- Gänssler, P. & Stute, W. (1977). *Wahrscheinlichkeitstheorie*. Berlin: Springer.
- Ghiselli, E. E. (1963). Moderating effects and differential reliability and validity. *Journal of Applied Psychology*, **47**, 81-86.
- Goodman, L. A. & Kruskal, W. H. (1979). *Measures of association for cross-classifications*. New York: Springer.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Härtter, E. (1987). *Wahrscheinlichkeitsrechnung, Statistik und mathematische Grundlagen. Begriffe, Definitionen*. Göttingen: Vandenhoeck & Ruprecht.
- Hager, W. (1987). Grundlagen einer Versuchsplanung zur Prüfung empirischer Hypothesen in der Psychologie. In G. Lüer (Hrsg.), *Allgemeine experimentelle Psychologie* (S. 43-264). Stuttgart: Fischer.
- Hager, W. (1992). Eine Strategie zur Entscheidung über psychologische Hypothesen. *Psychologische Rundschau*, **43**, 18-29.
- Hambleton, R. H. & Swaminathan, H. (1985). *Item response theory*. Boston: Kluwer-Nijhoff.
- Hartung, J., Elpelt, B. & Klösener, K.-H. (1989). *Statistik. Lehr- und Handbuch der angewandten Statistik* (7. Auflage). München: Oldenbourg.
- Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances*. Baltimore: Johns Hopkins University Press.
- Hinderer, K. (1980). *Grundbegriffe der Wahrscheinlichkeitstheorie*. Berlin: Springer.

- Holling, H. (1989). Psychische Beanspruchung durch Wartezeiten in der Mensch-Computer-Interaktion. Berlin: Springer.
- Jöreskog, K. G. (1979). Statistical estimation of structural equation models in longitudinal-developmental investigations. In J. R. Nesselroade & P. B. Baltes (Eds.), **Longitudinal research in the study of behavior and development** (pp.303-352). New York: Academic Press.
- Jöreskog, K. G. & Sörbom, D. (1977). Statistical models and methods for analysis of longitudinal data. In D.J. Aigner & A. S. Goldberger (Eds.), **Latent variables in socio-economic models** (pp.285-325). Amsterdam: North-Holland.
- Jöreskog, K. G. & Sörbom, D. (1989). **Lisrel 7. A guide to the program and its applications** (2nd edition). Chicago, IL: SPSS Inc.
- Kenny, D. A. & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. **Psychological Bulletin**, **96**, 201-210.
- Knoche, N. (1990). **Modelle der empirischen Pädagogik**. Mannheim: B.I.-Wissenschaftsverlag.
- Knoke, D. & Burke, P. J. (1980). **Log-linear models**. Beverly Hills, CA: Sage.
- Kolmogoroff, A. (1970). **Grundbegriffe der Wahrscheinlichkeitsrechnung**. Berlin: Springer. (Nachdruck der Originalarbeit von 1933).
- Krantz, D.H., Luce, R.D., Suppes, P. & Tversky, A. (1971). **Foundations of measurement. Vol I. Additive and polynomial representations**. New York: Academic Press.
- Krauth, J. & Lienert, G.A. (1973). **KFA - Die Konfigurationsfrequenzanalyse**. Freiburg: Alber.
- Krickeberg, K. & Ziezold, H. (1979). **Stochastische Methoden**. Berlin: Springer.
- Kubinger, K. (1988). **Moderne Testtheorie**. Weinheim: Psychologie Verlags Union.
- Langeheine, R. (1982). Log-lineare Modelle. In J. van Koolwijk & M. Wieken-Mayser (Hrsg.), **Techniken der empirischen Sozialforschung. Band 8. Kausalanalyse** (S.122-195). München: Oldenbourg.
- Langeheine, R. & Rost, J. (Eds.). (1988). **Latent trait and latent class models**. New York: Plenum.
- Langeheine, R. & van de Pol, E (1990a). Veränderungsmessung bei kategorialen Daten. **Zeitschrift für Sozialpsychologie**, **21**, 88-100.
- Langeheine, R. & van de Pol, F. (1990b). A unifying framework for Markov modeling in discrete space and discrete time. **Sociological Methods & Research**, **18**, 416-441.
- Lehmann, G. (1983). Testtheorie. In H. Feger & J. Bredenkamp (Hrsg.), **Messen und Testen. Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie Z: Forschungsmethoden der Psychologie, Band 3 (S. 427-543)**. Göttingen: Hogrefe.
- Lewis, C. (1986). Test theory and Psychometrika: The past twenty-five years. **Psychometrika**, **51**, 11-22.
- Lienert, G.A. (1961). **Testaufbau und Testanalyse**. Weinheim: Beltz.
- Loeve, M. (1977). **Probability theory I** (4th edition). New York: Springer.
- Loeve, M. (1978). **Probability theory II** (4th edition). New York: Springer.

- Lord, F. M. & Novick, M. R. (1968). **Statistical theories of mental test scores**. Reading, MA: Addison-Wesley.
- Luce, R.D., Krantz, D., Suppes, P. & Tversky, A. (1990). **Foundations of measurement. Vol. III. Representation, axiomatization, and invariance**. San Diego: Academic Press.
- Mill, J. S. (1862). **System der deductiven und inductiven Logik. Erster Theil** (Übersetzung von J. Schiel. 2. deutsche, nach der fünften des Originals erweiterte Auflage. Original erschienen 1843). Braunschweig: Vieweg.
- Möbus, C. & Nagl, W. (1983). Messung, Analyse und Prognose von Veränderungen. In J. Bredenkamp & H. Feger (Hrsg.), **Hypothesenprüfung - Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie I: Forschungsmethoden der Psychologie, Band 5** (239-470). Göttingen: Hogrefe.
- Moosbrugger, H. (1981). Zur differentiellen Validität bei nichtlinearen Test-Kriteriumsbeziehungen. **Zeitschrift für Differentielle und Diagnostische Psychologie, 2**, 219-234.
- Müller, P. H. (Hrsg.) (1975). **Lexikon der Stochastik** (2. Auflage). Berlin: Akademie-Verlag.
- Mulaik, S.A. (1986). Factor analysis and Psychometrika: Major developments. **Psychometrika, 51**, 22-33.
- Novick, M. R. (1966). The axioms and principal results of classical test theory. **Journal of Mathematical Psychology, 3**, 1-18.
- Penner, L.A. & Wymer, W.E. (1983). The moderator variable approach to behavioral predictability: Some of the variables some of the time. **Journal of Research in Personality, 17**, 339-353.
- Plachky, D. (1981). **Stochastik II**. Wiesbaden: Akademische Verlagsgesellschaft.
- Poppen K. R. (1984). **Logik der Forschung** (8. Auflage). Tübingen: Mohr. (1. Auflage 1934).
- Rasch, G. (1980). **Probabilistic models for some intelligence and attainment tests** (2nd edition). Chicago: University of Chicago Press. (1. Auflage erschienen 1960 in Kopenhagen: Nielsen & Lydiche).
- Renyi, A. (1977). **Wahrscheinlichkeitsrechnung**. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Roberts, F. (1979). **Measurement theory**. Reading, MA: Addison-Wesley.
- Roskam, E.E. & Suck, R. (1987) (Eds.). **Pro gress in Mathematical Psychology - I**. Amsterdam: North-Holland.
- Rost, J. (1988). **Quantitative und qualitative probabilistische Testtheorie**. Bern: Huber.
- Rost, J. (1991). A logistic mixture distribution model for polytomous item responses. **British Journal of Mathematical and Statistical Psychology, 44**, 75-92.
- Rost, J. & Langeheine, R. (1991). Mischverteilungsmodelle: Die Methodologie der kommenden Jahre. In D. Frey (Hrsg.), **Bericht über den 37. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1990** (S.590-596). Göttingen: Hogrefe.
- Rost, J. & Spada, H. (1983). Die Quantifizierung von Lerneffekten anhand von Testdaten. **Zeitschrift für Differentielle und Diagnostische Psychologie, 4**, 29-49.

- Rost, J. & Strauß, B. (1992). Psychometrics and test theory-recent developments. *German Journal of Psychology*, **16**, 91-119.
- Saris, W. & van Meurs, A. (1990). *Evaluation of measurement instruments by meta-analysis and multitrait-multimethod* studies. Amsterdam: North-Holland.
- Saris, W. E. & Stronkhorst, L. E. (1984). Causal *modelling in nonexperimental research*. Amsterdam: Sociometric Research Foundation.
- Saunders, D. R. (1956). Moderator variables in prediction. *Educational und Psychological Measurement*, **16**, 209-222.
- Schmitt, M. J. (1990). *Konsistenz als Persönlichkeitseigenschaft!* Berlin: Springer.
- Schmitt, M. J. & Steyer, R. (1990) Beyond intuition and classical test theory: A reply to Epstein. *Methodika*, **4**, 101-107.
- Schmitt, M. & Götz-Baltes, B. (1992). Common and uncommon moderator concepts: Comment on Wermuth's „Moderating Effects in Multivariate Normal Distributions“. *Methodika*, **6**, 1-4.
- Sneed, J. D. (1971). *The logical structure of mathematical physics*. Dordrecht: Reidel.
- Spohn, W. (1980). Stochastic independence, causal independence, and shieldability. *Journal of Philosophical Logic*, **9**, 73-99.
- Spohn, W. (1983). Deterministic and probabilistic reasons and causes. *Erkenntnis*, **19**, 371-396.
- Spohn, W. (1990). Direct and indirect causes. *Topoi*, **9**, 125-145.
- Spohn, W. (1991). A reason for explanation: Explanations provide stable reasons. In W. Spohn, B.C. van Fraassen & B. Skyrmes (Eds.), *Existence and explanation* (pp. 165-196). Dordrecht: Kluwer.
- Spohn, W. (im Druck). On Reichenbach's principle of the common cause. In W.C. Salmon & G. Wolters (Eds.), *Logic, language, and the structure of scientific theories*. Pittsburgh, PA: University of Pittsburgh Press.
- Stegmüller, W. (1979). *The structuralist view of theories*. Berlin: Springer.
- Stegmüller, W. (1983). *Erklärung, Begründung, Kausalität* (Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band 1). Berlin: Springer.
- Stegmüller, W. (1986). *Theorie und Erfahrung: 3. Teilband. Die Entwicklung des neuen Strukturalismus seit 1973* (Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band II). Berlin: Springer.
- Stephan, E. (1990). Zur *logischen* Struktur *psychologischer Hypothesen*. Berlin: Springer.
- Steyer, R. (1987). Konsistenz und Spezifität: Definition zweier zentraler Begriffe der Differentiellen Psychologie und ein einfaches Modell zu ihrer Identifikation. *Zeitschrift für Differentielle und Diagnostische Psychologie*, **8**, 245-258.
- Steyer, R. (1988). *Experiment, Regression und Kausalität. Die logische Struktur kausaler Regressionsmodelle*. Unveröff. Habilitationsschrift, Universität Trier, FB I - Psychologie.
- Steyer, R. (1989). Models of classical psychometric test theory as stochastic measurement models: Representation, uniqueness, meaningfulness, identifiability, and testability. *Methodika*, **3**, 25-60.

- Steyer, R. (1992). **Theorie kausale; Regressionsmodelle**. Stuttgart: Gustav Fischer Verlag.
- Steyer, R. & Eid, M. (1993). **Messen und Testen**. Berlin: Springer.
- Steyer, R. & Schmitt, M.J. (1990a). The effects of aggregation across and within occasions on consistency, specificity, and reliability. **Methodika**, **4**, 58-94.
- Steyer, R. & Schmitt, M. J. (1990b). Latent state-trait models in attitude research. **Quality and Quantity**, **24**, 427-445.
- Steyer, R. & Schmitt, M.J. (1993a). An introduction to latent state-trait theory. In R. Steyer, H. Gräser, & K.F. Widaman (Eds.), **Consistency and specificity: Latent state-trait models in Differential Psychology** (pp. 1-19). New York: Springer.
- Steyer, R. & Schmitt, M. J. (1993b). Models of latent state-trait theory. In R. Steyer, H. Gräser, & K. E. Widaman (Eds.), **Consistency and specificity: Latent state-trait models in Differential Psychology** (pp. 21-50). New York: Springer.
- Steyer, R., Wender, K. E & Widaman, K.F. (Eds.). (1992). Proceedings of the 7th European Psychometric Meeting. Stuttgart: Gustav Fischer Verlag.
- Suppes, P. (1970). **A probabilistic theory of causality**. Amsterdam: North-Holland.
- Suppes, P. (1981). Scientific causal talk: A reply to Martin. **Theory and Decision**, **13**, 363-380.
- Suppes, P. & Zinnes, J.L. (1963). Basic measurement theory. In: R. D. Luce, R. R. Bush & E. Galanter (Eds.), **Handbook of Mathematical Psychology. Vol. Z** (pp. 3-76). New York: Wiley.
- Tack, W. H. (1980). Zur Theorie psychometrischer Verfahren. Formalisierung der Erfassung von Situationsabhängigkeit und Veränderung. **Zeitschrift für Differentielle und Diagnostische Psychologie**, **1**, 87-108.
- Thurstone, L. L. (1931). **The reliability and validity of tests**. Ann Arbor, MI: Edwards Brothers.
- Torgerson, W.S. (1986). Scaling and Psychometrika: Spatial and alternative representations of similarity data. **Psychometrika**, **51**, 57-63.
- Tutz, G. (1989). **Latent Trait-Modelle für ordinale Beobachtungen**. Berlin: Springer.
- Viertl, R. (1990). **Einführung in die Stochastik**. Berlin: Springer.
- Westermann, R. (1987). **Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie**. Berlin: Springer.
- Westmeyer, H. (Hrsg.) (1989). Psychological theories from a structuralist point of view. New York: Springer.
- Westmeyer, H. (Hrsg.) (1992). **The structuralist program in psychology: Foundations and applications**. Toronto: Hogrefe-Huber.
- Wermuth, N. (1989). Moderating effects in multivariate normal distributions. **Methodika**, **3**, 74-93.
- Wottawa, H. (1980). Grundriß **der Testtheorie**. München: Juventa Verlag.
- Zimmerman, D. W. (1975). Probability spaces, Hilbert spaces, and the axioms of test theory. **Psychometrika**, **40**, 395-412.
- Zimmerman, D. W. & Williams, R. H. (1977). The theory of test validity and correlated errors of measurement. **Journal of Mathematical Psychology**, **16**, 135-152.

## Statistische Entscheidung

***Axel Ostmann und Joachim Wutke***

### 1. **Einleitung**

In der empirischen und experimentellen Psychologie werden in der Regel nach der Datenerhebung und der Datenaufbereitung spezielle Verfahren angewandt, um bezüglich der formulierten statistischen Hypothesen zu statistisch gestützten Entscheidungen zu gelangen. Dabei gilt unverändert, was bereits Bredenkamp (1972, S. 9) feststellte: „Der Signifikanztest ist das bei empirisch arbeitenden Psychologen beliebteste und am häufigsten verwendete Verfahren der beurteilenden Statistik.“ Vielen Anwendern gilt der Signifikanztest, von Bredenkamp (1972, S.159) als „eine Regel, mit der statistische Hypothesen angenommen oder abgelehnt werden“ interpretiert, als unproblematisches Standardverfahren. Dabei wird gern übersehen, daß der Signifikanztest nicht das einzige Verfahren zur Unterstützung statistischer Entscheidungen ist, „daß es ‚den‘ Signifikanztest nicht gibt“ (Hager & Westermann, 1983, S.71) und daß jedes statistische Entscheidungsverfahren mit einer Reihe erheblicher Probleme zu rechnen hat. So bemerkt etwa Stegmüller (1973, S. 1): „Es besteht bis zum heutigen Tag eine ungeheure Kluft zwischen logischen und wissenschaftstheoretischen Analysen von Begriffen der Prüfung, der Bestätigung und der Bewährung von Hypothesen auf der einen Seite, und von Fachleuten im Gebiet der mathematischen Statistik angestellten Untersuchungen über diese Themenkreise auf der anderen Seite.“ Wenngleich wir auf die begrifflichen und wissenschaftstheoretischen Probleme der statistischen Bewährung von Hypothesen nur am Rande eingehen und auf die Arbeiten von Hacking (1965), Lenzen (1974) und Stegmüller (1973) verweisen, wollen wir hier eine Reihe von klassischen Versuchen vorstellen, welche die Kluft zwischen statistischen Entscheidungen und konkreten statistischen Verfahren zu schließen versuchen.

Im Kontext der Einführung statistischer Grundbegriffe (Abschnitt 2) soll auf Probleme hingewiesen werden, die es bezüglich des Zusammenhangs zwischen empirischen Daten und Hypothesen gibt und die sich bezüglich der Begrün-

dung der Wahl eines geeigneten statistischen Modells zur Entscheidungshilfe ergeben. Nach der Vorstellung der ursprünglichen Idee des Signifikanztests von R.A. Fisher (Abschnitt 3) wird das am häufigsten angewandte Entscheidungsverfahren, das Hypothesentesten nach Neyman und Pearson dargestellt (Abschnitt 4). In einigen psychologischen Lehrbüchern zur Statistik wird unter dem Namen „Signifikanztest“ ein Verfahren eingeführt, welches man, wie wir später sehen werden, eigentlich als „Testen von Nullhypothesen nach Neyman und Pearson“ bezeichnen müßte (Abschnitt 5). Wir werden auch dafür eintreten, nicht vom „statistischen Schließen“ zu sprechen; mittels eines statistischen Modells schließt man nicht in der Art eines logischen Schlusses, man entscheidet sich. Dieser Tatbestand ist den in Abschnitt 6 vorgestellten Bayesianern allerdings geläufig, denn der sich auf das Bayessche Theorem stützende Ansatz betont insbesondere die Vorläufigkeit aller statistischen Entscheidungen.

Mehrfach bereits gab es „Signifikanztest-Kontroversen“ (siehe etwa Morrison und Henkel, 1970; Witte, 1989), und weil es in diesem immerwährenden Streit zwischen Signifikanzlern und Bayesianern oftmals um weltanschauliche Positionen geht, erscheint uns eine Rückbesinnung auf die problemgeschichtliche Ausgangslage der unterschiedlichen statistischen Entscheidungsverfahren wichtiger als die Vorstellung spezieller Techniken.

Aus wissenschaftstheoretischer Sicht ist es das Ziel der experimentellen Forschung in der Psychologie, „gesetzesförmige Aussagen von hypothetischem Charakter zu formulieren“ (Herrmann 1979, S. 17). Der hypothetische, das heißt vorläufige Charakter der gesetzesförmigen Aussagen in der Psychologie resultiert auch aus der Verwendung statistischer Entscheidungsverfahren, mit denen man zu einer Bewertung der aus den wissenschaftlichen Hypothesen abgeleiteten statistischen Hypothesen kommt. Statistische Hypothesen sind weder verifizierbar noch falsifizierbar, eben deshalb muß man sich - stets nur vorläufig - entscheiden. In der Testtheorie nach Neyman und Pearson gibt es deshalb bei der Entscheidung bezüglich statistischer Hypothesen zwei mögliche Gefahren: Die irrtümliche Annahme einer eigentlich falschen Hypothese und die irrtümliche Verwerfung einer eigentlich richtigen Hypothese - sowohl Annahme wie Verwerfung sind mit der Möglichkeit eines Irrtums behaftet, eine zweifelsfreie Absicherung gegen beide Irrtumsmöglichkeiten gibt es nicht. Für statistische Hypothesen gilt (leider) nicht das von Popper (1966, S. 13) für wissenschaftliche Theorien geforderte Abgrenzungskriterium zur Metaphysik, das Kriterium der Falsifizierbarkeit. „Ein empirisch wissenschaftliches System muß an der Erfahrung scheitern können.“ Als Gegner der Induktion hielt Popper den Schluß von den durch die Erfahrung bestätigten besonderen Beobachtungsaussagen auf die Theorie für logisch unzulässig. Als Vertreter einer deduktionistischen Methodologie fordern etwa Erdfelder und Bredenkamp (siehe Kapitel 2) eine implikative Beziehung zwischen psychologischer

und statistischer Hypothese; letztere soll aus ersterer deduziert werden können. Da Deduktionen jedoch nicht gehaltserweiternd sind, bleibt die Frage, an welcher Stelle die zum Testen statistischer Hypothesen notwendigen probabilistischen Annahmen eingeführt werden; am konsequentesten wäre wohl die Verwendung probabilistischer Theorien und stochastischer Modelle.

Wer statistische Entscheidungen trifft, sollte sich darüber im klaren sein, daß er sich einer probabilistischen und/oder entscheidungstheoretischen Sicht des Problems der Bewährung von Hypothesen angeschlossen hat. Er muß das Problem wissenschaftlicher Gewißheit bescheidener betrachten, etwa so, wie Lakatos (1974, S. 93) es bezüglich des Übergangs von einem wissenschaftlichen Determinismus Newtonscher oder Kantscher Provenienz zu einem „neuzeitlichen Probabilismus“ formuliert hat: „Der Probabilismus wurde von einer Reihe von Philosophen entwickelt, die der Ansicht waren, daß wissenschaftliche Theorien, obwohl gleichermaßen unbeweisbar, dennoch verschiedene Grade von Wahrscheinlichkeit in bezug auf die vorhandene Evidenz besäßen. Wissenschaftliche Redlichkeit verlangt demnach weniger, als man gedacht hat: Sie besteht darin, daß man nur hochwahrscheinliche Theorien vorbringt oder, noch bescheidener, daß man für jede wissenschaftliche Theorie die Erfahrungsdaten und die Wahrscheinlichkeit der Theorie im Lichte dieser Daten spezifiziert.“ (Es ist anzumerken, daß der von Lakatos verwendete Begriff der „Wahrscheinlichkeit der Theorie“ mit dem mathematischen Wahrscheinlichkeitsbegriff wenig gemein hat). Eine Einführung in die Geschichte des Probabilismus und seine Bedeutung für die moderne Wissenschaft geben Gigerenzer et al. (1991).

Diese Arbeit möchte dem Anwender statistischer Entscheidungsverfahren aufzeigen, welche impliziten und expliziten Annahmen mit jedem konkreten Verfahren verbunden sind, angefangen von Annahmen über den „state of the world“, mathematischen Vorannahmen und Annahmen über die Rolle von Hypothesen im Forschungsprozeß. Wir wollen zeigen, daß statistische Entscheidungen kein „triviales Standardritual“ am Ende einer Datenerhebung sein sollten. Ob es uns allerdings gelingt, solch hartnäckigen Unsinn aus der Welt zu schaffen wie etwa den, mittels einer „Signifikanz“ sei eine Hypothese zu beweisen, ist zweifelhaft - zu viele Versuche vor uns sind bereits gescheitert. Obwohl wir über statistische Entscheidungsverfahren berichten, weisen wir an einigen Stellen darauf hin, daß aus der Konfrontation einer zunächst wissenschaftlichen und dann (ja nicht notwendig) statistischen Hypothese mit den Daten nicht unbedingt die Anwendung eines statistischen Entscheidungsverfahrens resultieren muß; man könnte sich mit anderen, etwa deskriptiven Techniken begnügen. Sorgfältige Analyse klug aggregierter Rohdaten kann mehr Erkenntnisse bringen als ein möglicherweise belangloses signifikantes Ergebnis. Auch muß man sich ja nicht gleich entscheiden; man könnte sich auch zu einer „Urteilsenthaltung (als) adäquateste Reaktion“ (Stegmüller 1973,



S. 3) entschließen. Vielleicht lassen eine Vergrößerung der empirischen Indizienbasis oder eine verbesserte Datenerhebung später eine angemessenere Entscheidung zu.

## 2. Statistische Grundbegriffe

Im Kontext statistischer Entscheidungen, mißverständlicherweise oft auch Inferenzstatistik genannt, sollen Aussagen getroffen werden, mittels derer Charakteristika der erhobenen empirischen Daten in spezieller Weise verallgemeinert werden sollen. Die - kurz so genannte - Statistik bedient sich dabei eines probabilistischen Kalküls. Sind gewisse methodologische Vorannahmen und wahrscheinlichkeitstheoretische Oberhypothesen bezüglich der Herkunft der Daten gemacht, so kann ein empirisches Datum  $x$  als Realisation einer Zufallsvariable  $X$  interpretiert werden. **Zufallsvariable** sind in der Regel reellwertige Funktionen auf einem Wahrscheinlichkeitsraum  $\Omega$ , den wir im folgenden allerdings unspezifiziert lassen wollen (vgl. Steyer, Kap. 15). Der Wahrscheinlichkeitsraum soll uns dazu dienen, das Eintreten eines Datums  $x$  mit dem Ziehen eines speziellen Zufalls  $\omega$  zu unterlegen. Dieser Zufall führt bei der Prozedur, dem Einzelexperiment  $X$  im Sinne der Erhebung dieses einen Datums, eben zu  $X(\omega) = x$ . Der Begriff „Experiment“ wird hier technisch gebraucht und verstanden als „Ziehen eines Datums im Rahmen eines speziellen Zufallsmodells“; die in der empirischen Psychologie übliche Forschungsmethode ist hier nicht gemeint. Der Hinweis, daß mit der Anwendung des Kalküls unterstellt wird, daß Daten durch einen wie auch immer gearteten Zufall generiert sind, ist uns so wichtig, daß wir an vielen Stellen und in den meisten Formeln dieses Abschnittes zur Erinnerung  $\omega$  mitnotieren. Wie wir später noch begründen werden, hegen wir den Verdacht, daß aus der Unterschlagung dieser zentralen Idee der Zufallsvariablen manche Fehlinterpretation statistischer Aussagen folgen kann. Die Idee, Daten auch als vom Zufall abhängig zu konzipieren, hat etwa zur Konsequenz, auch den Ausgang eines statistischen Tests als vom Zufall abhängig anzunehmen.

Zufallsvariablen werden charakterisiert durch ihre Verteilung, die man als **Verteilungsfunktion**  $F_x$  mit  $F_x(x) = p(X \leq x) = p(\{\omega; X(\omega) \leq x\})$  angibt. Wenn möglich, wird die Verteilungsfunktion im stetigen Fall über ihre Dichte und im diskreten Fall über ihre Wahrscheinlichkeitsfunktion angegeben, die jeweils mit  $f_x$  bezeichnet wird. In der psychologischen Forschung wird im stetigen Fall oft die Annahme gemacht, ein Datum sei durch eine normalverteilte Zufallsvariable erzeugt, im diskreten Fall nimmt man öfters eine binomialverteilte oder hypergeometrisch verteilte Zufallsvariable an. Innerhalb des probabilistischen Kalküls werden Aussagen darüber, „woher die Daten kommen“, durch eine Spezifikation der Verteilung von  $X$  gemacht. Diese Spezifikation

besteht in der Regel in der exakten Angabe einer ganz speziellen Verteilung. Man sagt dann, das Datum  $x$  ist aus einer Population mit eben dieser Verteilung gezogen. Zwei Populationen sind gleich, wenn die entsprechenden Zufallsvariablen identisch verteilt sind, man schreibt dann  $X \sim Y$ .

Im Normalfall interessiert nicht nur ein einzelnes Datum, üblicherweise werden in einer empirischen Untersuchung nacheinander mehrere Daten  $x_1, \dots, x_n$  gewonnen. In der Regel wird unterstellt, daß alle Daten unabhängige Realisierungen derselben Zufallsvariablen  $X$  sind; man kann sie als Realisierung einer **Stichprobe**  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$  ansehen, also ist  $\mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)) = (x_1, \dots, x_n)$ . Dabei ist eine Stichprobe vom Umfang  $n$  ein  $n$ -Vektor von unabhängigen und identisch verteilten Zufallsvariablen (also  $X_i \sim X$ ,  $i \in \{1, \dots, n\}$ ). Einen konkreten Datensatz kann man als Vektor  $x = (x_1, \dots, x_n)$  im sogenannten **Stichprobenraum** darstellen.

Ist die Verteilung von  $X$  bekannt, so läßt sich die Verteilung einer zugehörigen Stichprobe  $X$  errechnen. Normalerweise ist die Ausgangsverteilung, die Verteilung von  $X$ , jedoch nicht bekannt. Ist wenigstens bekannt oder wird zumindest als Oberhypothese vorausgesetzt, daß die Verteilung der Zufallsvariablen  $X$  Element einer parametrisierten Menge von Verteilungen ist, so spricht man von einem parametrisierten Wahrscheinlichkeitsmodell für  $X$ . **Parameter** sind Kennzahlen von Verteilungen. Diese Vorannahme eines parametrisierten Wahrscheinlichkeitsmodells ist zumeist notwendig oder zumindest hilfreich, um trotz Unkenntnis der genauen Verteilung weitere Aussagen treffen zu können. In einem solchen Wahrscheinlichkeitsmodell werden spezielle Verteilungen durch die Angabe eines oder mehrerer Parameter identifiziert, die Verteilung von  $X$  also etwa durch einen entsprechenden Parametervektor  $\theta = (\theta_i)_{i \in I}$ . Eine Menge von Verteilungen heißt  $n$ -parametrisch, falls man zur Identifizierung  $n$  Parameter benötigt. Die konkrete Wahl der Parameter weist meist eine gewisse Willkür auf oder folgt Konventionen; in der Psychologie wird als typisches Wahrscheinlichkeitsmodell oft angenommen,  $X$  sei normalverteilt mit (unbekanntem) Erwartungswert  $\mu$  und (unbekannter) Varianz  $\sigma^2$ , oder  $X$  sei binomialverteilt mit (bekanntem)  $N$  und (unbekanntem)  $p$ . In beiden Verteilungsfamilien lassen sich auch spezifische Verteilungen durch Angabe zweier Zahlen identifizieren, nur müssen dies eben nicht notwendig Erwartungswert und Varianz sein; zur Identifizierung einer Normalverteilung etwa könnte man auch zwei beliebige, aber verschiedene Quartile oder Centile verwenden, beides Kennwerte, die in der Vergangenheit in der Intelligenzforschung durchaus verwendet wurden.

Bei Vorliegen eines solchen **Wahrscheinlichkeitsmodells** besteht eine naheliegende Standardaufgabe darin, aufgrund von Daten zu beurteilen, welcher spezifische Parametervektor  $\theta = (\theta_i)_{i \in I}$  zur Beschreibung von  $X$  adäquat ist, zudem sollte spezifiziert werden, mit welchen Unsicherheiten oder möglichen

Fehlern ein solches Urteil verbunden ist. Bei einer solchen Aufgabe, die sich zudem meist nur mit einer kleineren Anzahl von Parametern beschäftigt (in der Psychologie werden häufig nur die ersten zwei Momente, gelegentlich die ersten vier, betrachtet), spricht man von einer Aufgabe der **parametrischen Statistik**.

Die Frage nach einer zu den Daten passenden Beschreibung der Zufallsvariablen  $X$  stellt sich natürlich in ähnlicher Weise, wenn kein spezifiziertes Wahrscheinlichkeitsmodell vorliegt, wenn z.B. infolge mangelnder Skalenqualität eine Verteilung aus der Menge gängiger parametrisierter Verteilungen nicht zur Verfügung steht. Hier hätte eine mögliche Parametrisierung die Komplexität der Angabe der gesamten Verteilungsfunktion und nicht nur einzelner Parameter. In diesen Fällen spricht man von einer Aufgabe der **nichtparametrischen Statistik**.

**Statistische Hypothesen** sind Spezifikationen der Verteilung einer Zufallsvariablen. Sie spezifizieren die Verteilung entweder vollständig oder (meistens) teilweise, etwa durch Angabe der Parameter der Verteilung. Erwartungswert, Varianz, Median oder Interquartilsabstand sind hierfür bekannte Beispiele. In einem parametrischen Wahrscheinlichkeitsmodell ist eine statistische Hypothese deshalb identifizierbar mit einer Teilmenge  $0_0$  der Menge  $\Theta$  aller möglichen Parametervektoren  $\theta = (\theta_i)_{i \in I}$  des entsprechenden Modells. Eine oft benutzte Annahme ist von der Form: „ $E(X) = \mu$  und  $X$  normalverteilt“. Die Normalverteilungsannahme kann als Wahrscheinlichkeitsmodell oder **Oberhypothese** aufgefaßt werden. Aus der zweiparametrischen Hypothesenmenge  $\Theta$  mit Erwartungswert und Varianz als Parameter kann die eben formulierte Hypothese mit der Menge  $0_0 = \{(l, @, ), e, \sim \mathbb{R}^+\}$  identifiziert werden. Unter der Annahme der Gültigkeit dieser Hypothese kann man dann auch Aussagen wie etwa

$$p(\omega; X(\omega) \in C) < 0,05 \quad (2.1)$$

erhalten, wobei  $C$  eine bestimmte Menge im Stichprobenraum ist. Eine solche Aussage besagt, daß für das gegebene Wahrscheinlichkeitsmodell die Wahrscheinlichkeit dafür, Stichprobendaten aus dem Bereich  $C$  zu erhalten, für alle Verteilungen mit dem Erwartungswert  $\mu$  kleiner als 0,05 ist. Diese Aussage hat eine für die üblichen statistischen Verfahren typische Form, es werden Aussagen über die Wahrscheinlichkeit von Daten gemacht unter der Annahme der Richtigkeit bestimmter Oberhypothesen.

Wir gehen im folgenden davon aus, daß die Wahl des speziellen Wahrscheinlichkeitsmodells begründet und expliziert wird. Dieser Schritt der Wahl einer Oberhypothese wird bedauerlicherweise von vielen Anwendern als vollkommen unproblematisch angesehen, in einer Vielzahl von Fällen wird konven-

tionell auf ein Normalverteilungsmodell zurückgegriffen. Dies kann angemessen sein, muß es aber nicht.

Wir gehen ferner davon aus, daß in geeigneter, auf das gewählte Wahrscheinlichkeitsmodell bezugnehmenderweise Daten  $x = (x_1, \dots, x_n)$  erhoben werden, die sich als Realisation einer Stichprobe der Größe  $n$  auffassen lassen. Es ist danach die Frage, wie die erhobenen Daten verwendet werden können, um etwas über die Verteilung der Zufallsvariablen und ihrer Parameter zu erfahren. Schon in der deskriptiven Statistik werden Datensätze  $x$  in geeigneter Weise komprimiert und verrechnet, arithmetisches Mittel, Median, Range, (empirische) Varianz und viele andere Kennziffern sind hierfür Beispiele. Zu jeder Kennziffer gehört eine Rechenvorschrift oder Funktion  $t = t_n: \mathbb{R}^n \rightarrow \mathbb{R}$ , die den erhobenen Daten  $x$  eine Kennziffer  $t(x)$  zuweist. Mit einer solchen Funktion versucht man, die in den Daten vorhandene Information in geeigneter Weise aufzuschließen. Verknüpft man nun das oben vorgestellte Wahrscheinlichkeitsmodell mit einer solchen Funktion  $t$ , so wird der Wert von  $t$  abhängig vom Zufall: es entsteht eine sogenannte Stichprobenfunktion oder Statistik  $T = T_n$ , also

$$T(\omega) = T_n(\omega) = t(X_1(\omega), \dots, X_n(\omega)) \quad (2.2)$$

Statistiken sind damit ebenfalls Zufallsvariablen und werden wie diese durch ihre Verteilung charakterisiert. Spezifiziert eine Hypothese die Verteilung von  $X$  vollständig, so läßt sich auch die Verteilung einer Statistik  $T_n$  vollständig bestimmen. Generell gilt, daß Verteilungsspezifikationen für  $X$  mit Verteilungsspezifikationen für  $T_n$  einhergehen. Ist die Verteilung von  $X$  spezifiziert, so können Ausdrücke wie  $p(T_n \leq t)$  im Prinzip errechnet werden. Ist eine Hypothese  $H$  gegeben, so werden auch Ausdrücke wie  $p(T_n \leq t \mid H)$  verwandt; diese Schreibweise wird insbesondere von bayesianisch orientierten Statistikern präferiert. Das Zeichen  $\mid$  wird dann nicht wie üblich im strengen Sinne zur Notation einer bedingten Wahrscheinlichkeit verwandt. Eine bedingte Wahrscheinlichkeit wurde ein entsprechendes Wahrscheinlichkeitsmodell voraussetzen, etwa eines, in dem nicht nur  $X$ , sondern auch die Hypothese  $H$  als Zufallsvariablen vorkommen, und für die zudem eine gemeinsame Verteilung existiert. Wir wollen hier jedoch das Zeichen  $\mid$  - etwas ungenau, aber hinreichend präzisierbar - als „unter der Annahme der Gültigkeit beziehungsweise Richtigkeit von“ lesen, wie es die meisten Bayesianer auch tun.

Im Kontext statistischer Entscheidungen werden Statistiken verwendet, um Parameter zu schätzen und Hypothesen zu testen. Dabei wird allerdings oftmals übersehen, daß die Auswahl einer geeigneten Statistik durchaus ein Problem sein kann. Die Statistiklehrbücher machen am Beispiel der Stichprobenvarianz darauf aufmerksam, daß die Übertragung der Rechenvorschrift, mit der man einen Parameter einer Verteilung bestimmt, in eine Rechenvorschrift

für eine Statistik keineswegs ein Verfahren darstellt, das automatisch einen „guten“ **Schätzer** erzeugt. Schon Anfänger lernen, daß man bei erwünschter Erwartungstreue in der Formel für die Stichprobenvarianz durch  $(n-1)$  dividiert. Die Probleme entstehen dadurch, daß jede Statistik, da sie eine Zufallsvariable ist, ihrer Verteilung nach beurteilt werden muß. Bezüglich der Beantwortung der Frage, was nun eine „gute“ Statistik ist, was ein „guter“ Schätzer ist und was eine „gute“ Testgröße ist, werden üblicherweise eine Reihe von Kriterien herangezogen, von denen die wichtigsten auf R.A. Fisher zurückgehen.

Statistiken sollen der Reduktion von Daten dienen. Eine erste naheliegende Forderung an eine Statistik ist demnach, daß bei dieser Reduktion von Daten in eine Statistik möglichst kein Informationsverlust auftreten soll. Im idealen Fall ließen sich die Daten vollständig aus der Statistik rekonstruieren. Diese Eigenschaft „kein Informationsverlust“ wird im Begriff der Suffizienz einer Statistik präzisiert, Fisher hielt dieses Kriterium für zentral. Ist ein parametrisiertes Wahrscheinlichkeitsmodell gegeben, so nennt man eine Statistik  $T$  **suffizient**, falls die unter den Werten  $t$  der Statistik bedingten Dichten oder Wahrscheinlichkeiten  $f_x(x \mid T = t)$  jeweils für alle  $\theta \in \Theta$  übereinstimmen. Diese bedingten Größen werden dann bereits durch die Partition des Stichprobenraums in die Bereiche  $\{x; T = t\}$  der Daten gleichen Wertes  $t$  der Statistik induziert (vgl. etwa Lindgren 1976, S. 224f.). Deshalb kann man in entsprechender Weise suffiziente Partitionen des Stichprobenraums definieren. Es läßt sich dann zeigen, daß die gesamte aus den Daten verfügbare Information, die zur Identifikation von  $\theta$  verwendet werden kann, in der Angabe der speziellen Menge der Partition oder - anders ausgedrückt - in der Angabe  $T = t$  enthalten ist.

Uns ist es an dieser Stelle wichtig festzustellen, daß durch den Übergang vom Datensatz zum Wert einer Statistik die Daten reduziert werden und daß deshalb die Wahl einer geeigneten Statistik begründet werden sollte. Es sollte insbesondere gezeigt werden, daß die gewählte Statistik keinen oder einen geringeren Informationsverlust aufweist als andere, daß die gewählte Statistik also die suffizienteste ist.

Verwendet man die Statistik  $T$  als Schätzer für den Parameter  $\theta$ , so wird diese **Schätzung** bezüglich einer Reihe von Eigenschaften beurteilt. Eine Statistik  $T$  bzw.  $T_n$  (mit  $n$  als Stichprobengröße) heißt:

- **erwartungstreu** oder unverfälscht (unbiased) falls  $E(T) = \theta$ . Ein bekanntes Beispiel, einen Schätzer erwartungstreu zu machen, ist die Korrektur des Nenners bei der Schätzung der Stichprobenvarianz,
- **asymptotisch erwartungstreu** falls  $\lim_n E(T_n) = \theta$ ,

- **konsistent** falls  $\lim_n p(|T_n - \theta| > \epsilon) = 0$  für alle  $\epsilon$ . Die relative Häufigkeit ist ein bekanntes Beispiel für einen konsistenten Schätzer für den Parameter  $p$ ,
- **effizient**, falls  $T$  eine erwartungstreue Statistik ist und unter allen anderen erwartungstreuen die geringste Varianz hat. Schon Anfänger lernen, daß bezüglich des ersten Moments das arithmetische Mittel eine geringere Varianz hat als der Median und demgemäß „effizienter“ ist.

Nicht in jedem Wahrscheinlichkeitsmodell stehen für jeden Parameter Schätzer mit allen wünschenswerten Eigenschaften zur Verfügung. So gibt es etwa bei der Familie der hypergeometrischen Verteilungen nicht für jeden Parameter erwartungstreue Schätzer (vgl. etwa Hartung 1985, S.207). In jedem Anwendungsfall sollte überlegt werden, welche Güteeigenschaften eines Schätzers notwendig und wünschenswert sind.

Nun wollte man nicht nur wissen, wie „gut“ ein Schätzer ist, man wollte auch wissen, wie „genau“ die aus den Daten erhaltene Schätzung eines Parameters ist. Die Präzisierung und Lösung dieser Aufgabe ohne bayesianische Methoden (dort ist der Parameter eine Zufallsvariable) gelang Neyman (1935). Seine Modifikation der Schätz Aufgabe besteht darin, mittels zweier Statistiken  $A$  und  $B$  aus den Daten  $(X_1, X_2, \dots, X_n)(\omega) = (x_1, x_2, \dots, x_n)$  Intervallgrenzen  $a = A(\omega)$  und  $b = B(\omega)$  zu bestimmen, innerhalb derer der (nichtstochastische, feste aber unbekannte) Parameter  $\theta$  „wahrscheinlich“ liegt; das heißt: Das Paar  $(A, B)$  soll die Eigenschaft haben, daß  $p(A \leq \theta \leq B) = 1 - \alpha$  mit kleinem  $\alpha$ . Die Realisierung  $[a, b]$  heißt dann das aufgrund der Daten und der Schätzer  $A$  und  $B$  bestimmte **Konfidenzintervall** zum vorgegebenen Niveau  $1 - \alpha$ . Das Verfahren zur Bestimmung von Konfidenzintervallen weist große Ähnlichkeit mit dem im nächsten Abschnitt vorgestellten Signifikanztest auf, dem Komplement des Konfidenzintervalls hier entspricht dort der kritische Bereich, in dem  $H_0$  verworfen werden soll, dem Komplement  $\alpha$  des (Konfidenz-)Niveaus  $1 - \alpha$  hier entspricht dort das Signifikanzniveau  $\alpha$ ;  $\alpha$  steht also jedesmal für die maximal tolerierte Irrtumswahrscheinlichkeit.

Verwendet man eine Statistik  $T$  zum Testen, so soll eine Entscheidung darüber getroffen werden, welche von mehreren Handlungsmöglichkeiten aufgrund des Wahrscheinlichkeitsmodells und aufgrund der aus den Daten verfügbaren Informationen als adäquat angesehen werden kann. Ein Test besteht (aus vereinfachter heutiger Sicht) aus einer Statistik  $T$  und einer Entscheidungsregel  $d: \mathbb{R} \rightarrow A$ , wobei  $A$  die Menge der Handlungsalternativen darstellt. Aus Fisherscher Sicht wurde sich etwa der bekannte t-Test darstellen lassen als t-Verteilung, abhängig von den Freiheitsgraden, und als Entscheidungsregel, unter welchen Bedingungen die Nullhypothese beibehalten wird oder nicht beibehalten („verworfen“) wird.

Wird nun  $T = t$  realisiert, so legt die **Entscheidungsregel**  $d(t)$  fest, was ein Entscheider „tun soll“, welches Element  $a$  aus  $A$  er wählen soll. Der Modaloperator „soll“ weist darauf hin, daß ein statistischer Entscheider gehalten ist, sich aus Gründen wissenschaftlicher Redlichkeit an die konventionell vereinbarten Entscheidungsregeln zu halten. Die Elemente  $a$  von  $A$  werden meist als Entscheidung für eine Teilmenge  $0_a$  der Parametermenge  $\Theta$  des Wahrscheinlichkeitsraumes interpretiert. Da die Teilmenge  $0_a$  eine Hypothese  $H$  darstellt, kann man die Handlungsalternative  $a$  in diesem Falle mit der Hypothese  $H = 0_a$  identifizieren. Die Menge der Handlungsalternativen  $A$  kann aber auch weitere, nicht durch Parametermengen identifizierbare Elemente enthalten, etwa ein Element  $u$  für die Alternative „keine Entscheidung treffen“ oder Elemente für „Modell präzisieren“ oder „weitere Daten erheben“; die letzte Alternative ist etwa in der Entscheidungsregel für sequentielle Testverfahren enthalten.

üblicherweise wird unterschieden zwischen einem

- komplementären Alternativtest mit  $A = \{H_0, H_1\}$ , wobei  $H_1$  als Komplement von  $H_0$  im Parameterraum definiert ist, und einem
- Entscheidbarkeitstest mit  $A = \{H_0, u, H_1\}$ , wobei  $H_0$  und  $H_1$  sich hier in der Regel nicht zum ganzen Parameterraum ergänzen; deshalb verbleibt hier das Element  $u$  für „keine Entscheidung“ in der Menge der Handlungsalternativen  $A$ .

**Hypothesen**, die nur aus einem Parameter bestehen, heißen einfache Hypothesen oder Punkthypothesen, Ein gebräuchlicher Anwendungsfall in der Psychologie ist ein komplementärer Alternativtest mit der Punkthypothese  $H_0 = (0.)$  (auch „zweiseitiger Test“ genannt) oder der unspezifischen Bereichshypothese  $H_0 = \{\theta \in \mathbb{R}, \theta \leq 0\}$  („einseitiger Test“). Sind alle Hypothesen einfach, so heißt der Test einfach. Man spricht vom „einfachen Alternativtest“ für  $A = \{H_0, H_1\}$  mit  $H_i = \{e_i\}$ ,  $i = 0, 1$ , und vom „einfachen Entscheidbarkeitstest“ für  $A = \{H_0, u, H_1\}$ , mit  $H_i = \{e_i\}$ . Einfache Tests haben für parametrische Wahrscheinlichkeitsmodelle den Charakter der Überprüfung einer **Effektgröße**; diese wird ausgedrückt in der Differenz  $\mu_0 - \mu_1$ . Wie wir später sehen werden, eröffnet der einfache Entscheidbarkeitstest zusätzlich die Möglichkeit einer Erweiterung zu einem sequentiellen Test.

Oft hat man aus inhaltlichen Erwägungen Grund, eine Hypothese, sie heiße Nullhypothese oder einfach  $H_0$ , besonders auszuzeichnen. Der Bereich  $K = \{t; d(t) \neq H_0\}$  führt zum Verwerfen von  $H_0$  und heißt **kritischer Bereich**; entsprechend heißt  $C = \{x; d(t(x)) \neq H_0\}$  kritischer Bereich im Stichprobenraum. Ein Alternativtest  $(T, d, A)$  kann damit auch durch  $(T, K)$  oder  $C$  charakterisiert werden.

Halten wir fest: Ein Test, aufgefaßt als Prozedur des statistischen Entscheidens, ist charakterisiert durch

- die Teststatistik  $T$  sowie die Verteilung von  $T$ ,
- die Menge der Handlungsalternativen  $A$ ,
- die Entscheidungsregel  $d$ .

Im Sinne dieser allgemeinen Formulierung kann man das Schätzen eines Parameters  $\theta$  auch als Spezialfall des Testens auffassen:  $A$  ist im Falle des Schätzens die Menge aller möglicher Parameterwerte bzw. die entsprechende feinste Partition dieser Menge.

Da die Verteilungen vieler Statistiken oftmals nur aufwendig exakt zu berechnen sind, verwenden Anwender und Statistiker seit Beginn der Entwicklung statistischer Entscheidungsverfahren durch Karl Pearson gern in Tabellenform zugängliche Verteilungen oder weichen auf leichter zu handhabende Grenzverteilungen aus, falls diese gute Näherungen darstellen, etwa weil die Stichprobe einen großen Umfang hat. Die  $\chi^2$ -Verteilung bei Fragen des Goodness-of-fit oder für die Verteilung von Varianzen folgt oft diesem pragmatischen Rationale. Die Entscheidungsregel wird oft so modifiziert, daß bestimmte Grenzverteilungen verwendet werden können, etwa indem man dem Test  $(T, A, d)$  eine Statistik hinzufügt, die aus Transformation von  $T$  hervorgeht und die eben diese in Tabellen oder in der verwendeten Software verfügbare Grenzverteilung hat. Dazu kommen Überlegungen, wann eine aufgrund der Grenzverteilung getroffene Entscheidung eine hinreichend gute Näherung darstellt. In der Praxis wird diese Frage oft nicht gesondert untersucht; meistens greifen Anwender hierbei auf „Faustregeln“ zurück, die in Lehrbüchern zu finden sind. In der Regel geben solche Faustregeln an, ab welcher Stichprobengröße eine befriedigende Näherung als erreicht betrachtet werden kann.

Wie alle Entscheidungen können auch statistische Entscheidungen falsch sein. Da die Teststatistik eine Zufallsvariable ist, kann man die Wahrscheinlichkeit von Fehlentscheidungen im Prinzip berechnen. Als **a-Fehler** oder Fehler der 1. Art bezeichnet man jene Entscheidung, bei der  $H_0$  nach der Entscheidungsregel des Tests verworfen wird, obwohl  $H_0$  eigentlich zutrifft. Die Wahrscheinlichkeit für dieses Ereignis, oder genauer, die Wahrscheinlichkeit für das Verwerfen von  $H_0$  bei Vorliegen einer Verteilung aus  $H_0$  ist zumindest kleiner gleich

$$\alpha = \sup_{\theta \in H_0} p_{\theta}(\{\omega; T(\omega) \in K\}) \quad (2.3)$$

Es ist das Verdienst von Neyman und Pearson, in kritischer Abgrenzung zu Fisher darauf aufmerksam gemacht zu haben, daß nicht nur das irrtümliche Verwerfen, sondern auch die irrtümliche Annahme von  $H_0$  einen Fehler darstellt. Dieser Fehler wird als **B-Fehler** oder Fehler 2. Art bezeichnet. Wissenschaftler sollten sich bemühen, auch für diesen Fehler eine Obergrenze ein-



zuhalten (vergleiche Kapitel 14 dieses Bandes). Ausgehend von dieser Idee zweier möglicher Fehler haben Neyman und Pearson (1933) Gütekriterien vorgeschlagen. Als Hilfsmittel zur Formulierung dieser Kriterien haben sie (für statistische Tests) zunächst eine Gütefunktion (power function) definiert, die wir im folgenden für den Alternativtest vorstellen.

Die **Gütefunktion**  $\pi$  (eigentlich  $\pi_n$ ) auf einer parametrisierten Verteilungsfamilie gibt an, wie groß die Wahrscheinlichkeit ist, bei Vorliegen einer mit  $\theta$  identifizierten Verteilung und der Stichprobengröße  $n$  gegen die Hypothese  $H_0$  zu entscheiden; anders ausgedrückt: es wird die Größe der kritischen Region angegeben, gemessen in Wahrscheinlichkeit. Für den Fehler 1. Art gilt dann

$$\alpha = \sup_{\theta \in H_0} \pi(\theta) \tag{2.4}$$

Das Vorliegen eines  $\beta$ -Fehlers oder Fehlers 2.Art hat dann höchstens die Wahrscheinlichkeit

$$\beta = \max_{\theta \in H_1} 1 - \pi(\theta) \tag{2.5}$$

Verwendet man einen komplementären Alternativtest, so ist bei Stetigkeit von  $\pi$ , die im allgemeinen vorausgesetzt werden kann,  $\beta = 1 - \alpha$ . Die Größen  $\alpha$  und  $\beta$  ergänzen sich hier jeweils zu 1; wird der eine Fehler kleiner, so wird der andere Fehler größer. Wer beide Fehler kontrollieren möchte, sollte entweder keine komplementären Alternativtests benutzen oder müßte sich mit hohen Fehlerwahrscheinlichkeiten zufrieden geben, nur wurde eine Befund-sicherung etwa mit  $\alpha = \beta = 0,50$  wahrscheinlich wenig Akzeptanz erlangen.

Mittels der Gütefunktion  $\pi = r_n$  lassen sich für einen Test folgende **Güteeigenschaften** definieren; ein Test heißt

- zum Niveau  $\alpha$ , wenn  $\pi(\theta) \leq \alpha$  für alle  $\theta \in H_0$  gilt,
- unverfälscht (unbiased), wenn  $\pi(\theta) \geq \alpha$  für alle  $\theta \in H_1$  gilt,
- konsistent, wenn  $\lim_{n \rightarrow \infty} \pi(\theta) = 1$  für alle  $\theta \in H_1$  gilt.

Bei der Definition eines Tests werden alle Hypothesen gleichbehandelt, die Zuordnung der Namen  $H_0$  und  $H_1$  ist eigentlich willkürlich, wenngleich sich in der Psychologie gewisse Konventionen durchgesetzt haben; so wird als  $H_0$  oft jene Punkthypothese bezeichnet, die einen „Nulleffekt“ postuliert, gelegentlich wird jene Hypothese, die man gern zurückweisen (to nullify) wurde, als  $H_0$  bezeichnet. Leider wird die folgenreiche Entscheidung, welche Hypothese als  $H_0$  betrachtet werden soll, nur selten begründet (siehe etwa Neyman 1950, S. 262f.). Nach Festlegung der Hypothesen wird bezüglich der Eigenschaft „zum Niveau  $\alpha$ “ die Obergrenze für einen tolerierbaren Fehler der

1. Art festgelegt. Wenn, wie in der Praxis meistens üblich, „zum Niveau  $\alpha$ “ klein gewählt wird, ist der kritische Bereich klein und die Wahrscheinlichkeit für die Entscheidung zugunsten von  $H_0$  groß, weil  $1 - \alpha$  groß ist. Viele Anwender bewerten die Kontrolle des Fehlers 2. Art als nachrangig oder sogar völlig verzichtbar, andere hingegen betonen ihre Wichtigkeit. So fordert etwa Bredenkamp (1980, S.22): „ $\alpha$ ,  $\beta$  und die Effektgröße müssen vor dem Experiment festgelegt worden sein, so daß auch der benötigte Stichprobenumfang ... geschätzt werden kann.“ Dieser Forderung wird in der Praxis nur selten entsprochen. Ein möglicher Grund für den Verzicht auf die Kontrolle des  $\beta$ -Fehlers kann allerdings auch sein, daß das irrtümliche Festhalten an  $H_0$  im Forschungskontext als unerheblich bewertet wird: Man hat etwas übersehen, aber das macht nichts.

Üblicherweise wird die Obergrenze des  $\alpha$ -Fehlers, das sogenannte **Signifikanzniveau**, niedrig festgelegt. Wenn der maximale Fehler 1. Art festgelegt ist, soll nun auch der  $\beta$ -Fehler kontrolliert werden, d.h. man strebt an, den  $\beta$ -Fehler so klein wie möglich zu halten. Dies geschieht durch Auswahl eines sogenannten **besten Tests**. Tests mit selben Niveau  $\alpha$  lassen sich nämlich bezüglich ihrer Güte vergleichen. Es sei  $H_0 = \{e_0\}$ . Die Tests  $u$  und  $v$  haben die Gütefunktion  $r_u$  bzw.  $r_v$ . Dann heißt Test  $u$  gleichmäßig besser als Test  $v$ , falls  $r_u(\theta) \geq r_v(\theta)$  für alle  $\theta \in H_1$ . Wenn  $H_1$  eine Punkthypothese ist, so gibt das Neyman-Pearson-Lemma Bedingungen für die Existenz eines besten Tests zum vorgegebenen Niveau  $\alpha$  (vgl. Lindgren 1976, S.299-302). Dem Anwender begegnet das Problem des besten Tests zum Beispiel, wenn entschieden werden soll, ob zur statistischen Sicherung von Mittelwertsunterschieden der  $t$ -Test oder der  $U$ -Test angewandt werden soll.

Die Güteeigenschaft der Unverfälschtheit eines Tests verlangt, daß der  $\alpha$ -Fehler keinesfalls größer wird als der  $\beta$ -Fehler. Diese Eigenschaft hat zum Beispiel der einseitige Gaußtest, sie ist aber keineswegs trivial, wie wir weiter unten (Seite 34) an einem Beispiel von Neyman und Pearson sehen können, die dort im Hypothesenraum statt des Mittelwertes (wie beim Gaußtest) die Varianz variieren. Auch der zweiseitige Kolmogoroff-Smirnov-Test ist verfälscht (siehe etwa Büning & Trenkler, 1978, S. 90). Die Summe  $\alpha + \beta$  beider Fehler ist bei Unverfälschtheit kleiner gleich 1. Der beste denkbare Test würde natürlich beide Fehler ganz vermeiden, wurde also gleichzeitig  $\alpha = 0$  und  $\beta = 0$  erfüllen. Dieser ideale Test existiert natürlich in der Praxis nicht, jedoch kann man fordern, daß ein Test sich mit wachsendem Stichprobenumfang diesem Idealfall stetig annähert. Diese Eigenschaft ist die obengenannte Konsistenz.

Anhand zweier Probleme aus der Geschichte der Mathematik, die mit den Namen Arbuthnot und Bayes verbunden sind, soll ein grundlegender Unterschied zwischen den bayesianischen und den nicht-bayesianischen Zugängen zur Statistik aufgezeigt werden. Entscheidend ist dabei die unterschiedliche

Richtung des statistischen Denkens: Einmal wird dabei ein Bezug von den Hypothesen zu den Daten hergestellt (gelegentlich wird dies „statistischer Schluß“ genannt), das andere Mal wird ein Rückbezug von den Daten zu den Hypothesen vorgenommen (gelegentlich „statistischer Rückschluß“ genannt). Bei der Darstellung beider Probleme wird eine Zufallsvariable  $X$  betrachtet, die nur zweier Werte fähig ist: 0 und 1. Mit  $p_0$  sei  $p(X = 0)$  bezeichnet.

Arbuthnot stellte sich 1711 die Aufgabe zu zeigen, daß die Wahrscheinlichkeit einer Knabengeburt nicht  $1/2$  ist. Er will das Ergebnis seiner Untersuchung für ideologische Zwecke nutzen, nämlich für einen Gottesbeweis; aber das soll hier nicht interessieren. In unserer Schreibweise sei  $X_i(\omega) = 0$ , falls die  $i$ -te Beobachtung eine Knabengeburt ist, und  $X_i(\omega) = 1$ , falls sie eine Mädchengeburt ist. Arbuthnot stellt die Hypothese  $p_0 = 1/2$  auf, allerdings nur um zu zeigen, daß diese als extrem unwahrscheinlich abzulehnen ist. Als Datensatz benutzt Arbuthnot Geburtsregister, die sich über einen Zeitraum von 82 Jahren erstrecken. Als Statistik wird die Anzahl der Jahre verwendet, in dem innerhalb dieses Zeitraums ein Knabenüberschuß vorliegt. Nun wurden in allen 82 Jahren mehr Knaben als Mädchen geboren, und die Statistik  $T$  nimmt einen solch extremen Wert an, daß Arbuthnot seine Vermutung gestützt glaubt, eben daß die Wahrscheinlichkeit einer Knabengeburt nicht gleich  $1/2$  ist.

Drei Aspekte erscheinen aus heutiger Sicht an diesem Lösungsansatz unbefriedigend. Zum einen scheint die Auswahl der Statistik abhängig gewesen zu sein vom gewünschten Resultat. Zum zweiten liefert Arbuthnot keine von den Daten unabhängige Begründung für seine Verwerfungsregel. Und schließlich unterzieht er nicht die Hypothese einer strengen Prüfung, an die er eigentlich glaubt, sondern er prüft die Hypothese, die er gern verwerfen möchte. Eine ausführliche Darstellung dieses frühen statistischen Entscheidungsversuchs findet sich bei Hacking (1966) sowie bei Freudenthal und Steiner (1966, S. 180).

Das zweite historische Beispiel geht zurück auf Bayes und ist posthum 1764 veröffentlicht worden. Bayes hat sein Problem weitgehend befriedigend gelöst. Hier beschreibt  $X$  den Ausgang eines Wurfexperiments.  $X_i() = 0$  bedeute hier, daß der  $i$ -te Wurf eines Gegenstandes in den linken Teil eines zweigeteilten Zielfeldes fällt, und  $X_i() = 1$  bedeute, daß der  $i$ -te Wurf in den rechten Teil des Zieles fällt. Das Ziel kann man sich als Rechteck vorstellen, die Teilung als vertikale Linie, die das Rechteck in zwei Teile schneidet. Nun sei jedoch unbekannt, wo diese Teilung liegt, wo also das linke Feld aufhört und wo das rechte Feld anfängt. Bayes nimmt nun an, daß die Teilung des Zielfeldes in „links“ und „rechts“ ebenfalls durch ein Zufallsexperiment zustande genommen ist, formal also durch eine Zufallsvariable  $Y$ . Wenn das der Fall ist und wir  $p_0 = Y()$  setzen, betrachten wir auch den Parameter  $p_0$  der Ver-

teilung der Zufallsvariablen  $X$  als Realisierung einer Zufallsvariablen, nämlich  $Y$ . Es ist nun nach dem **Bayesschen Theorem** möglich, aus den Daten  $(x_1, \dots, x_n)$  (den Würfeln) und der Verteilung von  $Y$  (bestimmt durch die Grenze zwischen links und rechts) auf die Verteilung von  $X$  zu schließen. Es ist nämlich mit  $x = (x_1, \dots, x_n)$

$$f_Y(p_0) | X = \frac{f_Y(p_0) f_X(x | p_0)}{\int f_Y(\theta) f_X(x | \theta) d\theta} \quad (2.6)$$

Da  $Y = p_0$  als Hypothese über  $X$  aufgefaßt werden kann, läßt sich diese Formel auch wie folgt verkürzt schreiben (- für proportional):

$$\begin{aligned} p(\text{Hypothese} | \text{Daten}) &\sim p(\text{Hypothese}) p(\text{Daten} | \text{Hypothese}) \\ \text{oder nochmals verkürzt als} \\ p(H | D) &\sim p(H) p(D | H) \end{aligned} \quad (2.7)$$

Der Ausdruck  $p(H | D)$  ist bei gegebenen Daten  $x$  eine bis auf einen konstanten Faktor gegebene Funktion, die als **Likelihoodfunktion** bezeichnet wird. Für verschiedene Realisierungen der Stichprobe ist  $p(H | D)$  eine mehrdimensionale Statistik, denn sie weist nach Normierung jedem Zufall  $\omega$  eine Funktion zu.

Zum Zwecke statistischer Entscheidungen kann man die Formel des Bayesschen Theorems wie folgt interpretieren: die Verteilung von  $Y$ , angebar durch die Werte der Wahrscheinlichkeitsdichte für Hypothesen der Form  $Y = p_0$ , wird berücksichtigt durch eine zunächst beliebig angenommene Wahrscheinlichkeit  $p(H)$ . Der Ausdruck  $p(D | H)$  steht für die Wahrscheinlichkeit der Daten unter der Annahme der Gültigkeit der Hypothese und  $p(H | D)$  ist die nach dem Vorliegen der Daten erhaltene „Rückschluß“-Wahrscheinlichkeit, die aufgrund der Daten revidierte Wahrscheinlichkeit der Hypothese.

Damit haben wir die beiden statistischen „Schließ“-Richtungen vorliegen: Während Arbuthnot die Wahrscheinlichkeit der Daten unter seiner Hypothese  $p(D | H)$  betrachtet hat (und wegen ihrer Unwahrscheinlichkeit auch seine Hypothese verworfen hat), untersucht Bayes die Wahrscheinlichkeit der Hypothese unter gegebenen Daten,  $p(H | D)$ . In diesem Konzept wird die Wahrscheinlichkeit als Stützmaß für die Gültigkeit von Hypothesen interpretiert. Zumindest auf den ersten Blick unbefriedigend am Bayesschen Ansatz ist, daß die volle Spezifikation der Verteilung von  $Y$  verlangt wird, was zudem voraussetzt, daß man die Verteilung von  $X$  als durch einen Zufallsprozeß zustande gekommen aufzufassen bereit ist. In der Folgezeit wurde vor allem die oft verwendete Annahme der Gleichwahrscheinlichkeit für  $Y$  problematisiert. Dies war einer der Gründe, warum der Bayessche Ansatz für lange Zeit in Mißkredit geriet, denn wenn der Wertebereich von  $Y$  die volle Zahlengerade

(kontinuierlich und unbeschränkt) ist, so ist wegen mangelnder Integrierbarkeit in klassischer Weise eine Gleichwahrscheinlichkeit überhaupt nicht definierbar.

### 3. *Der Signifikanztest nach Fisher*

Auf den englischen Statistiker R.A. Fisher gehen bedeutende Neuentwicklungen der mathematischen Statistik zurück wie etwa die eben kurz skizzierte Schätztheorie für Parameter oder die Varianzanalyse. Für uns von besonderem Interesse ist der **Signifikanztest**, dessen Rationale bei heutigen statistischen Entscheidungen häufig Verwendung findet. Am Beispiel der Lösung eines sehr alten praktischen Problems kann man aufzeigen, daß wesentliche Momente des statistischen Entscheidens auch schon vor Fishers Präzisierung des Signifikanztests intuitiv plausibel waren. Stigler (1977) beschreibt die Prozedur „the trial of the pyx“, mittels derer seit dem 12. Jahrhundert in der königlich britischen Münze täglich nach einem Zufallsprinzip einige Münzen aus der laufenden Serie entnommen und in einer besonderen Büchse (pyx) gesammelt wurden. In einem Vertrag zwischen der Krone und der Münze waren Toleranzen für das mittlere Gewicht einer Münze vereinbart. Stigler vergleicht diese Toleranzen mit der uns heute vertrauten **2σ-Regel**. Bei Kontrollen in unregelmäßigen Abständen wurde der Inhalt der Büchse gewogen, da nun aber die zulässige Toleranz für eine einzelne Münze mit  $n$ , der Anzahl der Münzen in der Büchse, und nicht etwa mit III;- multipliziert wurde, soll es in knapp 800 Jahren nur zu einer einzigen Strafe wegen Unterschreitung des vereinbarten Gewichts gekommen sein.

Als erste moderne statistische Testprozedur, die von praktischen Anwendungsfällen unabhängig konzipiert war, gilt der von Karl Pearson im Jahr 1900 entwickelte „**goodness-of-fit-test**“. Pearson betrachtete, anders als seine Vorgänger, nicht nur einzelne Abweichungen zwischen einer theoretisch erwarteten und einer empirisch beobachteten Häufigkeitsverteilung, auch nahm er nicht die Summe der Beträge der Abweichungen, vielmehr schlug er die Teststatistik

$$\chi^2_{\text{beob}} = S \frac{(m' - m)^2}{m'} \tag{3.1}$$

als Maß für die Beurteilung der Abweichung beider Verteilungen vor (hier in der Originalschreibweise von Pearson mit  $m'$  als theoretischer und  $m$  als beobachteter Häufigkeit in einem Intervall und  $S$  als Summation über alle Intervalle; vgl. Kendall & Stuart, 1973, S.436). Sowohl die Berechnung der  $\chi^2$ -Verteilung als auch die Tabellierung der Wahrscheinlichkeit von  $p(\chi^2 < \chi^2_{\text{beob}})$  für bis zu 20 Intervalle und bis zu 15 Werten von  $\chi^2$  durch Pearson waren

förderlich für die rasche Verbreitung dieses Verfahrens, denn die Arbeit der Anwender wurde durch diese Hilfen ungemein erleichtert. Die Testprozedur ging von der Annahme aus, daß bei einem „guten Fit“ die Unterschiede zwischen der theoretischen Verteilung - die hier gleichsam den Status einer Nullhypothese annimmt - und den empirisch erhobenen Daten gering ist, daß also  $\chi^2_{\text{beob}}$  klein ist und demgemäß  $p$  groß. Wenn nunmehr  $\chi^2_{\text{beob}}$  sehr groß ist und deshalb  $p$  sehr klein, so ist unter der Annahme der Angemessenheit der theoretischen Verteilung ein sehr seltenes Ereignis eingetreten. Aber ist nicht auch ein extrem kleiner Wert von  $\chi^2_{\text{beob}}$  im Rahmen der üblichen Meßfehlertheorie kaum zu erklären oder ebenfalls ein sehr seltenes Ereignis? Schon Pearson hat den Gedanken verfolgt, daß auch „zu gute“ Daten Zweifel erlauben.

In seinem Vorläufer eines Testmodells hat Pearson jedoch weder so etwas wie ein festes Signifikanzniveau noch eine klare Entscheidungsregel vorgegeben, die Entscheidung, ob Empirie und Theorie hinreichend gut übereinstimmen, wurde in das Ermessen des einzelnen Forschers gelegt. Allerdings wollte Pearson die Entscheidung erleichtern durch die Angabe von „odds“, von Wettquotienten gegen den Zufall. Wenn etwa aus der Konfrontation beider Häufigkeitsverteilungen ein  $p = 0,043$  resultierte, so wurde gesagt, die „odds“ dafür, daß die Daten von der theoretischen Verteilung abweichen, sind 23:1.

Die entscheidende Weiterentwicklung dieser Gedanken erfolgte durch Fisher in den zwanziger Jahren (siehe Fisher 1951, 1959) mit dem Signifikanztest. Sein Verfahren wird in Abgrenzung zur Testtheorie von Neyman und Pearson gelegentlich auch „Nullhypothesen-Testen“ genannt (letztgenannter Egon S. Pearson war ein Sohn von Karl Pearson; die Beschäftigung mit Grundfragen des statistischen Entscheidens war also Familientradition). Das Rationale des Vorschlags von Fisher läßt sich gut demonstrieren an einem Beispiel, am „problem of the lady tasting tea“ (Fisher, 1951, S. 11; Neyman, 1950, S. 272-94; siehe auch Hagen & Seifert, 1979, S. 190), gleichzeitig lassen sich einige kritische Punkte aufzeigen.

Die Geschichte geht so: Eine ungenannte Lady behauptet, sie hätte ein spezielles geschmackliches Unterscheidungsvermögen, sie könne nämlich sicher herauschmecken, ob in eine Tasse

- (a) zuerst die Milch und dann der Tee oder
- (b) zuerst der Tee und dann die Milch

eingefüllt worden ist. Fisher, hier in der Rolle eines skeptischen Geschmacksphysiologen, bezweifelt dieses Vermögen und formuliert die folgende Nullhypothese  $H_0$ : die Lady besitzt das von ihr behauptete Unterscheidungsvermögen nicht. Diese Nullhypothese  $H_0$  gilt es zu testen. Hierzu schlägt Fisher eine spezielle Versuchsanordnung vor: es werden jeweils vier Tassen nach bei-

den Modalitäten (a) und (b) hergestellt; alle 8 Tassen werden dann der Lady in zufälliger Reihenfolge zur Geschmacksprobe vorgesetzt. Die Lady, über diese Versuchsanordnung informiert, soll nun die 8 Tassen in zwei Teilmengen zu je vier Tassen aufteilen, wobei die eine Teilmenge nur solche nach (a) hergestellte und die andere nur solche nach (b) hergestellte Tassen enthalten darf.

Bezüglich der Frage, bei welchem empirischen Ergebnis nun ein skeptischer Wissenschaftler seine Nullhypothese verwirft (und dann vielleicht der Lady glaubt), schlägt Fisher seinen Signifikanztest vor. Dies in einer Art und Weise, die durchaus der Position von Popper in der Beschreibung von Lakatos (1974, S. 90) vergleichbar ist: „Die intellektuelle Redlichkeit besteht nicht darin, daß man versucht, seine Position fest zu verankern oder sie durch Beweis (oder ‚wahrscheinlich machen‘) zu begründen - die intellektuelle Redlichkeit besteht vielmehr darin, daß man jene Bedingungen genau festlegt, unter denen man gewillt ist, die eigene Position aufzugeben.“ Fisher setzt das in die folgenden Forderungen um:

- (1) Man stelle (zusätzlich zu Vorannahmen, die als bewährt gelten können, nur) eine einzige Hypothese auf, bei deren Verwerfung in der Regel bisheriges Wissen zur Revision anstünde: die **Nullhypothese**  $H_0$ .
- (2) Man bestimme die unter der Annahme der Gültigkeit von  $H_0$  für diese spezielle Versuchsanordnung resultierende Zufallsvariable  $X$  und ihre Verteilung. In unserem Beispiel resultiert für die beiden Modalitäten zur Einordnung der Teetassen in (a) und (b) eine Zufallsvariable  $X$  mit den Ausprägungen  $x = 0, 1, 2, 3$  bzw. 4 richtigen Zuordnungen. Wir bezeichnen die Werte  $x$  der Zufallsvariablen  $X$  „Anzahl richtiger Zuordnungen“ im folgenden als „Treffer“ sowie die Wahrscheinlichkeiten von  $X = x$  mit  $p_x$ . Die Wahrscheinlichkeitsverteilung unter der Annahme der Gültigkeit von  $H_0$  ist eine hypergeometrische, die im folgenden tabelliert ist.

Tabelle 3.1: Lady’s tea dilemma, 1.Version

Treffer	0	1	2	3	4
$P_x$	$\frac{1}{70}$	$\frac{16}{70}$	$\frac{36}{70}$	$\frac{16}{70}$	$\frac{1}{70}$

- (3) Man lege einen **kritischen Bereich** fest, der diejenigen Realisationen der Zufallsvariablen enthält, die nicht mit der Nullhypothese kompatibel erscheinen. Hier etwa könnte jemand vorschlagen,  $H_0$  sei zu verwerfen, wenn die Lady vier richtige Zuordnungen je Modalität getroffen hat. Nach Zerlegung des Ereignisraums in zwei disjunkte Teilmengen ergäbe sich dann folgende Verwerfungsregel:
  - Verwerfe  $H_0$ , wenn die Lady jeweils vier richtige Zuordnungen getroffen hat.

Nur wenn die Lady keinen Fehler macht, wird die Nullhypothese verworfen. Dann ist unter der Annahme der Gültigkeit von  $H_0$  ein extremes Ereignis eingetreten, welches eine bedeutsame Abweichung („significant discrepancy“) von  $H_0$  anzeigt. Nur unter der Bedingung, daß sich von ihm kontrolliert und unter seinen Augen etwas extrem Unwahrscheinliches ereignet hat, ist der skeptische Wissenschaftler bereit, seine Hypothese aufzugeben. Mit den ironischen Worten von Neyman (1952, S.43) könnte man es auch so ausdrücken: Wenn  $H_0$  zutrifft, so ist das, was der Wissenschaftler im Falle der Abweichung beobachtet, ein Wunder. Da wir heutzutage aber nicht mehr an Wunder glauben, entschließen wir uns, lieber nicht mehr an die Nullhypothese zu glauben. Doch zurück zum Beispiel; selbst bei Anwendung obiger Verwerfungsregel wäre sich ein skeptischer Wissenschaftler nicht sicher, ob die Verwerfung von  $H_0$  zweifelsfrei richtig wäre, wenn die Lady jeweils vier Treffer gehabt hätte. Denn dies Ereignis ist unter der Annahme der Gültigkeit von  $H_0$  nicht unmöglich, mit einer Wahrscheinlichkeit von  $p = 1/70 = 0,0143$  ist es eben nur sehr unwahrscheinlich.

- (4) Man wähle ein **Signifikanzniveau**. Fisher hat keinen eindeutigen und rational begründeten Vorschlag gemacht, wie genau der durch die Verwerfungsregel festgelegte kritische Bereich bestimmt werden soll, allerdings hat er vorgeschlagen, fairerweise vor einer Untersuchung festzulegen, welche Wahrscheinlichkeit eines Fehlers maximal toleriert werden kann. Diese Wahrscheinlichkeit bestimmt die Größe des kritischen Bereichs, und legt man diesen an den Rand der Zufallsverteilung, ist damit eine Regel zur Verwerfung der  $H_0$  eindeutig formulierbar. Da Wissenschaftler bezüglich der Verwerfung ihrer Hypothese (von deren Richtigkeit sie eigentlich überzeugt waren) skeptischer sein sollten als Menschen im Alltag und sie deshalb ihre Nullhypothese länger beibehalten sollten (Wissenschaftler sind eben nicht so leicht zu überzeugen), schlägt Fisher als übliches Signifikanzniveau 5 % vor (1951, S. 13), in heutiger Schreibweise  $\alpha \leq 0,05$ . Wie Wendt (1966) in Untersuchungen zeigen konnte, haben Menschen im Alltag in vielen Situationen ein „subjektives Verlässlichkeitsniveau“ um 0,20. In der Sprache englischer Wettbüros sind wir im Alltag oft geneigt, uns schon dann für die Verwerfung einer gehegten Vermutung zu entscheiden, wenn die „odds“ 1:4 gegen einen Irrtum stehen, der Wissenschaftler entscheidet sich erst dann gegen seine  $H_0$ , wenn es 1:19 gegen einen Irrtum steht. Gelegentlich wird dieser Sachverhalt übertragen in das Sprachspiel, ein Wissenschaftler möchte sich höchstens einmal bei 20 Entscheidungen irren, aber diese Aussage ist genaugenommen falsch. In der Psychologie haben sich heute als Konventionen die Signifikanzniveaus  $\alpha \leq 0,05$ ,  $\alpha \leq 0,01$  und  $\alpha \leq 0,001$  eingebürgert.
- (5) Man verwerfe die Nullhypothese genau dann, wenn ein extremes Ereignis eine bedeutsame Abweichung von  $H_0$  anzeigt und wenn die Wahrschein-



lichkeit, die **Verwerfung von  $H_0$**  irrtümlich vorzunehmen, höchstens so groß ist wie das Signifikanzniveau.

Im Falle des „Lady’s tea dilemma“ wurde die Wahl eines Niveaus  $\alpha \leq 0,05$  bedeuten, daß der Lady im Versuch kein einziger Zuordnungsfehler unterlaufen dürfte. Man könnte nun aber argumentieren, daß die Lady, selbst wenn sie das von ihr behauptete Vermögen hätte, ja auch zufällig, gleichsam „aus Versehen“, im Experiment einen Fehler machen könnte. Die Entscheidungsregel des Wissenschaftlers läßt Fehler zu, da dürfte sich die Lady durchaus auch einmal irren. Wollte nun ein großzügiger Wissenschaftler der Lady einen „Zufallsfehler“ konzidieren, so hätte er zwei Handlungsmöglichkeiten. Zum einen könnte er seine Nullhypothese bereits verwerfen, wenn die Lady mindestens 3 richtige Zuordnungen trifft, er müßte in diesem Fall aber bereit sein, einen Fehler von  $p = 17/70 \approx 0,243$  zu tolerieren. Zum zweiten könnte er einen Vorschlag von Fisher aufgreifen und seinen Versuchsplan sensitivieren, indem im Experiment die Anzahl der Geschmacksproben auf 12, je zweimal 6, erhöht wird. Wie zuvor (vgl. Tab.3.1) ist die Wahrscheinlichkeitsverteilung eine hypergeometrische, jedoch mit anderen Parametern. Die folgende Tabelle zeigt die entsprechenden „Trefferwahrscheinlichkeiten“.

Tabelle 3.2: Lady’s tea dilemma, 2.Version

Treffer	0	1	2	3	4	5	6
$P_x$	$\frac{1}{924}$	$\frac{36}{924}$	$\frac{225}{924}$	$\frac{400}{924}$	$\frac{225}{924}$	$\frac{36}{924}$	$\frac{1}{924}$

In diesem von Fisher als „more sensitive“ charakterisierten Versuchsplan können quantitativ kleinere Abweichungen von der Nullhypothese entdeckt werden, und die Verwerfungsregel könnte in diesem Fall lauten:

- Verwerfe  $H_0$ , wenn die Lady fünf oder sechs richtige Zuordnungen trifft.

Die Wahrscheinlichkeit dafür, die Nullhypothese irrtümlich zu verwerfen, ist dann  $p_5 + p_6 = 37/924 \approx 0,04 < \alpha \leq 0,05$ . Bei Wahl des üblichen Signifikanzniveaus könnte man jetzt der Lady einen „Fehler“ konzidieren und dennoch  $H_0$  verwerfen.

Insgesamt erscheint Fishers Vorschlag zur Bewertung der Nullhypothese - einmal abgesehen von einer gewissen Willkürlichkeit bei der Wahl des Signifikanzniveaus und damit der Verwerfungsregel - schlüssig und sowohl dem skeptischen Wissenschaftler als auch der Lady gerecht zu werden. Wie aber ein Einwand von Neyman (1950, S. 272) zeigt, wirft dieses eigentlich triviale Beispiel an unvermuteter Stelle Probleme auf. Denn wenn man die experimentelle Versuchsanordnung zur Überprüfung der Fähigkeit der Lady etwas

modifiziert und ihr nach Zufall Teeproben serviert, die jeweils mit  $p(a) = p(b) = 0,50$  hergestellt wurden, so können die einzelnen Teeproben als Bernoulli-Versuche aufgefaßt werden, und das dem Versuch zugrunde liegende Zufallsmodell ist jetzt ein Urnenmodell „mit Zurücklegen“. Statt einer hypergeometrischen Verteilung resultiert nun eine Binomialverteilung als Zufallsverteilung unter  $H_0$ . In diesem Fall wurden in der ersten Versuchsordnung von Fisher vier richtige Zuordnungen der Lady zu einer Irrtumswahrscheinlichkeit von  $p = 1/16 = 0,0625$  führen, falls man aufgrund des empirischen Ergebnisses die Nullhypothesen verwerfen wurde (bei der hypergeometrischen Verteilung war  $p = 0,0143$ ). Oder anders herum gesagt: Bei vorgegebenen Signifikanzniveau von  $\alpha \leq 0,05$  wurden hier vier richtige Zuordnungen nicht zur Verwerfung von  $H_0$  führen.

Wie schon Morrison und Henkel (1970) aufgezeigt haben, finden sich in den üblichen Statistiklehrbüchern wenig Hinweise auf diese komplexen Zusammenhänge zwischen experimenteller Versuchsplanung, resultierenden Zufallsverteilungen und den letztendlichen statistischen Entscheidungen. Manche Autoren scheinen der Meinung zu sein, einzig die Wahl des Signifikanzniveaus könne eine Entscheidung bezüglich einer statistischen Hypothese beeinflussen.

Nun sollte am Beispiel von „lady’s tea dilemma“ nur das Grundprinzip der Signifikanztestung demonstriert werden; in die Praxis der Statistik eingegangen sind natürlich Fishers Neuentwicklungen wie etwa der F-Test, der z-Test oder der „exact-probability-test“ sowie seine Weiterentwicklungen des t-Tests oder des  $\chi^2$ -Tests. Dabei hat seine Bevorzugung von effizienten, erwartungstreuen und möglichst suffizienten Schätzern und seine Präferenz für aus der Normalverteilung abgeleiteten Zufallsverteilungen bis heute Wirkung in der Anwendung statistischer Entscheidungsverfahren. Allerdings hat der Signifikanztest nach Fisher nicht nur von den Bayesianern heftige Kritik erfahren, auch Neyman und Pearson waren überhaupt nicht mit dem Vorgehen von Fisher einverstanden. Von Neyman ist überliefert, daß er den Signifikanztest für „schlechter als nutzlos“ hielt. Vor allem zwei Punkte sollen erwähnt werden:

- (a) Kritik am Nullhypothesen-Testen: Fisher kennt nur eine Hypothese, demgemäß kennt er nur einen möglichen Fehler, den  $\alpha$ -Fehler, und er spricht nur davon, ab wann die Hypothese als widerlegt („disproved“) gilt. Aus den in der Einleitung genannten Gründen wurden wir nicht mehr von einer Widerlegung, sondern nur noch von einer Verwerfung dieser Hypothese sprechen. Die meisten statistischen Anwender testen heute (mindestens) zwei konkurrierende Hypothesen simultan und würden Stegmüller zustimmen in seiner Behauptung, „isolierte statistische Hypothesen können niemals einer adäquaten theoretischen Beurteilung

unterzogen werden. Wir werden versuchen, die These zu begründen, daß die Beurteilung einer statistischen Hypothese nur erfolgen kann in bezug auf eine Klasse von Alternativhypothesen, die mit der zur Diskussion stehenden Hypothese konkurrieren.“ (1973, S. 23) Wenngleich diese Kritik zutrifft, so muß man Fisher dennoch zugute halten, daß für ihn, der seine Statistik oft zur Suche neuer Anbaumethoden in der Landwirtschaft einsetzte, die Testung seiner Nullhypothese im Forschungsprozeß einen ganz speziellen Stellenwert besaß. Fisher hat Nullhypothesen formuliert, die aus dem aktuellen Wissensstand begründet waren. Im Falle der Lady hätte die Verwerfung von  $H_0$  chemische oder physiologische Forschungen ausgelöst, denn ein solches Unterscheidungsvermögen wäre mit bekannten Theorien nicht vereinbar gewesen. Fisher hatte nach der Verwerfung der Nullhypothese keinen „statistisch gesicherten Effekt“, sondern eine interessante inhaltliche Fragestellung, der er sich hätte widmen müssen. Anwender, die heute übliche Signifikanztests nicht zur Effektsicherung, sondern zur zufallskritischen Absicherung benutzen, handeln häufig in diesem Sinne. Hier wird der Test als „screening“ in einem frühen Stadium der Theoriebildung eingesetzt um abzuschätzen, was einer genaueren Untersuchung lohnt.

- (b) Defizite der rationalen Begründung: Fisher konzentrierte sich bei der Auswahl der von ihm betrachteten Parameter auf die besten Schätzer, etwa auf das arithmetische Mittel und die Varianz, nur ist nicht zu begründen, daß in jedem Anwendungsfall die besten Schätzer auch die besten Tester sind. Es sind durchaus Fälle denkbar, in denen statt des arithmetischen Mittels der Median oder der Modus oder statt der Varianz der Range oder eine Centildifferenz inhaltlich geeignetere Kennwerte wären. Unbestreitbar sind die Gütekriterien, die Fisher in seiner Schätztheorie entwickelt und benutzt hat, einleuchtend und überzeugend; die Wahl der Parameter und der Prüfstatistiken ist dadurch jedoch nicht eindeutig bestimmt. Zweifelsohne war die Auswahl leicht zu berechnender Parameter und die Tabellierung von Prüfverteilungen sehr „benutzerfreundlich“, nur birgt dies eben auch die Gefahr einer schematischen Anwendung. Auch wenn das vorgegebene Verfahren für den selbst zu entscheidenden Fall gar nicht so gut paßt, etwa wegen anderer Oberhypothesen oder wegen Fragestellungen, für die die gängigen Parameter gar nicht optimal sind, wird das Standardverfahren gerechnet, etwa angemessenere **nicht-parametrische Verfahren** werden meist nicht erwogen. Sowohl die Höhe des Signifikanzniveaus als auch die Platzierung des daraus resultierenden Verwerfungsbereichs an den Rand als auch die hälftige Aufteilung dieses Bereichs bei zweiseitiger Fragestellung sind letztlich konventionell, wenngleich nicht ohne Plausibilität. Obwohl Fisher selbst versucht hat, mit seiner „fiducial probability“ ein quantitatives Maß zur Stützung von Hypothesen zu entwickeln (siehe Hacking 1965, S. 133; Stegmüller 1973,

S.258), so bleibt festzuhalten, daß nach der Durchführung eines Signifikanztests nur ein **klassifikatorischer Stützungsbegriff** Anwendung finden kann: Die Entscheidungen können nur lauten „die  $H_0$  wird verworfen“ oder „die  $H_0$  wird nicht verworfen“.

#### **4. Die Testtheorie von Neyman und Pearson**

Mit den Namen Neyman und Pearson wird gerne eine dem Fisherschen Signifikanztest überlegene und dem Bayesianischen Denken entgegengesetzte Theorie vom statistischen Entscheiden verbunden. Diese Sicht ist zwar nicht ganz falsch, jedoch zumindest recht unvollständig. Neyman und Pearson nehmen in ihren Arbeiten oftmals explizit die Ansätze von Fisher auf, waren allerdings stets um mathematische Präzisierung und um Anreicherungen bemüht - wobei sie des öfteren in Gegensatz zu Fisher gerieten. Vor allem Fishers Schätztheorie und seine und anderer Arbeiten zu Stichprobenverteilungen und ihrer Momente war Anreiz zu weiterführenden Arbeiten. Mitte der zwanziger Jahre war erreicht, was Karl Pearson einst programmatisch verlangt hatte, die Statistik war nicht mehr auf normalverteilte oder binomialverteilte Populationen beschränkt. Aber mit der Ausarbeitung vieler neuer statistischer Techniken entstand das Problem, Kriterien dafür zu finden, welches Verfahren bei einer Fragestellung am angemessensten ist. Das ist deshalb keine nur akademische Frage für Theoretiker, weil unterschiedliche Techniken offensichtlich zu verschiedenen Ergebnissen und damit zu unterschiedlichen Entscheidungen führten - und diese Frage geht auch Praktiker an. Fisher etwa vertrat den Standpunkt, die Likelihood beim Schätzen zu verwenden, aber weder die Definition der Likelihood noch ihre theoretischen Eigenschaften noch darauf aufbauende Prinzipien wie „maximum likelihood“ wurden von ihm klar dargelegt. Während E. A. Pearson als Reaktion daraufhin vorschlug, Brüche zwischen Likelihoods zu bilden und das „likelihood-ratio“-Kriterium für Entscheidungen zu benutzen, war für Neyman Fishers Lob der Likelihood inkonsequent, weil Fisher gleichzeitig die „Methoden der inversen Wahrscheinlichkeit“ (die Bayesianischen Methoden der „rückschließenden“ Statistik) vehement ablehnte. Neyman und Pearson betrachteten auch nach dem Entwurf ihrer eigenen Testtheorie im Jahre 1926 immer wieder alternative Zugänge zum statistischen Entscheidungsproblem und versuchten, sie integrativ zu verarbeiten. Ihre Maxime „keep doors open“ galt nicht zuletzt auch dem Bayesianismus (Neyman & Pearson, 1928, 1933b).

Im folgenden werden in etwas modernerer Fassung die wesentlichen Aspekte der Neyman-Pearson-Testtheorie vorgestellt, wie sie hauptsächlich zwischen 1928 und 1933 entstand. Eine neuere und lesenswerte Darstellung findet sich bei Neyman (1950).

In ihrer Testtheorie präzisieren Neyman und Pearson (1926) den Begriff der **statistischen Hypothese** in einer Weise, die der statistisch-formalen Bearbeitung zugänglich ist. Sie unterscheiden zwischen einfachen Hypothesen (also  $H = \{\theta\}$ ) und zusammengesetzten Hypothesen und verlangen statt einer einzelnen (Fisherschen Null-) Hypothese die Angabe eines Alternativenraumes  $\Theta$ . Damit wird die Prüfung einer einzigen (denkbaren) Hypothese - in einer wohlbestimmten Theorie können ja nicht verschiedenen Hypothesen bezüglich des gleichen Sachverhalts gleichzeitig Gültigkeit beanspruchen - ersetzt durch eine Entscheidung zwischen alternativen Möglichkeiten; diese Entscheidung allerdings kann falsch sein. Konsequenterweise wird der Begriff des statistischen Schließens ersetzt zugunsten des Begriffs der statistischen Entscheidung. Die „Signifikanz“ der Daten im Sinne ihrer Unwahrscheinlichkeit unter der bisher für zutreffend gehaltenen Hypothese ist für Neyman und Pearson kein Entscheidungskriterium. Erstens ist bestenfalls nur eine der beiden Irrtumsmöglichkeiten (nämlich der Fehler 1. Art) über die Wahrscheinlichkeit ihres Auftretens kontrolliert, zweitens gibt es kein eindeutiges Kriterium, welcher Bereich mit Wahrscheinlichkeit zur Verwerfung von  $H_0$  führt, und drittens sind wahrscheinlichkeitstheoretische Aussagen über die „Signifikanz“ der Daten weder möglich noch sinnvoll, denn seltene Ereignisse geschehen ja auch - wie etwa Lottogewinne oder Erdbeben. Begründete Aussagen sind nur über die Güte der Entscheidungsregel möglich. Deshalb haben Neyman und Pearson die Begriffe des  $\alpha$ -Fehlers, des  $\beta$ -Fehlers sowie die **Teststärkefunktion**  $\pi$  präzisiert und definiert. Darauf aufbauend haben sie Gütekriterien für Tests zur Verfügung gestellt. Leider finden in der psychologischen Literatur Überlegungen bezüglich der **Teststärke** unverändert geringe Beachtung. So berichten Sedlmeier und Gigerenzer (1989), daß in nur 2 von 64 betrachteten Untersuchungen die Teststärke überhaupt erwähnt wurde, bestimmt wurde sie nie. Ferner berichten sie, daß sich seit der klassischen Arbeit von Cohen (1962) nur wenig verändert hat. Interpretiert man die Teststärke im Sinne der Wahrscheinlichkeit, ein signifikantes Ergebnis zu erhalten, wenn wirklich ein Effekt vorhanden ist, so ist die Teststärke für mittelgroße Effekte nach Schätzungen von Sedlmeier und Gigerenzer im Median von 0,46 im Jahre 1960 (Cohen) auf 0,37 im Jahre 1984 abgesackt.

Um eine Hypothese  $\theta_0 \in \Theta$  zu testen, formulieren Neyman und Pearson das Problem wie folgt um: Es gilt, im Stichprobenraum Konturen  $\Phi(\mathbf{x}) = \text{konstant}$  zu finden (mögliche Kandidaten sind etwa Iso-Likelihoodflächen), so daß

- 1. die Wahrscheinlichkeit unter  $H_0 = \theta_0$ , daß die Stichprobe  $\Sigma = X$  innerhalb dieser Konturen liegt, gleich  $\alpha$  ist, und
- 2. diese Wahrscheinlichkeit unter jedem anderen  $\theta \in \Theta$  maximal ist.

Die hiermit angestrebte Eigenschaft ist die **Unverfälschtheit** des Tests, die zusammen mit  $\Phi$  ein Kriterium für die Wahl der kritischen Region im Stich-

Probenraum liefert. So schreibt etwa Neyman in einem Brief an Pearson (Pearson 1966, S. 18), „using such contours and rejecting  $H_0$  when  $\Sigma$  is inside  $\Phi = \text{const}$ , we are sure that a true hypothesis is rejected with a frequency less than  $\alpha$ , and that if  $H_0$  is false and the true hypothesis is, say  $H'$ , then most often the observed sample will be inside  $\Phi = \text{const}$  and hence  $H_0$  will be rejected“.

In fast allen damals untersuchten Fällen konnten die Konturen der Likelihoodfunktion zur Lösung der obigen Aufgabe benutzt werden. Aber schon für die Hypothese  $H_0 = N(0;1)$ , das heißt für die Hypothese, daß die Daten aus einer normalverteilten Population mit Mittelwert 0 und Varianz 1 stammen, berichtet Neyman (Pearson 1966, S. 18), daß sich die Likelihoodkonturen nicht verwenden lassen, denn danach wurde  $N(0;1,1)$  öfter verworfen als  $H_0$ . Es war also zu klären, unter welchen Bedingungen sich die Likelihoodfunktion zur Konstruktion eines Tests verwenden läßt.

Für den einfachen Alternativtest  $\Theta = \{f_0, f_1\}$  ist sie nach dem Neyman-Pearson-Lemma sogar für die Konstruktion eines besten Tests verwendbar. In unserer Notation lautet das **Neyman-Pearson-Lemma** wie folgt: Der Test mit der kritischen Region  $C_k = \{x; f(x|e_0) = k f(x|e_1)\}$  ist bester Test, wobei  $k \geq 0$  so gewählt werden muß, daß das entsprechende erreicht wird. Oft läßt sich dann die kritische Region auch für eine andere, leichter berechenbare Statistik angeben, deren Verwendung man den Vorzug gibt.

Testet man eine einfache Hypothese gegen eine zusammengesetzte, so muß es einen uniform besten Test, also einen Test, der gegen jede einfache Alternative bester Test ist, nicht geben. Das gilt erst recht, wenn nicht nur  $H_1$ , sondern auch  $H_0$  zusammengesetzt ist. In Analogie zu der oben genannten kritischen Region kann man Tests mit der kritischen Region

$$C_k = \{x; \sup_{\theta_0 \in H_0} f(x|\theta_0) = k \sup_{\theta_1 \in H_1} f(x|\theta_1)\} \quad (4.1)$$

bilden (deren Eigenschaften jedoch im Einzelfall zu prüfen sind). Da  $L(\theta) = f(x|\theta)$  oder in anderer Notation  $L(\theta|x)$  die schon genannte Likelihoodfunktion darstellt, kann der Wert  $k$  als ein Maß der Angemessenheit alternativer Hypothesen aufgefaßt werden. Ein solcher Test heißt Likelihood-Quotienten-Test. Wir werden ihm in Abwandlung weiter unten wieder begegnen (S. 37).

Bei Neyman und Pearson bleibt die Bestimmung eines adäquaten  $\alpha$ -Wertes offen: Dieser kann nicht aus mathematischen Kriterien abgeleitet werden, sondern muß aus forschungslogischer und substanzwissenschaftlicher Sicht begründet werden. Nicht die Statistiker, die Fachwissenschaftler sind bezüglich dieser Frage in die Pflicht genommen. Manche Kritiker sind der Meinung,

daß eine solche Begründung für die Obergrenze eines  $\alpha$ -Fehlers bei Fragestellungen der Grundlagenforschung oft schwerfallen wurde, sie lehnen zumindest eine unrelativierte Verwendung von Neyman-Pearson-Tests ab. Andere plädieren für eine strenge Anwendung von Neyman-Pearson-Tests, so etwa in der Psychologie Bredenkamp (1972), der als Verfahren einen um die Angabe einer Effektgröße  $\delta$  modifizierten Entscheidbarkeitstest vorschlägt:

- (1) Der Experimentator spezifiziere als Abweichung von  $H_0$  einen Effekt, d.h. er legt eine für den Forschungszweck „praktisch bedeutsame“ Differenz  $\delta = \delta_1 - \delta_0$  fest; damit wird auch der Parameterbereich für eine Indifferentzone festgelegt.
- (2) Neben der Kontrolle des  $\alpha$ -Fehlers erachtet Bredenkamp die Kontrolle des  $\beta$ -Fehlers für ebenso wichtig. Deshalb wähle man ein kleines  $\beta$ ; wenn in der Population ein Effekt in der postulierten Größe vorliegt, wird man ihn mit Wahrscheinlichkeit  $(1-\beta)$  entdecken können; kleinere Abweichungen sind praktisch wenig bedeutsam.
- (3) Danach ergibt sich ein weiterer Vorteil; aus dem postulierten Effekt  $\delta$  und dem maximal tolerierten Fehler 2.Art kann der notwendige Mindestumfang der zu untersuchenden Stichprobe hergeleitet werden und damit der Aufwand für die Untersuchung abgeschätzt werden.

Natürlich ist die exakte Festlegung von  $\beta$  und  $\delta$  im gewissen Sinne willkürlich und die meisten psychologischen Theorien sind für solche genauen Effektprognosen nicht spezifiziert genug. Bredenkamp bemerkt aber zu recht, daß auch eine Nichtfestlegung eine implizite Festlegung ist, daß eine explizite Festlegung Kritik und Wiederholbarkeit erleichtert.

Als Integrationsversuch im Sinne von Neyman und Pearson schlägt Witte (1980, 1989, 1991) ein vierstufiges, streng hierarchisches Prüfverfahren vor, bei dem die jeweils nächste Stufe nur ausgeführt werden sollte, wenn die Prüfung bis dahin erfolgreich verlaufen ist:

- (1) Zur „Prüfung der Datenerhebungssituation“ schlägt Witte vor, zwei konkurrierende einfache Hypothesen  $\theta_0$  und  $\theta_1$  aufzustellen (es können auch präzise Intervalle sein, so Witte 1991), sowie „Fehlertoleranzen“  $\sim 1 - \beta = \alpha$  zu wählen (beides etwa  $\leq 0,05$ ). Mittels eines geeigneten Neyman-Pearson-Tests (etwa dem t-Test bei Mittelwertshypothesen) bestimme man die minimale Stichprobengröße  $n$ , bei der beide Fehlertoleranzen eingehalten werden können. Scheitert diese erste Prüfung, etwa am zu kleinen  $n$  der eigenen Stichprobe, so sollte nicht weiter geprüft werden, „dann sollte jedoch der beobachtete Effekt publiziert werden“.
- (2) Im zweiten Schritt der „Hypothesenprüfung“ wird die Bildung eines Likelihood-Quotienten vorgeschlagen. Wenn etwa

$$\frac{L(\theta_1 | x)}{L(\theta_2 | x)} \geq \Phi = \frac{1 - \beta}{\alpha} \quad (4.2)$$

so wurde  $\delta_1$  vorläufig akzeptiert. Kommt man hier zu keiner Entscheidung, „so ist die Theorie in den gewählten empirischen Bereichen nicht zur Erklärung oder Prognose heranzuziehen“ (Witte 1991).

- (3) Bei der „Hypothesenqualifikation“ soll geprüft werden, ob sich die Akzeptierung von  $\delta_1$  nicht allein darauf stützt, daß  $L(\theta_0 | x)$  so klein ist.  $L(w, |x)$  wird deshalb in Relation gesetzt zur maximalen

$$\frac{L(\theta_1 | x)}{\max L(\theta_i | x)} \geq Q_c = 1 - \sqrt{(1 - \beta)\alpha} \quad (4.3)$$

„Die Idee ist hierbei nicht nur, die Stützung der Hypothese [hier  $e_1$ ] von einer bestimmten anderen abhängig zu machen [hier  $S_0$ ], sondern von der empirisch plausibelsten unter Berücksichtigung von Fehlerschwankungen [hier  $\max L(\theta_i | x)$ ]“ (Witte 1989) (eckige Klammern hinzugefügt).

- (4) Zur letzten Prüfung der „Effektqualifikation“ wird der beobachtete Effekt in der Stichprobe daraufhin beurteilt, „ob er auf Datenebene (nicht im Wahrscheinlichkeitsmodell) von Bedeutung ist:  $\eta_c^- \geq 0,10$  mag eine denkbare Konvention sein“ (Witte 1989). Scheitert Prüfschritt 3, soll man weiter über die Parameter nachdenken, scheitert Schritt 4, ist die Theorie eventuell zu global formuliert.

Wittes Vorschlag zur Integration mehrerer statistischer Entscheidungsmodelle in ein einziges Verfahren ist von Diepgen (1991) kritisiert worden. Die Verwendung der Testtheorie von Neyman und Pearson nicht als Entscheidungsverfahren, sondern „nur“ zur Bestimmung des notwendigen Stichprobenumfangs hält er für unplausibel, die Deutung der (hypothetischen) Fehlerwahrscheinlichkeiten als „Fehlertoleranzen“ ist ihm suspekt. Ein gewichtiges Argument bringt Diepgen gegen Wittes Vorschlag, in Schritt 2 beim Likelihood-Quotienten eine Grenzgröße  $\Psi$  einzuführen und diese auch noch an  $\alpha$  und  $\beta$  zu koppeln. Diepgen vermutet hier wohl nicht zu Unrecht, daß die letztendliche Willkürlichkeit der Festsetzung einer Obergrenze für  $\alpha$  und  $\beta$  in der Neyman-Pearson-Testtheorie hier ersetzt wird durch die Willkürlichkeit der Festsetzung von  $\Psi$ , obwohl doch in der Likelihood-Stützungs-Philosophie „hypothetische Wahrscheinlichkeiten für das Treffen richtiger oder falscher Entscheidungen . . . überhaupt nicht der Maßstab sind“. (Diepgen 1991). Willkürlich ist natürlich auch die Festsetzung von  $Q_c$  in Schritt 3. Bezüglich des letzten Schrittes zur Effektqualifikation wird bemängelt, daß hier nur Stichprobendaten betrachtet werden, obwohl es in einer Inferenzstrategie nur um Aussagen über Effektstärken in der Population gehen sollte.



Für übliche Anwender erscheint der vierstufige Vorschlag von Witte, der alles enthält, was in der Statistik „gut und teuer“ ist, als sehr anspruchlich und aufwendig. Zudem stimmen wir Diepgen in seiner Beurteilung zu, daß dieses Verfahren die Subjektivität des Forschers nur anders einfließen als als die anderen Modelle, bei Witte müßten immerhin 5 Größen,  $\alpha$ ,  $\beta$ ,  $\Psi$ ,  $Q_c$  und  $\eta_c^2$ , letztlich konventionell oder nur plausibel festgesetzt werden.

Nicht nur für diesen Fall sondern bei statistischen Entscheidungen generell mögen aber nun sowohl die Kosten für die Datenerhebung als auch die Kosten für mögliche Fehlentscheidungen von Interesse sein.

Bei einem Test, der bei der Regelung des Hochschulzugangs helfen soll, kann es sein, daß neben relativ geringen Kosten für die Annahme eigentlich ungeeigneter Bewerber die Maxime tritt, möglichst keinen geeigneten Bewerber abzulehnen. Unter diesen Voraussetzungen wäre man hier bei einer statistischen Entscheidung bereit, eine relativ hohe Wahrscheinlichkeit für den Zugang eigentlich ungeeigneter Bewerber zu akzeptieren. Bei einem Test, der bezüglich der Auswahl für einen Pilotenlehrgang helfen soll, sieht die Interessenlage natürlich ganz anders aus: Hier gilt es nach Möglichkeit zu vermeiden, auch nur einen eigentlich ungeeigneten Bewerber aufzunehmen.

Von Wald (1950) wurde die Theorie von Neyman und Pearson erweitert um die explizite Berücksichtigung solcher Kosten. Schon Neyman und Pearson haben mit dem Entscheidbarkeitstest versucht, Fehlentscheidungen mittels Einführung von Indifferenzzonen zu vermindern. Wald geht nun diesen Weg konsequent weiter und entwickelt allgemeinere Entscheidungsregeln, die auch die Größe der Stichprobe beachten. In diesen sequentiellen Tests werden dann solange Daten erhoben, bis anhand von Daten und Entscheidungsregel zu vertretbaren Experimentalkosten die Kosten möglicher Fehlentscheidungen vertretbar klein sind. Das Risiko  $r(d, \theta)$  der Verwendung der Entscheidungsregel bei Vorliegen der durch  $\theta$  charakterisierten Verteilung ist die Summe des dann erwarteten Verlustes bei Verwendung der Entscheidungsregel  $d$  und den zu erwartenden Experimentalkosten. Wald (1950) untersucht insbesondere zwei Familien von Entscheidungsregeln, die er als Bayesianische Regel und als Minimaxregel vorstellt. Bei der bayesianischen Regel ist es das Ziel, das bezüglich einer gegebenen Apriori-Verteilung  $q$  erwartete Risiko durch die Wahl von  $d_b$  zu minimieren, also

$$E_q r(d_b, \cdot) = \min_d E_q r(d, \cdot) \tag{4.4}$$

Die Minimax-Regel  $d_{mm}$  rechnet mit dem ungünstigsten Fall, minimiert also das maximale Risiko, also

$$\max_{\theta} r(d_{mm}, \theta) = \min_d \max_{\theta} r(d, \theta) \tag{4.5}$$

Die Minimaxregel ist eine Bayesianische Regel für die ungünstigste Apriori-Verteilung; hat man keine Informationen über diese Verteilung oder ist sie unbestimmbar, kann diese Regel in diesem Falle als rational gelten. Diepgen (1987) bemerkt zu den Vorteilen des Ansatzes von Wald: „Die Sequentialstatistik (oder sequentielle Testung) - ähnlich wie die Bayes-Statistik - macht es möglich, auch während des Experiments für die Entscheidungsfindung zu lernen. Als Folge davon . . . ist die Stichprobengröße z.T. über 50 % geringer als die entsprechenden notwendigen Stichprobengrößen bei konventionellen Signifikanztests.“ Sequentielle Testverfahren haben unter Psychologen nur wenige Anhänger gefunden, vielleicht deshalb, weil Kosten, egal ob für die Datenerhebung oder für mögliche Fehlentscheidungen, in der Grundlagenforschung nur selten diskutiert werden. Diepgen äußert sich hier sehr deutlich: „Besonders ärgerlich, geradezu schon mysteriös, erscheint mir die weitgehende Ignoranz der deutschen Psychologie gegenüber Sequentialstatistik, obwohl diese Statistik zentrale Probleme des herkömmlichen Signifikanztests löst, nämlich die Problematik der  $\beta$ -Fehler-Kontrolle und der praktischen Bedeutsamkeit, und dies auch noch mit dem ökonomischen Vorteil wesentlich kleinerer Stichproben.“

### **5. Eine Variante des Signifikanztests in der Psychologie**

Es gibt einen Typ von Signifikanztestung in der psychologischen Literatur, dem wir wegen der relativen Häufigkeit der Verwendung einen eigenen Abschnitt widmen. Man könnte diese „hybrid theory“ (Gigerenzer et al. 1991, S. 106) ironisch als „Nullhypothesentestung nach Neyman und Pearson“ oder als „Signifikanztest nach Fisher mit Alibi-H,“ nennen. Gigerenzer (1986) beschreibt und deutet das Vorgehen wie folgt: „In den anspruchsvolleren Lehrbüchern geht dabei Neyman und Pearsons Theorie, meist anonym, als das ‚Überich‘ der psychologischen Forschung ein. Hier wird das Ziehen von Zufallsstichproben aus einer definierten Population betont, es wird gelehrt, das Signifikanzniveau und den Stichprobenumfang so festzulegen, daß die erwünschte Macht ( $1-\beta$ ) des Tests erreicht wird, und es werden keine Wahrscheinlichkeitsaussagen über den Bestätigungsgrad von Hypothesen oder über Einzelergebnisse erlaubt. Das Fishersche „Ego“ aber bestimmt weitgehend, wie in der realen Forschung vorgegangen wird - wenn auch manchmal mit Schuldgefühlen, die „Regeln“ verletzt zu haben. Dort werden kaum Zufallsstichproben gezogen, selten Punktalternativen spezifiziert, um den Fehler 2. Art bestimmbar zu machen, und es werden sogar Aussagen über die Wahrscheinlichkeit getroffen, mit der eine Nullhypothese widerlegt sei.“ Und vielleicht, um zu ergänzen, spielt jene üble Forschungspraxis, bei der nur „Signifikanzen“ zählen, hier die Rolle des „Es“, des Antriebes. Das von Gigerenzer zutreffend charakterisierte Verfahren sieht dann etwa so aus:

- (1) Es wird nur eine  $H_0$  als Punkthypothese spezifiziert. Beide falschen Ety-mologien des Begriffs Nullhypothese kommen hier zusammen: Meist wird damit ein „Null-Effekt“ behauptet, eigentlich soll sie verworfen, nul-lifiziert werden.
- (2) Die Alternativhypothese  $H_1$  wird als Rest des Hypothesenraumes nicht näher spezifiziert, man erwartet „irgendeinen Effekt“. Gigerenzer be-merkt richtig, „eine Vorgehensweise, der Neyman und Pearson nie zuge-stimmt haben“.
- (3) Gelegentlich wird ein Signifikanzniveau vorher festgelegt, meistens wer-den „Signifikanzen“ hinterher, z.B. durch \*\* angezeigt. Hager und We-stermann (1982) stellten fest, daß nur in 7 von 76 Untersuchungen das Signifikanzniveau vorher festgelegt war.
- (4) Überlegungen zur Kontrolle des  $\beta$ -Fehlers, zur Größe des erwarteten Ef-fekts, zur erwarteten Richtung, zur Stichprobengröße oder zur Teststärke werden nicht berichtet und wohl auch nicht angestellt.

Eine ähnliche Auflistung dieser üblichen Anwendung gibt Wottawa (1990). Er betont zudem: „Das prinzipielle Problem dieses Vorgehens wird deutlich, wenn man sich vor Augen hält, daß (unter den üblichen Stetigkeitsannahmen) die  $H_0$  unabhängig von allen empirischen Befunden für die jeweilige Popula-tion nicht gültig ist, und zwar als Folge ihrer punktförmigen Formulierung.“

Bei vielen Anwendern beliebt ist auch die Aufwertung einer statistischen Ent-scheidung in den Status eines logischen Schlusses. So schreibt etwa Wendt (1983, S. 485): „Logisch gesehen folgt die Neyman-Pearsonsche Entsch-eidungstheorie dem Schlußprinzip des **Modus tollens** ...“ Im Sinne dieser „Lo-gik“ sollen aus  $H_0$  gewisse Stichprobendaten „folgen“ („aus  $H_0$  folgt  $D_0$ “, so Wendt); wenn sich diese Daten aber nicht ergeben („nun aber nicht  $D_0$ “), so darf man als Konklusion schließen, daß  $H_0$  nicht gilt („also nicht  $H_0$ “); man darf dann sogar schließen, daß nunmehr  $H_0$  gilt („da aber nicht  $H_0$  folgt  $H_1$ “ - dieser letzte „Schluß“ gehört übrigens nicht mehr zum modus tollens). Nun muß aber leider festgestellt werden, daß die Interpretation einer statistischen Entscheidungsregel als Syllogismus aus mehreren Gründen unsinnig ist: Daten „folgen“ nicht aus Hypothesen, sondern sie lassen sich unter der Annahme der Gültigkeit auch der Oberhypothesen mathematisch ableiten. Daten be-kommen nicht den Wahrheitswert „falsch“, sondern sie sind höchstens im Lichte einer Hypothese sehr unwahrscheinlich; wir „schließen“ nicht auf „nicht  $H_0$ “ und dann auf  $H_1$  - wir entscheiden uns nach vorher festgelegten Regeln. Bei manchen Autoren drängt sich der Verdacht auf, daß sie bezüglich ihrer - möglicherweise falschen - statistischen Entscheidungen durch Unter-legung einer solchen Schluß-„Logik“ Unfehlbarkeit suggerieren wollen.

Ob das eben beschriebene „Gegen-He-Testen“ (mit anschließendem Bericht der „signifikanten Effekte“) überhaupt forschungslogisch sinnvoll ist, kann in

den meisten Fällen bezweifelt werden. Zum einen: „auch ganz geringe Abweichungen können zu einem statistisch signifikanten Ergebnis führen, wenn nur die Zahl  $n$  der Beobachtungen genügend groß ist“ (Hager und Westermann 1982). Zum anderen: „Das Finden einer statistischen Signifikanz ist das vielleicht unwichtigste Attribut eines guten Experiments; sie reicht nie hin, um behaupten zu können, daß eine Theorie sich in brauchbarer Weise bewährt hat, daß eine sinnvolle empirische Tatsache festgestellt worden ist oder daß ein Experimentalbericht veröffentlicht werden sollte.“ (Lykken 1968)

Auf ein methodologisches Paradox speziell dieser Art von Hypothesenüberprüfung hat bereits Meehl (1967) hingewiesen: Während in der Physik durch verbesserte experimentelle Anordnungen und Erhöhung der Meßgenauigkeit die bislang als bewährt geltenden Theorien unter Druck geraten und modifiziert oder ersetzt werden müssen, passiert in der Psychologie, die den Signifikanztest solcherart einsetzt, eher das Gegenteil: Die Theorien zeigen die Tendenz, sich zu immunisieren. Sie stehen bei dieser Art Nullhypothesentestung ja auch eigentlich nie ernstlich zur Disposition. Lakatos (1974, S. 170, Fußnote) nimmt diesen Sachverhalt zum Anlaß, polemisch zu fragen, „ob die Funktion von statistischen Techniken in den Sozialwissenschaften nicht vor allem darin besteht, daß sie einen Mechanismus liefern, der Scheinbestätigung und den Anschein von ‚wissenschaftlichem Fortschritt‘ an Stellen produziert, wo sich in Wirklichkeit nur pseudointellektueller Mist anhäuft“.

Die Diskussion um das Nullhypothesentesten ist alt und seit langem heftig, schon Roozeboom (1960) konnte feststellen, daß die Methode wegen ihrer Unangemessenheit nachdrücklich kritisiert worden ist - nur, warum hat sich nichts geändert? Eine Reihe von möglichen Gründen zählt Gigerenzer (1986) auf, er vermutet komplizierte Wechselbeziehungen zwischen der Forschungspraxis und dieser Art von Signifikanztest, denn es zeigt sich, daß „der Forscher weniger für ein sorgfältig angelegtes experimentelles Design verstärkt wird als für den Exorzismus von Nullhypothesen“ und daß „die verbreitete falsche Interpretation von ‚signifikant‘ als ‚replizierbar‘ dazu verführt, Experimente eher nicht zu wiederholen“. Nicht ganz schuldlos an dieser Verwechslung von „signifikant“ mit „replizierbar“ oder mit „damit konnte gezeigt werden“ oder gar mit „nunmehr ist bewiesen“ dürfte die fatale Metapher vom *modus tollens* sein. Aber nicht nur die Forschungspraxis, auch die Ausbildungspraxis scheint am Beharrungsvermögen dieses Verfahrens beteiligt zu sein, denn in den meisten Statistiklehrbüchern wird eine kritische Diskussion praktisch nicht erwähnt (Leiser 1982), statistische Entscheidungsverfahren werden als im Grunde unproblematisch angesehen, meistens werden sie in Grundkursen im ersten Semester angeboten. Oft wird hier auch der Eindruck erweckt, die angewandten Verfahren selbst wurden entscheiden (insbesondere bei EDV-gestützten Verrechnungen, etwa durch SPSS), die Rolle des für seine Entscheidung verantwortlichen Forschers bleibt unerwähnt.

Wie bereits gesagt tendieren viele Anwender dazu, erst einmal abzuwarten, „welche von der Fülle getesteter Haupteffekte, Interaktionen oder Korrelationen signifikant werden“ (Gigerenzer 1986) - „Erklärungen“ für die Signifikanzen werden posthoc gesucht. Wenn ein Jäger mit einem Schrotgewehr auf einen Schwarm Vögel zielt und uns hinterher erzählt, er habe genau die heruntergefallenen auch treffen wollen, so wäre dies Jägerlatein.

Viele dieser Anwender neigen auch „zur üblichen Fehlinterpretation des Signifikanztests“, nämlich dazu, „p als Maß für Signifikanz zu nehmen“ (Bakan, 1966). Wie schon Witte (1980, S.53) feststellte, sagt das Signifikanzniveau nichts über einen Effekt aus. Eine Fehlinterpretation besteht oftmals darin, nach Berechnung der Prüfgröße einer Tabelle den korrespondierenden Wahrscheinlichkeitswert zu entnehmen und eine der folgenden Zuordnungen zu treffen: signifikant (\*), hochsignifikant (\*\*), oder höchst signifikant (\*\*\*). Damit wird das Modell des statistischen Tests überinterpretiert, p wird als komparatives Stützmaß angesehen, „hochsignifikant“ ist signifikanter als „signifikant“, und solcherart nachgewiesenen Effekten kann man noch mehr trauen. Die „\*“ legen die Assoziation nahe, die Befunde sollten durch Michelin-Sterne ausgezeichnet werden. Aber mehr als ein klassifikatorischer Stützbegriff kann aus den Theorien von Fisher (verwerfen/nicht verwerfen) oder Neyman und Pearson (annehmen/eventuell nicht entscheiden/ablehnen) nicht abgeleitet werden. Wer Aussagen des Typs „ist besser gestützt als“ anstrebt, muß sich anderer Konzepte bedienen. Man könnte in diesem Fall etwa unter Verwendung des Likelihoodkonzepts und der Likelihoodregel die Likelihoods der Daten unter der Annahme der Richtigkeit zweier konkurrierender Hypothesen vergleichen. Unter Beachtung der Axiome von Koopman ließe sich dann ein **komparativer Stützbegriff** entwickeln (siehe Stegmüller 1973, S. 84).

Wir wollen dem Signifikanztest nun nicht jedwede Berechtigung im Forschungsprozeß absprechen, in einem frühen Stadium der Theorienbildung als „screening“-Prozedur und/oder zur zufallskritischen Absicherung kann er wertvolle Dienste leisten. Auf diesen letzten Aspekt weist etwa Mittenecker (1983) hin: „Erkenntnis im Alltag (geht) in gleicher Weise vor, nur unter Anwendung von schlecht oder nicht kontrollierten ‚Statistiken‘: Das Herausbilden von Vorurteilen und ‚Aberglauben‘ war und ist das jahrtausendealte Ergebnis vieler Faktoren, unter anderem sicher aber auch des Nichtbeachtens des Zufalls in Stichproben und Beobachtungen.“ Es sollte sich allerdings in der Psychologie einbürgern, daß „Signifikanz“ für eine Hypothese kein hinreichendes Kriterium (vielleicht ein notwendiges) dafür sein kann, als Aussage „in den Bestand des Wissens, der eine Wissenschaft ausmacht, aufgenommen zu werden“ (Wendt 1983, S. 471). Auf die Gefahr für eine Wissenschaft, wenn einzig der Signifikanztest in der Rolle von Ockhams Rasiermesser entscheidet, was in den Bestand aufgenommen wird und was nicht, hat bereits Popper (1966, S. 198) hingewiesen. In einer Fußnote beklagt er sich über Carnap; die-

ser hatte das Poppersche Konzept „Grad der Bewährung“ uminterpretiert als „Grad der Bestätigung“ und damit als ein Synonym für eine Wahrscheinlichkeit. Nur hatte Popper mit seinem „Grad der Bewährung“ für eine Hypothese viel mehr gemeint als eine wahrscheinlichkeitstheoretische Betrachtung der Relation zu den Daten. Zur Bewährung gehörte neben dieser Betrachtung auch die Aufklärung des Zusammenhangs zwischen Hypothese und Theorie, die begriffliche Eindeutigkeit der Hypothese sowie die Bewährung in bezug auf inhaltlich echt konkurrierende Hypothesen. Nur solche Hypothesen sollten als vorläufig bewährt gelten, die einen fairen Wettkampf bestanden hatten.

## 6. Hypothesenprüfung nach Bayes

Spätestens seit der klassischen Arbeit von Edwards, Lindman und Savage (1963) gelten die „Bayesianer“ neben der Signifikanztestung nach Fisher und der Testtheorie nach Neyman und Pearson als dritte große statistische Schule. Das Bayessche Verfahren zur Prüfung von Hypothesen endet jedoch nicht notwendig mit einer statistischen Entscheidung für oder gegen eine Hypothese, oft genügt die Angabe von Aposteriori-Wahrscheinlichkeiten oder von Wettquotienten, den uns schon von Karl Pearson bekannten odds. Die bayesianische Statistik befaßt sich mit „der graduellen Stützung oder Schwächung von Hypothesen auf Grund von Wahrscheinlichkeiten. Die beiden kategorialen Entscheidungsmöglichkeiten ‚für‘ oder ‚gegen‘ eine Hypothese, ‚Akzeptierung‘ oder ‚Verwerfung‘ einer Hypothese, die bei einem Test im Vordergrund stehen, spielen im Bayes-Ansatz eher eine sekundäre Rolle“ (Kleiter 1981, S. 13). In einer klassisch einfachen Schreibweise, in der „basic form“ (Edwards et al. 1963, S. 198), kann man das Theorem von Bayes, für diesen Zweck interpretiert als Satz zur Revision subjektiver Wahrscheinlichkeiten, schreiben als

$$p(H|D) = \frac{p(H)p(D|H)}{p(D)} \quad (6.1)$$

Natürlich liegt das für Hypothesenprüfung interpretierte Theorem von Bayes in mannigfaltiger Form auch für stetige Fälle von Hypothesen, von Daten oder von Prüfstatistiken vor (siehe etwa Kleiter 1981). Die problematischste Größe in obiger Formel ist  $p(H)$ , die Apriori-Wahrscheinlichkeit einer Hypothese vor der Betrachtung neuer Daten. Zunächst ist anzumerken, daß für einen Forscher in die Schätzung von  $p(H)$  drei Informationsquellen eingehen können:

- (1) Wenn keinerlei empirische Indizien bezüglich der Hypothese bekannt sind, geht in  $p(H)$  nur die subjektive Evidenz des Forschers in seine Hypothese ein (oder eine Wette, siehe weiter unten).

- (2) Wenn empirische Indizien aus eigenen oder fremden Quellen vorliegen, können diese in die Schätzung von  $p(H)$  eingehen.
- (3) Wenn in einer Voruntersuchung bereits eine Hypothesenbewertung mittels des Bayesschen Theorems vorgenommen wurde, so kann die daraus resultierende Aposteriori-Wahrscheinlichkeit als neue Apriori-Wahrscheinlichkeit in die Nachfolgeuntersuchung eingehen. Diese Revision kann beliebig oft wiederholt werden, und dies wird von Bayesianern auch als gewichtiger Vorteil ihres Verfahrens genannt, erspart es ihnen doch die Problematik von Metaanalysen, wie sie im Gefolge von Signifikanztests üblich sind.
- (4) Wenn zu einer Hypothese mehrere stochastisch unabhängige Datensätze vorliegen und diese Daten auch sonst unter vergleichbaren Bedingungen erhoben wurden, kann in einer simultanen Verrechnung die Apriori-Wahrscheinlichkeit revidiert werden (Beispiele hierzu siehe etwa Kleiter, 1981, S. 127):

$$p(D_1, \dots, D_n | H) = p(D_1 | H)p(D_2 | H) \dots p(D_n | H) \tag{6.2}$$

$p(H | D)$  als Aposteriori-Wahrscheinlichkeit oder „inverse“ Wahrscheinlichkeit ist die vom Forscher „im Lichte der Daten“ revidierte Wahrscheinlichkeit,  $p(H | D)$  ist die Größe, in die nach Bayesianischer Auffassung das „Lernen aus Erfahrung“ eingeht. „Das Bayes-Theorem . . . ist eine Regel, wie die Unsicherheit über eine statistische Hypothese auf Grund einlangender Daten rational revidiert wird“ (Kleiter, 1981, S. 14). Die Größe  $D$  bezeichnet hier, anders als etwa bei der Signifikanztestung, nicht nur den Wert eines Parameters oder den Wert einer statistischen Prüfgröße für diesen Parameter, sondern  $D$  kann die in allen Daten enthaltene Information in Form der gesamten Verteilung bezeichnen.

Auch dieser Sachverhalt, bei der Hypothesenprüfung nicht notwendig nur auf wenige Parameter zurückgreifen zu müssen, wird von Bayesianern als gewichtiger Vorteil ihres Verfahrens angeführt.  $p(D | H)$  ist die schon von Fisher (1959, S.68) bekannte Likelihood, die „Wahrscheinlichkeit“ der Daten unter der Hypothese.

Einer der zentralen Kritikpunkte am bayesianischen Ansatz betrifft den zugrundeliegenden „subjektiven“ Wahrscheinlichkeitsbegriff, im Gegensatz zum sonst in der Statistik üblichen objektiven und meistens frequentistischen Wahrscheinlichkeitsbegriff. Allerdings, „daß es sich . . . bei dem Begriffspaar personelle (d.h. subjektive) Wahrscheinlichkeit und statistische Wahrscheinlichkeit . . . um zwei voneinander verschiedene, wissenschaftlich wichtige und exakt durchführbare Deutungen des Wahrscheinlichkeitskalküls handelt, wird keinesfalls allgemein anerkannt“ (Stegmüller 1973, S. 27). Wie schon zwischen Fisher und Neyman und Pearson, so tun sich zwischen Bayesianer und Nicht-

Bayesianern schier unüberwindliche weltanschauliche Differenzen bezüglich dieses Themas auf, die etwa Stegmüller (1973, S.41) dadurch zu überwinden sucht, daß er in Anlehnung an Braithwaite vorschlägt, Wahrscheinlichkeiten weder als frequentistisches noch als subjektives Konzept einzuführen, sondern als theoretischen Begriff, als eine nicht explizit definierbare theoretische Disposition physikalischer Systeme. Wir wollen diesen Grundlagenstreit hier nicht weiter verfolgen, können allerdings konstatieren, daß der Bayessche Ansatz wegen dieses Subjektivismus entschiedene Ablehnung erfährt. So bemerkt etwa Bredenkamp (1972, S.149): "... diese Kombination (von persönlichen Hypothesenwahrscheinlichkeiten und statistischen Wahrscheinlichkeiten) ist nur sinnvoll, wenn die subjektiven Wahrscheinlichkeiten den Axiomen und Theoremen der mathematischen Wahrscheinlichkeitslehre genügen. Oftmals summieren sich die persönlichen Wahrscheinlichkeiten nicht zu 1." Dazu soll allerdings bemerkt werden, daß uns nichts daran hindert, für ein normatives subjektives Wahrscheinlichkeitskalkül ein von der Axiomatik der Wahrscheinlichkeitsräume begründet abweichendes Kalkül einzuführen, wie es etwa neuerdings unter dem Label „belief functions“ (vgl. Shafer 1982) nicht ohne Erfolg geschieht.

Rützel (1979) vermutet, die Apriori-Wahrscheinlichkeiten seien als subjektive Überzeugungsstärken nicht neutral und es gäbe keine intersubjektiv eindeutigen Kriterien zu ihrer Festsetzung. Er wurde deshalb bayesianisches Hypothesentesten als „privates Hypothesentesten“ bezeichnen, und als solches „ist es unverbindlich und für den öffentlichen wissenschaftlichen Gebrauch abzulehnen“. Auch Witte (1980, S. 40) sieht diese Gefahr: „Zum einen werden subjektive Wahrscheinlichkeiten a priori bezüglich der Wahrscheinlichkeit der Hypothese zugelassen, zum anderen wird die Testtheorie ausgebaut zu einer Entscheidungstheorie, d.h. es werden den möglichen Handlungen Nutzenwerte zugeordnet... Beide Erweiterungen bringen wegen der Subjektivität der einzuführenden Größen erhebliche Schwierigkeiten mit sich. Sie erschweren den wissenschaftlichen Diskurs.“

Tholey (1982) vermutet als ein Motiv für diese Art der Ablehnung des Bayesianismus „offenbar eine völlig irrationale . . . Angst vor allem, was den Anschein des Subjektivismus hat“. Die Bayesianer führen gegen diese Vorwürfe, mittels eines subjektiven Wahrscheinlichkeitsbegriffs ließen sich statistische Hypothesenprüfungen nicht rational vornehmen, gewichtige Gegenargumente ins Feld. Zum einen wenden sie ein, daß es relativ gleichgültig sei, welche Aprioris man zunächst einsetzt, denn nach dem „principle of stable estimation“ (siehe etwa Wendt 1983, S. 501) werden auch sehr schlecht geschätzte Aprioris durch einigermaßen aussagekräftige Daten sehr schnell zu erheblich besseren Hypothesenwahrscheinlichkeiten  $p(H \mid D)$  korrigiert. Zum anderen verweisen Bayesianer gern auf die Effizienz ihres Verfahrens im Vergleich zu Menschen im Alltag, die sich als konservative Informationsverarbeiter erwei-



sen. So kommen Coombs, Daves und Tversky (1975, S. 176f.) bezugnehmend auf Untersuchungen von Ward Edwards zu dem Schluß, „Menschen sind konservative Wahrscheinlichkeitsschätzer in dem Sinne, daß ihre Schätzungen um ein Beträchtliches weniger extrem sind als die nach der Bayesschen Regel berechneten“.

Wir möchten hier auf ein interessantes Phänomen hinweisen: Für Fisher war der den Signifikanztest nutzende skeptische Wissenschaftler „konservativ“, er hielt länger an seiner Nullhypothese fest als ein Mensch im Alltag; für Bayesianer sind die Menschen im Alltag „konservativ“, sie nutzen die in den Daten enthaltene Information nicht so effektiv wie das Bayessche Theorem.

Als weiteres Gegenargument gegen den Vorwurf des Subjektivismus führen die Bayesianer an, daß mit  $p(H)$  nicht unbedingt der subjektive, private Glaubensgrad eines Forschers gemeint ist, daß man vielmehr den hier verwendeten subjektiven Wahrscheinlichkeitsbegriff in einem normativen Sinn interpretieren kann, denn „subjektive Wahrscheinlichkeit bedeutet hier nicht den empirisch zu ermittelnden Glaubensgrad einer Person, sondern denjenigen Glaubensgrad, den eine rationale Person besitzen sollte, . . . man bezieht sich meist auf den Begriff eines fairen und rationalen Wettsystems“. (Tholey 1981). Ein Beispiel für diese Idee der Operationalisierung durch Wetten gibt Kleiter (1981, S. 117); diese Darstellung entspricht der „odd-form“ des Theorems von Bayes, eine Form, die auch Edwards et al. (1963, S.218) diskutieren. Es lägen zwei konkurrierende Hypothesen  $H_i$  vor,  $i = 1,2$ . Aus der Anwendung des Bayesschen Theorems

$$p(H_i) | D = \frac{p(H_i)p(D | H_i)}{p(D)} \tag{6.3}$$

und nach Betrachtung der beiden Quotienten

$$\lambda_0 = p(H_1) / p(H_2) \tag{6.4}$$

$$\lambda_n = p(H_1 | D) / p(H_2 | D) \tag{6.5}$$

läßt sich als Quotient der Bayes-Theoreme für  $H_1$  und  $H_2$  schreiben:

$$\lambda_n = \lambda_0 L \tag{6.6}$$

mit  $L$  als Likelihood-Quotient  $p(D | H_1)/p(D | H_2)$ , der angibt, um wieviel wahrscheinlicher die Daten unter  $H_1$  sind als unter  $H_2$ . Man kann also sagen: die „posterior odds“ sind gleich dem Produkt aus „prior odds“ mal „likelihood ratio“. Man geht davon aus, „daß der Wert (Nutzen) der Konsequenz zahlenmäßig auf einer Intervallskala bestimmt ist, . . . daß man auch den Wetten einen Wert zuordnen kann und daß dieser Wert eine Reihe von Eigenschaften aufweist, insbesondere, daß er unter bestimmten Voraussetzungen additiv ist“.

Solche Axiomatisierungen von Nutzen- und Wettsystemen liegen vor, etwa bei Kleiter (1981, S. 42) oder bei Luce und Krantz (1971), die ein System von bedingten Entscheidungen axiomatisieren. In Krantz, Luce, Suppes und Tversky (1971, S.373) geben sie ein Beispiel für solche bedingten Entscheidungen: angenommen, jemand hat sich entschlossen, an einem bestimmten Tag von New York nach Boston zu fahren und hat dafür das Flugzeug, den Zug oder das Auto ins Auge gefaßt. Jede Möglichkeit zieht nun aber bestimmte Risiken mit bestimmten Konsequenzen nach sich, und diese können sich von Fahrt zu Fahrt ändern. Wichtig ist, daß jede Wahl zwischen den Fahrmöglichkeiten die möglichen Risiken einschränkt auf jene, die dem gewählten Transportmittel eigen sind; die Bewertung jeder bedingten Entscheidung braucht auch nur jene Risiken in Rechnung zu stellen. Wer etwa ans Flugzeug denkt, muß folgende mögliche Ausgänge beachten: pünktliche Ankunft, unpünktliche Ankunft, Streichung des Fluges, Absturz usw. Einige dieser möglichen Ausgänge können nur für einen einzigen Flug zutreffen, andere können typisch sein für alle Fahrten mit diesem Transportmittel, etwa die Streichung eines Fluges. Andererseits schließt die Wahl eines Transportmittels manche Risiken aus, ein Flugzeug kann etwa nicht unterwegs liegenbleiben wie ein Auto. Dieses Beispiel mag als Illustration der Situation eines Bayesianers dienen, der sich zwischen verschiedenen Hypothesen, verbunden mit verschiedenen Aprioris und Aposterioris, und verschiedenen Entscheidungskonsequenzen bewegt.

Wir können die Debatte über die Angemessenheit des Bayesschen Ansatzes nicht beenden, allerdings ist Witte (1980, S.40) zuzustimmen, der zu bedenken gibt: „Gleichzeitig jedoch können (die subjektiven Wahrscheinlichkeiten und die Nutzenwerte für die Handlungen) sehr nützliche Erweiterungen sein, weil man die gesamte Subjektivität des Forschers abbilden kann in eine Entscheidungsstrategie, denn diese subjektiven Elemente sind nicht völlig zu eliminieren.“ Richtig, denn mit der Wahl eines Versuchsplans, eines Zufallsmodells einer Prüfstatistik und einer Entscheidungsregel gehen auch bei Signifikanztestern willkürliche und letztlich subjektive Momente ein, die nur deshalb als solche wenig auffallen, weil sie als Konventionen von den meisten Anwendern geteilt werden.

## **7. *Schlußbemerkungen***

Wir haben drei klassische Zugänge zum Problem der Hypothesenprüfung mittels statistischer Modelle diskutiert und dabei dem in der Psychologie häufig verwandten Signifikanztest besonderes Augenmerk geschenkt. Nur, ist es in einer empirischen Wissenschaft überhaupt notwendig, soviel Statistik zu treiben? Bredenkamp (1980, S. 23) meint ja, denn “. . . leider ist es so, daß in der Psychologie mit wenigen Ausnahmen nicht auf Statistik verzichtet werden kann“. Den Grund sieht Witte (1989) im Entwicklungsstand der Psychologie:

„So wie viele Theorien konstruiert sind, ist der klassische Signifikanztest die einzig mögliche Überprüfungs­methode und besser als reine Spekulation, weil man zumindest Zufallsschwankungen nicht als Stützung interpretiert.“ Nicht nur statistische Entscheidungen werden befürwortet, speziell der Signifikanztest wird präferiert. So schreibt etwa Westermann (1987, S. 124): „Selbstverständlich können außer dem Signifikanztest auch andere Ansätze zur Generierung von Kriterien über empirische Hypothesen herangezogen werden. Insbesondere ist hier an die Likelihood-Tests und die Bayes-Statistik zu denken. Der Signifikanztest ist jedoch m. E. beiden Alternativen vorzuziehen.“

Eindeutige Stellungnahmen also, aber für welchen Signifikanztest? Für den klassischen Nullhypothesentest nach Fisher? Damit ließen sich keine Effekte sichern. Für den modifizierten Neyman-Pearson-Test, wie etwa Bredenkamp ihn vorgeschlagen hat? Oder für jenen seltsamen Zwitter, der sich in der Psychologie so großer Beliebtheit erfreut? Methodiker plädieren immer wieder für ein Testen nach Neyman und Pearson, denn „um einen Experimentalaufbau angemessen beurteilen zu können, ist (es) notwendig,  $\alpha$ - und  $\beta$ -Fehler anzugeben sowie ein Effektmaß“ (Witte 1989). Dabei sollten  $\alpha$  und  $\beta$  als kleine Werte festgelegt werden (Bredenkamp 1972, S. 149). Es bleibt zunächst unerklärlich, warum viele Anwender diesen methodischen Maximen nicht folgen wollen. Dabei eröffnet gerade der modifizierte Signifikanztest eine weitere methodologische Möglichkeit, denn wie nicht nur Bredenkamp (1980, Vorwort S. VI) fordert, sind „experimentelle Untersuchungen entgegen den üblichen Gepflogenheiten von vorneherein so zu planen, daß psychologische Hypothesen falsifiziert werden können. Dieser Gesichtspunkt ist deshalb so wichtig, weil die Nullhypothese häufig der wissenschaftlichen Hypothese widerspricht, so daß auf ihr Zutreffen erkannt werden können muß, damit eine psychologische Hypothese, die statistisch überprüft wird, überhaupt falsifizierbar ist.“ Und, wie erinnerlich, kann  $H_0$  in seinem Modell mit Wahrscheinlichkeit  $1-\beta$  angenommen werden. Wir würden allerdings vorschlagen, den Begriff der Falsifikation der Logik zu überlassen und hier lieber davon zu sprechen, die psychologische Hypothese zu verwerfen. Die Frage ist nur, wieviele Anwender sich überhaupt dem von Popper begründeten Falsifikationismus verpflichtet fühlen, die meisten scheinen lieber für ihre wissenschaftliche Hypothese (meist umgesetzt als statistische  $H_1$ ) kämpfen zu wollen als gegen sie, und nicht wenige Anwender sind wohl eigentlich „implizite Verifikationisten“, die mit jeder „Signifikanz“ den Fundus gesicherten Wissens erweitern wollen - diesen Anwendern kommt der Zwitter gerade recht. Wie aber Gigerenzer et al. (1991, S. 107) richtig bemerken, haben weder Fisher noch Neyman und Pearson je davon gesprochen, daß eine Nullhypothese etwa auf dem 0,01 -Niveau zurückgewiesen werden könnte.

Eine ebenfalls von Methodikern erhobene Forderung, „bei der Prüfung von Theorien und Hypothesen . . . sowohl die wissenschaftliche Signifikanz in

Form von Effektgrößen wie die statistische Signifikanz in Form von Wahrscheinlichkeitsaussagen zu berücksichtigen“ (Westermann & Hager, 1984), findet ebenfalls kaum Gehör. Schon Bakan (1966) stellte die Frage, „how much of a difference makes a difference for what?“, die Frage nämlich, wie groß ist der Effekt und wie groß ist der Gewinn für die Theorie. Viele Begriffe (praktische Signifikanz, praktische Bedeutsamkeit, wissenschaftliche Signifikanz, wissenschaftliche Bedeutsamkeit, „material significance“) für diesen Aspekt und etliche Verfahren zur numerischen Abschätzung liegen vor (siehe etwa Bredenkamp 1970; Hager 1983), nur werden sie selten eingesetzt. Hager und Westermann (1982) vermuten als Grund: „Es scheint wohl zwischen den Herausgebern und Lesern ein Gentlemen's Agreement zu bestehen, daß nie Maße der erklärten Varianz verlangt werden, sondern stets nur Signifikanztests. Warum? Vielleicht zum Teil deshalb, weil Maße der erklärten Varianz für alle so peinlich wären.“ Wenn dem so wäre, stünde es schlecht um die von Lakatos angeführte intellektuelle Redlichkeit, denn dann ginge es vornehmlich um die Immunisierung der eigenen Theorien.

Wir haben versucht darzulegen, daß statistische Entscheidungen stets dort sinnvoll sind, wo in begründeter und problematisierbarer Weise ein statistisches Modell an die Daten gelegt werden kann und wo in begründeter und problematisierbarer Weise Entscheidungen getroffen werden können. Insbesondere Befürwortern der Signifikanztestung sei angeraten, einige Punkte zu bedenken:

Das Problem, welcher Parameter ein inhaltlich angemessener Indikator für ein psychologisches Konstrukt ist, muß wissenschaftlich, nicht statistisch begründet werden.

In das Modell des Signifikanztests geht Hintergrundwissen in Form akzeptierter statistischer Oberhypothesen und stochastischer Unabhängigkeitsannahmen ein. Diese sind zwar zum Teil selbst wieder statistisch prüfbar, zum Beispiel die Annahme der stochastischen Unabhängigkeit mit dem Runs-Test und die Annahme der Normalverteilung mit dem Kolmogoroff-Smirnoff-Test (nur gehen in diese Verfahren natürlich auch wieder Oberhypothesen ein). Viele Anwender enden mit ihren Überlegungen dort, wo schon „Student“ angefangen hatte: Wenn man schon eine Verteilungsannahme machen muß, dann nimmt man wenigstens eine Verteilung, die gut bekannt und tabelliert ist. In problematischen Fällen, in denen die übliche Normalverteilungsannahme offensichtlich verletzt scheint, spricht wenig dagegen, erst einmal einige ihrer Parameter zu schätzen und gegenüber diesen Vorannahmen bayesianisch hinzuzulernen.

Die Wahl des statistischen Zufallsmodells, gegen das man sich zufallskritisch absichern will, ist nicht trivial. Wie am Einwand von Neyman bezüglich der hypergeometrischen und der Binomialverteilung aufgezeigt, gibt es komplexe - und leider nur wenig beachtete - Zusammenhänge

zwischen experimenteller Versuchsanordnung, resultierenden Zufallsvariablen und statistischen Entscheidungen.

- Die statistische Entscheidung, die nach den vorher festzulegenden Entscheidungsregeln getroffen wird, ist nicht nur bezüglich ihrer möglichen Irrtümern und deren möglichen Kosten zu problematisieren, sondern auch bezüglich der Theorie, aus der die wissenschaftlichen Hypothesen abgeleitet sind. So bemerkt etwa Deppe (1977, S. 162), daß es zwischen wissenschaftlichen Hypothesen und statistischen Hypothesen (d.h. solchen, die das beobachtbare Verhalten von Zufallsvariablen zum Gegenstand haben) zwei Übersetzungen geben muß: „Die Übersetzung der wissenschaftlichen Hypothesen in statistische, und die Rückübersetzung des Ergebnisses der Prüfung statistischer Hypothesen in ein die wissenschaftlichen Hypothesen betreffendes Ergebnis.“ Dieses Problem der Rückübersetzung der statistischen Entscheidungen in die Theorie wurde deutlich bei der Diskussion der vorher zu postulierenden oder zumindest hinterher zu berichtenden Effektstärken.

Die Problematik der überwiegend mit parametrischen Verfahren arbeitenden Tests haben wir aufgezeigt. Es fällt auf, daß in der Psychologie die oftmals naheliegenden nichtparametrischen Verfahren weitgehend ungenutzt bleiben. Tests für zentrale Tendenzmaße und Dispersionsmaße stehen ausreichend zur Verfügung (siehe etwa Büning und Trenkler 1978; Bortz, Lienert und Boehnke 1990). Mit Ausnahme des Kolmogoroff-Smirnoff-Tests haben die gebräuchlichsten nichtparametrischen Teststatistiken die Normalverteilung oder die  $\chi^2$ -Verteilung als Asymptote; allerdings lassen sich für manche Tests Güteeigenschaften nicht herleiten. Bei vielen Fragestellungen wäre es hilfreich, die Möglichkeiten der Schätztheorie zu nutzen und insbesondere Konfidenzintervalle anzugeben, eine Forderung, die Rozeboom schon 1960 stellte: „... wann immer möglich, sollte der grundlegende statistische Bericht in Form von Konfidenzintervallen erfolgen.“

Wir hatten auf die strukturellen Ähnlichkeiten zwischen der Angabe solcher Intervalle und der Signifikanztestung nach Fisher hingewiesen. Bei vielen Publikationen wäre zudem zu fragen, ob der „basic statistical report“ (Rozeboom) nicht ohnehin besser in Form einer wohlauflbereiteten deskriptiven Statistik erfolgen sollte. Leider ist die Ansicht weitverbreitet, die ritualisierte Angabe von (\*\*\*) zur Charakterisierung von Signifikanzen sei in jedem Fall informativer als eine kluge Darstellung von Konfidenzintervallen oder Rohdaten, aber der Bericht von Rohdaten scheint unmodern geworden zu sein. „Der Streit um den Signifikanztest wird eigentlich stellvertretend geführt als Streit über Theorienbildung. . . . Wer eine ‚bessere‘ schließende Statistik will, muß präzisere Theorien formulieren“, schreibt Witte (1989) über die „im wahrsten Sinne des Wortes provinziell geführte Signifikanztestkontroverse“ in der Zeitschrift für Sozialpsychologie, und mit Witte können wir resümieren,

„seit der Kontroverse um 1970 haben sich keine neuen Diskussionspunkte ergeben“.

## **Literatur**

- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, **66**, 423-37.
- Bickel, P.J. (1969). A distribution-free version of the Smirnov 2-sample test in the p-variate case. *Annals of Mathematical Statistics*, **40**, 1-23.
- Bortz, J., Lienert, G.A. & Boehnke, K. (1990). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.
- Bredenkamp, J. (1970). über Maße der praktischen Signifikanz. *Zeitschrift für Psychologie*, **177**, 310-8.
- Bredenkamp, J. (1972). *Der Signifikanztest in der psychologischen Forschung*. Frankfurt: Akademische Verlagsanstalt.
- Bredenkamp, J. (1980). *Theorie und Planung psychologischer Experimente*. Darmstadt: Steinkopff.
- Büning, H. & Trenkler, G. (1979). Nichtparametrische *statistische Methoden*. Berlin: De Gruyter.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: a review. *Journal of Abnormal and Social Psychology*, **65**, 145-53.
- Coombs, C. H., Daves, R. M. & Tversky, A. (1975). *Mutbemetische Psychologie*. Weinheim: Beltz.
- David, F. N. (ed.) (1966). *Research papers in statistics. Festschrift for J. Neyman*. London: Wiley.
- Deppe, W. (1977). *Formale Modelle in der Psychologie*. Stuttgart: Kohlhammer.
- Diepgen, R. (1991). Inkonsistentes zur Signifikanzproblematik. Ein Kommentar zu Witte (1989). *Psychologische Rundschau*, **42**, 29-33.
- Edwards, W., Lindman, H. & Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, **70**, 193-242.
- Fisher, R. A. (1951). *The design of experiments*. Edinburgh: Oliver and Boyd.
- Fisher, R.A. (1959). *Statistical methods and scientific inference*. London: Oliver and Boyd.
- Freudenthal, H. & Steiner, H. G. (1966). Aus der Geschichte der Wahrscheinlichkeitstheorie und der mathematischen Statistik. In H. Behnke (Hrsg.), *Grundzüge der Mathematik Band 4: Praktische Methoden und Anwendungen* (S. 149-95). Göttingen: Vandenhoeck und Ruprecht.
- Gigerenzer, G. (1986). Wissenschaftliche Erkenntnis und die Funktion der Inferenz-Statistik. Anmerkungen zu E. Leiser. *Zeitschrift für Sozialpsychologie*, **17**, 183-9.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J. & Krüger, L. (1991). *The Empire of Chance. How probability changed science and everyday life*. Cambridge: Cambridge University Press.

- Haagen, K. & Seifert, H. G. (1979). **Methoden der Statistik für Psychologen**. Band II. Stuttgart: Kohlhammer.
- Hacking, I. (1965). **Logic of Statistical Inference**. Cambridge: University Press.
- Hager, W. (1983). über univariate Maße der wissenschaftlichen Signifikanz. **Zeitschrift für Psychologie**, *191*, 295-309.
- Hager, W. & Westermann, R. (1982). Die Elle - 10 Jahre danach. **Zeitschrift für Sozialpsychologie**, *13*, 250-2.
- Hager, W. & Westermann, R. (1983). Planung und Auswertung von Experimenten. In J. Bredekamp & H. Feger (Hrsg.). **Hypothesenprüfung. Enzyklopädie der Psychologie, Serie 1, Band 5** (S.24-238). Göttingen: Hogrefe.
- Hartung, J. (1985). **Statistik, Lehr- und Handbuch der angewandten Statistik**. München: R. Oldenbourg.
- Herrmann, T. (1979). **Psychologie als Problem**. Stuttgart: Klett-Cotta.
- Kendall, M. G. & Quart, A. (1973). **The advanced theory of statistics. Vol. 2: Inference and relationship**. London: Griffin.
- Kleiter, G. D. (1981). **Bayes-Statistik**. Berlin: de Gruyter.
- Krantz, D. H., Luce, R. D., Suppes, P. & Tversky, A. (1971). **Foundations of measurement**. New York: Academic Press.
- Lakatos, I. (1974). Falsifikation und die Methodologie wissenschaftlicher Forschungsprogramme. In I. Lakatos & A. Musgrave (Hrsg.), **Kritik und Erkenntnisfortschritt** (S. 89-190). Vieweg: Braunschweig.
- Leiser, E. (1982). Wie funktioniert sozialwissenschaftliche Statistik?. **Zeitschrift für Sozialpsychologie**, *13*, 125-39.
- Lenzen, W. (1974). **Theorien der Bestätigung wissenschaftlicher Hypothesen**. Stuttgart: Frommann-Holzboog.
- Lindgren, B.W. (1976). **Statistical Theory**. New York: Macmillan.
- Luce, R.D. & Krantz, D.H. (1971). Conditional expected utility. **Econometrica**, *39*, 253-71.
- Lykken, D.T. (1968). Statistical significance in psychological research. **Psychological Bulletin**, *70*, 151-9.
- Meehl, P.E. (1967). Theory testing in psychology and physics: a methodological paradox. **Philosophy of Science**, *34*, 103-15.
- Mittenecker, E. (1983). Anmerkungen und Stellungnahme zu E. Leiser: Wie funktioniert die sozialwissenschaftliche Statistik? **Zeitschrift für Sozialpsychologie**, *14*, 68-71.
- Morrison, D. E. & Henkel, R. E. (Hrsg.) (1970). **The significance controversy**. Chicago: Aldine.
- Morrison, D. E. & Henkel, R. E. (1970). Significance test in behavioral research: skeptical conclusions and beyond. In D.E. Morrison & R.E. Henkel (Hrsg.), **The significance controversy**. Chicago: Aldine.
- Neyman, J. (1935). On the problem of confidence intervals, **Annals of Mathematical Statistics**, *6*, 111.

- Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability, *Philosophical Transactions, A* **236**, 333-80.
- Neyman, J. (1943). On the problem of testing hypotheses. *Annals of Mathematical Statistics*, **14**, 238-52.
- Neyman, J. (1950). *First Course in Probability and Statistics*. New York: Holt.
- Neyman, J. (1952). *Lectures and conferences on mathematical statistics and probability*. Washington: Department of Agriculture.
- Neyman, J. und Pearson, E.S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, **20A**, 175-240, 263-94.
- Neyman, J. und Pearson, E.S. (1933). On the problem of the most efficient of statistical hypotheses. *Philosophical Transactions, A* **231**, 289-337.
- Neyman, J. und Pearson, E.S. (1933). The testing of statistical hypotheses in relation to probabilities a priori, *Proceedings of the Cambridge Philosophical Society*, **29**, 492-510.
- Pearson, E. S. (1966). *The Neyman-Pearson Story: 1926-34*. In F. N. David (Hrsg.), Research papers in statistics (S. 1-23). London: Wiley.
- Popper, K. R. (1966). *Logik der Forschung*. Tübingen: Mohr.
- Rozeboom, W. W. (1960). The fallacy of the nullhypothesis significance test. *Psychological Bulletin*, **57**, 416-28.
- Rützel, E. (1979). Bayessches Hypothesentesten und warum die Bayesianer Bias-ianer heißen sollten. *Archiv für Psychologie*, **131**, 211-32.
- Shafer, G. (1982). Belief functions and parametric models. *Journal of the Royal Statistical Society, B* **44**, 322-52.
- Sigel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill 1956.
- Stegmüller, W. (1969). *Wissenschaftliche Erklärung und Begründung*. Berlin: Springer.
- Stegmüller, W. (1973). *Personelle und statistische Wahrscheinlichkeit. 2. Halbband: Statistisches Schließen, statistische Begründung, statistische Analyse*. Berlin: Springer.
- Stigler, S. (1977). Eight centuries of sampling inspection: the trial of the pyx. *Journal of the American Statistical Association*, **72**, 493-500.
- Tholey, P. (1982). Signifanztest und Bayessche Hypothesenprüfung. *Archiv für Psychologie*, **134**, 319-42.
- von Mises, R. (1942). On the correct use of Bayes' formula. *Annals of Mathematical Statistics*, **13**, 156-63.
- von Mises, R. (1964). *Mathematical Theory of Probability and Statistics*. New York: Academic Press.
- Wald, A. (1947). *Sequential Analysis*. New York: Wiley.
- Wald, A. (1950). *Statistical Decision Functions*. New York: Wiley.
- Wendt, D. (1966). Versuche zur Erfassung des subjektiven Verlässlichkeitsniveaus. *Zeitschrift für Psychologie*, **172**, 40-81.



- Wendt, D. (1983). Statistische Entscheidungstheorie und Bayes-Statistik. In J. Bredenkamp & H. Feger (Hrsg.), **Hypothesenprüfung. Enzyklopädie der Psychologie, Themenbereich B, Serie 1, Band 5** (S.471-529). Göttingen: Hogrefe.
- Westermann, R. (1987). **Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie**. Berlin: Springer.
- Westermann, R. & Hager, W. (1984). Zur Verwendung von Effektgrößen in der theorie-orientierten Sozialforschung. **Zeitschrift für Sozialpsychologie, 15**, 159-66.
- Witte, E. (1980). **Signifikanztest und statistische Inferenz**. Stuttgart: Enke.
- Witte, E. (1989). Die „letzte“ Signifikanztestkontroverse und daraus abzuleitende Konsequenzen. **Psychologische Rundschau, 40**, 76-84.
- Witte, E. (1991). Antworten auf die „Bemerkungen“ von Diepgen. **Psychologische Rundschau, 42, 34-7**.
- Wottawa, H. (1990). Einige Überlegungen zu (Fehl-)Entwicklungen der psychologischen Methodenlehren. **Psychologische Rundschau, 41**, 84-107.

## Autoren-Register

- Abel, U. 513, 548  
Abelson, R.P. 204, 247  
Achinstein, P. 433  
Adams, E. W. 590, 598  
Aebli, H. 199, 243  
Aertsen, A. 192, 195  
Agassi, J. 400, 422, 424  
Agazzi E. 283, 289  
Agnew, N.M. 6, 42  
Aiserman, M.A. 349, 386  
Aitkin, M. 640, 644  
Albert, D. 645  
Albert, H. 254, 289, 314,  
317, 337, 390, 391, 424  
Alich, L.-M. 272, 289  
Allen, J. 204, 243  
Allert, T. 118, 134  
Amico, R.P. 445  
Andersen, E.B. 17, 42,  
661, 685, 687  
Anderson, D. 640, 644  
Anderson, G. 55, 92  
Anderson, J.R. 203, 204,  
215, 217, 239, 243,  
296, 337, 367, 387,  
404, 424  
Anderson, N.H. 568, 598  
Andersson, G. 259, 269,  
282f, 286, 289, 399,  
424  
Andres, J. 639, 644, 648  
Anstis, S. 161, 191  
Anzai, Y. 203, 243  
Arbuckle, J. 572, 598  
Ardila, R. 58f, 65, 92  
Aristoteles 168, 347, 387  
Arminger, G. 526, 539f,  
548, 553  
Armstrong, B. D. M. 304,  
337  
Aronson, E. 508ff, 526,  
534f, 539, 548  
Ash, R.A. 658, 687  
Ashby, F.G. 640, 647  
Ashby, W.R. 368, 369, 387  
Asher, L. 196  
Atkinson, J. W. 255, 289,  
296f, 313, 337, 397,  
404, 410f, 419, 424  
Atkinson, R.C. 195f, 425  
Austin, J.L. 116, 130  
Austin, J.T. 653, 688  
Ax, A. F. 543, 548  
Baars, B.J. 321, 337  
Baton, F. 419, 431  
Baddeley, A. 607, 644  
Baird, J.C. 569, 598  
Bakan, D. 625, 644, 732,  
734  
Balestra, D.J. 443  
Balzer, W. 261f, 290,  
332, 334, 337f, 430,  
457, 650, 687, 688  
Bandura, A. 528, 548  
Barber, T.X. 66, 92  
Barker, P. 283, 286, 291  
Barker, R. G. 53, 92  
Barker, S. 433  
Barlow, H.B. 165, 167,  
175, 192  
Baron, J. 17, 42  
Baron, R. 172, 196  
Barr, W.F. 322, 338, 413,  
425  
Bartley III, W. W. 400,  
423, 425  
Bartussek, D. 669, 688  
Bassok, M. 227, 243  
Batchfelder, W. H. 635,  
639, 644, 647  
Battig, W.F. 404, 425  
Bauer, H. 656, 658, 661,  
663, 665, 688  
Baumrind, 518  
Bayertz, K. 283, 290  
Beatty, J. 734  
Behnke, H. 734  
Bell, D.E. 11, 42  
Bern, D.J. 416, 425  
Bennett, B.M. 174, 192  
Bentler, P.M. 484, 506,  
540, 549, 653, 688  
Berger, P.L. 108, 130  
Bergmann, J.R. 103, 120,  
131  
Berkeley 145, 148  
Berkowitz, L. 400, 410,  
425  
Bernouilli 169, 564  
Bertalanffy, L. 355, 356,  
387  
Berwick, R.C. 204, 243  
Bickel, P. J. 734  
Birch, D. 297, 337, 404  
Birdsall, T. G. 10, 44  
Birkett, 513  
Bischof, N. 67, 92, 297,  
338, 349, 354, 358,  
361, 362, 387  
Bishop, Y. M. M. 661, 688  
Blake, A. 196  
Bless, H. 522f, 543, 549,  
554  
Blettner, M. 513, 555  
Block, N. 321, 338  
Bock, R.D. 580, 599  
Boehm, G. 100, 131f  
Boehnke, K. 733, 734  
Boesken, W.H. 526, 549  
Boguslaw, R. 368, 387  
Böhme, G. 560, 598  
Bohnen, A. 426  
Bohner, G. 543, 549  
Bohnsack, R. 103, 131  
Bollen, K.A. 682, 688  
Borg, I. 70, 92  
Borgman, C. L. 287, 292  
Boring, E.G. 143f, 192  
Bortz, J. 20, 21, 41, 42,  
351, 361, 387, 478,  
506, 733, 734

- Bosch, K. 658, 688  
 Bourbaki, N. 562, 599  
 Bower, G.H. 276f, 291  
 Box, G.E. 514, 549  
 Boyd, R. 170f, 192  
 Brachman, R. 204, 243  
 Braddick, O.J. 192, 197  
 Bradley, R.A. 9, 42  
 Bradshaw, 451  
 Braitenberg, V. 192  
 Braithwaite, R. B. 368, 369, 387  
 Brandtstädter, J. 291  
 Brauer, W. 216, 243  
 Breidenkamp, J. 22, 40, 41, 63, 76, 82, 85f, 92, 198, 226, 230, 457, 462, 481, 491, 496, 501f, 505f, 508f, 513f, 549, 605, 608ff, 620, 625, 627ff, 633, 638, 644, 649, 667, 669, 672, 688, 694, 706, 719, 728, 730f, 734, 735, 736  
 Brehm, J. W. 397, 425  
 Breiman, L. 658, 688  
 Brentano F. 143, 570, 599  
 Breuer, E. 251, 261, 290  
 Brewer, M. 539, 548  
 Bridgman P. W. 56, 92, 309, 338  
 Brinberg, D. 82, 92  
 Briskman, L.B. 257, 283, 290  
 Brocke, B. 50, 92, 272, 290  
 Brown, J.S. 221, 228, 243, 248, 688  
 Browne, M.W. 683  
 Brozek, J. 144, 192  
 Brunswick, E. 137, 157, 159f, 192  
 Buchner, A. 76f, 92, 627, 646  
 Bude, H. 112, 131  
 Buffart, H.F. 194  
 Bugelski, B.R. 607, 645  
 Bühler, K. 55f, 92, 100, 109, 131, 137, 157, 159  
 Buhmann, R. 102, 131  
 Bunge, M. 58f, 65, 92, 269, 272ff, 290, 306, 322, 331, 338, 393f, 400, 405, 425  
 Büning, H. 706, 733, 734  
 Bunt, A.A. 573, 600  
 Buring, I.E. 528f, 551  
 Burke, P. J. 661, 690  
 Burnstein 404  
 Burton, R.R. 229, 243  
 Bush, R.R. 23, 42, 206, 245  
 Buss, A.H. 544, 549  
 Butterfield, E. C. 265, 283, 286, 290, 426  
 Calder, B.J. 530, 551  
 Campbell, D.T. 82f, 86f, 92f, 440, 459, 512, 521, 526f, 534f, 549, 555, 605, 645  
 Campbell, N.R. 563  
 Cannell, C.F. 549  
 Cantril, H. 535, 551  
 Carbonell, J.G. 242, 243  
 Carlsmith, J.M. 508, 518, 535, 539, 543, 548f, 608, 646  
 Carnap, R. 42, 81f, 105, 131, 306f, 338, 396, 425, 429, 434, 438, 444, 455, 480, 506, 725  
 Carrier, M. 139, 192, 236, 243, 320, 338  
 Cartwright-Smith, J. 542, 552  
 Cassirer, E. 108, 131, 174, 192  
 Chalmers, A.F. 13, 18, 43, 311, 338, 509, 549  
 Chang, J.J. 173, 194  
 Chapanis, A. 368, 378, 384, 387  
 Chapman, L.J. 534, 549  
 Chapman, M. 60, 92  
 Chapman, D.T. 534, 549  
 Chassein, B. 522, 554f  
 Cheek, J.M. 669, 688  
 Chi, M. T. H. 204, 227, 243, 248  
 Chomsky, N. 118, 131, 204, 215, 243  
 Christensen, L. 545, 549  
 Church, A. 215, 243  
 Churchland, P.M. 319, 338, 394, 425  
 Churchland, P. S. 166, 192  
 Claparède, E. 61, 92  
 Clark 404  
 Clement, J. 454  
 Cliff, N. 599  
 Coen, C. W. 192  
 Cohen, J. 497, 506, 627, 629f, 632, 645, 717, 734  
 Cohen, L.J. 431f, 434f, 437f, 441, 449, 461  
 Cohen, M.A. 172, 180, 192  
 Cohen, M.R. 589, 599  
 Cohen, R.S. 426  
 Cohen, Y. 433, 449  
 Coker, M.C. 551  
 Collingwood, R. G. 331, 338  
 Collins, A.M. 203, 204, 208, 244  
 Colonius, H. 575, 578, 599, 687, 688  
 Comte, A. 57, 93  
 Conway, M. 523, 553  
 Cook, D.A. 521, 528f, 537, 551  
 Cook, T.D. 82, 93, 530, 549f  
 Cooley, C.H. 99, 131  
 Coombs, C.H. 15, 20, 43, 48, 52, 69f, 93, 326, 338, 560, 572, 601, 729, 734  
 Corwin, J. 636, 638, 647  
 Cox, J. W. R. 433  
 Craik, K. J. W. 170, 192  
 Crane, D. 287, 290  
 Cronbach, L.J. 637, 645  
 Cronbach, N.J. 313, 338  
 Crott, H. 265, 290

- Crowne, D. 525, 549  
 Cummins, R. 328, 338  
  
 Daiser, W. 217, 244  
 Dalbert, C. 526, 549  
 Damerow, P. 13, 14, 43  
 Danziger, K. 58f, 93  
 Darlington, R.B. 669, 688  
 Daston, L. 734  
 Daugman, J.G. 171, 192  
 Daves, R.M. 729, 734  
 David, F.N. 734  
 Davis, E. 204, 244  
 Dawes, R. M. 15, 43, 69, 326, 338  
 De Groot, A.D. 7, 12, 26, 43  
 De Gruijter, D.N.M. 654, 688  
 De Olivera, M.B. 446  
 Dearing, B.E. 79f, 94  
 Debreu, G. 564, 599  
 DeGreef, E. 687, 688  
 Del Rosario, M.L. 530, 551  
 Demets, D.L. 513, 550  
 Demokrit 401  
 Dennett, D.C. 139, 161, 171, 192  
 Deppe, W. 206, 244, 643, 645, 733, 734  
 Descartes, R. 13, 14, 43, 127f, 131, 561  
 DeValois, K.K. 175, 192  
 Diederich, W. 687, 688  
 Diepgen, R. 720, 722, 734  
 Diercks, S.R. 551  
 Dilthey, W. 102, 116f, 131  
 Dingler, H. 24, 43, 54, 93  
 Dittrich, K. 547, 549  
 Dollard, J. 315  
 Dörner, D. 16, 38, 66f, 76f, 81, 91f, 227, 244, 344, 358, 374, 378, 388, 479, 506, 519, 535f, 549  
 Dreistadt, R. 454  
 Dreyfus, H. 238, 244  
 Drösler, J. 567, 599  
  
 Duhem, P. 179f, 193, 408, 425, 643f  
 Dunbar, K. 451  
 Dupr{, J. 197  
 Durkheim, E. 128, 131  
 Dux, G. 134  
 Dykstra, R.L. 602  
  
 Eagly 528  
 Ebbinghaus, H. 257, 290  
 Eberle, T.S. 99, 131  
 Edgington, E. S. 513f, 550, 610, 645  
 Edidin, A. 446  
 Edler, L. 513, 555  
 Edwards, W. 726, 729, 734  
 Efron, B. 81, 93, 642, 645  
 Egan, J.P. 10, 43  
 Ehrhardt, K. J. 374, 387  
 Eid, M. 654, 658, 661, 663, 665, 676, 678, 681, 682, 693  
 Eimer, M. 328f, 338  
 Einhorn, J.H. 338  
 Einstein, A. 410  
 Ekstrand, B.R. 411, 427  
 Eller, F. 332  
 Ellsworth, P.C. 508, 548  
 Elpelt, B. 689  
 Engelkamp, J. 38, 43  
 Epstein, W. 54, 93  
 Erdfelder, E. 22, 35, 40, 41, 76, 82, 85f, 226f, 230, 457, 649, 667, 669, 672, 676, 688  
 Erdfelder, E. 605, 615f, 618, 627, 635, 638f, 645f  
 Ericsson, K.A. 55, 60f, 65, 93  
 Eron, L.D. 529, 552  
 Escolana, S.K. 298, 338  
 Essler, W.K. 251, 290, 430, 434f, 625, 645  
 Estes, W. K. 286, 290  
 Euklid 562  
 Eysenck, M. W. 55, 62, 66f, 93  
  
 Fagot, R.F. 590, 598  
 Fahrenkrug, H. 532, 550  
 Fallon, A.E. 543, 553  
 Falmagne, J.-C. 137, 146, 193, 567, 575f, 578f, 581, 585, 590, 594, 599, 620, 645, 676, 688  
 Faraday, M. 417  
 Farrell, E. 286, 290  
 Faßnacht, G. 51f, 93  
 Faul, F. 627, 633, 646  
 Fazio, R.H. 521f, 550f  
 Feather, N.T. 300, 338, 420, 425  
 Fechner, G.T. 137, 140ff, 144f, 146, 150, 156, 165, 168f, 192f, 558, 561, 564, 566, 569f, 574ff, 599  
 Feger, H. 51f, 55f, 70, 93, 198ff, 508, 550, 645f, 735, 736  
 Festinger, L. 296, 298, 535, 543, 550, 608, 646  
 Feyerabend, P. 6, 43, 68, 94, 317, 337f, 391, 393f, 399, 407, 425  
 Fichter, M.M. 532, 550  
 Fienberg, S.E. 661  
 Fischer, G.H. 654, 676, 685, 687, 689  
 Fishburn, P.C. 582, 599  
 Fisher, R. A. 511f, 550, 610, 695, 701, 704f, 709f, 717, 722, 727, 729, 733, 734  
 Flick, U. 114, 131f  
 Flores, F. 238, 248  
 Fodor, J. 320, 339  
 Fodor, J. A. 118, 133, 161, 166, 193  
 Foppa, K. 67, 94  
 Formann, A. 676, 689, 358  
 Fox, W. C. 10, 44  
 Franklin 183, 193  
 Franz, H.-W. 131  
 Freitag, B. 317, 341  
 Frenz, H.-G. 530, 550  
 Freudenthal, G. 13, 43

- Freudenthal, H. 707, 734  
 Frey, D. 414, 425, 645, 647  
 Frey, S. 530, 550  
 Friedeburg, L.V. 134  
 Friedmann, L.M. 513, 550  
 Frings, M.S. 134  
 Funke, J. 76f, 92  
 Furberg, C.D. 513, 550  
  
 Gabriel, K.R. 610, 646  
 Gadamer, H.-G. 100, 102, 131f  
 Gadenne, V. 16, 38, 55f, 60f, 75, 82f, 82f, 94, 177f, 236, 251, 290, 312, 314, 321, 325, 329, 336, 339, 357, 358, 395, 414f, 425, 440, 457f, 481, 491, 506, 509f, 514f, 531, 537, 540, 550, 604f, 611, 613, 643, 646  
 Gage, N.L. 645  
 Gähnde, U. 687, 689  
 Gaifman, H. 443, 450, 454  
 Galanter, E. 206, 245, 321, 340, 352, 388, 567f, 574, 580  
 Galilei, G. 282  
 Ganes, P.A. 80, 94  
 Gänssler, P. 658, 689  
 Gärdenfors, P. 433, 449  
 Gardner, H. 203, 244  
 Garfield, E. 287, 290  
 Garner, W.R. 174, 193  
 Garrett, M. 197  
 Garz, D. 103, 132ff  
 Gaston, J. 287, 290  
 Geissler, H.G. 197  
 Genesereth, M. R. 205, 208, 244  
 Gentner, D. 286, 290  
 Gerhardt, U. 103, 132  
 Gerjets, P. 33, 39, 81, 84, 465, 509, 605, 651  
 Geyer, S. 531f, 550  
 Ghiselli, E.E. 669  
  
 Gholson, B. 283, 286, 291  
 Gibson, J. J. 137, 155, 159ff, 163, 172, 193  
 Giddens, A. 103, 132  
 Gigerenzer, G. 33, 43, 49, 67, 73, 94, 171, 193, 203, 205, 244, 276, 285f, 291, 366, 386, 387, 579, 599, 696, 717, 722f, 731, 734  
 Glaser, B.G. 111f, 132  
 Glaser, R. 227, 243  
 Glück, G. 51, 94  
 Glymour, C. 336, 339, 398, 425  
 Gödel, K. 215, 244  
 Goffman, E. 128, 132  
 Goldstein, I. 320, 339  
 Gollwitzer, P.M. 547, 550  
 Gonzales, M.H. 508, 548  
 Good, I. J. 80, 94  
 Goodman, L.A. 661, 689  
 Goodman, N. 441, 445f, 455  
 Gopnik, A. 60, 94  
 Gora, A. 191  
 Götz-Baltes, B. 668, 692  
 Gould, S. J. 171, 193, 559, 599  
 Grabitz, H. J. 397, 425  
 Grabitz-Gniech, G. 397, 425  
 Graham, N. 175, 188, 193  
 Grassmann, H. 179f, 193, 599  
 Grathoff, R. 103, 107, 132  
 Graubard, S.R. 151, 193  
 Graumann, C.F. 51, 55f, 93f, 281, 291, 292, 321, 339  
 Grawe, K. 35, 43  
 Green, D.M. 58, 94, 480, 507, 580, 637, 646  
 Greeno, J. 221, 248, 569, 602  
  
 Greve, W. 51f, 94  
 Grice, H.P. 522, 550  
 Griffin, D.R. 193  
 Groeben, N. 48, 56, 65, 94, 251, 291, 304, 317, 326, 341, 477, 485, 491f, 506, 604, 610, 646  
 Grondin, J. 100, 132  
 Grossberg, S. 172, 176, 180, 192f, 197  
 Grünbaum, A. 443  
 Grundmann, A. 56, 97  
 Gschneidinger, E. 523, 554  
 Gulliksen, H. 676, 677, 678, 689  
 Gundlach, H. 144, 192  
 Gurwitsch, A. 107, 132  
 Gussew, L.A. 386  
 Guthrie, W. K. C. 595, 600  
  
 Haagen, K. 489, 506, 710, 735  
 Habermas, J. 103, 126, 132, 134  
 Hacker, W. 270, 273, 291  
 Hacking, I. 183f, 193, 195, 694, 707, 715, 735  
 Hager, W. 87f, 94, 442, 457, 462, 478f, 505f, 487f, 494ff, 508, 551, 605, 609, 612ff, 618, 623, 627, 632, 635, 646f, 651, 689, 694, 723, 724, 732, 735, 737  
 Hambleton, R. H. 676, 688  
 Hanneman, R. A. 376, 383, 387  
 Hänni, R. 35, 43  
 Hanson, N.R. 185, 193  
 Harnatt, J. 493, 507  
 Hartigan, J.A. 80, 95  
 Härter, E. 658, 689  
 Hartung, J. 661, 689, 702, 735  
 Hartwig, F. 79f, 94  
 Hastorf, A.H. 535, 551

- Haugeland, J. 328, 339  
Hauptert, B. 103, 132  
Häubler, A. 531, 551  
Hayduk, L.A. 682, 689  
Hays, W.L. 495, 500, 507  
Heath, L. 530, 551  
Hebb, D.O. 268  
Heckhausen, H. 255, 257, 286, 291, 299, 316, 339  
Heckhausen, T. 69f, 94  
Heidbrink, H. 283, 291  
Heidegger, M. 102, 132  
Heider, F. 157, 159, 172, 194  
Heinze, T. 135  
Heise, E. 465  
Hejl, P.M. 259f, 291  
Hehing, I. 104, 132  
Hempel, C. G. 325f, 339, 434, 437, 441, 483, 507  
Henkel, R.E. 493, 507, 714, 695, 735  
Hennekens, C. H. 528f, 551  
Hennigan, 539, 530  
Henry, N. W. 639, 646  
Herbart 142, 168, 174  
Hering, E. 137, 143, 150, 157, 159, 165, 172, 194  
Herr, P.M. 524, 551  
Herrmann, C. 53, 87, 96  
Herrmann, T. 24, 38, 43, 50f, 76, 91f, 183f, 187, 191, 194, 202, 218, 244, 252, 255f, 261, 263, 265, 269, 271, 273, 275, 281, 284, 286, 291, 293, 315, 328f, 332, 339, 349, 358, 361, 387, 394f, 538, 551, 558, 634, 646, 695, 735  
Herrnstein, R. J. 195  
Hesse, M. 433f, 441, 446, 450, 454  
Hesse, M.B. 363, 387  
Heyer, D. 582, 585f, 600  
Hickey, T. 87, 95  
Hildenbrand, B. 103, 108, 132  
Hildreth, E.C. 204, 244  
Hilgard, E.R. 276, 277, **291**  
Hinde, J. 640, 644  
Hinderer, K. 658, 689  
Hintikka, J. 437  
Hippler, 522f, 554f  
Hirschauer, S. 108, **132**  
Hitzler, R. 35, 58, 66, **72**, 107f, 130, 132f, 227  
Hoagiin, D.C. 79, 95  
Hobbes, T. 169  
Hodapp, V. 540, 551  
Hofer, M. 476, 507  
Hoffman, D.D. 161, 174f, 192, 194, 196  
Hoffman, W.C. 174, 194  
Hoffmann, N. 294  
Hoffmeister, H. 531, 551  
Hofstätter, P. 361, 388  
Hogarth, R.M. 338  
Hölder, O. 563, 600  
Holland, J. 203, 244, 449ff  
Holland, P.W. 661  
Hollerbach, J.M. 203, 246  
Holling, H. 687, 690  
Holmann, E. W. 580, 600  
Holyoak, K.J. 203, 244, 449ff  
Holzkamp, K. 24, 43, 55, 88, 95  
Honer, A. 103, 107, 114, 132f  
Hookway, C. 192  
Hopcroft, J.E. 216, 244  
Hoppe, H. 433  
Hörmann, H. 281, 285, 291  
Hosmer, D.W. 539, 551  
Hsu, C.-E 610, 646  
Hubbard, M. 545f, 553  
Hubel, D.H. 176, 183, **195**  
Huber, O. 484, 507, 533, 551  
Hubert, L. 77, 95  
Hübner, K. 441  
Huesman, L.R. 529, 552  
Huh, C.L. 277, 303, **340, 404**  
Hultsch, D.F. 87, 95  
Hume, D. 84, 95, 145, 148, 432  
Hummell, H.J. 511, 551  
Hunter, J.S. 514, 549  
Hunter, W. G. 549  
Husserl, E. 107, 130, 133  
Hussy, W. 40, 41, 82, 87f, 254, 291, 457, 479, 484, 507, 516, 618, 635  
Hyland, M. 372, 374, 388  
Irlle, M. 253, 291, 397, 425, 544, 551  
Irmen, L. 648  
Irwin, M.E. 551  
Isefer, A. 648  
Iser, W. 102, 133  
Iverson, G. J. 620, 646  
Iverson, J. 582, 600  
Jackendoff, R. 316, 321, 340  
Jacob, F. 181, 194  
Jagodzinski, W. 687, 689  
James, W. 18, 43, 55, 162, 194  
Jarvie, I.C. 426  
Jason, G. J. 431  
Jauß, H.R. 102, 133  
Jepson, D. 453  
Johnson, E. J. 61, 96  
Johnson, N.L. 626, 646  
Johnson-Laird, P. N. 202f, 244, 373, 388  
Jones, G.E. 441f, 446  
Jones, L.V. 580, 599  
Jones, R.M. 204, 227, 248  
Jöreskog, K.G. 484, 507, 540, 551, 638f, 646, 682, 685, 690  
Josephson, W.L. 539, 551  
Judd, C.M. 669, 690  
Julesz, B. 173, 183, 194

- Kächele, H. 25, 43  
 Kahn, R.L. 549  
 Kahneman, D. 555  
 Kalish, D. 427  
 Kalke, W. 151, 153, 194  
 Kallmeyer, W. 119, 133  
 Kant, I. 57, 95, 696  
 Kantowitz, B.H. 196  
 Katz, D. 150, 185, 194  
 Katz, J.J. 118, 133  
 Kaufmann, L. 59, 95  
 Keane, M.T. 67, 93  
 Keith-Spiegel, P. 542, 551  
 Kelvin 180  
 Kemeny, J.G. 432, 434  
 Kemke, C. 174, 194  
 Kendall, M.G. 709, 735  
 Kendlner, T.S. 87, 95  
 Kennedy, J.J. 539, 551  
 Kenny, D.A. 669, 690  
 Kepler, J. 410  
 Keppler 177  
 Kern, D. 555  
 Keupp, H. 114, 131f  
 Klages, H. 267f, 271, 291  
 Klahr, D. 451  
 Kleck, R.E. 542, 552  
 Kleene, S.C. 215, 244  
 Kleiner, B. 80, 95  
 Kleiter, G.D. 726, 727,  
 729, 730, 735  
 Klir, J. 373, 388  
 Klix, F. 207, 245, 286,  
 291  
 Klopfer, B. 515, 551  
 Klösener, K.-H. 689  
 Klumpp, G. 523, 554  
 Klusemann, H. W. 135  
 Kluwe, R.H. 61, 95  
 Knauth, B. 103, 133  
 Knoblauch, H. 103, 108,  
 133  
 Knoche, N. 678, 690  
 Knoke, D. 661, 690  
 Knorr-Cetina, K. 103,  
 108, 114, 133  
 Knowles, E. 526, 551  
 Kocher, G.P. 551  
 Kodratoff, Y. 242, 245  
 Koenderink, J. J. 161,  
 175, 194  
 Koeppler, K. 485, 491,  
 507  
 Koh, K. 451  
 Kohler, I. 171, 194  
 Köhler, W. 172  
 Kohlmeier, L. 526, 531,  
 551  
 Kolmogoroff, A. 651, 690  
 Konau, E. 118, 134  
 König, R. 126, 133  
 Koocher, 542  
 Kopernikus, N. 412  
 Kordig, D.R. 283, 291  
 Kosslyn, S.M. 203, 246  
 Kotz, S. 626, 646  
 Kraiker, C. 332, 340  
 Kraimer, K. 103, 132ff  
 Krambeck, J. 118, 134  
 Krämer, S. 169, 194, 215,  
 245  
 Krantz, D.H. 164, 175,  
 186, 194, 196f, 453,  
 564, 568f, 571, 580,  
 587, 590, 600, 603,  
 626, 647, 676, 691,  
 730, 735  
 Krapp, A. 476, 507  
 Krauth, J. 661, 690  
 Krickeberg, K. 658, 690  
 Kriz, J. 283, 291  
 Krohn, W. 599  
 Krohne, H. W. 263ff,  
 272, 274, 291  
 Kronthaler, E. 432, 434,  
 447  
 Krüger, L. 734  
 Kruglanski, A. W. 398,  
 426  
 Krull, W. 562, 600  
 Krumhansl, C.L. 145,  
 195, 567, 576  
 Kruse, L. 542, 552  
 Kruse, P. 260, 293  
 Kruskal, W.H. 661, 689  
 Kubinger, K. 654, 676, 690  
 Kubovy, M. 172, 187,  
 195f  
 Kuhn, T.S. 184, 187,  
 190, 195, 282f, 285,  
 292, 317, 336, 340,  
 398f, 407f, 414, 426,  
 453f, 587, 600  
 Kuipers, T. A. F. 317,  
 340, 437  
 Kulkarni, D. 451  
 Külpe, O. 314, 340  
 Kummer, W. 434  
 Kumpf, M. 542, 552  
 Kunda, Z. 453  
 Kyburg, H.E. 433, 437  
 La Mettrie, J.O. de 169,  
 195  
 Lachman, J.L. 265, 282,  
 292, 426  
 Lachman, R. 265, 276,  
 282f, 286, 292, 410,  
 426  
 Laird, J. E. 204, 215, 245  
 Lakatos, I. 67, 81f, 86,  
 95, 252, 263, 283, 292,  
 407, 412, 418f, 425,  
 435ff, 605, 609, 646,  
 696, 711, 724, 735  
 Laming, D.R. J. 147,  
 195, 580, 600  
 Land 183  
 Lane, N.R. 440  
 Lane, S.A. 440  
 Langeheine, J. 639, 647  
 Langeheine, R. 661, 676,  
 685, 690, 691  
 Langer, S. K. 108, 133  
 Langley, P.A. 451  
 Lantermann, E.D. 67,  
 93, 519, 535f, 549  
 Lanzetta, J.T. 542, 552  
 Larimer, J. 572, 598  
 Larsen, R. J. 544, 552  
 Lashley, K. 166  
 Lasniak, H. 203, 246  
 Lau, T. 103, 108, 133  
 Laudan, L. 283, 292,  
 398, 426, 431  
 Lazarsfeld, P.F. 639, 646  
 Leahey, T.H. 348, 355,  
 359, 388

- Leeper, R. 54, 95  
 Leeuwenberg, E. L. J. 194  
 Lefkowitz, M.M. 529, 552  
 Lehmann, G. 678, 690  
 Leibniz 142, 168f  
 Leiser, E. 724, 735  
 Lemeshow, S. 539, 551  
 Lenzen, W. 438, 441, 444, 694, 735  
 Leontjew, A.N. 270, 273, 292  
 Lepper, M.R. 480f, 507, 545f, 553  
 Levelt, W. J. M. 276, 292, 573, 579, 600  
 Levesque, H. 204, 243  
 Levi, I. 435f, 449  
 Levine, M.V. 63f, 95, 576, 600  
 Levinson, A. 441  
 Lewin, K. 174, 296, 340  
 Lewis, C. 653, 676, 690  
 Lewis, D. 331, 340  
 Lewis, M.W. 227, 243  
 Lewontin, R.C. 171, 193  
 Liebmann, S. 184, 195  
 Lienert, G.A. 633, 647, 661, 676, 690, 733, 734  
 Lihi, W. 521, 553  
 Lindgren, B.W. 701, 706, 735  
 Lindman, H. 726, 734  
 Lindzey, G. 195  
 Lingle, J.H. 523, 552  
 Link, S.W. 197  
 Litwin, G.H. 300, 337, 411, 424  
 Livingstone, M. S. 176, 183, 195  
 Locke 145, 148  
 Loeve, M. 658, 690  
 Loftus, E.F. 204, 244  
 Lord, F.M. 9, 43, 676, 677, 678, 691  
 Lucas, R. J. 170, 195  
 Luce, R.D. 8f, 43, 137, 145, 164, 195ff, 206, 245, 564, 567f, 571, 574f, 576ff, 590ff, 600, 603, 626, 640, 647, 676, 690, 691, 730, 735  
 Lück, H.E. 287, 291f  
 Luckmann, B. 107, 133  
 Luckmann, T. 100, 103, 106ff, 111, 124, 130, 133f  
 Lüders, C. 64, 95, 103, 133, 207, 246, 646f  
 Lundeen, E. J. 551  
 Lykken, D.T. 724, 735  
 Lyons, W. 55f, 95, 340  
 Mac Cormack, E.R. 168, 195  
 Mach, E. 143, 150, 157ff, 165, 182, 195  
 Machina, K.F. 430  
 Mackie, J.L. 330, 340, 441  
 Makov, U.E. 639, 647  
 Malewski, A. 325, 340, 398, 416, 426  
 Malpas, J. 211, 245  
 Mandler, G. 321, 340  
 Manns, M. 53, 96  
 Marcovici, S. 620, 646  
 Mark, M.M. 526, 552  
 Markov, A. A. 215, 245  
 Marley, A.A. 580  
 Marlowe, D. 525, 549  
 Marquard, O. 104  
 Marr, D. 137, 152f, 158, 161, 181, 195, 204, 217, 245  
 Martin, G. 562, 601  
 Martin, L.L. 522, 554  
 Martinez, T. 174, 196  
 Marx, M. 369, 388  
 Masterman, M. 282, 292  
 Matthiesen, U. 103, 134  
 Maturana, H.R. 260, 292, 317, 340  
 Mausfeld, R. 34, 42, 73f, 89, 163, 177, 181, 195, 560, 566, 585f, 596f  
 McClelland, D. C. 255, 268, 292, 297, 313, 340  
 McClelland, G. H. 572, 601  
 McClelland, J. A. 206, 245, 247  
 McClelland, J. L. 451  
 McCorquodale, K. 303, 314f, 340  
 McGrath J.E. 82, 92  
 McGuigan, F.J. 484, 507  
 McLaughlin, P. 13, 43  
 Mead, G.H. 100, 119, 124, 134, 286, 292  
 Meehl, P.E. 87, 96, 303, 313f, 338, 340, 491, 497, 507, 637, 645, 724, 735  
 Mehler, J. 197  
 Meja, V. 135  
 Mergenthaler, E. 25, 43  
 Merkel, J. 570, 601  
 Merton, R.K. 287, 292  
 Merz, F. 552  
 Metz, G. 520, 552  
 Metzger, W. 276, 278, 292, 552  
 Meyer, W.U. 255, 268, 292  
 Middleton, D. 10, 44  
 Milgram, S. 518, 543f, 552  
 Mill, J.S. 419, 431, 508, 552, 669, 691  
 Miller, D. 404, 423, 426, 443f  
 Miller, G.A. 315, 321, 340, 352, 388  
 Mills, J. 519, 548  
 Minsky, M. 204, 215, 245  
 Mittelst&, J. 139, 192, 236, 243, 320, 338, 561, 601  
 Mittenecker, E. 725, 735  
 Möbus, C. 207, 245, 685, 691  
 Mohyeldin Said, K.A. 151, 195f  
 Möller, H. 40f, 82, 87f, 457, 484, 497, 506f, 516, 618, 635  
 Mondriaan 183



- Montada, 265, 292  
 Moosbrugger, H. 669, 691  
 Moravcsik, J.M. 206, 225, 245  
 Morgans, L. 404  
 Morgenstern, O. 564  
 Morrison, D.E. 493, 507, 714, 735  
 Mosteller, F. 23, 42, 79, 95f  
 Moulines, C.U. 430, 457, 650, 687, 688  
 Moulton, R. W. 301, 340  
 Mulaik, S.A. 653, 691  
 Müller, G. 522, 540, 554  
 Müller, G.E. 143, 196  
 Müller, P.H. 665, 691  
 Müller-Doohm, S. 103, 134  
 Mura, A. 443  
 Murray, D.J. 33, 43  
 Murray, H.A. 297, 316f, 340, 404, 412, 419f, 423, 425  
 Musgrave, A. 283, 292, 646, 735  
 Muthig, P. 485, 491, 507  
  
 Nagel, E. 7f, 29, 43, 433, 589, 599  
 Nag], W. 685, 691  
 Narens, L. 564, 571, 574, 582, 590f, 594f, 599, 601f  
 Nassen, U. 100, 34  
 Naß, E. 531, 551  
 Nebel, A. 522f, 552, 555  
 Neches, R. 206, 245  
 Neel, A.F. 349, 388  
 Neisser, U. 534, 552  
 Nell, K. 332  
 Nelson, R. J. 26f, 43, 44  
 Neville, J. W. 551  
 Newell, A. 7, 29, 44, 203f, 213f, 217f, 220, 237, 245f, 286, 292, 320, 340, 385, 388, 451f  
 Newton, I. 179, 282, 401, 412, 419, 696  
 Newton-Smith, W. H. 195f  
 Neyman J. 507, 609, 611, 694, 702, 704f, 710f, 716f, 731, 732, 735, 736  
 Niederée, R. 163, 195, 562, 565, 575, 582, 586, 602  
 Niiniluoto, I. 437f, 441  
 Nilsson, N.J. 205, 208, 244  
 Nisbett, R.E. 57, 62, 96, 203, 244, 404, 426, 432, 448ff, 462, 480, 507  
 Noack, H. 543, 553  
 Noma, E. 569, 598  
 Norman, D. A. 296, 341  
 Novick, M.R. 9, 43, 676, 677, 678, 691  
 Nowack, W. 539, 552  
 Nüse, R. 317, 341  
 Nuwer, K. 172, 196  
  
 Occam, W. 403  
 Oelkers, J. 135  
 Oerter, R. 265, 292  
 Oevermann, U. 103, 111, 118, 120f, 134  
 Ohlsson, S. 227, 246  
 Oldenbürger, H.-A. 78, 96  
 Opp, K.D. 484, 507  
 Opwis, K. 25, 36, 199, 203, 207, 223, 246, 451  
 Orne, M. 515, 552  
 Ortony, A. 168, 192, 196  
 Osherson, S. 203, 246  
 Ostmann, A. 33, 40, 41, 42, 86, 609, 615, 684  
 Ostrom, T.M. 523f, 552f  
 Oswald, M. 55f, 59f, 63, 96, 320f, 329, 339f, 395, 414, 425  
  
 Pachella, R. G. 178, 196  
 Pähler, K. 418, 426  
 Palacios, A. 198  
 Palermo, D. S. 283, 293  
 Palm, G. 195  
 Palmer, S.E. 28f, 44  
 Papert, S. 320  
 Parducci, A. 524, 553  
 Parker, S. 565, 602  
 Paul, W.J. 44  
 Pawlow, J.P. 268, 359, 371, 388, 410f, 426  
 Pearson, E.S. 507, 609, 611, 695, 716f, 731, **736**  
 Pearson, K. 704, 709, 710, 716  
 Penner, L.A. 669, 691  
 Penrose, R. 151, 170, **196, 238, 246**  
 Pentland, A.P. 196  
 Perry, C. 441f  
 Petefi, J.S. 293  
 Peterson, W. W. 10, 44  
 Pfanzagl, J. 565, 574, 590, 595, 602  
 Pfeifer, E. 378, 387  
 Pinker, S. 204, 246  
 Plachky, D. 658, 691  
 Planck, M. 140  
 Plateau, M. J. 565, 570, 602  
 Plato 168  
 Plessner, H. 107, 134  
 Plötzner, R. 223, 246  
 Pocock, S.J. 553  
 Podlech, A. 293  
 Poggio, T. 161, 196  
 Pollock, J.L. 430  
 Pomerantz, J.R. 172, 195f  
 Pongratz, L.J. 55, 60, 96  
 Poppen K. 7, 8, 44, 54f, 75, 81f, 86, 96, 126, 134, 254, 292, 314, 325, 336, 341, 343f, 388, 396f, 399f, 407f, 411, 419f, 426, 429, 432, 437, 440f, 443f, 455, 461, 604, 647, 649, 691, 695, 725, 731, 736  
 Porter, T. 734

- Posner, M.I. 29, 44, 203,  
**246**
- Post, E. 215, 246
- Prakash, C. 174, 192
- Prell, S. 476, 507
- Pribram, K.H. 172, 196,  
321, 340, 352, 388
- Price, L. 515, 553
- Pritsch, M. 543, 553
- Pugh, A.L. 384, 388
- Putnam, H. 139, 150f,  
166, 170, 196, 320,  
341, 441
- Pyke, S. W. 6, 42
- Pylyshyn, Z.W. 152,  
161, 170, 190, 193,  
196, 213, 217, 237,  
238, 246, 285, 292,  
320, 341
- Pythagoras 150
- Quillian, M.R. 204,  
208, 246
- Quine, W.V. 157, 196,  
643f
- Radnitzky, G. 426
- Raiffa, H. 11, 42
- Ramachandran, V. S. 155,  
161, 181, 186, 196
- Ramsey, F.P. 564, 602
- Ramsey, J.O. 594, 602
- Rapaport, D. 303, 341
- Rasch, G. 654, 655, 691
- Ratliff, F. 164, 196
- Readhead, M. 177, 196
- Reber, A.S. 79, 96
- Reese, H. W. 257, 286,  
**292**
- Reeves, B. 287, 292
- Rehm, J. 521, 531f, 539,  
548f, 550, 553
- Rehm, R. 30, 40, 52,  
**622, 624**
- Reichardt, W. 175, 196
- Reichenbach, H. 49, 66f,  
**75, 81f, 96, 431, 434,**  
**446**
- Reichertz, J. 103, 108,  
134
- Reid, T. **148**
- Reimann, P. **227, 229,**  
**243, 247**
- Renn, J. **13, 43**
- Renyi, A. 658, 691
- Rescher, N. 171, 196
- Restle, F. 569, 602
- Reumann, S.R. 161, 196
- Rice, R.E. 287, 292
- Richards, W.A. 175, 194
- Richardson, G. P. 384,  
**388**
- Riecken, 535
- Riedel, M. 100, 134
- Riedwyl, H. 80, 96
- Riefer, D.M. 635, 639,  
644, 647
- Riemann, G. 103, 134
- Riemersma, J.B. 573, 600
- Risteau, C.A. 193
- Ritchie, B.F. 427
- Rittenauer-Schatka, H.  
523, 554
- Ritter, H. 174, 196
- Rivadulla, A. 435
- Roberts, F.S. 578, 593,  
602, 676, 691
- Robertson, T. 602
- Robinson, R. 590, 598
- Rock, I. 54, 93
- Rodin, J. 542, 553
- Rogers, H. 215, 247
- Rogosa, D. 530, 553
- Rohracher, H. 55
- Röhrs, A. 56, 97
- Rollman, G.B. 542, 553
- Rorty, R. 139, 185, 197,  
319, 341
- Rosenbaum, Z. 593
- Rosenbloom, P. S. 204,  
**215, 245**
- Rosenkrantz, R. D. 433
- Rosenthal, R. 515, 553
- Roskam, E.E. 20, 44,  
47f, 69f, 96, 560, 567,  
602, 687, 691
- Rosonoer, L.I. 386
- Ross, D. 528, 543, 548
- Ross, L. 404, 426
- Ross, L.D. 542, 545f, 553
- Ross, S.A. 528, 548
- Rössner, L. 272, 289
- Rost, J. 639, 647, 653,  
654, 655, 676, 685,  
690f
- Roth, E. 353, 388
- Roth, G. 317, 341
- Rothacker, E. 276, 292
- Rothman, K. 532, 553
- Rozeboom, W. W. 595f,  
602, 724, 733, 736
- Rozin, P. 543, 553
- Rucci, A.J. 286, 292
- Rumelhart, D. E. 206,  
245, 247, 296, 340, 451
- Runesone, S. 172, 197
- Russo, J.E. 61, 65, 96
- Rutherford 410
- Rützel, E. 728, 736
- Sahner, H. 260, 261, 287,  
**293**
- Salancik, G.R. 523, 553
- Saimon, W.C. 326, 332f,  
**421, 426,** 430f, 435,  
**441, 445f**
- Saris, W. 682, 692
- Sarnoff, I. 543, 554
- Saunders, D.R. 669, 692
- Savage, L.J. 726, 734
- Schachter, S. 535, 554
- Schafroth, M. 80, 96
- Schank, R.P. 204, 239,  
**247**
- Scheele, B. 65, 94
- Scheerer, E. 144, 197,  
**287, 293**
- Schefe, P. 206, 247
- Scheler, M. 109, 134
- Schilpp, P.A. 426
- Schleiermacher, F. D. E.  
102, 134
- Schlesinger, G. 447
- Schlesselman, J. J. 532,  
**554**
- Schlick, M. 315, 341
- Schlürer, J. 547, 549
- Schmid, M. 398, 427

- Schmitt, M. J. 655, 668,  
669, 685, 692, 693
- Schmolensky, P. 206, 247
- Schneider, B. 565, 602
- Schneirla, T.C. 361, 388
- Schoemaker, P. J. H. 15f,  
44
- Schreier, M. 317, 341
- Schröder, P. 133
- Schroeder-Heister, P.  
561, 601
- Schröer, N. 120, 134
- Schuler, H. 542, 546, 554
- Schult, T. J. 227, 247
- Schulten, K. 174, 196
- Schultz, J. 77, 95
- Schultze, J. 53, 96
- Schulz, T. 485, 491, 507
- Schurz, G. 336, 341
- Schütz, A. 100, 103ff,  
111, 122ff, 134f
- Schütze, F. 103, 135
- Schwager, K. W. 75, 96
- Schwartz, E.L. 192
- Schwarz, N. 521ff, 543,  
549f, 554f
- Schwenkmezger, P. 526,  
549
- Scott, 564
- Scriven, M. 645
- Searle, J. R. 116, 135,  
155, 170, 197, 234,  
238, 239, 247, 321,  
341, 377, 388
- Sechrest, L. 521, 555
- Seifert, H.G. 506, 710,  
735
- Sehen, R. 11, 44
- Selz, O. 227, 247
- Semmer, N. 35, 43
- Sergejew, J. 344, 388
- Shafer, G. 728, 736
- Shapere, D. 254, 255,  
267, 270, 275, 277, 293
- Shapiro, E. 211, 247
- Shaw, J.C. 203, 245, 451
- Shepard, R.N. 152, 160f,  
172, 179, 197
- Sherman, S.J. 524, 551
- Shiffrin, R.M. 410, 425
- Shirley, E.S. 433
- Shye, S. 197
- Sibicky, M.E. 551
- Sieber, J.E. 542, 554
- Sievers, W. 642, 647
- Sigel, S. 736
- Simmel, G. 102, 135
- Simon, D.P. 229, 247
- Simon, H.A. 26, 29, 44,  
60f, 65, 93, 203, 229,  
237, 243, 245, 246,  
247, 286, 292, 320,  
340, 385, 388, 451f
- Simons, A. 523, 554
- Sinclair, R.C. 526, 552
- Singer, W. 172, 197
- Sinnett, L.M. 544, 552
- Sirovich, L. 164, 196
- Skinner, B.F. 277, 285,  
293, 315, 329, 341, 413
- Sleigh, A.C. 192, 197
- Sloman, A. 161, 197
- Smart, J.J.C. 181, 197
- Smirnowa, I.M. 386
- Smith, A. F. M. 639, 647
- Smith, B.C. 203, 213, 247
- Smith, E.E. 203, 244, 246
- Smolensky, P. 152, 197
- Smolka-Koerdt, G. 133
- Sneed, J.D. 38, 44, 261,  
262, 290, 293, 332,  
341, 430, 457, 650,  
687, 688, 692
- Snodgrass, J.G. 636, 638,  
647
- Sober, E. 171, 197
- Soeffner, H.-G. 35, 58,  
66, 72, 103, 108f, 111,  
121, 130, 134f, 227
- Solomon, R.L. 526, 554
- Sörbom, D. 484, 507,  
540, 551, 638f, 646,  
682, 685, 690
- Spada, H. 25, 36, 199,  
223, 229, 246, 247,  
451, 685, 691
- Spangenberg, P.M. 133
- Spence, K. W. 416, 425,  
427
- Sperling, G. 173, 175, 197
- Spinoza 141
- Spiro, K. 196
- Spohn, W. 437, 676, 692
- Srubar, I. 109, 135
- Stachowiak, H. 96, 205,  
247, 252f, 269, 271f,  
289, 293
- Stanley, J.C. 82f, 86, 92,  
459, 512, 526f, 530,  
534f, 549, 605, 645
- Stapf, K.-H. 645
- Steffen, 528
- Steger, H. 133
- Stegmühen W. 17, 261,  
268f, 271, 283, 293,  
325f, 329, 332, 341,  
430f, 435, 438f, 444,  
447, 457, 676, 687,  
689, 692, 615, 647,  
694, 696, 714, 715,  
725, 727, 728, 736
- Stehr, N. 135
- Steiger, J.H. 540, 554
- Stein, D. 565, 602
- Steiner, H.G. 707, 734
- Steinleitner, M. 521, 553
- Stelzl, I. 540f, 554
- Stemmer, N. 433, 449
- Stephan, E. 255, 261,  
263, 294, 313, 332,  
337, 342, 687, 692
- Stephens, D.L. 61, 96
- Sterling, L. 211, 247
- Stetter, F. 216, 247
- Stevens, S.S. 137, 144f,  
150, 155, 197, 564,  
568f, 576, 589, 597,  
602f
- Steyer, R. 674f, 678, 681,  
682, 685, 687f, 692,  
693
- Steyer, R. 9, 40, 41, 82f,  
85, 460, 619, 622, 624,  
647, 649, 654, 655,  
658, 661, 663f, 671,  
672
- Stigler, S. 709, 736
- Stiles, W.S. 179, 198
- Störig, H.J. 14, 44

- Strack 30, 40, 52, 622,  
**624**, 523f, 543, 549,  
**552**, 554f
- Strauss, A.L. 111f, 128,  
 132, 135
- Strauß, B. 653, 676, 692
- Strawson, P.E. 447
- Stronkhorst, L. E. 682,  
**692**
- Strube, G. 206, 247, 451,  
**579**, **599**
- Stuart, A. 709, 735
- Student, 514, 555
- Stumpf, M. 223, 246
- Stute, W. 658, 689
- Suck, R. 687
- Sukale, M. 332, 342
- Suppe, F. 283, 293f, 296,  
**332**, **342**, **434**
- Suppes, P. 164, 196f,  
 332f, 342, 559, 564,  
 573f, 576ff, 586, 603,  
 626, 647, 676, 690,  
 691, 693, 730, 735
- Swaminathan, H. 676,  
**689**
- Swets, J.A. 58, 94, 637,  
**646**
- Swijtink, Z.G. 32, 44,  
**734**
- Szagan, G. 265, 294
- Tack, W.H. 3, 12, 17,  
 44, 48, 73, 96, 137,  
 177, 198, 206, 247,  
 369, 388, 567, 571,  
 577, 603, 639, 647,  
 678, 693
- Tal, A.A. 386
- Tarski, A. 430
- Teegen, F. 56, 97
- Terry, M.E. 9, 42
- Textor, M.R. 542, 555
- Thagard, P. 203, 244,  
**432**, 448f, 454f, 462
- Thissen, D. 80, 97
- Tholey, P. 728, 729, 736
- Thomas, H. 89f, 97
- Thompson, E. 198
- Thorndike, L.L. 277
- Thurstone, L. L. 569,  
 575ff, 585, 676, 693
- Tillmann-Bartylla, D. 133
- Titchener, E.B. 55
- Titterington, D.M. 639,  
 641, 647
- Tolman, E.C. 296, 303,  
 319, 342, 342, 416, 427
- Torgerson, W.S. 653, 693
- Toulmin, S. 283, 294,  
 315, 333, 342
- Townsend, J.T. 197, 640,  
**647**
- Traxel, W. 553, 388
- Traxel, W. 511, 555
- Trenkler, G. 706, 733, 734
- Troscianko, T. 196
- Tschan, F. 35, 43
- Tseelon, E. 5, 44
- Tukey, J. W. 96f
- Tukey, W. 79f, 95
- Tuomela, R. 337, 342, 441
- Turing, A. 150f, 154,  
 168, 215, 235, 247,  
 377, 388
- Turner, J.C. 537, 555
- Tutz, G. 654, 693
- Tversky, A. 11, 15, 42,  
 43, 69, 197, 326, 338,  
 465, 523, 555, 564,  
 603, 626, 647, 676,  
 690f, 729, 730, 734,  
 735
- Tweney, R.D. 286, 292
- Ueckert, H. 332, 342
- Ullman, G. 294
- Ullman, J. D. 216, 244
- Ullman, S. 161, 198, 204,  
 244
- Ulrich, H. 291
- Underwood, B. J. 411,  
 427
- Ungeheuer, G. 135
- Upshaw, H.S. 524, 553
- Uttal, W.R. 179, 190, 198
- Valach, M. 373, 388
- van Acker, P.O.F.C. 10,  
 44
- van Buggenhaut, J. 687
- van de Pol, F. 685, 690
- van den Daele, W. 599
- van der Kamp, L.J. T.  
 654, 688
- van Fraasen, B. 315
- van Fraasen, B. C. 198,  
 595
- van Meter, D. 10, 44
- van Meurs, A. 692
- van Santen, J. P. H. 175,  
 197
- van Wright, G.H. 432
- VanLehn, K. 204, 221,  
 225, 227, 228, 231,  
 243, 247, 248
- Varela, F.J. 198
- Vartanian, A. 169, 198
- Viale, R. 195f
- Viertl R. 658, 693
- von Cranach, M. 67, 94
- von der Malsburg, C.  
 165, 195
- von Eye, A. 484, 507,  
 645
- von Fritz, K. 139, 193,  
 595, 599
- von Glasersfeld, E. 317,  
 339
- von Goethe, J. W. 401
- von Helmholtz, H. 143,  
 149, 158, 160, 165,  
 169, 179, 194, 563, 600
- von Humboldt, W. 102
- von Kardorff, E. 114,  
 131f
- von Kries, J. 143, 146f,  
 194, 564, 600
- von Kutschera, F. 403,  
 426, 430, 433f, 437f,  
 445, 448
- von Metraux, A. 132,  
 287, 292
- von Mises, R. 736
- von Neumann, J. 564
- von Rosenstiel, L. 114,  
 131f
- von Savigny, E. 256, 292
- von Savigny, K. F. 101
- von Strien, P. 287, 294

- von Winckelmann, J. 135  
 Vonderach, G. 103, 135  
 Voß, A. 103, 108, 135  
  
 Wagner, D. 555  
 Wahrendorf, J. 513, 555  
 Wainer, H. 80, 97  
 Wald, A. 609, 647, 721, 736  
 Walder, L.O. 529, 552  
 Walker, E. C.T. 197  
 Wallach, H. 54, 97  
 Wandell, B. 577, 603  
 Wandschneider, D. 355, 388  
 Wann, T.W. 321, 342  
 Wartofsky, M. W. 426  
 Watkins, J. 422, 427  
 Webb, E.J. 521, 555  
 Webb, J.C. 169, 198  
 Weber, M. 103, 106, 112, 135, 142  
 Wegenast, K. 135  
 Wegener, H. 388  
 Weidner, 525, 555  
 Weimer, W.B. 283, 294  
 Weiner, B. 255, 258, 294, 299, 342, 421, 427  
 Weingart, P. 287, 294  
 Weiskrantz, L. 62, 97  
 Wellens, T. R. 526, 552  
 Wendel, H. J. 317, 342, 395, 427  
 Wender, K.F. 676, 693  
 Wendt, D. 712, 723, 725, 728, 736, 737  
  
 Wenger, E. 229, 248  
 Wentura, D. 51f, 94  
 Wermuth, N. 668, 693  
 Wertheimer, M. 451  
 Wessels, M.G. 367, 388  
 Westermann, R. 33, 39, 81, 261, 263, 271, 294, 332f, 337, 342, 430, 442, 457ff, 480f, 507, 487, 494f, 508f, 551, 605, 609, 612ff, 618, 632, 646f, 651, 687, 693f, 723, 724, 731f, 735, 737  
 Westmeyer, H. 53, 96, 251, 261, 272f, 294, 304, 308, 326, 332, 337f, 342, 457, 477, 485, 506, 604, 610, 634, 646f, 687, 693  
 Wettersten, J.R. 411, 427  
 Wharton, J.D. 530, 551  
 Whewell 419  
 Wickens, T.D. 639, 647  
 Widaman, K.F. 676, 693  
 Wilcox, R.R. 514, 555  
 Wilkes, K.V. 195f  
 Wilks, S.S. 640f, 647  
 Will, U. 432, 441, 449  
 Williams, L.P. 418, 427  
 Williams, R. H. 678, 693  
 Willmes, K. 611, 633, 648  
 Wilson, A. 97  
 Wilson, T.D. 57, 62f, 96  
 Winer, B.J. 514, 555  
 Winkelmann, K. 332  
  
 Winograd, T. 204, 238, 248  
 Witte, E. 695, 719f, 725, 728, 730, 731, 733, 737  
 Wolf, H. 195  
 Wolff, s. 103, 114, 131, 133  
 Wolfle, L.M. 653  
 Woodward, J. 186, 198  
 Wottawa, H. 678, 693, 723, 737  
 Wright, F.T. 602  
 Wright, H.E. 53, 92  
 Wundt, W. 55f, 60f, 97, 100, 136, 143, 184, 283, 319, 342, 584, 603  
 Wutke, J. 33, 40, 41, 42, 86, 609, 615, 684  
 Wymer, W.E. 669, 691  
 Wyss, D. 365, 388  
 Wyszecki, G. 179, 198  
  
 Young 149  
  
 Zebrowitz, L.A. 535, 555  
 Zelen, M. 543, 555  
 Ziezold, H. 658  
 Zimbardo, P. G. 542f, 553f  
 Zimmermann, D. W. 678, 693  
 Zimmermann, E. 184  
 Zinnes, J.L. 676, 693  
 Zytkow, 451

# Sach-Register

- Alpha-Fehler (s. Fehler erster Art)
- Beta-Fehler (s. Fehler zweiter Art)
- Abduktion 431
- Abgrenzungskriterium 400
- Abhängigkeit zwischen Zufallsvariablen 659
- Abhängigkeit, korrelative 662f
  - , nichtdeterministische 651
  - , regressive 662ff
  - , schwache kausal regressive 674
  - , starke kausal regressive 674f
  - , stochastische 652ff
- Adäquatheit, empirische 315f, 389f, 399, 405f, 423
- Adäquatheitsbedingung 443f
- Algorithmus 214
- Allgemeine Methodenlehre 3ff
- Allgemeines Lineares Modell (ALM) (s. Modell, allgemeines lineares)
- Alpha-Lambda-System 437
- Analogie 363, 365f, 431, 450, 454
- Analogieschluß (s. Analogie)
- Analytizität der c-Funktion 435
- Anfangsbedingung 325
- Annahmekenner 252ff
- Ansätze, pragmatische 429
- Antwortformat 522ff
- Anwendung, intendierte (s. Strukturalismus)
- Anwendungsbezug 87f
- Äquilibrium, reflektives 455
- Äquivalenzbeziehung 496
- Äquivalenzprinzip 596f
- Argument, deduktives 430, 464
  - , induktives 429f, 464
- Argumentation, pragmatische 446f
- Aussage, gesetzesartige 302, 304, 695
  - , je-desto 305
  - , universelle 304
  - , synthetische 476
  - , wenn-dann 304f
- Aussagenkonzeption 296, 302, 305, 308, 332, 336
- Austauschbeziehung 265ff
- Automatentheorie 349
- Automorphismus einer Struktur 591
- Axiom 295, 299f, 302, 306, 335, 391f, 572
- Axiome, meßtheoretische 586
- Bayes-Theorem 437, 612, 708, 726f, 729
- Bedeutsamkeit 73f, 626f, 500
- Bedeutungsüberschuß 309
- Begriff, Beobachtungs- 308f, 310f
  - , theoretischer 308, 310f, 315, 318, 320, 372
- Begründungszusammenhang 49f, 66ff, 75ff
- Behauptungen, empirische 439
- Behaviorismus 235, 319, 349
- Beliebigkeit 109
- Beobachtbarkeit 311
- Beobachtung 47ff
  - , experimentelle 185
  - , Fremd- 66ff
  - , heuristische 66ff
  - , Selbst- (s. Introspektion)
  - , wissenschaftliche 50ff
- Beobachtungsaussagen 408f, 423
- Beobachtungsfehler 582, 584, 587
- Beobachtungsprädikate 586
- Beobachtungsprotokoll 51ff, 68ff
- Beobachtungswissen 432
- berechenbar 213ff
- Berechnung 151
- Berechnungsuniversalität 213f, 216
- Berechnigungsfaktoren 460
- Bereich, kritischer 703, 711
- Beschreibung 402
- Bestätigung 406, 418f, 422
  - , Paradoxien der 433, 449
  - , deduktive 443f
- Bestätigungsgrad 356, 406
- Bestätigungstendenz 414, 418
- Bestätigung, empirische 389, 418f
  - , induktive 418
  - , induktive 434, 443f, 464
  - , irrtümliche 417

- Bestätigungsgrad 356  
 Bewährung 407, 418, 421  
   -> einer Hypothese 429, 433, 441  
   -, einer Theorie 429, 441  
     vorläufige 407  
 Bewährungsgrad 442  
 Bewußtsein 59ff, 319f  
 Bisektionsverfahren 565  
 Bisymmetriestrukturen 565  
 Black Box 345ff, 402  
 Bootstrap-Methode 642  
 Bradley-Terry-Luce-Skalierung 9  
 Brückenannahme 188ff
- Ceteris-distributionibus-paribus-Klausel 670  
 Ceteris-paribus-Bedingung 31, 189, 327, 408, 442, 460, 652, 669  
 Chinesisches Zimmer 238f  
 Code, perzeptueller 149, 164, 165, 179, 556  
 cognitive science (s. Kognitionswissenschaft)  
 computational theory, Ebene der 152ff, 161, 181, 185, 186  
 Computermetapher 367  
 Computermodell 200ff  
 constant-utility-Modell 575  
 cross-lagged panel design 530
- Daten, empirische 35, 47ff, 489f  
   -, sozialwissenschaftliche 104, 110, 113ff, 127  
 Datenanalyse, exploratorische 76ff  
   -, konfirmatorische 76ff  
 Datenerhebung, nicht-standardisierte 114f  
   -, qualitative 113  
   -, standardisierte 114f  
 Datenmodell 493  
 Datenrestriktion 19f, 35f, 201  
 Datenschutz 546f  
 Deduktion 696  
 Definition, operationale 309f, 634  
 Dekomponierbarkeit, additive 672f  
 Denken 160  
 Denkprozesse 151f  
 Determinationskoeffizient 663  
 Deutung 106ff, 126
- Differenzmethode 431  
 Differenzstrukturen 565  
 Dissonanztheorie 608  
 Disziplin, angewandte 8  
 Disziplin, reine 8  
 Domainproblem 255, 267, 270  
 Doppel-Blind-Versuch 515f  
 Duhem-Quine-Problem 643f  
 Duhem-Quine-These 442f
- Effekte, experimentelle 184  
 Effektgröße 500, 627f, 633, 703, 719, 732  
 Effizienz kognitiver Systeme 454  
 Eindeutigkeit 74  
 Einfachheit 15, 232, 373, 389, 391, 403f  
 Einflüsse, externale 258, 260, 279ff, 283, 286ff  
   -, internale 279ff  
   -, politisch/administrativ-organisatorische 281  
 Einheitlichkeit, semantische 389, 394f  
 Einzelbeobachtung 344  
 Einzelfallanalyse 112, 223, 229, 534ff  
 Empfehlungen, methodologische 410f  
 Einzelfallbestätigung 437f  
 Empfindung 142f, 148  
   -, Definition 146f  
 Empfindungsmessung 142f, 147  
 Empfindungsschwelle (s. Schwelle)  
 Empirie 486  
 Empirismus, logischer 487  
 Entdeckungszusammenhang 49f, 66ff, 75ff  
 Entscheidung, statistische 615, 635, 694ff  
   -, unter Risiko 433  
 Entscheidungsexperiment 231, 416  
 Entscheidungsregel 703  
 Ereignis, singuläres 325  
 Ereignisse, mögliche 655f  
 Erkenntnis, empirische 12f  
   -, rationale 12ff  
 Erkenntnisfortschritt 406, 416, 421  
 Erkenntnistheorie 344  
 Erklären (und Verstehen) 112  
 Erklärung 295, 323f, 329f, 399, 402f, 418f  
   -, Alternativ- 510ff

- , deduktiv-nomologische (DN) 324f
- , dispositionelle 328
- , genetische 328
- , Instantiierungs- 328
- , naturgesetzliche 400
- , statistische 326f
- , teleologische 328
- , unvollständige 327
- Erklärungsbegriff, pragmatischer 261
- Erklärungskraft 309, 389, 391, 396, 399, 402, 405, 419f, 422
- Erlebnisaspekt 320f
- Erlebnisdiskription 56ff
- Erwartung, bedingte 662f
- Erwartungswert 677f
- Erwünschtheit, soziale 525
- Essentielle tau-Äquivalenz, Bedeutsamkeit 680
- , Eindeutigkeit 680
- , Existenz 679f
- , Identifizierbarkeit 682f
- , Testbarkeit 681f
- Ethik psychologischer Untersuchungen 528, 541ff
- Evidenz, Prinzip der totalen 439
- Evolutionstheorie 157, 164, 171
- Exhaurieren 24
- Experiment 24f, 76, 183ff, 510ff, 608, 610, 622, 631, 697
- Experimentelle Beobachtung 185
- Effekte 184
- Experiment vs. Fallstudie (s. Fallstudie)
- Experiment mit Meßwiederholung 607
- Explanandum 324
- Explanans 324
- Explizitheit 203, 381f
- Fairneß (einer Untersuchung, Prüfung) (s. Untersuchung)
- Fall-Kontroll-Studie 53Iff
- Fallibilismus 317, 407, 422
- Fallstudie 535ff
- Falsifikation 226, 228, 399, 406f, 414, 418f, 422f, 440, 442, 497, 731
- Falsifikationsgrad 400, 408
- Falsifikationskriterium 413
- Falsifikation als Abgrenzungskriterium 400
- Falsifikation, irrtümliche 417
- Falsifikationismus 407, 411
- , dogmatischer 605
- methodologischer 605, 609
- Falsifizierbarkeit 390, 404, 409f, 695
- Fechner-Modell 576f
- Fechnersches Gesetz 142, 146
- Fehlbarkeit 394
- Fehler 570, 572, 575f, 582, 677f
- erster Art 609f, 616, 627f, 630ff, 641f, 704
- zweiter Art 609f, 616, 627f, 630ff, 641, 704
- begriff, numerisch-geometrisch inspirierter 573
- behandlung 587
- kontrolle 186, 190
- konzept, externes 584
- konzept, theorieexternes 582
- problem 570
- risiko 493f, 499, 504
- Wahrscheinlichkeit 493, 495f, 505
- Formalisierung 209f, 305f
- Forschungsprogramm 38, 25Iff, 538ff
- Forschungsstrategie 347
- Fourieranalyse 175
- Frustrations-Aggressions-Hypothese 643f
- Funktion, psychometrische 568, 576, 581, 585
- , psychophysikalische 593f
- Funktionalismus 150, 152, 170, 236ff, 320
- , probabilistischer 159
- , Turing-Maschinen- 151
- Gedankenexperiment 187, 238
- Gehalt 396f, 410
- , empirischer 396, 400f, 411f, 484
- , semantischer 487
- , theoretischer 421
- Gehaltsüberschuß 398f, 419
- Geltungsprüfung, empirische 224ff
- Generalisierung 453
- Gesetz 146, 181, 587, 303f, 308, 334, 586
- , deterministisches 305, 326
- , Koexistenz- 305
- , probabilistisches 305
- , statistisches 326



- , Sukzessions- 305
- , Webersches 574, 588, 592
- Gesetzeshypothese 304, 396f, 398
- Gesetzmäßigkeit 304, 398
- Gestaltpsychologie 157, 158, 166
- Gleichung, persönliche 584
- Grassmann-Struktur (Farbwahrnehmung) 175, 180, 588, 597
- Größenschätzung (magnitude estimation) 561, 568f, 571, 597
- Grue-Paradox 433
- Grundannahmen 295, 302
- Gültigkeit (psychologischer Schlüsse)
- Güteeigenschaften eines Tests 705
- Gütefunktion 705
- Handlungsprotokoll 105, 116
- Handlungssituation 105
- Hermeneutik 98, 100, 126ff
  - , Alltags- 123f
  - , sozialwissenschaftliche 98, 127, 129
  - , wissenschaftliche 101f
- Heuristik 16, 167, 177, 644
- hidden assumptions 9f
- Hilfsannahme 312f, 335, 392, 396, 408f, 412f, 423, 644
- Hilfshypothesen 312, 401, 410
- Hintergrundannahmen 431, 442
- Hintergrundwissen 312, 414f, 483, 485, 487, 492
- Homogenitätsannahme 354
- Hypothese 318, 475, 604, 700, 703
  - , allgemeinpsychologische 606, 625
  - , beschränkte universelle 478, 486
  - , einfache 717
  - , empirisch-inhaltliche 479, 488, 502f, 635
  - , Forschungs- 478
  - , Generierung 480
  - , Hilfs- 611
  - , inhaltliche 478, 489, 490f, 499
  - , Kausal- 484
  - , lokalisierende (bestimmte) Existenz- 477, 485
  - , Null- 703, 710f
  - , Ober- 697, 699
  - , operationale 478
  - , pseudo-singuläre 477, 485
  - , psychologische 604ff, 626ff, 634, 640, **642**
  - , Punkt- 703
  - , quasi-universelle 477f, 486, 489
  - , singuläre 477, 485
  - , statistische 477, 489, 604ff, 624, 626ff, 634, 640, 642, 699, 717
  - , statistische Aggregats- 624f
  - , statistische Ober- 615
  - , Test- 490, 494, 498
  - , theoretisch-inhaltliche 479, 486, 502
  - , unbeschränkte universelle 478, 486
  - , unbestimmte Existenz- 477, 485
  - , unvollständige 327
  - , Wahrscheinlichkeits- 486
  - , wissenschaftliche 475f, 501f
  - , zusammengesetzte 717
  - , Zusammenhangs- 484
- Hypothesengenerierung 353
- Hypothesenprüfung 604f, 632, 643f
- Hypothesenprüfung, deduktivistische 605f
- Idealisierung 15, 181, 188ff, 321f, 418, 586
- Idealtypus 111f
- Identifizierbarkeit 636f, 639, 682f
- Immunisierungsstrategien 401
- Immunität 336
- Implementation 211f, 220
- Implikationsbeziehung 491f
- Indifferenzprinzip 435
- Indikator 308, 314
- Induktion 406, 418, 422, 429f, 465, 695
  - , Begründbarkeit der 445f
  - , eliminative 431
  - , enumerative 430f, 439, 450
  - , praktische Rechtfertigung der 445f
  - , psychologische Analysen der 450ff
- Induktion, strukturalistische Analyse der 457ff
- Induktionskonzeption 450f
- Induktionsprinzip 431
- Induktionsproblem 429, 432f, 445, 464f
- Induktionsregel 446
- Induktivismus 406, 419, 421
  - , starker 441
- Inferenz 210, 212, 429
  - , ampliative 430

- Informationsbegriff 151, 154
- Informationsgehalt 225, 302, 308, 317
- Informationsverarbeitung, 28, 202ff
- Informationsverarbeitung, perceptuell-kognitive 150ff, 156f
- Inkommensurabilität 398
- Inkonsistenz, logische 392, 395
- Innovation, wissenschaftliche 267
- Instanzenbestätigung 438f
- Instrumentalismus 313f, 315f, 336, 392f, 407, 424
- Intentionalität 239, 321
- Interaktion (s. Wechselwirkung)
- Intermodalvergleich (cross modality matching) 568f, 571
- Interpretation 99f, 105, 110f, 116f, 130, 306
  - , partielle 306
  - , sequentielle 118ff
  - , epistemische 33
  - , ontische 33
- Interpretierbarkeit, operative 213
- Introspektion 56ff, 321, 352ff
- Invarianten 598
- Invarianz 140, 160, 588f
  - , Forderungen 594, 597
  - , Konzepte 592, 595
  - , Kriterien 598
  - , Prinzipien 594
  - betrachtung 177, 181, 594
  - hypothese des verbalen Lernens 606f
- Isomorphie 369
- Item-Response-Funktion 584
  
- Kalkül 306, 369
- Kasuistik 534ff
  - Computer- 535ff
- Kausalanalyse 539ff
  - aussagen 329
  - gesetz 329f
  - interpretation 424
- Kausalität 305, 329f
  - , schwache 674
  - , starke 674
- Kern, formaler (s. a. Strukturalismus) 458
- Kognitionswissenschaft 29, 139, 150, 191, 203f
- Kognitive Architektur 29, 215
- Kognitive Modellierung 203
  - Kohortenstudie 531ff
  - Kompensation 351
  - Kompetenzmodell 241
  - Komplexität von Systemen 26f
  - Komplexitätsreduktion 27f
    - , theorie 26
    - , vergleich 27
  - Koifidenzintervall 702, 733
  - Konfundierung 670f
  - Konnektionismus 172, 174, 206, 451
  - Konsistenz 221, 232
    - , externe 393
    - , interne 393
    - , logische 389, 391f, 395, 406
    - , semantische 391, 394f
  - Konstanz, Farb- 149, 156, 163
  - Konstrukt, hypothetisches 314
    - , theoretisches 372
  - Konstruktionen erster Ordnung 104, 202
    - zweiter Ordnung 36, 104, 112, 202
  - Konstruktivismus, radikaler 317
  - Konstruktvalidierung 313f
    - Validität 637
  - Kontexteinflüsse 522ff
  - Kontrolltechniken 509ff, 624
    - , statistische 539ff
  - Konzept, random-utility 582
  - Konzeptualisierung 207f
  - konzeptuelle Replikation 611
  - Korrelationskoeffizient 665
  - Kriterium 484
  - Kryptodeterminismus 361
  - Künstliche Intelligenz 152, 161, 162, 187, 234ff, 320
    - , schwache These der 234f
    - , starke These der 234f
  - Künstliche Sehsysteme 161, 188
  
  - Lambda-Kontinuum 437
  - Lautes Denken 56ff
  - Lebenswelt 107ff
  - Lebenswelt, Analyse der 108ff
    - , Deskription der 108
  - Leib-Seele-Problem 138f, 141f, 144, 150, 191, 561
  - Leistungsmotivation 297ff
  - Lernen, Wahrscheinlichkeits- 160
  - Likelihood 716, 725
    - Quotienten-Test 718

- Ratio-Kriterium 716
- , maximum 716
- Logik der Forschung 429
  - , deduktive 428
  - , induktive 424, 429, 434f, 448
- Manipulationsgrad 20f
- Markoff-Modell 639
- matching (s. Parallelisierung)
- Materialismus, eliminativer 319f
  - , funktionaler 320
- Mathematisierung 305f
- Mathesis Universalis 561
- meaningful parametrization 595
- meaningfulness 589, 594ff
  - , non- 595, 597
  - , konditionale 598
- measurement, conjoint 568
  - , functional 568
  - , random-conjoint 579
- Mechanisierung des Geistes 169
- Meehlsches Paradoxon 661
- Mehrebenenbetrachtung 217ff
- mentalistisch 236ff, 319
- Merkmalsvariabilität 449
- Messen, additiv-verbundenes 568
  - , extensives 564
  - , fundamentales 563
  - , abgeleitete 563
- Messung, additiv-verbundene 564, 571, 576, 579
  - , fundamentale 563f, 566
  - , Index- 73f
  - , per Definition 73f
  - , per fiat 73f
  - , Repräsentationstheorie der 74f
  - , Transformationsansatz der 590
- Meßfehler 143, 582
  - problem 650, 677f, 685
  - variable 678
- Meßmodell 643f
  - , stochastisches 683, 685f
- Metamerie 164, 179
- Metapher 167ff, 176f, 181, 284ff
- Metaphysik 400
- Metaprinzip 15f, 177
- Metatheorie 390
- Methoden, empirische 17ff
  - , Kontinuum der induktiven 435f
- Methodenlehre, allgemeine 4f
- Methodenlehre, spezifische 4f
- Methodenspezifität, Problem der 685f
- Methodenwahl 256ff
- Methodologie 4ff, 34f, 358, 389f
  - , deduktivistische 441, 445, 649
- Milieu (s. Lebenswelt)
- Mischverteilung 639
- Modell 8ff, 177, 306, 333, 367ff, 615
  - , algorithmisches 204
  - , allgemeines lineares 351, 627
    - essentiell tau-äquivalenter Variablen
  - , 679ff
  - , formales 205f, 233, 368
  - , ideales 322
  - , idiographisches 222
  - , individualisiertes 222, 229
  - , mathematisch-numerisches 200f, 205f, 233, 307
  - , mentales 202, 452f
  - , natürlich-sprachliches 205
  - , prototypisches 222f
  - , random-utility-- 575f, 581
  - , stochastisches 634ff, 640, 643f, 649ff, 654, 658f
    - symbolisches 205
  - anpassungstest 636, 640
  - bildung 202
  - daten 224f, 229f
  - experimente 228
  - geltungstest 636, 641, 643
- Modellierung 200ff
- Modelltheorie 205, 369
- Modellvergleichende Argumentation 231
- Moderatormodell 667f
- Modus tollens 723
- Multiple Determiniertheit, Problem der 650f, 659
- Naturgesetz 353, 414
- Naturwissenschaft 355
- Neobehaviorismus 319
- Neopositivismus 308f
- Neurale Netze (s. Konnektionismus)
- Neurophysiologie 157, 163, 165ff, 172, 175, 182
- Netz, semantisches 207ff
- Neyman-Pearson-Lemma 718
- nomologisch 295, 402

- nomothetisch 189
- non-statement-view (s. Strukturalismus)
- Nonparametrische Verfahren (s. Statistik, nichtparametrische)
- Normalbedingung 188
- Normativität 11f
  - , ideale 11
  - , praktische 11
- Nullhypothesentesten 722f
  
- Objektivität 51ff, 56
- one-shot case study 534
- Operationalisierbarkeit 483f
- Operationalisierung 308f, 309, 312, 325, **401**, 408f, 412, 417, 487f, 496, 502, 611, 634, 644
- Optik, inverse 150, 161
- Optimalität 15f
  
- Parallelisierung 533
- Parameter 698
- Passung von Personen- und Modelldaten 224f
- Performanzmodell 213
- Philosophie der normalen Sprache 447
- Placebo 515f
- Population, Problem heterogener 686
- Positivismus 354f
- Postulate 295f, 302
- power function (s. Gütefunktion)
- Prädikatenlogik 209ff, 306
- Prädiktor 484
- Praxisbezug (s. Anwendungsbezug)
- Präzision 203, 381f
- Prinzip der kritischen Prüfung 414f
- Prinzipien, methodologische 389f
- Probabilisierung 585, 587
- Probabilismus 696
- Problem P 252ff
- Problem, -Derivat 259, 263f, 268f, 270ff, 277f
  - , Indizierungs- 70f
  - , Induktions- 84ff
  - , Interpretations- 71f
  - , Klassifikations- 69f
  - , Quantifizierungs- 72ff
- Problemverschiebung, degenerative 439
- Produktionensystem 452
- Prognose, technologische 88
  
- Programm-Typen 269ff
- Programmnetz 265f, 265ff, 287
- PROLOG 211
- Prozedur, analytische 355ff
- Prozeßmodell 201, 213, 241
- Prüfbarkeit 303, 325, 401, 404, 405, 413
  - empirische 389, 401, 411
- Prüfbarkeitsgrad 400
- Prüfung 418f
  - , empirische 413f
  - , Prinzip der kritischen 414f
  - , Strenge der 413, 418
  - , wohlwollende 417
- Psychoanalyse 365
- Psychologie, Allgemeine 606
  - , Diagnostische 639
  - , Differentielle 639
  - , Entwicklungs- 639
  - , Sozial- 606
- Psychologiespezifität 5f
- Psychophysik 36, 137ff
  - , äußere 141, 144
  - , computistische 153ff
  - , funktionalistische 137, 149, 153ff, 157, 162ff
  - , innere 141, 144
  - , klassische 149
    - methodologische Probleme 138f
  - , skalenorientierte 138, 162f
  - , Traditionen 137ff, 156
  
- Qualia 138, 179, 182, 185
- Quantität 561
  
- Raben-Paradox (s. Grue-Paradox)
- Randbedingung 325, 398
- Randomisierung 85ff, 510ff, 607, 622ff, 670
- Randomisierungsmodell 514
- Randomisierungstests 514, 609ff, 632f, **642**
- Rationalismus, klassischer 16
  - , kritischer 253f, 312, 314, 407, 414, 418
- Rationalität 423, 428
- Rationalitätsbedingungen 434
- Reaktionsfunktion 350
- Reaktionszeit 178, 190
- Reaktivität 520f, 526

- Realisierung 658  
 Realismus 313f, 336, 394  
   -, experimenteller 516ff, 537  
   -, kritischer 314  
 Realität 486  
 Rechtfertigung, apriorische 432  
   -, deduktive 432  
   -, induktive 432  
 Reduktionismus 28f  
   -, Neuro- 165f, 176  
 Reduktionssatz 487  
 Reduktionssatz, bilateraler 634  
 Referenzklasse 449  
 Regeln, methodologische 389f, 393,  
   397, 422  
 Regelhierarchie 453  
 Regression 662f  
   -, isotone 572  
 Regressionshypothese 664  
 Reiz 162  
   -, distaler 152  
   -, ökologischer 163  
   -, proximaler 148, 152, 159, 160  
   -, Schlüssel- 163  
 Reizbestimmung 155  
 Reiz-Reaktionspsychologie 349  
 Rekonstruktion 105, 109, 111f, 117,  
   121ff, 202ff  
 Relation, empirische 71ff  
 Relativität 109  
 Reliabilität 628f, 678  
 Replizierbarkeit 51ff, 56  
 Repräsentation 151f  
   -, Fechnersche 574  
   -, additive 572  
 Repräsentationshypothese 213  
 Repräsentativität 83  
  
 S-O-R-Schema 349  
 S-R-Theorie 349  
 Schätzung 700, 704  
   -, „gute“ 701  
   -, asymptotisch erwartungstreue 701  
   -, effiziente 702  
   -, erwartungstreue 701  
   -, konsistente 702  
 Schluß, unbewußter 158, 160, 169  
 Schwelle 42, 144, 168  
 Semiordnung 568, 573, 581, 592  
  
 Sensitivitätsgesetz 594  
 Sequenzanalyse 117ff  
 Signalentdeckungstheorie (SDT) 58ff  
 Signifikanzniveau 706, 712  
 Signifikanztest 493f, 497f, 504, 513ff,  
   694, 709ff, 722ff, 725, 731  
 Simulation 200ff, 367f, 376f, 383f  
 Simultankontrolle 500  
 Sinn 101f, 104, 122ff  
 Sinnesdaten-Theorie 148  
 Sinnkonsistenz 121f  
 Sinnkriterium, empirisches 309  
 Situationale Spezifität, Problem der 685  
 Skala 589  
   -, psychophysikalische 145ff  
 Skalengkonstruktion 566f  
 Skalenniveau 558, 588, 590, 591, 593,  
   625f, 680  
   -, Ordinal- 634  
 Skalenniveaus, Theorie der 589  
 Skalenniveauf 680  
 Skalierung, direkte 72f  
 Skeptizismus 423  
 Sparsamkeitsprinzip 403  
 Sprache, Beobachtungs- 17f, 296  
   -, theoretische 296  
 Standardisierung 519  
 statement view (s. Aussagenkonzeption)  
 Statistik 700  
   -, deskriptive 700  
   -, nichtparametrische 632ff, 699, 715  
   -, parametrische 699  
   -, suffiziente 701  
 Statistiken, robuste 81  
 Sternchen-Daten 228  
 Stichprobe 698  
 Stichprobenraum 698  
 Stichprobenumfang 609, 611, 627f, 631ff  
 Störeinfluß (s. Störfaktor)  
 Störfaktor 30f, 82ff, 412f, 415f, 512ff,  
   607f, 623f  
 Störgröße (s. Störfaktor)  
 Störvariable (s. Störfaktor)  
 Strenge (einer Untersuchung, Prüfung)  
   (s. Untersuchung)  
 Stress-Index 572, 581  
 Streß, physischer und psychischer 542ff  
 Struktur, additiv-verbundene 564f, 572f,  
   578, 580, 582

- , deduktive 302
- , extensive 574, 591
- Strukturalismus 296, 332f, 390, 407, 408f, 413, 424, 457f
- , metatheoretischer 261ff, 271
- Strukturgleichungsmodelle 638f
- Strukturinvarianz 591
- Strukturmodell 373, 643
- Stützungsbegriff, klassifikatorischer 716
- , komparativer 725
- Subjektivität 321
- Suffizienz 228, 231f, 701
- Surplus 374f
- Symmetrie 177, 588, 594
- Symmetrie, Konzepte 591
- System, deduktives 295f, 302f
- , informationsverarbeitendes 320
- , soziales 252, 259ff, 265f, 280f, 284
- , formales axiomatisches 7f
- , k-dimensionales 437
- , logisch-formales 215
- , symbolverarbeitendes 202, 214
- , wissensbasiertes 36f, 205ff
- T-theoretisch 334
- Tatsachen, neue 419f
- Täuschung der Versuchsteilnehmer 544ff
- Teleologie 171
- Test 700
- , bester 706
- , Entscheidbarkeits- 703
- , komplementärer Alternativ- 703
- , konsistenter 705
- , sequentieller 721
- , statistischer 615
- , unverfälschter 705, 717
- , zum Niveau Alpha 705
- Teststärke 717
- Teststärkefunktion 717
- Text 115f
- Theorem 299f, 302, 306, 335
- Theorie 37f, 183, 295ff, 368ff
- , Ästhetik 177, 595
- , Begriff 178
- , geometrische orientiert 179ff
- , mechanistisch orientiert 180f
- , Hilfs- 459
- , oberflächliche 225
- , tiefe 225
- Theorie-Empirie-Überbrückungsproblem 606, 611, 634f, 637
- Theoriebewertung 38f
- Theoriebildung 343ff
- Theorieelement 458, 461f
- Theorieentwicklung 227, 241
- Theorieimport 267ff, 270ff, 274f, 277f
- Theorienaustausch 278
- Theoriennetz 263f, 335
- Theorienschema 368
- Theorieprüfung 344f
- Thurstone-Modell 575, 577
- Thurstone-Skala 588
- Tiefe 389, 401f
- TOTE-Schema 352
- Transformation, zulässige 588, 590f, 593
- Transitivität, schwache stochastische 578
- True-Score-Variablen 678
- Turing-Kriterium (s. Turing-Test)
- Turing-Maschine 150f, 154, 168, 170, 216
- Turing-Test 235f, 376f
- Übereinstimmungsmethode 431
- Überprüfbarkeit 476, 483
- Umkehrschluß, unzulässiger 430
- Umwelt, Invarianten der 163f
- Unabhängigkeit, regressive 663f
- , stochastische 660f
- Uniformitätsprinzip 441f, 446
- vermutung 456, 460
- Undeterminiertheit 225, 373
- Untersuchung, Fairneß 87, 605, 609, 613ff, 618, 620, 622f, 627f, 631f
- , Strenge 87, 605, 609, 613ff, 618, 620, 622, 627f, 632
- Unvollständigkeit 321, 323, 401, 415f
- , formaler Systeme 216
- Urbild 367, 368ff
- Validierung 221ff, 385
- Validität 21f, 41, 87ff, 500, 501f, 605, 613ff, 618ff, 623
- , Ableitungs- 502, 503, 504
- , externe 21, 23f, 82ff, 459
- , interne 21f, 82ff, 501
- , ökologische 501, 519, 537
- , Populations- 501
- , Situations- 501, 612
- , statistische 505

- , Variablen- 502
- Validitätskriterien 498
- Variable, abhängige 21, 484, 487, 501
  - , intervenierende 315
  - , konfundierte 32
  - , stochastische (s. Zufallsvariable) unabhängige 21, 484, 487, 501
- Varianzanalyse 607, 609, 630f
- Varianzanteil 682f
- Variationsbreite 438
- Verhaltensmessung 520ff
- Verhaltensmodell 373
- Verhaltensprotokolle 346
- Verifikation 406, 418, 440f, 497
- Verlässlichkeit 438
- Vermutungen, empirische 457ff
- Verstehen 98ff, 104, 105f, 126, 128, 130, 239, 353, 359
  - , Fremd- 98ff
  - , Selbst- 98f
  - , sozialwissenschaftliches 103, 111f, 128
  - , wissenschaftliches 101ff
- Versuchsperson 189
- Versuchsplan, interindividueller 525ff
  - , intraindividueller 525ff
  - , nicht-experimenteller 527ff
- Versuchsplanung 622
- Verteilung 657f
- Verteilungsfunktion 697
- Verwendungszusammenhang 445
- Vorgehen, deduktives 480
  - , hypothetico-deduktives 483
  - , induktives 480
  - , rationales 13f
- Vorhersage 490
- Vorhersage, statistische 490f, 494f, 498f, 503f
- Vorsichtigkeitsindex 435f
- Wahrheit 101, 313f, 317, 389, 392, 406f, 418, 421f
  - , mögliche 422
- Wahrnehmung 148, 160
  - , Aufgabenanalyse der -sfunktionen 162, 164
  - , Meßinstrumentkonzeption der 148ff, 155, 161, 172
  - , Täuschung der 158
  - , Unterdeterminiertheit der 158, 161
- Vagheit der 161
- Wahrnehmungsprozesse 152
- Wahrnehmungspsychologie 148
- Wahrscheinlichkeit 33f, 657f
  - , Aposteriori- 727
  - , Apriori- 726
  - , bedingte 660, 700
  - , frequentistische 727
  - , induktive 326, 406, 421
  - , logische 434f
  - , objektive 433
    - subjektive 726f, 433
- Wahrscheinlichkeitsmaß 655f
- Wahrscheinlichkeitsmodell 698f
- Wahrscheinlichkeitsraum 655
- Webersches Gesetz 142
- Wechselwirkung 356, 358
- Welten, mögliche 331
- Wert, wahrer 677f
- Wettssystem, faires 434
- Widerlegbarkeit 484
- Widerspruchsfreiheit 203, 221, 232, 303, 382, 389, 391f, 483
- Wiederholbarkeit (s. Replizierbarkeit)
- Wissen 218f, 222
- Wissensbasis 211
- Wissenschaft, analytische 355ff
  - empirische 7ff
- Wissenschaftsmode 257ff, 284ff
- Wissenschaftsparadigma 282ff
- Wissenschaftssystem 265f, 278, 280, 283f, 287
- Wissensebene 217f
- Wissensrepräsentation 204, 207
- Young-Helmholtz-Theorie des Farbensehens 149
- Zahl 561
- Zeichen 100, 108, 115
- Zufallsexperiment 435, 653ff
- Zufallsfehler 190, 514
- Zufallsstichprobe 83ff
- Zufallsvariable 657f, 697
- Zuordnung, adäquate 503
  - , suffiziente 503
- Zuordnungsregeln 309, 311
- Zusatzannahme 300, 408, 442
- Zustandsfunktion 351

## Autoren-Register

- Abel, U. 513, 548  
Abelson, R.P. 204, 247  
Achinstein, P. 433  
Adams, E. W. 590, 598  
Aebli, H. 199, 243  
Aertsen, A. 192, 195  
Agassi, J. 400, 422, 424  
Agazzi E. 283, 289  
Agnew, N.M. 6, 42  
Aiserman, M.A. 349, 386  
Aitkin, M. 640, 644  
Albert, D. 645  
Albert, H. 254, 289, 314, 317, 337, 390, 391, 424  
Alich, L.-M. 272, 289  
Allen, J. 204, 243  
Allert, T. 118, 134  
Amico, R.P. 445  
Andersen, E.B. 17, 42, 661, 685, 687  
Anderson, D. 640, 644  
Anderson, G. 55, 92  
Anderson, J.R. 203, 204, 215, 217, 239, 243, 296, 337, 367, 387, 404, 424  
Anderson, N.H. 568, 598  
Andersson, G. 259, 269, 282f, 286, 289, 399, 424  
Andres, J. 639, 644, 648  
Anstis, S. 161, 191  
Anzai, Y. 203, 243  
Arbuckle, J. 572, 598  
Ardila, R. 58f, 65, 92  
Aristoteles 168, 347, 387  
Arminger, G. 526, 539f, 548, 553  
Armstrong, B. D. M. 304, 337  
Aronson, E. 508ff, 526, 534f, 539, 548  
Ash, R.A. 658, 687  
Ashby, F.G. 640, 647  
Ashby, W.R. 368, 369, 387  
Asher, L. 196  
Atkinson, J. W. 255, 289, 296f, 313, 337, 397, 404, 410f, 419, 424  
Atkinson, R.C. 195f, 425  
Austin, J.L. 116, 130  
Austin, J.T. 653, 688  
Ax, A. F. 543, 548  
Baars, B.J. 321, 337  
Baton, F. 419, 431  
Baddeley, A. 607, 644  
Baird, J.C. 569, 598  
Bakan, D. 625, 644, 732, 734  
Balestra, D.J. 443  
Balzer, W. 261f, 290, 332, 334, 337f, 430, 457, 650, 687, 688  
Bandura, A. 528, 548  
Barber, T.X. 66, 92  
Barker, P. 283, 286, 291  
Barker, R. G. 53, 92  
Barker, S. 433  
Barlow, H.B. 165, 167, 175, 192  
Baron, J. 17, 42  
Baron, R. 172, 196  
Barr, W.F. 322, 338, 413, 425  
Bartley III, W. W. 400, 423, 425  
Bartussek, D. 669, 688  
Bassok, M. 227, 243  
Batchfelder, W. H. 635, 639, 644, 647  
Battig, W.F. 404, 425  
Bauer, H. 656, 658, 661, 663, 665, 688  
Baumrind, 518  
Bayertz, K. 283, 290  
Beatty, J. 734  
Behnke, H. 734  
Bell, D.E. 11, 42  
Bern, D.J. 416, 425  
Bennett, B.M. 174, 192  
Bentler, P.M. 484, 506, 540, 549, 653, 688  
Berger, P.L. 108, 130  
Bergmann, J.R. 103, 120, 131  
Berkeley 145, 148  
Berkowitz, L. 400, 410, 425  
Bernouilli 169, 564  
Bertalanffy, L. 355, 356, 387  
Berwick, R.C. 204, 243  
Bickel, P. J. 734  
Birch, D. 297, 337, 404  
Birdsall, T. G. 10, 44  
Birkett, 513  
Bischof, N. 67, 92, 297, 338, 349, 354, 358, 361, 362, 387  
Bishop, Y. M. M. 661, 688  
Blake, A. 196  
Bless, H. 522f, 543, 549, 554  
Blettner, M. 513, 555  
Block, N. 321, 338  
Bock, R.D. 580, 599  
Boehm, G. 100, 131f  
Boehnke, K. 733, 734  
Boesken, W.H. 526, 549  
Boguslaw, R. 368, 387  
Böhme, G. 560, 598  
Bohnen, A. 426  
Bohner, G. 543, 549  
Bohnsack, R. 103, 131  
Bollen, K.A. 682, 688  
Borg, I. 70, 92  
Borgman, C. L. 287, 292  
Boring, E.G. 143f, 192  
Bortz, J. 20, 21, 41, 42, 351, 361, 387, 478, 506, 733, 734



- Bosch, K. 658, 688  
 Bourbaki, N. 562, 599  
 Bower, G.H. 276f, 291  
 Box, G.E. 514, 549  
 Boyd, R. 170f, 192  
 Brachman, R. 204, 243  
 Braddick, O.J. 192, 197  
 Bradley, R.A. 9, 42  
 Bradshaw, 451  
 Braitenberg, V. 192  
 Braithwaite, R. B. 368, 369, 387  
 Brandtstädter, J. 291  
 Brauer, W. 216, 243  
 Breidenkamp, J. 22, 40, 41, 63, 76, 82, 85f, 92, 198, 226, 230, 457, 462, 481, 491, 496, 501f, 505f, 508f, 513f, 549, 605, 608ff, 620, 625, 627ff, 633, 638, 644, 649, 667, 669, 672, 688, 694, 706, 719, 728, 730f, 734, 735, 736  
 Brehm, J. W. 397, 425  
 Breiman, L. 658, 688  
 Brentano F. 143, 570, 599  
 Breuer, E. 251, 261, 290  
 Brewer, M. 539, 548  
 Bridgman P. W. 56, 92, 309, 338  
 Brinberg, D. 82, 92  
 Briskman, L.B. 257, 283, 290  
 Brocke, B. 50, 92, 272, 290  
 Brown, J.S. 221, 228, 243, 248, 688  
 Browne, M.W. 683  
 Brozek, J. 144, 192  
 Brunswick, E. 137, 157, 159f, 192  
 Buchner, A. 76f, 92, 627, 646  
 Bude, H. 112, 131  
 Buffart, H.F. 194  
 Bugelski, B.R. 607, 645  
 Bühler, K. 55f, 92, 100, 109, 131, 137, 157, 159  
 Buhmann, R. 102, 131  
 Bunge, M. 58f, 65, 92, 269, 272ff, 290, 306, 322, 331, 338, 393f, 400, 405, 425  
 Büning, H. 706, 733, 734  
 Bunt, A.A. 573, 600  
 Buring, I.E. 528f, 551  
 Burke, P. J. 661, 690  
 Burnstein 404  
 Burton, R.R. 229, 243  
 Bush, R.R. 23, 42, 206, 245  
 Buss, A.H. 544, 549  
 Butterfield, E. C. 265, 283, 286, 290, 426  
 Calder, B.J. 530, 551  
 Campbell, D.T. 82f, 86f, 92f, 440, 459, 512, 521, 526f, 534f, 549, 555, 605, 645  
 Campbell, N.R. 563  
 Cannell, C.F. 549  
 Cantril, H. 535, 551  
 Carbonell, J.G. 242, 243  
 Carlsmith, J.M. 508, 518, 535, 539, 543, 548f, 608, 646  
 Carnap, R. 42, 81f, 105, 131, 306f, 338, 396, 425, 429, 434, 438, 444, 455, 480, 506, 725  
 Carrier, M. 139, 192, 236, 243, 320, 338  
 Cartwright-Smith, J. 542, 552  
 Cassirer, E. 108, 131, 174, 192  
 Chalmers, A.F. 13, 18, 43, 311, 338, 509, 549  
 Chang, J.J. 173, 194  
 Chapanis, A. 368, 378, 384, 387  
 Chapman, L.J. 534, 549  
 Chapman, M. 60, 92  
 Chapman, D.T. 534, 549  
 Chassein, B. 522, 554f  
 Cheek, J.M. 669, 688  
 Chi, M. T. H. 204, 227, 243, 248  
 Chomsky, N. 118, 131, 204, 215, 243  
 Christensen, L. 545, 549  
 Church, A. 215, 243  
 Churchland, P.M. 319, 338, 394, 425  
 Churchland, P. S. 166, 192  
 Claparède, E. 61, 92  
 Clark 404  
 Clement, J. 454  
 Cliff, N. 599  
 Coen, C. W. 192  
 Cohen, J. 497, 506, 627, 629f, 632, 645, 717, 734  
 Cohen, L.J. 431f, 434f, 437f, 441, 449, 461  
 Cohen, M.A. 172, 180, 192  
 Cohen, M.R. 589, 599  
 Cohen, R.S. 426  
 Cohen, Y. 433, 449  
 Coker, M.C. 551  
 Collingwood, R. G. 331, 338  
 Collins, A.M. 203, 204, 208, 244  
 Colonius, H. 575, 578, 599, 687, 688  
 Comte, A. 57, 93  
 Conway, M. 523, 553  
 Cook, D.A. 521, 528f, 537, 551  
 Cook, T.D. 82, 93, 530, 549f  
 Cooley, C.H. 99, 131  
 Coombs, C.H. 15, 20, 43, 48, 52, 69f, 93, 326, 338, 560, 572, 601, 729, 734  
 Corwin, J. 636, 638, 647  
 Cox, J. W. R. 433  
 Craik, K. J. W. 170, 192  
 Crane, D. 287, 290  
 Cronbach, L.J. 637, 645  
 Cronbach, N.J. 313, 338  
 Crott, H. 265, 290

- Crowne, D. 525, 549  
 Cummins, R. 328, 338  
  
 Daiser, W. 217, 244  
 Dalbert, C. 526, 549  
 Damerow, P. 13, 14, 43  
 Danziger, K. 58f, 93  
 Darlington, R.B. 669, 688  
 Daston, L. 734  
 Daugman, J.G. 171, 192  
 Daves, R.M. 729, 734  
 David, F.N. 734  
 Davis, E. 204, 244  
 Dawes, R. M. 15, 43, 69, 326, 338  
 De Groot, A.D. 7, 12, 26, 43  
 De Gruijter, D.N.M. 654, 688  
 De Olivera, M.B. 446  
 Dearing, B.E. 79f, 94  
 Debreu, G. 564, 599  
 DeGreef, E. 687, 688  
 Del Rosario, M.L. 530, 551  
 Demets, D.L. 513, 550  
 Demokrit 401  
 Dennett, D.C. 139, 161, 171, 192  
 Deppe, W. 206, 244, 643, 645, 733, 734  
 Descartes, R. 13, 14, 43, 127f, 131, 561  
 DeValois, K.K. 175, 192  
 Diederich, W. 687, 688  
 Diepgen, R. 720, 722, 734  
 Diercks, S.R. 551  
 Dilthey, W. 102, 116f, 131  
 Dingler, H. 24, 43, 54, 93  
 Dittrich, K. 547, 549  
 Dollard, J. 315  
 Dörner, D. 16, 38, 66f, 76f, 81, 91f, 227, 244, 344, 358, 374, 378, 388, 479, 506, 519, 535f, 549  
 Dreistadt, R. 454  
 Dreyfus, H. 238, 244  
 Drösler, J. 567, 599  
  
 Duhem, P. 179f, 193, 408, 425, 643f  
 Dunbar, K. 451  
 Dupr{, J. 197  
 Durkheim, E. 128, 131  
 Dux, G. 134  
 Dykstra, R.L. 602  
  
 Eagly 528  
 Ebbinghaus, H. 257, 290  
 Eberle, T.S. 99, 131  
 Edgington, E. S. 513f, 550, 610, 645  
 Edidin, A. 446  
 Edler, L. 513, 555  
 Edwards, W. 726, 729, 734  
 Efron, B. 81, 93, 642, 645  
 Egan, J.P. 10, 43  
 Ehrhardt, K. J. 374, 387  
 Eid, M. 654, 658, 661, 663, 665, 676, 678, 681, 682, 693  
 Eimer, M. 328f, 338  
 Einhorn, J.H. 338  
 Einstein, A. 410  
 Ekstrand, B.R. 411, 427  
 Eller, F. 332  
 Ellsworth, P.C. 508, 548  
 Elpelt, B. 689  
 Engelkamp, J. 38, 43  
 Epstein, W. 54, 93  
 Erdfelder, E. 22, 35, 40, 41, 76, 82, 85f, 226f, 230, 457, 649, 667, 669, 672, 676, 688  
 Erdfelder, E. 605, 615f, 618, 627, 635, 638f, 645f  
 Ericsson, K.A. 55, 60f, 65, 93  
 Eron, L.D. 529, 552  
 Escolana, S.K. 298, 338  
 Essler, W.K. 251, 290, 430, 434f, 625, 645  
 Estes, W. K. 286, 290  
 Euklid 562  
 Eysenck, M. W. 55, 62, 66f, 93  
  
 Fagot, R.F. 590, 598  
 Fahrenkrug, H. 532, 550  
 Fallon, A.E. 543, 553  
 Falmagne, J.-C. 137, 146, 193, 567, 575f, 578f, 581, 585, 590, 594, 599, 620, 645, 676, 688  
 Faraday, M. 417  
 Farrell, E. 286, 290  
 Faßnacht, G. 51f, 93  
 Faul, F. 627, 633, 646  
 Fazio, R.H. 521f, 550f  
 Feather, N.T. 300, 338, 420, 425  
 Fechner, G.T. 137, 140ff, 144f, 146, 150, 156, 165, 168f, 192f, 558, 561, 564, 566, 569f, 574ff, 599  
 Feger, H. 51f, 55f, 70, 93, 198ff, 508, 550, 645f, 735, 736  
 Festinger, L. 296, 298, 535, 543, 550, 608, 646  
 Feyerabend, P. 6, 43, 68, 94, 317, 337f, 391, 393f, 399, 407, 425  
 Fichter, M.M. 532, 550  
 Fienberg, S.E. 661  
 Fischer, G.H. 654, 676, 685, 687, 689  
 Fishburn, P.C. 582, 599  
 Fisher, R. A. 511f, 550, 610, 695, 701, 704f, 709f, 717, 722, 727, 729, 733, 734  
 Flick, U. 114, 131f  
 Flores, F. 238, 248  
 Fodor, J. 320, 339  
 Fodor, J. A. 118, 133, 161, 166, 193  
 Foppa, K. 67, 94  
 Formann, A. 676, 689, 358  
 Fox, W. C. 10, 44  
 Franklin 183, 193  
 Franz, H.-W. 131  
 Freitag, B. 317, 341  
 Frenz, H.-G. 530, 550  
 Freudenthal, G. 13, 43

- Freudenthal, H. 707, 734  
 Frey, D. 414, 425, 645, 647  
 Frey, S. 530, 550  
 Friedeburg, L.V. 134  
 Friedmann, L.M. 513, 550  
 Frings, M.S. 134  
 Funke, J. 76f, 92  
 Furberg, C.D. 513, 550  
  
 Gabriel, K.R. 610, 646  
 Gadamer, H.-G. 100, 102, 131f  
 Gadenne, V. 16, 38, 55f, 60f, 75, 82f, 82f, 94, 177f, 236, 251, 290, 312, 314, 321, 325, 329, 336, 339, 357, 358, 395, 414f, 425, 440, 457f, 481, 491, 506, 509f, 514f, 531, 537, 540, 550, 604f, 611, 613, 643, 646  
 Gage, N.L. 645  
 Gähnde, U. 687, 689  
 Gaifman, H. 443, 450, 454  
 Galanter, E. 206, 245, 321, 340, 352, 388, 567f, 574, 580  
 Galilei, G. 282  
 Garnes, P.A. 80, 94  
 Gänssler, P. 658, 689  
 Gårdenfors, P. 433, 449  
 Gardner, H. 203, 244  
 Garfield, E. 287, 290  
 Garner, W.R. 174, 193  
 Garrett, M. 197  
 Garz, D. 103, 132ff  
 Gaston, J. 287, 290  
 Geissler, H.G. 197  
 Genesereth, M. R. 205, 208, 244  
 Gentner, D. 286, 290  
 Gerhardt, U. 103, 132  
 Gerjets, P. 33, 39, 81, 84, 465, 509, 605, 651  
 Geyer, S. 531f, 550  
 Ghiselli, E.E. 669  
  
 Gholson, B. 283, 286, 291  
 Gibson, J. J. 137, 155, 159ff, 163, 172, 193  
 Giddens, A. 103, 132  
 Gigerenzer, G. 33, 43, 49, 67, 73, 94, 171, 193, 203, 205, 244, 276, 285f, 291, 366, 386, 387, 579, 599, 696, 717, 722f, 731, 734  
 Glaser, B.G. 111f, 132  
 Glaser, R. 227, 243  
 Glück, G. 51, 94  
 Glymour, C. 336, 339, 398, 425  
 Gödel, K. 215, 244  
 Goffman, E. 128, 132  
 Goldstein, I. 320, 339  
 Gollwitzer, P.M. 547, 550  
 Gonzales, M.H. 508, 548  
 Good, I. J. 80, 94  
 Goodman, L.A. 661, 689  
 Goodman, N. 441, 445f, 455  
 Gopnik, A. 60, 94  
 Gorea, A. 191  
 Götz-Baltes, B. 668, 692  
 Gould, S. J. 171, 193, 559, 599  
 Grabitz, H. J. 397, 425  
 Grabitz-Gniech, G. 397, 425  
 Graham, N. 175, 188, 193  
 Grassmann, H. 179f, 193, 599  
 Grathoff, R. 103, 107, 132  
 Graubard, S.R. 151, 193  
 Graumann, C.F. 51, 55f, 93f, 281, 291, 292, 321, 339  
 Grawe, K. 35, 43  
 Green, D.M. 58, 94, 480, 507, 580, 637, 646  
 Greeno, J. 221, 248, 569, 602  
  
 Greve, W. 51f, 94  
 Grice, H.P. 522, 550  
 Griffin, D.R. 193  
 Groeben, N. 48, 56, 65, 94, 251, 291, 304, 317, 326, 341, 477, 485, 491f, 506, 604, 610, 646  
 Grondin, J. 100, 132  
 Grossberg, S. 172, 176, 180, 192f, 197  
 Grünbaum, A. 443  
 Grundmann, A. 56, 97  
 Gschneidinger, E. 523, 554  
 Gulliksen, H. 676, 677, 678, 689  
 Gundlach, H. 144, 192  
 Gurwitsch, A. 107, 132  
 Gussew, L.A. 386  
 Guthrie, W. K. C. 595, 600  
  
 Haagen, K. 489, 506, 710, 735  
 Habermas, J. 103, 126, 132, 134  
 Hacker, W. 270, 273, 291  
 Hacking, I. 183f, 193, 195, 694, 707, 715, 735  
 Hager, W. 87f, 94, 442, 457, 462, 478f, 505f, 487f, 494ff, 508, 551, 605, 609, 612ff, 618, 623, 627, 632, 635, 646f, 651, 689, 694, 723, 724, 732, 735, 737  
 Hambleton, R. H. 676, 688  
 Hanneman, R. A. 376, 383, 387  
 Hänni, R. 35, 43  
 Hanson, N.R. 185, 193  
 Harnatt, J. 493, 507  
 Hartigan, J.A. 80, 95  
 Härter, E. 658, 689  
 Hartung, J. 661, 689, 702, 735  
 Hartwig, F. 79f, 94  
 Hastorf, A.H. 535, 551

- Haugeland, J. 328, 339  
 Hauptert, B. 103, 132  
 Häußler, A. 531, 551  
 Hayduk, L.A. 682, 689  
 Hays, W.L. 495, 500, 507  
 Heath, L. 530, 551  
 Hebb, D.O. 268  
 Heckhausen, H. 255,  
 257, 286, 291, 299,  
 316,339  
 Heckhausen, T. 69f, 94  
 Heidbrink, H. 283, 291  
 Heidegger, M. 102, 132  
 Heider, F. 157, 159, 172,  
 194  
 Heinze, T. 135  
 Heise, E. 465  
 Hejl, P.M. 259f, 291  
 Hehing, I. 104, 132  
 Hempel, C. G. 325f, 339,  
 434, 437, 441, 483, 507  
 Henkel, R.E. 493, 507,  
 714, 695, 735  
 Hennekens, C. H. 528f,  
 551  
 Hennigan, 539, 530  
 Henry, N. W. 639, 646  
 Herbart 142, 168, 174  
 Hering, E. 137, 143, 150,  
 157, 159, 165, 172, 194  
 Herr, P.M. 524, 551  
 Herrmann, C. 53, 87, 96  
 Herrmann, T. 24, 38, 43,  
 50f, 76, 91f, 183f, 187,  
 191, 194, 202, 218,  
 244, 252, 255f, 261,  
 263, 265, 269, 271,  
 273, 275, 281, 284,  
 286, 291, 293, 315,  
 328f, 332, 339, 349,  
 358, 361, 387, 394f,  
 538, 551, 558, 634,  
 646, 695, 735  
 Herrnstein, R. J. 195  
 Hesse, M. 433f, 441,  
 446, 450, 454  
 Hesse, M.B. 363, 387  
 Heyer, D. 582, 585f, 600  
 Hickey, T. 87, 95  
 Hildenbrand, B. 103,  
 108, 132  
 Hildreth, E.C. 204, 244  
 Hilgard, E.R. 276, 277,  
**291**  
 Hinde, J. 640, 644  
 Hinderer, K. 658, 689  
 Hintikka, J. 437  
 Hippler, 522f, 554f  
 Hirschauer, S. 108, **132**  
 Hitzler, R. 35, 58, 66,  
**72**, 107f, 130, 132f, 227  
 Hoagiin, D.C. 79, 95  
 Hobbes, T. 169  
 Hodapp, V. 540, 551  
 Hofer, M. 476, 507  
 Hoffman, D.D. 161,  
 174f, 192, 194, 196  
 Hoffman, W.C. 174, 194  
 Hoffmann, N. 294  
 Hoffmeister, H. 531, 551  
 Hofstätter, P. 361, 388  
 Hogarth, R.M. 338  
 Hölder, O. 563, 600  
 Holland, J. 203, 244,  
 449ff  
 Holland, P.W. 661  
 Hollerbach, J.M. 203, 246  
 Holling, H. 687, 690  
 Holmann, E. W. 580, 600  
 Holyoak, K.J. 203, 244,  
 449ff  
 Holzkamp, K. 24, 43,  
 55, 88, 95  
 Honer, A. 103, 107, 114,  
 132f  
 Hookway, C. 192  
 Hopcroft, J.E. 216, 244  
 Hoppe, H. 433  
 Hörmann, H. 281, 285,  
 291  
 Hosmer, D.W. 539, 551  
 Hsu, C.-E 610, 646  
 Hubbard, M. 545f, 553  
 Hubel, D.H. 176, 183,  
**195**  
 Huber, O. 484, 507, 533,  
 551  
 Hubert, L. 77, 95  
 Hübner, K. 441  
 Huesman, L.R. 529, 552  
 Huh, C.L. 277, 303,  
**340, 404**  
 Hultsch, D.F. 87, 95  
 Hume, D. 84, 95, 145,  
 148, 432  
 Hummell, H.J. 511, 551  
 Hunter, J.S. 514, 549  
 Hunter, W. G. 549  
 Husserl, E. 107, 130, 133  
 Hussy, W. 40, 41, 82,  
 87f, 254, 291, 457,  
 479, 484, 507, 516,  
 618, 635  
 Hyland, M. 372, 374, 388  
 Irle, M. 253, 291, 397,  
 425, 544, 551  
 Irmen, L. 648  
 Irwin, M.E. 551  
 Isefer, A. 648  
 Iser, W. 102, 133  
 Iverson, G. J. 620, 646  
 Iverson, J. 582, 600  
 Jackendoff, R. 316, 321,  
 340  
 Jacob, F. 181, 194  
 Jagodzinski, W. 687, 689  
 James, W. 18, 43, 55,  
 162, 194  
 Jarvie, I.C. 426  
 Jason, G. J. 431  
 Jauß, H.R. 102, 133  
 Jepson, D. 453  
 Johnson, E. J. 61, 96  
 Johnson, N.L. 626, 646  
 Johnson-Laird, P. N.  
 202f, 244, 373, 388  
 Jones, G.E. 441f, 446  
 Jones, L.V. 580, 599  
 Jones, R.M. 204, 227, 248  
 Jöreskog, K.G. 484, 507,  
 540, 551, 638f, 646,  
 682, 685, 690  
 Josephson, W.L. 539, 551  
 Judd, C.M. 669, 690  
 Julesz, B. 173, 183, 194

- Kächele, H. 25, 43  
 Kahn, R.L. 549  
 Kahneman, D. 555  
 Kalish, D. 427  
 Kalke, W. 151, 153, 194  
 Kallmeyer, W. 119, 133  
 Kant, I. 57, 95, 696  
 Kantowitz, B.H. 196  
 Katz, D. 150, 185, 194  
 Katz, J.J. 118, 133  
 Kaufmann, L. 59, 95  
 Keane, M.T. 67, 93  
 Keith-Spiegel, P. 542, 551  
 Kelvin 180  
 Kemeny, J.G. 432, 434  
 Kemke, C. 174, 194  
 Kendall, M.G. 709, 735  
 Kendlner, T.S. 87, 95  
 Kennedy, J.J. 539, 551  
 Kenny, D.A. 669, 690  
 Kepler, J. 410  
 Keppler 177  
 Kern, D. 555  
 Keupp, H. 114, 131f  
 Klages, H. 267f, 271, 291  
 Klahr, D. 451  
 Kleck, R.E. 542, 552  
 Kleene, S.C. 215, 244  
 Kleiner, B. 80, 95  
 Kleiter, G.D. 726, 727, 729, 730, 735  
 Klir, J. 373, 388  
 Klix, F. 207, 245, 286, 291  
 Klopfer, B. 515, 551  
 Klösener, K.-H. 689  
 Klumpp, G. 523, 554  
 Klusemann, H. W. 135  
 Kluwe, R.H. 61, 95  
 Knauth, B. 103, 133  
 Knoblauch, H. 103, 108, 133  
 Knoche, N. 678, 690  
 Knoke, D. 661, 690  
 Knorr-Cetina, K. 103, 108, 114, 133  
 Knowles, E. 526, 551  
 Kocher, G.P. 551  
 Kodratoff, Y. 242, 245  
 Koenderink, J. J. 161, 175, 194  
 Koeppler, K. 485, 491, 507  
 Koh, K. 451  
 Kohler, I. 171, 194  
 Köhler, W. 172  
 Kohlmeier, L. 526, 531, 551  
 Kolmogoroff, A. 651, 690  
 Konau, E. 118, 134  
 König, R. 126, 133  
 Koocher, 542  
 Kopernikus, N. 412  
 Kordig, D.R. 283, 291  
 Kosslyn, S.M. 203, 246  
 Kotz, S. 626, 646  
 Kraiker, C. 332, 340  
 Kraimer, K. 103, 132ff  
 Krambeck, J. 118, 134  
 Krämer, S. 169, 194, 215, 245  
 Krantz, D.H. 164, 175, 186, 194, 196f, 453, 564, 568f, 571, 580, 587, 590, 600, 603, 626, 647, 676, 691, 730, 735  
 Krapp, A. 476, 507  
 Krauth, J. 661, 690  
 Krickeberg, K. 658, 690  
 Kriz, J. 283, 291  
 Krohn, W. 599  
 Krohne, H. W. 263ff, 272, 274, 291  
 Kronthaler, E. 432, 434, 447  
 Krüger, L. 734  
 Kruglanski, A. W. 398, 426  
 Krull, W. 562, 600  
 Krumhansl, C.L. 145, 195, 567, 576  
 Kruse, L. 542, 552  
 Kruse, P. 260, 293  
 Kruskal, W.H. 661, 689  
 Kubinger, K. 654, 676, 690  
 Kubovy, M. 172, 187, 195f  
 Kuhn, T.S. 184, 187, 190, 195, 282f, 285, 292, 317, 336, 340, 398f, 407f, 414, 426, 453f, 587, 600  
 Kuipers, T. A. F. 317, 340, 437  
 Kulkarni, D. 451  
 Külpe, O. 314, 340  
 Kummer, W. 434  
 Kumpf, M. 542, 552  
 Kunda, Z. 453  
 Kyburg, H.E. 433, 437  
 La Mettrie, J.O. de 169, 195  
 Lachman, J.L. 265, 282, 292, 426  
 Lachman, R. 265, 276, 282f, 286, 292, 410, 426  
 Laird, J. E. 204, 215, 245  
 Lakatos, I. 67, 81f, 86, 95, 252, 263, 283, 292, 407, 412, 418f, 425, 435ff, 605, 609, 646, 696, 711, 724, 735  
 Laming, D.R. J. 147, 195, 580, 600  
 Land 183  
 Lane, N.R. 440  
 Lane, S.A. 440  
 Langeheine, J. 639, 647  
 Langeheine, R. 661, 676, 685, 690, 691  
 Langer, S. K. 108, 133  
 Langley, P.A. 451  
 Lantermann, E.D. 67, 93, 519, 535f, 549  
 Lanzetta, J.T. 542, 552  
 Larimer, J. 572, 598  
 Larsen, R. J. 544, 552  
 Lashley, K. 166  
 Lasniak, H. 203, 246  
 Lau, T. 103, 108, 133  
 Laudan, L. 283, 292, 398, 426, 431  
 Lazarsfeld, P.F. 639, 646  
 Leahey, T.H. 348, 355, 359, 388

- Leeper, R. 54, 95  
 Leeuwenberg, E. L. J. 194  
 Lefkowitz, M.M. 529, 552  
 Lehmann, G. 678, 690  
 Leibniz 142, 168f  
 Leiser, E. 724, 735  
 Lemeshow, S. 539, 551  
 Lenzen, W. 438, 441, 444, 694, 735  
 Leontjew, A.N. 270, 273, 292  
 Lepper, M.R. 480f, 507, 545f, 553  
 Levelt, W. J. M. 276, 292, 573, 579, 600  
 Levesque, H. 204, 243  
 Levi, I. 435f, 449  
 Levine, M.V. 63f, 95, 576, 600  
 Levinson, A. 441  
 Lewin, K. 174, 296, 340  
 Lewis, C. 653, 676, 690  
 Lewis, D. 331, 340  
 Lewis, M.W. 227, 243  
 Lewontin, R.C. 171, 193  
 Liebmann, S. 184, 195  
 Lienert, G.A. 633, 647, 661, 676, 690, 733, 734  
 Lihi, W. 521, 553  
 Lindgren, B.W. 701, 706, 735  
 Lindman, H. 726, 734  
 Lindzey, G. 195  
 Lingle, J.H. 523, 552  
 Link, S.W. 197  
 Litwin, G.H. 300, 337, 411, 424  
 Livingstone, M. S. 176, 183, 195  
 Locke 145, 148  
 Loeve, M. 658, 690  
 Loftus, E.F. 204, 244  
 Lord, F.M. 9, 43, 676, 677, 678, 691  
 Lucas, R. J. 170, 195  
 Luce, R.D. 8f, 43, 137, 145, 164, 195ff, 206, 245, 564, 567f, 571, 574f, 576ff, 590ff, 600, 603, 626, 640, 647, 676, 690, 691, 730, 735  
 Lück, H.E. 287, 291f  
 Luckmann, B. 107, 133  
 Luckmann, T. 100, 103, 106ff, 111, 124, 130, 133f  
 Lüders, C. 64, 95, 103, 133, 207, 246, 646f  
 Lundeen, E. J. 551  
 Lykken, D.T. 724, 735  
 Lyons, W. 55f, 95, 340  
 Mac Cormack, E.R. 168, 195  
 Mach, E. 143, 150, 157ff, 165, 182, 195  
 Machina, K.F. 430  
 Mackie, J.L. 330, 340, 441  
 Makov, U.E. 639, 647  
 Malewski, A. 325, 340, 398, 416, 426  
 Malpas, J. 211, 245  
 Mandler, G. 321, 340  
 Manns, M. 53, 96  
 Marcovici, S. 620, 646  
 Mark, M.M. 526, 552  
 Markov, A. A. 215, 245  
 Marley, A.A. 580  
 Marlowe, D. 525, 549  
 Marquard, O. 104  
 Marr, D. 137, 152f, 158, 161, 181, 195, 204, 217, 245  
 Martin, G. 562, 601  
 Martin, L.L. 522, 554  
 Martinez, T. 174, 196  
 Marx, M. 369, 388  
 Masterman, M. 282, 292  
 Matthiesen, U. 103, 134  
 Maturana, H.R. 260, 292, 317, 340  
 Mausfeld, R. 34, 42, 73f, 89, 163, 177, 181, 195, 560, 566, 585f, 596f  
 McClelland, D. C. 255, 268, 292, 297, 313, 340  
 McClelland, G. H. 572, 601  
 McClelland, J. A. 206, 245, 247  
 McClelland, J. L. 451  
 McCorquodale, K. 303, 314f, 340  
 McGrath J.E. 82, 92  
 McGuigan, F.J. 484, 507  
 McLaughlin, P. 13, 43  
 Mead, G.H. 100, 119, 124, 134, 286, 292  
 Meehl, P.E. 87, 96, 303, 313f, 338, 340, 491, 497, 507, 637, 645, 724, 735  
 Mehler, J. 197  
 Meja, V. 135  
 Mergenthaler, E. 25, 43  
 Merkel, J. 570, 601  
 Merton, R.K. 287, 292  
 Merz, F. 552  
 Metz, G. 520, 552  
 Metzger, W. 276, 278, 292, 552  
 Meyer, W.U. 255, 268, 292  
 Middleton, D. 10, 44  
 Milgram, S. 518, 543f, 552  
 Mill, J.S. 419, 431, 508, 552, 669, 691  
 Miller, D. 404, 423, 426, 443f  
 Miller, G.A. 315, 321, 340, 352, 388  
 Mills, J. 519, 548  
 Minsky, M. 204, 215, 245  
 Mittelst&, J. 139, 192, 236, 243, 320, 338, 561, 601  
 Mittenecker, E. 725, 735  
 Möbus, C. 207, 245, 685, 691  
 Mohyeldin Said, K.A. 151, 195f  
 Möller, H. 40f, 82, 87f, 457, 484, 497, 506f, 516, 618, 635  
 Mondriaan 183

- Montada, 265, 292  
 Moosbrugger, H. 669, 691  
 Moravcsik, J.M. 206, 225, 245  
 Morgans, L. 404  
 Morgenstern, O. 564  
 Morrison, D.E. 493, 507, 714, 735  
 Mosteller, F. 23, 42, 79, 95f  
 Moulines, C.U. 430, 457, 650, 687, 688  
 Moulton, R. W. 301, 340  
 Mulaik, S.A. 653, 691  
 Müller, G. 522, 540, 554  
 Müller, G.E. 143, 196  
 Müller, P.H. 665, 691  
 Müller-Doohm, S. 103, 134  
 Mura, A. 443  
 Murray, D.J. 33, 43  
 Murray, H.A. 297, 316f, 340, 404, 412, 419f, 423, 425  
 Musgrave, A. 283, 292, 646, 735  
 Muthig, P. 485, 491, 507  
  
 Nagel, E. 7f, 29, 43, 433, 589, 599  
 Nag], W. 685, 691  
 Narens, L. 564, 571, 574, 582, 590f, 594f, 599, 601f  
 Nassen, U. 100, 34  
 Naß, E. 531, 551  
 Nebel, A. 522f, 552, 555  
 Neches, R. 206, 245  
 Neel, A.F. 349, 388  
 Neisser, U. 534, 552  
 Nell, K. 332  
 Nelson, R. J. 26f, 43, 44  
 Neville, J. W. 551  
 Newell, A. 7, 29, 44, 203f, 213f, 217f, 220, 237, 245f, 286, 292, 320, 340, 385, 388, 451f  
 Newton, I. 179, 282, 401, 412, 419, 696  
 Newton-Smith, W. H. 195f  
 Neyman J. 507, 609, 611, 694, 702, 704f, 710f, 716f, 731, 732, 735, 736  
 Niederée, R. 163, 195, 562, 565, 575, 582, 586, 602  
 Niiniluoto, I. 437f, 441  
 Nilsson, N.J. 205, 208, 244  
 Nisbett, R.E. 57, 62, 96, 203, 244, 404, 426, 432, 448ff, 462, 480, 507  
 Noack, H. 543, 553  
 Noma, E. 569, 598  
 Norman, D. A. 296, 341  
 Novick, M.R. 9, 43, 676, 677, 678, 691  
 Nowack, W. 539, 552  
 Nüse, R. 317, 341  
 Nuwer, K. 172, 196  
  
 Occam, W. 403  
 Oelkers, J. 135  
 Oerter, R. 265, 292  
 Oevermann, U. 103, 111, 118, 120f, 134  
 Ohlsson, S. 227, 246  
 Oldenbürger, H.-A. 78, 96  
 Opp, K.D. 484, 507  
 Opwis, K. 25, 36, 199, 203, 207, 223, 246, 451  
 Orne, M. 515, 552  
 Ortony, A. 168, 192, 196  
 Osherson, S. 203, 246  
 Ostmann, A. 33, 40, 41, 42, 86, 609, 615, 684  
 Ostrom, T.M. 523f, 552f  
 Oswald, M. 55f, 59f, 63, 96, 320f, 329, 339f, 395, 414, 425  
 Pachella, R. G. 178, 196  
 Pähler, K. 418, 426  
 Palacios, A. 198  
 Palermo, D. S. 283, 293  
 Palm, G. 195  
 Palmer, S.E. 28f, 44  
 Papert, S. 320  
 Parducci, A. 524, 553  
 Parker, S. 565, 602  
 Paul, W.J. 44  
 Pawlow, J.P. 268, 359, 371, 388, 410f, 426  
 Pearson, E.S. 507, 609, 611, 695, 716f, 731, **736**  
 Pearson, K. 704, 709, 710, 716  
 Penner, L.A. 669, 691  
 Penrose, R. 151, 170, **196, 238, 246**  
 Pentland, A.P. 196  
 Perry, C. 441f  
 Petefi, J.S. 293  
 Peterson, W. W. 10, 44  
 Pfanzagl, J. 565, 574, 590, 595, 602  
 Pfeifer, E. 378, 387  
 Pinker, S. 204, 246  
 Plachky, D. 658, 691  
 Planck, M. 140  
 Plateau, M. J. 565, 570, 602  
 Plato 168  
 Plessner, H. 107, 134  
 Plötzner, R. 223, 246  
 Pocock, S.J. 553  
 Podlech, A. 293  
 Poggio, T. 161, 196  
 Pollock, J.L. 430  
 Pomerantz, J.R. 172, 195f  
 Pongratz, L.J. 55, 60, 96  
 Poppen K. 7, 8, 44, 54f, 75, 81f, 86, 96, 126, 134, 254, 292, 314, 325, 336, 341, 343f, 388, 396f, 399f, 407f, 411, 419f, 426, 429, 432, 437, 440f, 443f, 455, 461, 604, 647, 649, 691, 695, 725, 731, 736  
 Porter, T. 734

- Posner, M.I. 29, 44, 203,  
**246**
- Post, E. 215, 246
- Prakash, C. 174, 192
- Prell, S. 476, 507
- Pribram, K.H. 172, 196,  
321, 340, 352, 388
- Price, L. 515, 553
- Pritsch, M. 543, 553
- Pugh, A.L. 384, 388
- Putnam, H. 139, 150f,  
166, 170, 196, 320,  
341, 441
- Pyke, S. W. 6, 42
- Pylyshyn, Z.W. 152,  
161, 170, 190, 193,  
196, 213, 217, 237,  
238, 246, 285, 292,  
320, 341
- Pythagoras 150
- Quillian, M.R. 204,  
208, 246
- Quine, W.V. 157, 196,  
643f
- Radnitzky, G. 426
- Raiffa, H. 11, 42
- Ramachandran, V. S. 155,  
161, 181, 186, 196
- Ramsey, F.P. 564, 602
- Ramsey, J.O. 594, 602
- Rapaport, D. 303, 341
- Rasch, G. 654, 655, 691
- Ratliff, F. 164, 196
- Readhead, M. 177, 196
- Reber, A.S. 79, 96
- Reese, H. W. 257, 286,  
**292**
- Reeves, B. 287, 292
- Rehm, J. 521, 531f, 539,  
548f, 550, 553
- Rehm, R. 30, 40, 52,  
**622, 624**
- Reichardt, W. 175, 196
- Reichenbach, H. 49, 66f,  
**75, 81f, 96, 431, 434,**  
**446**
- Reichertz, J. 103, 108,  
134
- Reid, T. **148**
- Reimann, P. **227, 229,**  
**243, 247**
- Renn, J. **13, 43**
- Renyi, A. 658, 691
- Rescher, N. 171, 196
- Restle, F. 569, 602
- Reumann, S.R. 161, 196
- Rice, R.E. 287, 292
- Richards, W.A. 175, 194
- Richardson, G. P. 384,  
**388**
- Riecken, 535
- Riedel, M. 100, 134
- Riedwyl, H. 80, 96
- Riefer, D.M. 635, 639,  
644, 647
- Riemann, G. 103, 134
- Riemersma, J.B. 573, 600
- Risteau, C.A. 193
- Ritchie, B.F. 427
- Rittenauer-Schatka, H.  
523, 554
- Ritter, H. 174, 196
- Rivadulla, A. 435
- Roberts, F.S. 578, 593,  
602, 676, 691
- Robertson, T. 602
- Robinson, R. 590, 598
- Rock, I. 54, 93
- Rodin, J. 542, 553
- Rogers, H. 215, 247
- Rogosa, D. 530, 553
- Rohracher, H. 55
- Röhrs, A. 56, 97
- Rollman, G.B. 542, 553
- Rorty, R. 139, 185, 197,  
319, 341
- Rosenbaum, Z. 593
- Rosenbloom, P. S. 204,  
**215, 245**
- Rosenkrantz, R. D. 433
- Rosenthal, R. 515, 553
- Roskam, E.E. 20, 44,  
47f, 69f, 96, 560, 567,  
602, 687, 691
- Rosonoer, L.I. 386
- Ross, D. 528, 543, 548
- Ross, L. 404, 426
- Ross, L.D. 542, 545f, 553
- Ross, S.A. 528, 548
- Rössner, L. 272, 289
- Rost, J. 639, 647, 653,  
654, 655, 676, 685,  
690f
- Roth, E. 353, 388
- Roth, G. 317, 341
- Rothacker, E. 276, 292
- Rothman, K. 532, 553
- Rozeboom, W. W. 595f,  
602, 724, 733, 736
- Rozin, P. 543, 553
- Rucci, A.J. 286, 292
- Rumelhart, D. E. 206,  
245, 247, 296, 340, 451
- Runesone, S. 172, 197
- Russo, J.E. 61, 65, 96
- Rutherford 410
- Rützel, E. 728, 736
- Sahner, H. 260, 261, 287,  
**293**
- Salancik, G.R. 523, 553
- Saimon, W.C. 326, 332f,  
**421, 426,** 430f, 435,  
**441, 445f**
- Saris, W. 682, 692
- Sarnoff, I. 543, 554
- Saunders, D.R. 669, 692
- Savage, L.J. 726, 734
- Schachter, S. 535, 554
- Schafroth, M. 80, 96
- Schank, R.P. 204, 239,  
**247**
- Scheele, B. 65, 94
- Scheerer, E. 144, 197,  
**287, 293**
- Schefe, P. 206, 247
- Scheler, M. 109, 134
- Schilpp, P.A. 426
- Schleiermacher, F. D. E.  
102, 134
- Schlesinger, G. 447
- Schlesselman, J. J. 532,  
**554**
- Schlick, M. 315, 341
- Schlürer, J. 547, 549
- Schmid, M. 398, 427



- Schmitt, M. J. 655, 668,  
669, 685, 692, 693
- Schmolensky, P. 206, 247
- Schneider, B. 565, 602
- Schneirla, T.C. 361, 388
- Schoemaker, P. J. H. 15f,  
44
- Schreier, M. 317, 341
- Schröder, P. 133
- Schroeder-Heister, P.  
561, 601
- Schröer, N. 120, 134
- Schuler, H. 542, 546, 554
- Schult, T. J. 227, 247
- Schulten, K. 174, 196
- Schultz, J. 77, 95
- Schultze, J. 53, 96
- Schulz, T. 485, 491, 507
- Schurz, G. 336, 341
- Schütz, A. 100, 103ff,  
111, 122ff, 134f
- Schütze, F. 103, 135
- Schwager, K. W. 75, 96
- Schwartz, E.L. 192
- Schwarz, N. 521ff, 543,  
549f, 554f
- Schwenkmezger, P. 526,  
549
- Scott, 564
- Scriven, M. 645
- Searle, J. R. 116, 135,  
155, 170, 197, 234,  
238, 239, 247, 321,  
341, 377, 388
- Sechrest, L. 521, 555
- Seifert, H.G. 506, 710,  
735
- Sehen, R. 11, 44
- Selz, O. 227, 247
- Semmer, N. 35, 43
- Sergejew, J. 344, 388
- Shafer, G. 728, 736
- Shapere, D. 254, 255,  
267, 270, 275, 277, 293
- Shapiro, E. 211, 247
- Shaw, J.C. 203, 245, 451
- Shepard, R.N. 152, 160f,  
172, 179, 197
- Sherman, S.J. 524, 551
- Shiffrin, R.M. 410, 425
- Shirley, E.S. 433
- Shye, S. 197
- Sibicky, M.E. 551
- Sieber, J.E. 542, 554
- Sievers, W. 642, 647
- Sigel, S. 736
- Simmel, G. 102, 135
- Simon, D.P. 229, 247
- Simon, H.A. 26, 29, 44,  
60f, 65, 93, 203, 229,  
237, 243, 245, 246,  
247, 286, 292, 320,  
340, 385, 388, 451f
- Simons, A. 523, 554
- Sinclair, R.C. 526, 552
- Singer, W. 172, 197
- Sinnett, L.M. 544, 552
- Sirovich, L. 164, 196
- Skinner, B.F. 277, 285,  
293, 315, 329, 341, 413
- Sleigh, A.C. 192, 197
- Sloman, A. 161, 197
- Smart, J.J.C. 181, 197
- Smirnowa, I.M. 386
- Smith, A. F. M. 639, 647
- Smith, B.C. 203, 213, 247
- Smith, E.E. 203, 244, 246
- Smolensky, P. 152, 197
- Smolka-Koerdt, G. 133
- Sneed, J.D. 38, 44, 261,  
262, 290, 293, 332,  
341, 430, 457, 650,  
687, 688, 692
- Snodgrass, J.G. 636, 638,  
647
- Sober, E. 171, 197
- Soeffner, H.-G. 35, 58,  
66, 72, 103, 108f, 111,  
121, 130, 134f, 227
- Solomon, R.L. 526, 554
- Sörbom, D. 484, 507,  
540, 551, 638f, 646,  
682, 685, 690
- Spada, H. 25, 36, 199,  
223, 229, 246, 247,  
451, 685, 691
- Spangenberg, P.M. 133
- Spence, K. W. 416, 425,  
427
- Sperling, G. 173, 175, 197
- Spinoza 141
- Spiro, K. 196
- Spohn, W. 437, 676, 692
- Srubar, I. 109, 135
- Stachowiak, H. 96, 205,  
247, 252f, 269, 271f,  
289, 293
- Stanley, J.C. 82f, 86, 92,  
459, 512, 526f, 530,  
534f, 549, 605, 645
- Stapf, K.-H. 645
- Steffen, 528
- Steger, H. 133
- Stegmühen W. 17, 261,  
268f, 271, 283, 293,  
325f, 329, 332, 341,  
430f, 435, 438f, 444,  
447, 457, 676, 687,  
689, 692, 615, 647,  
694, 696, 714, 715,  
725, 727, 728, 736
- Stehr, N. 135
- Steiger, J.H. 540, 554
- Stein, D. 565, 602
- Steiner, H.G. 707, 734
- Steinleitner, M. 521, 553
- Stelzl, I. 540f, 554
- Stemmer, N. 433, 449
- Stephan, E. 255, 261,  
263, 294, 313, 332,  
337, 342, 687, 692
- Stephens, D.L. 61, 96
- Sterling, L. 211, 247
- Stetter, F. 216, 247
- Stevens, S.S. 137, 144f,  
150, 155, 197, 564,  
568f, 576, 589, 597,  
602f
- Steyer, R. 674f, 678, 681,  
682, 685, 687f, 692,  
693
- Steyer, R. 9, 40, 41, 82f,  
85, 460, 619, 622, 624,  
647, 649, 654, 655,  
658, 661, 663f, 671,  
672
- Stigler, S. 709, 736
- Stiles, W.S. 179, 198
- Störig, H.J. 14, 44

- Strack 30, 40, 52, 622,  
**624**, 523f, 543, 549,  
**552**, 554f
- Strauss, A.L. 111f, 128,  
 132, 135
- Strauß, B. 653, 676, 692
- Strawson, P.E. 447
- Stronkhorst, L. E. 682,  
**692**
- Strube, G. 206, 247, 451,  
**579**, **599**
- Stuart, A. 709, 735
- Student, 514, 555
- Stumpf, M. 223, 246
- Stute, W. 658, 689
- Suck, R. 687
- Sukale, M. 332, 342
- Suppe, F. 283, 293f, 296,  
**332**, **342**, **434**
- Suppes, P. 164, 196f,  
 332f, 342, 559, 564,  
 573f, 576ff, 586, 603,  
 626, 647, 676, 690,  
 691, 693, 730, 735
- Swaminathan, H. 676,  
**689**
- Swets, J.A. 58, 94, 637,  
**646**
- Swijtink, Z.G. 32, 44,  
**734**
- Szagan, G. 265, 294
- Tack, W.H. 3, 12, 17,  
 44, 48, 73, 96, 137,  
 177, 198, 206, 247,  
 369, 388, 567, 571,  
 577, 603, 639, 647,  
 678, 693
- Tal, A.A. 386
- Tarski, A. 430
- Teegen, F. 56, 97
- Terry, M.E. 9, 42
- Textor, M.R. 542, 555
- Thagard, P. 203, 244,  
**432**, 448f, 454f, 462
- Thissen, D. 80, 97
- Tholey, P. 728, 729, 736
- Thomas, H. 89f, 97
- Thompson, E. 198
- Thorndike, L.L. 277
- Thurstone, L. L. 569,  
 575ff, 585, 676, 693
- Tillmann-Bartylla, D. 133
- Titchener, E.B. 55
- Titterington, D.M. 639,  
 641, 647
- Tolman, E.C. 296, 303,  
 319, 342, 342, 416, 427
- Torgerson, W.S. 653, 693
- Toulmin, S. 283, 294,  
 315, 333, 342
- Townsend, J.T. 197, 640,  
**647**
- Traxel, W. 553, 388
- Traxel, W. 511, 555
- Trenkler, G. 706, 733, 734
- Troscianko, T. 196
- Tschan, F. 35, 43
- Tseelon, E. 5, 44
- Tukey, J. W. 96f
- Tukey, W. 79f, 95
- Tuomela, R. 337, 342, 441
- Turing, A. 150f, 154,  
 168, 215, 235, 247,  
 377, 388
- Turner, J.C. 537, 555
- Tutz, G. 654, 693
- Tversky, A. 11, 15, 42,  
 43, 69, 197, 326, 338,  
 465, 523, 555, 564,  
 603, 626, 647, 676,  
 690f, 729, 730, 734,  
 735
- Tweney, R.D. 286, 292
- Ueckert, H. 332, 342
- Ullman, G. 294
- Ullman, J. D. 216, 244
- Ullman, S. 161, 198, 204,  
 244
- Ulrich, H. 291
- Underwood, B. J. 411,  
 427
- Ungeheuer, G. 135
- Upshaw, H.S. 524, 553
- Uttal, W.R. 179, 190, 198
- Valach, M. 373, 388
- van Acker, P.O.F.C. 10,  
 44
- van Buggenhaut, J. 687
- van de Pol, F. 685, 690
- van den Daele, W. 599
- van der Kamp, L.J. T.  
 654, 688
- van Fraasen, B. 315
- van Fraasen, B. C. 198,  
 595
- van Meter, D. 10, 44
- van Meurs, A. 692
- van Santen, J. P. H. 175,  
 197
- van Wright, G.H. 432
- VanLehn, K. 204, 221,  
 225, 227, 228, 231,  
 243, 247, 248
- Varela, F.J. 198
- Vartanian, A. 169, 198
- Viale, R. 195f
- Viertl R. 658, 693
- von Cranach, M. 67, 94
- von der Malsburg, C.  
 165, 195
- von Eye, A. 484, 507,  
 645
- von Fritz, K. 139, 193,  
 595, 599
- von Glasersfeld, E. 317,  
 339
- von Goethe, J. W. 401
- von Helmholtz, H. 143,  
 149, 158, 160, 165,  
 169, 179, 194, 563, 600
- von Humboldt, W. 102
- von Kardorff, E. 114,  
 131f
- von Kries, J. 143, 146f,  
 194, 564, 600
- von Kutschera, F. 403,  
 426, 430, 433f, 437f,  
 445, 448
- von Metraux, A. 132,  
 287, 292
- von Mises, R. 736
- von Neumann, J. 564
- von Rosenstiel, L. 114,  
 131f
- von Savigny, E. 256, 292
- von Savigny, K. F. 101
- von Strien, P. 287, 294

- von Winckelmann, J. 135  
 Vonderach, G. 103, 135  
 Voß, A. 103, 108, 135  
  
 Wagner, D. 555  
 Wahrendorf, J. 513, 555  
 Wainer, H. 80, 97  
 Wald, A. 609, 647, 721, 736  
 Walder, L.O. 529, 552  
 Walker, E. C.T. 197  
 Wallach, H. 54, 97  
 Wandell, B. 577, 603  
 Wandschneider, D. 355, 388  
 Wann, T.W. 321, 342  
 Wartofsky, M. W. 426  
 Watkins, J. 422, 427  
 Webb, E.J. 521, 555  
 Webb, J.C. 169, 198  
 Weber, M. 103, 106, 112, 135, 142  
 Wegenast, K. 135  
 Wegener, H. 388  
 Weidner, 525, 555  
 Weimer, W.B. 283, 294  
 Weiner, B. 255, 258, 294, 299, 342, 421, 427  
 Weingart, P. 287, 294  
 Weiskrantz, L. 62, 97  
 Wellens, T. R. 526, 552  
 Wendel, H. J. 317, 342, 395, 427  
 Wender, K.F. 676, 693  
 Wendt, D. 712, 723, 725, 728, 736, 737  
  
 Wenger, E. 229, 248  
 Wentura, D. 51f, 94  
 Wermuth, N. 668, 693  
 Wertheimer, M. 451  
 Wessels, M.G. 367, 388  
 Westermann, R. 33, 39, 81, 261, 263, 271, 294, 332f, 337, 342, 430, 442, 457ff, 480f, 507, 487, 494f, 508f, 551, 605, 609, 612ff, 618, 632, 646f, 651, 687, 693f, 723, 724, 731f, 735, 737  
 Westmeyer, H. 53, 96, 251, 261, 272f, 294, 304, 308, 326, 332, 337f, 342, 457, 477, 485, 506, 604, 610, 634, 646f, 687, 693  
 Wettersten, J.R. 411, 427  
 Wharton, J.D. 530, 551  
 Whewell 419  
 Wickens, T.D. 639, 647  
 Widaman, K.F. 676, 693  
 Wilcox, R.R. 514, 555  
 Wilkes, K.V. 195f  
 Wilks, S.S. 640f, 647  
 Will, U. 432, 441, 449  
 Williams, L.P. 418, 427  
 Williams, R. H. 678, 693  
 Willmes, K. 611, 633, 648  
 Wilson, A. 97  
 Wilson, T.D. 57, 62f, 96  
 Winer, B.J. 514, 555  
 Winkelmann, K. 332  
  
 Winograd, T. 204, 238, 248  
 Witte, E. 695, 719f, 725, 728, 730, 731, 733, 737  
 Wolf, H. 195  
 Wolff, s. 103, 114, 131, 133  
 Wolfle, L.M. 653  
 Woodward, J. 186, 198  
 Wottawa, H. 678, 693, 723, 737  
 Wright, F.T. 602  
 Wright, H.E. 53, 92  
 Wundt, W. 55f, 60f, 97, 100, 136, 143, 184, 283, 319, 342, 584, 603  
 Wutke, J. 33, 40, 41, 42, 86, 609, 615, 684  
 Wymer, W.E. 669, 691  
 Wyss, D. 365, 388  
 Wyszecki, G. 179, 198  
  
 Young 149  
  
 Zebrowitz, L.A. 535, 555  
 Zelen, M. 543, 555  
 Ziezold, H. 658  
 Zimbardo, P. G. 542f, 553f  
 Zimmermann, D. W. 678, 693  
 Zimmermann, E. 184  
 Zinnes, J.L. 676, 693  
 Zytkow, 451

# Sach-Register

- Alpha-Fehler (s. Fehler erster Art)
- Beta-Fehler (s. Fehler zweiter Art)
- Abduktion 431
- Abgrenzungskriterium 400
- Abhängigkeit zwischen Zufallsvariablen 659
- Abhängigkeit, korrelative 662f
  - , nichtdeterministische 651
  - , regressive 662ff
  - , schwache kausal regressive 674
  - , starke kausal regressive 674f
  - , stochastische 652ff
- Adäquatheit, empirische 315f, 389f, 399, 405f, 423
- Adäquatheitsbedingung 443f
- Algorithmus 214
- Allgemeine Methodenlehre 3ff
- Allgemeines Lineares Modell (ALM) (s. Modell, allgemeines lineares)
- Alpha-Lambda-System 437
- Analogie 363, 365f, 431, 450, 454
- Analogieschluß (s. Analogie)
- Analytizität der c-Funktion 435
- Anfangsbedingung 325
- Annahmekenner 252ff
- Ansätze, pragmatische 429
- Antwortformat 522ff
- Anwendung, intendierte (s. Strukturalismus)
- Anwendungsbezug 87f
- Äquilibrium, reflektives 455
- Äquivalenzbeziehung 496
- Äquivalenzprinzip 596f
- Argument, deduktives 430, 464
  - , induktives 429f, 464
- Argumentation, pragmatische 446f
- Aussage, gesetzesartige 302, 304, 695
  - , je-desto 305
  - , universelle 304
  - , synthetische 476
  - , wenn-dann 304f
- Aussagenkonzeption 296, 302, 305, 308, 332, 336
- Austauschbeziehung 265ff
- Automatentheorie 349
- Automorphismus einer Struktur 591
- Axiom 295, 299f, 302, 306, 335, 391f, 572
- Axiome, meßtheoretische 586
- Bayes-Theorem 437, 612, 708, 726f, 729
- Bedeutsamkeit 73f, 626f, 500
- Bedeutungsüberschuß 309
- Begriff, Beobachtungs- 308f, 310f
  - , theoretischer 308, 310f, 315, 318, 320, 372
- Begründungszusammenhang 49f, 66ff, 75ff
- Behauptungen, empirische 439
- Behaviorismus 235, 319, 349
- Beliebigkeit 109
- Beobachtbarkeit 311
- Beobachtung 47ff
  - , experimentelle 185
  - , Fremd- 66ff
  - , heuristische 66ff
  - , Selbst- (s. Introspektion)
  - , wissenschaftliche 50ff
- Beobachtungsaussagen 408f, 423
- Beobachtungsfehler 582, 584, 587
- Beobachtungsprädikate 586
- Beobachtungsprotokoll 51ff, 68ff
- Beobachtungswissen 432
- berechenbar 213ff
- Berechnung 151
- Berechnungsuniversalität 213f, 216
- Berechnigungsfaktoren 460
- Bereich, kritischer 703, 711
- Beschreibung 402
- Bestätigung 406, 418f, 422
  - , Paradoxien der 433, 449
  - , deduktive 443f
- Bestätigungsgrad 356, 406
- Bestätigungstendenz 414, 418
- Bestätigung, empirische 389, 418f
  - , induktive 418
  - , induktive 434, 443f, 464
  - , irrtümliche 417

- Bestätigungsgrad 356  
 Bewährung 407, 418, 421  
   -> einer Hypothese 429, 433, 441  
   -, einer Theorie 429, 441  
     vorläufige 407  
 Bewährungsgrad 442  
 Bewußtsein 59ff, 319f  
 Bisektionsverfahren 565  
 Bisymmetristrukturen 565  
 Black Box 345ff, 402  
 Bootstrap-Methode 642  
 Bradley-Terry-Luce-Skalierung 9  
 Brückenannahme 188ff
- Ceteris-distributionibus-paribus-Klausel 670  
 Ceteris-paribus-Bedingung 31, 189, 327, 408, 442, 460, 652, 669  
 Chinesisches Zimmer 238f  
 Code, perzeptueller 149, 164, 165, 179, 556  
 cognitive science (s. Kognitionswissenschaft)  
 computational theory, Ebene der 152ff, 161, 181, 185, 186  
 Computermetapher 367  
 Computermodell 200ff  
 constant-utility-Modell 575  
 cross-lagged panel design 530
- Daten, empirische 35, 47ff, 489f  
   -, sozialwissenschaftliche 104, 110, 113ff, 127  
 Datenanalyse, exploratorische 76ff  
   -, konfirmatorische 76ff  
 Datenerhebung, nicht-standardisierte 114f  
   -, qualitative 113  
   -, standardisierte 114f  
 Datenmodell 493  
 Datenrestriktion 19f, 35f, 201  
 Datenschutz 546f  
 Deduktion 696  
 Definition, operationale 309f, 634  
 Dekomponierbarkeit, additive 672f  
 Denken 160  
 Denkprozesse 151f  
 Determinationskoeffizient 663  
 Deutung 106ff, 126
- Differenzmethode 431  
 Differenzstrukturen 565  
 Dissonanztheorie 608  
 Disziplin, angewandte 8  
 Disziplin, reine 8  
 Domainproblem 255, 267, 270  
 Doppel-Blind-Versuch 515f  
 Duhem-Quine-Problem 643f  
 Duhem-Quine-These 442f
- Effekte, experimentelle 184  
 Effektgröße 500, 627f, 633, 703, 719, 732  
 Effizienz kognitiver Systeme 454  
 Eindeutigkeit 74  
 Einfachheit 15, 232, 373, 389, 391, 403f  
 Einflüsse, externale 258, 260, 279ff, 283, 286ff  
   -, internale 279ff  
   -, politisch/administrativ-organisatorische 281  
 Einheitlichkeit, semantische 389, 394f  
 Einzelbeobachtung 344  
 Einzelfallanalyse 112, 223, 229, 534ff  
 Empfehlungen, methodologische 410f  
 Einzelfallbestätigung 437f  
 Empfindung 142f, 148  
   -, Definition 146f  
 Empfindungsmessung 142f, 147  
 Empfindungsschwelle (s. Schwelle)  
 Empirie 486  
 Empirismus, logischer 487  
 Entdeckungszusammenhang 49f, 66ff, 75ff  
 Entscheidung, statistische 615, 635, 694ff  
   -, unter Risiko 433  
 Entscheidungsexperiment 231, 416  
 Entscheidungsregel 703  
 Ereignis, singuläres 325  
 Ereignisse, mögliche 655f  
 Erkenntnis, empirische 12f  
   -, rationale 12ff  
 Erkenntnisfortschritt 406, 416, 421  
 Erkenntnistheorie 344  
 Erklären (und Verstehen) 112  
 Erklärung 295, 323f, 329f, 399, 402f, 418f  
   -, Alternativ- 510ff

- , deduktiv-nomologische (DN) 324f
- , dispositionelle 328
- , genetische 328
- , Instantiierungs- 328
- , naturgesetzliche 400
- , statistische 326f
- , teleologische 328
- , unvollständige 327
- Erklärungsbegriff, pragmatischer 261
- Erklärungskraft 309, 389, 391, 396, 399, 402, 405, 419f, 422
- Erlebnisaspekt 320f
- Erlebnisdeskription 56ff
- Erwartung, bedingte 662f
- Erwartungswert 677f
- Erwünschtheit, soziale 525
- Essentielle tau-Äquivalenz, Bedeutsamkeit 680
- , Eindeutigkeit 680
- , Existenz 679f
- , Identifizierbarkeit 682f
- , Testbarkeit 681f
- Ethik psychologischer Untersuchungen 528, 541ff
- Evidenz, Prinzip der totalen 439
- Evolutionstheorie 157, 164, 171
- Exhaurieren 24
- Experiment 24f, 76, 183ff, 510ff, 608, 610, 622, 631, 697
- Experimentelle Beobachtung 185
- Effekte 184
- Experiment vs. Fallstudie (s. Fallstudie)
- Experiment mit Meßwiederholung 607
- Explanandum 324
- Explanans 324
- Explizitheit 203, 381f
- Fairneß (einer Untersuchung, Prüfung) (s. Untersuchung)
- Fall-Kontroll-Studie 53Iff
- Fallibilismus 317, 407, 422
- Fallstudie 535ff
- Falsifikation 226, 228, 399, 406f, 414, 418f, 422f, 440, 442, 497, 731
- Falsifikationsgrad 400, 408
- Falsifikationskriterium 413
- Falsifikation als Abgrenzungskriterium 400
- Falsifikation, irrtümliche 417
- Falsifikationismus 407, 411
- , dogmatischer 605
- methodologischer 605, 609
- Falsifizierbarkeit 390, 404, 409f, 695
- Fechner-Modell 576f
- Fechnersches Gesetz 142, 146
- Fehlbarkeit 394
- Fehler 570, 572, 575f, 582, 677f
- erster Art 609f, 616, 627f, 630ff, 641f, 704
- zweiter Art 609f, 616, 627f, 630ff, 641, 704
- begriff, numerisch-geometrisch inspirierter 573
- behandlung 587
- kontrolle 186, 190
- konzept, externes 584
- konzept, theorieexternes 582
- problem 570
- risiko 493f, 499, 504
- Wahrscheinlichkeit 493, 495f, 505
- Formalisierung 209f, 305f
- Forschungsprogramm 38, 25Iff, 538ff
- Forschungsstrategie 347
- Fourieranalyse 175
- Frustrations-Aggressions-Hypothese 643f
- Funktion, psychometrische 568, 576, 581, 585
- , psychophysikalische 593f
- Funktionalismus 150, 152, 170, 236ff, 320
- , probabilistischer 159
- , Turing-Maschinen- 151
- Gedankenexperiment 187, 238
- Gehalt 396f, 410
- , empirischer 396, 400f, 411f, 484
- , semantischer 487
- , theoretischer 421
- Gehaltsüberschuß 398f, 419
- Geltungsprüfung, empirische 224ff
- Generalisierung 453
- Gesetz 146, 181, 587, 303f, 308, 334, 586
- , deterministisches 305, 326
- , Koexistenz- 305
- , probabilistisches 305
- , statistisches 326

- , Sukzessions- 305
- , Webersches 574, 588, 592
- Gesetzeshypothese 304, 396f, 398
- Gesetzmäßigkeit 304, 398
- Gestaltpsychologie 157, 158, 166
- Gleichung, persönliche 584
- Grassmann-Struktur (Farbwahrnehmung) 175, 180, 588, 597
- Größenschätzung (magnitude estimation) 561, 568f, 571, 597
- Grue-Paradox 433
- Grundannahmen 295, 302
- Gültigkeit (psychologischer Schlüsse)
- Güteeigenschaften eines Tests 705
- Gütefunktion 705
- Handlungsprotokoll 105, 116
- Handlungssituation 105
- Hermeneutik 98, 100, 126ff
  - , Alltags- 123f
  - , sozialwissenschaftliche 98, 127, 129
  - , wissenschaftliche 101f
- Heuristik 16, 167, 177, 644
- hidden assumptions 9f
- Hilfsannahme 312f, 335, 392, 396, 408f, 412f, 423, 644
- Hilfshypothesen 312, 401, 410
- Hintergrundannahmen 431, 442
- Hintergrundwissen 312, 414f, 483, 485, 487, 492
- Homogenitätsannahme 354
- Hypothese 318, 475, 604, 700, 703
  - , allgemeinpsychologische 606, 625
  - , beschränkte universelle 478, 486
  - , einfache 717
  - , empirisch-inhaltliche 479, 488, 502f, 635
  - , Forschungs- 478
  - , Generierung 480
  - , Hilfs- 611
  - , inhaltliche 478, 489, 490f, 499
  - , Kausal- 484
  - , lokalisierende (bestimmte) Existenz- 477, 485
  - , Null- 703, 710f
  - , Ober- 697, 699
  - , operationale 478
  - , pseudo-singuläre 477, 485
  - , psychologische 604ff, 626ff, 634, 640, **642**
  - , Punkt- 703
  - , quasi-universelle 477f, 486, 489
  - , singuläre 477, 485
  - , statistische 477, 489, 604ff, 624, 626ff, 634, 640, 642, 699, 717
  - , statistische Aggregats- 624f
  - , statistische Ober- 615
  - , Test- 490, 494, 498
  - , theoretisch-inhaltliche 479, 486, 502
  - , unbeschränkte universelle 478, 486
  - , unbestimmte Existenz- 477, 485
  - , unvollständige 327
  - , Wahrscheinlichkeits- 486
  - , wissenschaftliche 475f, 501f
  - , zusammengesetzte 717
  - , Zusammenhangs- 484
- Hypothesengenerierung 353
- Hypothesenprüfung 604f, 632, 643f
- Hypothesenprüfung, deduktivistische 605f
- Idealisierung 15, 181, 188ff, 321f, 418, 586
- Idealtypus 111f
- Identifizierbarkeit 636f, 639, 682f
- Immunisierungsstrategien 401
- Immunität 336
- Implementation 211f, 220
- Implikationsbeziehung 491f
- Indifferenzprinzip 435
- Indikator 308, 314
- Induktion 406, 418, 422, 429f, 465, 695
  - , Begründbarkeit der 445f
  - , eliminative 431
  - , enumerative 430f, 439, 450
  - , praktische Rechtfertigung der 445f
  - , psychologische Analysen der 450ff
- Induktion, strukturalistische Analyse der 457ff
- Induktionskonzeption 450f
- Induktionsprinzip 431
- Induktionsproblem 429, 432f, 445, 464f
- Induktionsregel 446
- Induktivismus 406, 419, 421
  - , starker 441
- Inferenz 210, 212, 429
  - , ampliative 430

- Informationsbegriff 151, 154
- Informationsgehalt 225, 302, 308, 317
- Informationsverarbeitung, 28, 202ff
- Informationsverarbeitung, perceptuell-kognitive 150ff, 156f
- Inkommensurabilität 398
- Inkonsistenz, logische 392, 395
- Innovation, wissenschaftliche 267
- Instanzenbestätigung 438f
- Instrumentalismus 313f, 315f, 336, 392f, 407, 424
- Intentionalität 239, 321
- Interaktion (s. Wechselwirkung)
- Intermodalvergleich (cross modality matching) 568f, 571
- Interpretation 99f, 105, 110f, 116f, 130, 306
  - , partielle 306
  - , sequentielle 118ff
  - , epistemische 33
  - , ontische 33
- Interpretierbarkeit, operative 213
- Introspektion 56ff, 321, 352ff
- Invarianten 598
- Invarianz 140, 160, 588f
  - , Forderungen 594, 597
  - , Konzepte 592, 595
  - , Kriterien 598
  - , Prinzipien 594
  - betrachtung 177, 181, 594
  - hypothese des verbalen Lernens 606f
- Isomorphie 369
- Item-Response-Funktion 584
  
- Kalkül 306, 369
- Kasuistik 534ff
  - Computer- 535ff
- Kausalanalyse 539ff
  - aussagen 329
  - gesetz 329f
  - interpretation 424
- Kausalität 305, 329f
  - , schwache 674
  - , starke 674
- Kern, formaler (s. a. Strukturalismus) 458
- Kognitionswissenschaft 29, 139, 150, 191, 203f
- Kognitive Architektur 29, 215
- Kognitive Modellierung 203
- Kohortenstudie 531ff
- Kompensation 351
- Kompetenzmodell 241
- Komplexität von Systemen 26f
- Komplexitätsreduktion 27f
  - , theorie 26
  - , vergleich 27
- Koifidenzintervall 702, 733
- Konfundierung 670f
- Konnektionismus 172, 174, 206, 451
- Konsistenz 221, 232
  - , externe 393
  - , interne 393
  - , logische 389, 391f, 395, 406
  - , semantische 391, 394f
- Konstanz, Farb- 149, 156, 163
- Konstrukt, hypothetisches 314
  - , theoretisches 372
- Konstruktionen erster Ordnung 104, 202
  - zweiter Ordnung 36, 104, 112, 202
- Konstruktivismus, radikaler 317
- Konstruktvalidierung 313f
  - Validität 637
- Kontexteinflüsse 522ff
- Kontrolltechniken 509ff, 624
  - , statistische 539ff
- Konzept, random-utility 582
- Konzeptualisierung 207f
- konzeptuelle Replikation 611
- Korrelationskoeffizient 665
- Kriterium 484
- Kryptodeterminismus 361
- Künstliche Intelligenz 152, 161, 162, 187, 234ff, 320
  - , schwache These der 234f
  - , starke These der 234f
- Künstliche Sehsysteme 161, 188
  
- Lambda-Kontinuum 437
- Lautes Denken 56ff
- Lebenswelt 107ff
- Lebenswelt, Analyse der 108ff
  - , Deskription der 108
- Leib-Seele-Problem 138f, 141f, 144, 150, 191, 561
- Leistungsmotivation 297ff
- Lernen, Wahrscheinlichkeits- 160
- Likelihood 716, 725
  - Quotienten-Test 718



- Ratio-Kriterium 716
- , maximum 716
- Logik der Forschung 429
  - , deduktive 428
  - , induktive 424, 429, 434f, 448
- Manipulationsgrad 20f
- Markoff-Modell 639
- matching (s. Parallelisierung)
- Materialismus, eliminativer 319f
  - , funktionaler 320
- Mathematisierung 305f
- Mathesis Universalis 561
- meaningful parametrization 595
- meaningfulness 589, 594ff
  - , non- 595, 597
  - , konditionale 598
- measurement, conjoint 568
  - , functional 568
  - , random-conjoint 579
- Mechanisierung des Geistes 169
- Meehlsches Paradoxon 661
- Mehrebenenbetrachtung 217ff
- mentalistisch 236ff, 319
- Merkmalsvariabilität 449
- Messen, additiv-verbundenes 568
  - , extensives 564
  - , fundamentales 563
  - , abgeleitete 563
- Messung, additiv-verbundene 564, 571, 576, 579
  - , fundamentale 563f, 566
  - , Index- 73f
  - , per Definition 73f
  - , per fiat 73f
  - , Repräsentationstheorie der 74f
  - , Transformationsansatz der 590
- Meßfehler 143, 582
  - problem 650, 677f, 685
  - variable 678
- Meßmodell 643f
  - , stochastisches 683, 685f
- Metamerie 164, 179
- Metapher 167ff, 176f, 181, 284ff
- Metaphysik 400
- Metaprinzip 15f, 177
- Metatheorie 390
- Methoden, empirische 17ff
  - , Kontinuum der induktiven 435f
- Methodenlehre, allgemeine 4f
- Methodenlehre, spezifische 4f
- Methodenspezifität, Problem der 685f
- Methodenwahl 256ff
- Methodologie 4ff, 34f, 358, 389f
  - , deduktivistische 441, 445, 649
- Milieu (s. Lebenswelt)
- Mischverteilung 639
- Modell 8ff, 177, 306, 333, 367ff, 615
  - , algorithmisches 204
  - , allgemeines lineares 351, 627
    - essentiell tau-äquivalenter Variablen
  - , 679ff
  - , formales 205f, 233, 368
  - , ideales 322
  - , idiographisches 222
  - , individualisiertes 222, 229
  - , mathematisch-numerisches 200f, 205f, 233, 307
  - , mentales 202, 452f
  - , natürlich-sprachliches 205
  - , prototypisches 222f
  - , random-utility-- 575f, 581
  - , stochastisches 634ff, 640, 643f, 649ff, 654, 658f
    - symbolisches 205
  - anpassungstest 636, 640
  - bildung 202
  - daten 224f, 229f
  - experimente 228
  - geltungstest 636, 641, 643
- Modellierung 200ff
- Modelltheorie 205, 369
- Modellvergleichende Argumentation 231
- Moderatormodell 667f
- Modus tollens 723
- Multiple Determiniertheit, Problem der 650f, 659
- Naturgesetz 353, 414
- Naturwissenschaft 355
- Neobehaviorismus 319
- Neopositivismus 308f
- Neurale Netze (s. Konnektionismus)
- Neurophysiologie 157, 163, 165ff, 172, 175, 182
- Netz, semantisches 207ff
- Neyman-Pearson-Lemma 718
- nomologisch 295, 402

- nomothetisch 189
- non-statement-view (s. Strukturalismus)
- Nonparametrische Verfahren (s. Statistik, nichtparametrische)
- Normalbedingung 188
- Normativität 11f
  - , ideale 11
  - , praktische 11
- Nullhypothesentesten 722f
  
- Objektivität 51ff, 56
- one-shot case study 534
- Operationalisierbarkeit 483f
- Operationalisierung 308f, 309, 312, 325, **401**, 408f, 412, 417, 487f, 496, 502, 611, 634, 644
- Optik, inverse 150, 161
- Optimalität 15f
  
- Parallelisierung 533
- Parameter 698
- Passung von Personen- und Modelldaten 224f
- Performanzmodell 213
- Philosophie der normalen Sprache 447
- Placebo 515f
- Population, Problem heterogener 686
- Positivismus 354f
- Postulate 295f, 302
- power function (s. Gütefunktion)
- Prädikatenlogik 209ff, 306
- Prädiktor 484
- Praxisbezug (s. Anwendungsbezug)
- Präzision 203, 381f
- Prinzip der kritischen Prüfung 414f
- Prinzipien, methodologische 389f
- Probabilisierung 585, 587
- Probabilismus 696
- Problem P 252ff
- Problem, -Derivat 259, 263f, 268f, 270ff, 277f
  - , Indizierungs- 70f
  - , Induktions- 84ff
  - , Interpretations- 71f
  - , Klassifikations- 69f
  - , Quantifizierungs- 72ff
- Problemverschiebung, degenerative 439
- Produktionensystem 452
- Prognose, technologische 88
  
- Programm-Typen 269ff
- Programmnetz 265f, 265ff, 287
- PROLOG 211
- Prozedur, analytische 355ff
- Prozeßmodell 201, 213, 241
- Prüfbarkeit 303, 325, 401, 404, 405, 413
  - empirische 389, 401, 411
- Prüfbarkeitsgrad 400
- Prüfung 418f
  - , empirische 413f
  - , Prinzip der kritischen 414f
  - , Strenge der 413, 418
  - , wohlwollende 417
- Psychoanalyse 365
- Psychologie, Allgemeine 606
  - , Diagnostische 639
  - , Differentielle 639
  - , Entwicklungs- 639
  - , Sozial- 606
- Psychologiespezifität 5f
- Psychophysik 36, 137ff
  - , äußere 141, 144
  - , computistische 153ff
  - , funktionalistische 137, 149, 153ff, 157, 162ff
  - , innere 141, 144
  - , klassische 149
    - methodologische Probleme 138f
  - , skalenorientierte 138, 162f
  - , Traditionen 137ff, 156
  
- Qualia 138, 179, 182, 185
- Quantität 561
  
- Raben-Paradox (s. Grue-Paradox)
- Randbedingung 325, 398
- Randomisierung 85ff, 510ff, 607, 622ff, 670
- Randomisierungsmodell 514
- Randomisierungstests 514, 609ff, 632f, **642**
- Rationalismus, klassischer 16
  - , kritischer 253f, 312, 314, 407, 414, 418
- Rationalität 423, 428
- Rationalitätsbedingungen 434
- Reaktionsfunktion 350
- Reaktionszeit 178, 190
- Reaktivität 520f, 526

- Realisierung 658  
 Realismus 313f, 336, 394  
   -, experimenteller 516ff, 537  
   -, kritischer 314  
 Realität 486  
 Rechtfertigung, apriorische 432  
   -, deduktive 432  
   -, induktive 432  
 Reduktionismus 28f  
   -, Neuro- 165f, 176  
 Reduktionssatz 487  
 Reduktionssatz, bilateraler 634  
 Referenzklasse 449  
 Regeln, methodologische 389f, 393, 397, 422  
 Regelhierarchie 453  
 Regression 662f  
   -, isotone 572  
 Regressionshypothese 664  
 Reiz 162  
   -, distaler 152  
   -, ökologischer 163  
   -, proximaler 148, 152, 159, 160  
   -, Schlüssel- 163  
 Reizbestimmung 155  
 Reiz-Reaktionspsychologie 349  
 Rekonstruktion 105, 109, 111f, 117, 121ff, 202ff  
 Relation, empirische 71ff  
 Relativität 109  
 Reliabilität 628f, 678  
 Replizierbarkeit 51ff, 56  
 Repräsentation 151f  
   -, Fechnersche 574  
   -, additive 572  
 Repräsentationshypothese 213  
 Repräsentativität 83  
  
 S-O-R-Schema 349  
 S-R-Theorie 349  
 Schätzung 700, 704  
   -, „gute“ 701  
   -, asymptotisch erwartungstreue 701  
   -, effiziente 702  
   -, erwartungstreue 701  
   -, konsistente 702  
 Schluß, unbewußter 158, 160, 169  
 Schwelle 42, 144, 168  
 Semiordnung 568, 573, 581, 592  
  
 Sensitivitätsgesetz 594  
 Sequenzanalyse 117ff  
 Signalentdeckungstheorie (SDT) 58ff  
 Signifikanzniveau 706, 712  
 Signifikanztest 493f, 497f, 504, 513ff, 694, 709ff, 722ff, 725, 731  
 Simulation 200ff, 367f, 376f, 383f  
 Simultankontrolle 500  
 Sinn 101f, 104, 122ff  
 Sinnesdaten-Theorie 148  
 Sinnkonsistenz 121f  
 Sinnkriterium, empirisches 309  
 Situationale Spezifität, Problem der 685  
 Skala 589  
   -, psychophysikalische 145ff  
 Skalenkonstruktion 566f  
 Skalenniveau 558, 588, 590, 591, 593, 625f, 680  
   -, Ordinal- 634  
 Skalenniveaus, Theorie der 589  
 Skalenniveauf 680  
 Skalierung, direkte 72f  
 Skeptizismus 423  
 Sparsamkeitsprinzip 403  
 Sprache, Beobachtungs- 17f, 296  
   -, theoretische 296  
 Standardisierung 519  
 statement view (s. Aussagenkonzeption)  
 Statistik 700  
   -, deskriptive 700  
   -, nichtparametrische 632ff, 699, 715  
   -, parametrische 699  
   -, suffiziente 701  
 Statistiken, robuste 81  
 Sternchen-Daten 228  
 Stichprobe 698  
 Stichprobenraum 698  
 Stichprobenumfang 609, 611, 627f, 631ff  
 Störeinfluß (s. Störfaktor)  
 Störfaktor 30f, 82ff, 412f, 415f, 512ff, 607f, 623f  
 Störgröße (s. Störfaktor)  
 Störvariable (s. Störfaktor)  
 Strenge (einer Untersuchung, Prüfung) (s. Untersuchung)  
 Stress-Index 572, 581  
 Streß, physischer und psychischer 542ff  
 Struktur, additiv-verbundene 564f, 572f, 578, 580, 582

- , deduktive 302
- , extensive 574, 591
- Strukturalismus 296, 332f, 390, 407, 408f, 413, 424, 457f
- , metatheoretischer 261ff, 271
- Strukturgleichungsmodelle 638f
- Strukturinvarianz 591
- Strukturmodell 373, 643
- Stützungsbegriff, klassifikatorischer 716
- , komparativer 725
- Subjektivität 321
- Suffizienz 228, 231f, 701
- Surplus 374f
- Symmetrie 177, 588, 594
- Symmetrie, Konzepte 591
- System, deduktives 295f, 302f
- , informationsverarbeitendes 320
- , soziales 252, 259ff, 265f, 280f, 284
- , formales axiomatisches 7f
- , k-dimensionales 437
- , logisch-formales 215
- , symbolverarbeitendes 202, 214
- , wissensbasiertes 36f, 205ff
- T-theoretisch 334
- Tatsachen, neue 419f
- Täuschung der Versuchsteilnehmer 544ff
- Teleologie 171
- Test 700
- , bester 706
- , Entscheidbarkeits- 703
- , komplementärer Alternativ- 703
- , konsistenter 705
- , sequentieller 721
- , statistischer 615
- , unverfälschter 705, 717
- , zum Niveau Alpha 705
- Teststärke 717
- Teststärkefunktion 717
- Text 115f
- Theorem 299f, 302, 306, 335
- Theorie 37f, 183, 295ff, 368ff
- , Ästhetik 177, 595
- , Begriff 178
- , geometrische orientiert 179ff
- , mechanistisch orientiert 180f
- , Hilfs- 459
- , oberflächliche 225
- , tiefe 225
- Theorie-Empirie-Überbrückungsproblem 606, 611, 634f, 637
- Theoriebewertung 38f
- Theoriebildung 343ff
- Theorieelement 458, 461f
- Theorieentwicklung 227, 241
- Theorieimport 267ff, 270ff, 274f, 277f
- Theorienaustausch 278
- Theoriennetz 263f, 335
- Theorienschema 368
- Theorieprüfung 344f
- Thurstone-Modell 575, 577
- Thurstone-Skala 588
- Tiefe 389, 401f
- TOTE-Schema 352
- Transformation, zulässige 588, 590f, 593
- Transitivität, schwache stochastische 578
- True-Score-Variabel 678
- Turing-Kriterium (s. Turing-Test)
- Turing-Maschine 150f, 154, 168, 170, 216
- Turing-Test 235f, 376f
- Übereinstimmungsmethode 431
- Überprüfbarkeit 476, 483
- Umkehrschluß, unzulässiger 430
- Umwelt, Invarianten der 163f
- Unabhängigkeit, regressive 663f
- , stochastische 660f
- Uniformitätsprinzip 441f, 446
- vermutung 456, 460
- Undeterminiertheit 225, 373
- Untersuchung, Fairneß 87, 605, 609, 613ff, 618, 620, 622f, 627f, 631f
- , Strenge 87, 605, 609, 613ff, 618, 620, 622, 627f, 632
- Unvollständigkeit 321, 323, 401, 415f
- , formaler Systeme 216
- Urbild 367, 368ff
- Validierung 221ff, 385
- Validität 21f, 41, 87ff, 500, 501f, 605, 613ff, 618ff, 623
- , Ableitungs- 502, 503, 504
- , externe 21, 23f, 82ff, 459
- , interne 21f, 82ff, 501
- , ökologische 501, 519, 537
- , Populations- 501
- , Situations- 501, 612
- , statistische 505

- , Variablen- 502
- Validitätskriterien 498
- Variable, abhängige 21, 484, 487, 501
  - , intervenierende 315
  - , konfundierte 32
  - , stochastische (s. Zufallsvariable) unabhängige 21, 484, 487, 501
- Varianzanalyse 607, 609, 630f
- Varianzanteil 682f
- Variationsbreite 438
- Verhaltensmessung 520ff
- Verhaltensmodell 373
- Verhaltensprotokolle 346
- Verifikation 406, 418, 440f, 497
- Verlässlichkeit 438
- Vermutungen, empirische 457ff
- Verstehen 98ff, 104, 105f, 126, 128, 130, 239, 353, 359
  - , Fremd- 98ff
  - , Selbst- 98f
  - , sozialwissenschaftliches 103, 111f, 128
  - , wissenschaftliches 101ff
- Versuchsperson 189
- Versuchsplan, interindividueller 525ff
  - , intraindividueller 525ff
  - , nicht-experimenteller 527ff
- Versuchsplanung 622
- Verteilung 657f
- Verteilungsfunktion 697
- Verwendungszusammenhang 445
- Vorgehen, deduktives 480
  - , hypothetico-deduktives 483
  - , induktives 480
  - , rationales 13f
- Vorhersage 490
- Vorhersage, statistische 490f, 494f, 498f, 503f
- Vorsichtigkeitsindex 435f
- Wahrheit 101, 313f, 317, 389, 392, 406f, 418, 421f
  - , mögliche 422
- Wahrnehmung 148, 160
  - , Aufgabenanalyse der -sfunktionen 162, 164
  - , Meßinstrumentkonzeption der 148ff, 155, 161, 172
  - , Täuschung der 158
  - , Unterdeterminiertheit der 158, 161
- Vagheit der 161
- Wahrnehmungsprozesse 152
- Wahrnehmungspsychologie 148
- Wahrscheinlichkeit 33f, 657f
  - , Aposteriori- 727
  - , Apriori- 726
  - , bedingte 660, 700
  - , frequentistische 727
  - , induktive 326, 406, 421
  - , logische 434f
  - , objektive 433
    - subjektive 726f, 433
- Wahrscheinlichkeitsmaß 655f
- Wahrscheinlichkeitsmodell 698f
- Wahrscheinlichkeitsraum 655
- Webersches Gesetz 142
- Wechselwirkung 356, 358
- Welten, mögliche 331
- Wert, wahrer 677f
- Wettssystem, faires 434
- Widerlegbarkeit 484
- Widerspruchsfreiheit 203, 221, 232, 303, 382, 389, 391f, 483
- Wiederholbarkeit (s. Replizierbarkeit)
- Wissen 218f, 222
- Wissensbasis 211
- Wissenschaft, analytische 355ff
  - empirische 7ff
- Wissenschaftsmode 257ff, 284ff
- Wissenschaftsparadigma 282ff
- Wissenschaftssystem 265f, 278, 280, 283f, 287
- Wissensebene 217f
- Wissensrepräsentation 204, 207
- Young-Helmholtz-Theorie des Farbensehens 149
- Zahl 561
- Zeichen 100, 108, 115
- Zufallsexperiment 435, 653ff
- Zufallsfehler 190, 514
- Zufallsstichprobe 83ff
- Zufallsvariable 657f, 697
- Zuordnung, adäquate 503
  - , suffiziente 503
- Zuordnungsregeln 309, 311
- Zusatzannahme 300, 408, 442
- Zustandsfunktion 351