

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.

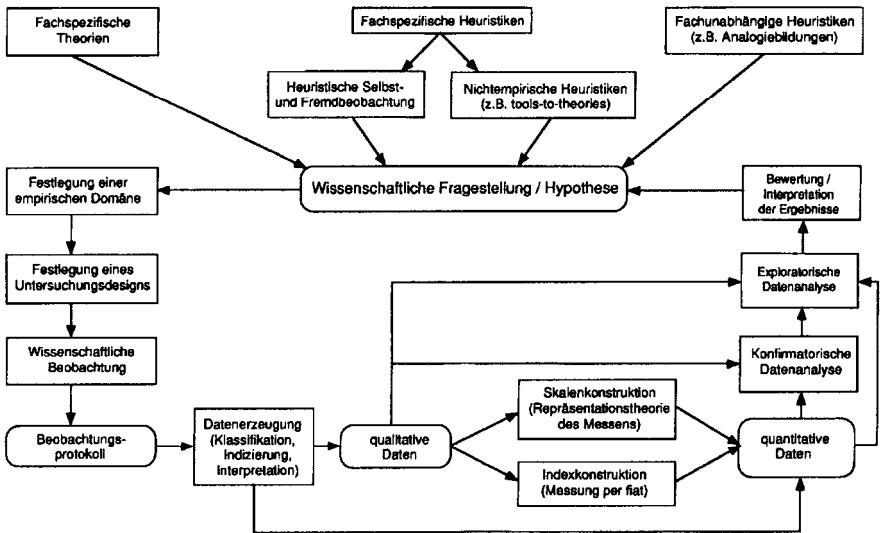


Abb. 1: Mögliche Sequenzen von Forschungsaktivitäten (in eckigen Kästen) und Forschungsprodukten (in abgerundeten Kästen) bei der Erzeugung und Verwendung empirischer Daten.

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 Vorsichtsmaßnahmen 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 Verwendungsintentionen - 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 hypothestestende 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) angestrenktes 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)¹. 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-

¹ 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 sind². 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 ist³. 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

3 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 vorgenommen⁴; 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 irrelevant⁵.

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 unmöglich⁶. 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 Theorienevaluation 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 unbekanntem 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 b 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 aus⁸. 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 ist⁹. 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

⁹ 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 werden¹⁰. 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

¹⁰ 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 kausaler¹¹ 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, beigerümpfüg 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_n und AV_n“ 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 werden¹².

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“¹³ 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.