

2. Kapitel

Beobachtung und Beschreibung von Erleben und Verhalten

Hubert Feger und Carl F. Graumann

1. Vorbemerkungen zu Thema und Terminologie

Wenn Erleben und Verhalten die zentralen Begriffe sind, die den Gegenstand der Psychologie umschreiben, und wenn vor jeder weiteren Datenanalyse Beobachtungen anzustellen und diese zu beschreiben sind, darf man dann nicht eine intensive und traditionsreiche Auseinandersetzung über die Beziehungen zwischen Beobachten, Beschreiben, Erleben und Verhalten erwarten? Bekanntlich ist der Diskussionsstand sehr unausgeglichene, und wir können dort keine Systematik berichten, wo der Forschungsstand zu viele unverbundene Probleme nur nebeneinander stellt.

Wir gehen davon aus, daß sowohl Erleben als auch Verhalten sowohl beobachtet als auch beschrieben werden können. Wenn es derjenige selbst ist, der erlebt und sich verhält, der dies beobachtet und beschreibt, so können wir von Selbstberichten und Selbstbeschreibungen sprechen (self report, self recording). Geschieht Beobachtung und Beschreibung durch einen oder mehrere andere, so können wir das als Fremdbeobachtung bezeichnen. Die erste systematische Schwierigkeit beginnt mit dem Einwand, ihr Erleben könne eine Person nur selbst beobachten. Demnach gehört zum Thema dieses Kapitels:

1. *Die* Beobachtung des eigenen *Erlebens*, meist, synonym, als Selbstbeobachtung und Erlebnisbeschreibung bezeichnet (wobei unklar bleibt, warum in der Literatur nicht auch durchgängig und synonym von Erlebnisbeobachtung und Selbstbeschreibung die Rede ist). Erkenntnis- und wissenschaftstheoretische Arbeiten sind in diesem Bereich relativ zahlreich; nur einige können wir streifen. Methodische Arbeiten sind hingegen ausgesprochen selten, sie finden sich nicht einmal gehäuft zur Blütezeit des „Introspektionismus“. Den Terminus Introspektion wollen wir nur dann für Selbstbeobachtung gebrauchen, wenn zugleich entsprechende bewußtseinstheoretische Positionen (S.U.) mitgemeint sind. Andererseits hat auch Skinners Hypothese, die Möglichkeit „private

events“ zu beschreiben, beruhe auf sozialem Lernen, ebenfalls nicht zu einer Serie gezielter Experimente geführt, die diesen Prozeß untersucht und die Verzahnung innerer Zustände und angebotenen Beschreibungsarsenal genauer analysiert hätten.

2. *Die Beobachtung des eigenen Verhaltens* (wobei strittig ist, wie berechtigt es bei Selbstberichten ist, zwischen Erleben und Verhalten scharf zu trennen).

3. *Die Beobachtung fremden Verhaltens* (gegenwärtig im therapeutischen Kontext oft als Verhaltens einschätzung akzentuiert und bezeichnet). Für die wissenschaftliche Gemeinschaft liegt stets nur das Produkt, nicht der Prozeß der Beobachtung in Form von Beschreibungen i. w. S. des Wortes vor. Empirische psychologische Arbeiten über den Zusammenhang zwischen Beschreiben und Beobachten sind selten; methodische Arbeiten über Verhaltensbeobachtung, besonders über eher technische Fragen, aber auch über die Psychologie der Verhaltensbeobachtung, sind in jüngerer Zeit häufiger. Das folgende Kapitel kann nicht mehr als eine Auswahl der in der Literatur bearbeiteten Thementele ansprechen.

2. Formen der Erlebnisbeschreibung

2.1 Selbstbeobachtung und Erlebnisbeschreibung als Methoden und Themen der Psychologie

Die Selbstbeobachtung ist dem Alltagsverstand bekannt, aber nicht immer ganz geheuer, noch ist es die Beobachtung, die einer über einen anderen anstellt.

Unter dem Titel „Von dem Beobachten seiner selbst“ schreibt ein einflußreicher und für das Schicksal der wissenschaftlichen Psychologie maßgeblicher Denker zur Zeit der vorletzten Jahrhundertwende:

„Das Bemerken (animadvertere) ist noch nicht ein *Beobachten* (observare) seiner selbst. Das letztere ist eine methodische Zusammenstellung der an uns selbst gemachten Wahrnehmungen, welche den Stoff zum *Tagebuch eines Beobachtens seiner selbst* abgibt und leichthin zu Schwärmerei und Wahnsinn hinführt.“

Der Autor und die Quelle, Immanuel Kant und seine „Anthropologie“ von 1800 (1. Aufl. 1798), schließen bei aller humanistischen Diktion und pathologischen Verweisung wohl jede Assoziation mit zeitgenössischer Wiederverwendung der Selbst-Beobachtung und des Tagebuchs als Mittel einer „humanistisch“ orientierten klinischen Psychologie zweifelsfrei aus. Wohl aber belegt das wohl unmißverständliche Zitat die Tatsache, daß der Selbstbeobach-

tung vor aller psychologisch-methodologischen Spezifizierung eine anthropologische Allgemeinheit zuzusprechen ist.

Man darf dabei aber nicht übersehen, daß, von der Beobachtung seiner selbst zu reden, nicht unbedingt und nicht einmal in erster Linie identisch ist mit dem, was anderswo und seit geraumer Zeit auch bei uns mit „*Introspektion*“ bezeichnet wird, von wo ausgehend der „Introspektionismus“ die (theoretische?) Richtung ist, von der kein Vertreter je existierte, der sich selbst so verstanden und benannt hätte, Introspektion, wörtlich „Innenschau“, meint, unter der doppelten Voraussetzung einer cartesischen Dichotomie von Innen- und Außenwelt als „kognitiver“ und „ausgedehnter“ Substanz und eines analog (oder parallel?) zum äußeren Auge gedachten inneren Auges, der Blick nach innen auf das eigene Bewußtsein, sofern - aber auch nur sofern - *Bewußtsein als rein Inneres* auffaßbar erscheint (vgl. hierzu Graumann 1966). Fassen wir hingegen Bewußtsein im phänomenologischen Sinne intentional, d.h. immer als Bewußtsein oder Erleben von etwas Inner- oder Außerweltlichem, dann ist alles „da draußen“ Wahrgenommene, Vorgestellte, Vermutete, Erinnernte etc. zwar als Erlebtes auch prinzipiell erlebnisdeskriptiv konstatierbar, aber nicht „introspezierbar“. Die vor allem in der behavioristischen Literatur der USA anzutreffende Gleichsetzung von „phänomenal“ und „introspektiv“ erscheint angesichts etwa der Wahrnehmung „blauer Berge“ und ähnlicher Erlebnisse unvertretbar.

Nun würde jemand, der die „blauen Berge“ aus der Ferne erlebt hat, nicht nur überrascht sein, wenn man ihm dieses Erlebnis als „Introspektion“ anrechnete; er würde ebenso lebhaft protestieren, wenn ihm dieser wundervolle Blick in die Weite als „Selbstbeobachtung“ attestiert würde. Vergißt man für einen Augenblick die - wie man sieht - nicht immer klärende Terminologie, dann beobachten wir uns doch am ehesten selbst, wenn wir anfangen, über uns zu stutzen und nachzudenken. Das kann durchaus - in einem ursprünglichen Sinne des Wortes Selbstbeobachtung - angesichts des allmorgendlichen Spiegelbildes passieren. Und der Verdacht, daß, relativ und zwar „negativ“ zum gewohnten Bild, heute Auge, Lippen oder das Ganze anders aussehe, kann zur Steigerung der Aufmerksamkeit beim Hinsehen führen, ohne daß nach den Spielregeln der Psychoanalyse bereits „Narzißmus“ zu diagnostizieren wäre. Auch dann würden wir ohne Bedenken von einer Selbstbeobachtung reden, wenn einer, der kritisiert wurde, weil er ständig „nich' wahr“ sagt, beim Reden darauf achtet und dann, solange er darauf achtet, es auch nicht hört, weil nicht mehr sagt. Denn Beobachten heißt ja nur so viel wie ein intensives Achten-auf, und Auf-sich-selbst-Achten tut man, wenn dazu Anlaß oder Grund bestehen und einem eine entsprechende Paß-auf-Instruktion von einem selbst oder anderen erteilt worden ist. In diesem allgemein verständlichen und klaren Sinn unterscheiden sich alltägliche und wissenschaftliche Selbstbeobachtung nicht; nicht einmal die Zweifel an der Tauglichkeit dieses Vorgehens! Denn, ob ich

das, was an meinem Verhalten zu beobachten ist, auch tatsächlich beobachte, bleibt immer offen, und ich erfahre nie verlässlich, ob es an der Beobachtungsabsicht oder am Ausbleiben des fraglichen Phänomens liegt, wenn ich das Erwartete nicht beobachte.

Jenseits dieser auch der Alltagserfahrung vertrauten Selbstbeobachtung, die, habitualisiert, wie schon Kant wußte, auch „krankhaft“ werden kann, liegt das Grübeln, die Reflexion über sich selbst. Warum habe ich nur dies getan, das gelassen? Was in aller Welt hat mich dazu gebracht, getrieben etc.? - Fragen, die man sich selbst über die Ursachen oder Gründe des eigenen (aber genauso: fremden) Tuns stellt. Diese Form der Retrospektion sollte als eine Form des Nachdenkens über sich selbst (und die Welt) nicht mit dem gleichen Begriff gefaßt werden wie die mehr oder minder unmittelbare Beobachtung eigenen Seins oder Tuns.

Wenn heute gerne unscharf vom „verbalen Report“ (statt von Selbstbeobachtung oder Erlebnisbeschreibung) die Rede ist, geht dieser Unterschied zwischen dem unmittelbaren Achten-auf und dem u.U. räsonierenden Nachdenken-über (mutmaßliche) Zusammenhänge verloren; es zählt nur das „verbale“ Resultat. Ein „verbaler Report“ als Antwort auf Warum-Fragen (vgl. Nisbett & Wilson 1977) kann sehr unterschiedliche Quellen haben; eine davon mag das eigene Erleben sein, beziehungsweise die Art und Weise, wie wir über unsere Erlebnisse, über unser Inneres, unsere Innerlichkeit, über das ganz Private etc. zu reden gelernt haben, sei es gegenüber nahestehenden Vertrauten, gegenüber Fremden oder auch gegenüber sich „persönlich“ gebenden Fremden wie Psychologen.

Damit bleibt als letztes und sicher nicht geringstes Problem die von Psychologen gerne und bewußt vollzogene Gleichsetzung von Beobachtung und Beschreibung, hier also von „Selbstbeobachtung“ und „Erlebnisbeschreibung“. Zwar lehrt uns die Anwendung der Datentheorie auf die Beobachtungsmethodik, daß eine Beobachtung, was ihre wissenschaftliche Güte betrifft, nie besser sein kann als ihre Kategorisierung. Aber dieser stolze Satz klingt weit weniger überzeugend, wenn wir kein verlässliches Verfahren aufweisen können, daß die Beziehung zwischen dem Phänomen und seiner Kategorisierung zu präzisieren gestattet. Falls es diese Beziehung überhaupt gibt zwischen „Anschauung“ und „Begriff“, wird es so viele Methodenklassen geben, wie Modalitäten dieser Beziehung denkbar sind. Die Flucht vor diesem Problem, das die Geschichte der abendländischen Philosophie mitgeprägt hat, besteht im Rückzug auf die Rede, auf den Diskurs: Bescheiden wir uns mit der Art und Weise, wie Leute über bestimmte Themen reden und analysieren diese Rede ohne Rekurs, sei es auf „Tatsachen“, sei es auf „Erlebnisse“.

Wenn dies einige unterscheidbare Modi der kognitiven Beschäftigung des Individuums mit sich selbst sind, dann sollten sie auch Themen einer Wissenschaft

darstellen, die das Erleben und Verhalten von Menschen zum Gegenstand hat. Tatsächlich reichen seit einiger Zeit, nimmt man nur Persönlichkeits- und Sozialpsychologie, umfangreiche Forschungsthemen von der Selbstaufmerksamkeit und Selbstwahrnehmung über die Selbstattribution und Selbstbeurteilung bis hin zur Selbstdarstellung im Alltagsleben. Sie alle verdanken ihre Existenz dem anthropologischen Sachverhalt der *Reflexivität des menschlichen Bewußtseins*. Das heißt, daß der einzelne nicht nur seine Umwelt und Mitwelt, sondern auch sich selbst in den verschiedenen Modalitäten der Wahrnehmung, Erinnerung, Antizipation, des Empfindens und Fühlens, des Denkens und Urteilens erfährt (erlebt) und anderen darüber direkt oder indirekt Aussagen machen kann.

Diese, sagen wir, doppelte Fähigkeit, sich selbst in verschiedenen Modalitäten zu erleben und darüber hinreichend differenzierende Aussagen zu machen, ergibt nicht nur eine ganze Klasse von problemgeladenen Forschungsthemen. Sie konstituiert auch die Klasse(n) von wissenschaftlichen Verfahren, die Aussagen des Menschen über sich selbst zum Ausgang nehmen. Diese Verfahren können sehr unterschiedlichen (wie diagnostischen, klinischen) Zwecken dienen. In diesem Abschnitt werden sie nur insofern behandelt, als sie *erlebnisdeskriptiv* sein bzw. Erlebnisdeskription ermöglichen sollen.

Dem Hinweis darauf, daß der kognitiv-sprachliche Rückzug auf sich selbst bzw. auf das eigene Erleben, der seit Beginn der wissenschaftlichen Psychologie als Forschungsmethode diente, inzwischen auch als Forschungsthema volle Aufmerksamkeit findet, kann man bereits entnehmen, daß am Kernproblem der Erlebnisdeskription, nämlich der Beziehung von Erlebnis und Deskription, noch gearbeitet wird.

2.2 Selbstbeobachtung und Experiment:

Die Begründung der wissenschaftlichen Psychologie

Blickt man auf die Anfänge der wissenschaftlichen Psychologie im neunzehnten Jahrhundert zurück, so wird erkennbar, daß die Begründung der Psychologie als Einzelwissenschaft eine methodologische war. Zumindest für Wilhelm Wundt gilt, daß er die Überzeugung, „daß die Fortschritte jeder Wissenschaft innig an die Fortschritte der Untersuchungsmethoden gebunden sind“ (Wundt 1862, xi), in einer für lange Zeit verbindlichen Weise in die Tat umgesetzt hat. Worin aber sah Wundt den für die Psychologie so dringend nötigen Fortschritt? Was immer schon, vorwissenschaftlich - und das heißt vor allem: philosophisch - Psychologie treiben ausmachte, war die Selbstwahrnehmung oder Selbstbeobachtung. Generationen von Philosophen, aber auch Seelenkundlern, hatten aus der unmittelbaren Erfahrung ihrer eigenen Empfindungen und Ideen, Affekte und Begierden und der Reflektion darüber ihre Psy-

chologie geschöpft. Nur war darüber Psychologie, zumindest im Vergleich zu der sich kräftig entwickelnden Naturwissenschaft, nicht zur Wissenschaft geworden. Gleichwohl, als Wundt dies im Sinne der Begründung einer Psychologie als Naturwissenschaft zu ändern sich anschickte, blieb auch für ihn gültig: „Alle Psychologie beginnt mit der Selbstbeobachtung, und diese bleibt zur Beurteilung der außer uns stehenden psychischen Erscheinungen immer ein unentbehrliches Hilfsmittel“ (a. a. O., xvi). Allerdings, so fügt Wundt hinzu, ist sie völlig unzureichend für die Entwicklungs- und Kausalanalyse der Erscheinungen.

Daß wir der Selbstbeobachtung geradezu auch bedürfen, um „außer uns stehende“ psychische Erscheinungen überhaupt verstehen zu können, vertrat auch - hier, wie so oft, wundtscher als Wundt - noch sechzig Jahre später E.B. Titchener (1914, 32), wenn er behauptet, daß wir „immer wieder auf die experimentelle Selbstbeobachtung zurückgreifen müssen“, wenn wir den Versuch machten, „die psychischen Prozesse eines Kindes oder eines Hundes oder eines Insekts“ zu verstehen, wie sie sich an äußeren Verhaltensmerkmalen zu erkennen geben. „Wir können uns keine Prozesse in einem anderen Bewußtsein vorstellen, die wir nicht in unserem eigenen finden.“ Fazit: „Experimental introspection . . . is the sole gateway to psychology.“

Auch William James, um einen der Wundt-Anhängerschaft unverdächtigen Mitbegründer der modernen Psychologie zu zitieren, beginnt sein Kapitel über die Forschungsmethoden der (ausdrücklich:) Naturwissenschaft Psychologie mit dem hervorgehobenen Satz: *Introspective observation is what we have to rely on first and foremost and always*“ und fügt hinzu, das Wort Introspektion bedürfe wohl keiner Definition, es meine selbstverständlich „looking into our own minds and reporting what we there discover. Every one agrees that we there discover states of consciousness“ (James 1890, I, 185).

Daß wir Bewußtsein irgendwelcher Art haben, nennt James das *inconcussum* in einer Welt, wo alles andere sich als bezweifelbar erwiesen hat. Entsprechend wird die Überzeugung, daß wir Bewußtsein haben, und auch, daß wir unser Bewußtsein von dem unterscheiden können, was Gegenstand dieses Bewußtseins werden kann, als „das grundlegendste aller Postulate der Psychologie“ bezeichnet (ebda.).

Wir nehmen zwar die Begründer der modernen Psychologie als Ausgangsbeispiele für die enge Verknüpfung von Psychologie und Selbstbeobachtung, legen aber Wert darauf, daß diese Konzeption, vor allem die von naturwissenschaftlicher Psychologie und Selbstbeobachtungsmethode, keine historische Episode geblieben ist. Bevor wir uns den Formen und Problemen der Selbstbeobachtungsmethode bzw. der Erlebnisdeskription zuwenden, die auch die Gegenwart beschäftigen, sei noch einer der einflußreichsten deutschsprachigen Psychologen der Nachkriegsjahre bemüht, der ebenfalls Psychologie als Na-

turwissenschaft vertrat, Hubert Rohrer. Für ihn läßt sich gleichwohl ein grundsätzlicher Unterschied zwischen den Forschungsmethoden der Psychologie und der übrigen Wissenschaften formulieren: „In allen anderen Wissenschaften hat man Dinge oder Vorgänge zu untersuchen, die außerhalb des eigenen Erlebens liegen, während man in der Psychologie gezwungen ist, die eigenen Erlebnisse zu beschreiben; das Objekt der Psychologie - das bewußte Erleben - findet man nirgendwo anders als in sich selbst.. *Die wichtigste Methode der Psychologie ist daher die Selbstbeobachtung*“¹⁾ (Rohrer 1963, 69 f.).

Allen, denen es um die Begründung oder Bewahrung der Psychologie als Wissenschaft, speziell als Naturwissenschaft zu tun war, ist die Problematik der Selbstbeobachtung vertraut gewesen. Sie hatte bereits manchen Denker, so vor allem Kant (1786; 1800) und Comte (1830-1842) dazu gebracht, Psychologie als Wissenschaft für unmöglich zu halten - immer unterstellt, daß sich eine den Namen Psychologie tragende Wissenschaft mit dem Psychischen (oder dem Bewußtsein oder dem Erleben) notwendig und letztlich zu beschäftigen habe.

Insofern sollte es nicht überraschen, bei den Befürwortern der Selbstbeobachtung als grundlegender Methode der Psychologie auch die kritischsten Aussagen über diese Methode zu finden. So hat es Sekundärkenner von Wundt immer wieder überrascht, neben dem im obigen Zitat zum Ausdruck kommenden Bekenntnis zur Selbstbeobachtung bei Wundt auch Sätze zu finden, wonach „Selbstbeobachtung, wenn wir das Wort Beobachtung im wissenschaftlichen Sinne verstehen, unmöglich ist. Je mehr wir uns anstrengen, uns selber zu beobachten, um so sicherer können wir sein, daß wir überhaupt gar nicht beobachten“ (Wundt, 1906, 196).

Was hier widersprüchlich erscheint, erklärt sich nicht daraus, daß sich Wundts Methodik-Konzept von seiner Heidelberger zu seiner Leipziger Zeit geändert hat. Das ist zwar auch der Fall (Graumann 1980), betrifft aber nicht die Selbstbeobachtung. Die scheinbar widersprüchlichen Aussagen sind, wie Blumenthal (1975) und Metge (1980) erneut gegen tradierte Mißverständnisse der Wundtschen Methodik demonstrieren, voll vereinbar. Man muß sich nur deutlich machen, daß Wundt von Anfang an die reine Selbstbeobachtung für methodisch wertlos, die mit dem experimentellen Verfahren gepaarte Selbstbeobachtung (besser allerdings: Selbstwahrnehmung) jedoch für unabdingbar ansah. Es war die nicht (hinreichend) kontrollierbare und damit letztlich nicht verifizierbare Selbstbeobachtung, die er anfänglich den vorwissenschaftlichen Psychologen (Wundt 1862; 1888), später den Denkpsychologen der Kälpe-Schule (s.U.) zum Vorwurf machte (Wundt 1907).

¹⁾ Bei Rohrer ist nur das letzte Wort hervorgehoben.

Um die für seine Psychologie unverzichtbare Selbstwahrnehmung in eine wissenschaftliche Methode zu transformieren, muß sie der Kontrolle des Experiments unterworfen werden, wobei unter Experiment, in der Tradition von Francis Bacon (1620; 1974), ganz zu Anfang und - wir meinen - doch in verbindlicher Weise ein dreifach kontrollierendes Verfahren verstanden wird (Wundt 1863, I, Vf.):

„Durch das Experiment erzeugen wir die Erscheinung künstlich aus den Bedingungen heraus, die wir in der Hand halten. Wir verändern diese Bedingungen und verändern dadurch in meßbarer Weise auch die Erscheinung.“

Da die Veränderung von Bedingungen die Wiederholung konstanter Bedingungen (und sei es nur für Kontrollzwecke), wie Wundt später selbst sah, zur Voraussetzung hat, läßt sich konsequenterweise von einem (durch Manipulierbarkeit, Wiederholbarkeit, Variierbarkeit und Meßbarkeit) vierfach kontrollierenden Verfahren reden. Allerdings hat Wundt, für den das Experiment eine Form der Beobachtung war, später das strengere Kriterium der Meßbarkeit durch das der (aufmerksamen) Beobachtbarkeit ersetzt (vgl. hierzu Graumann 1980, 76f.). Für die methodische Selbstbeobachtung ergab sich dann der folgende Regelkanon:

- (1) „Der Beobachter muß, wo möglich, in der Lage sein, den Eintritt des zu beobachtenden Vorgangs selbst bestimmen zu können.
- (2) Der Beobachter muß, soweit möglich, im Zustand gespannter Aufmerksamkeit die Erscheinungen auffassen und in ihrem Verlauf verfolgen.
- (3) Jede Beobachtung muß zum Zweck der Sicherung der Ergebnisse unter den gleichen Umständen mehrmals wiederholt werden können.
- (4) Die Bedingungen, unter denen die Erscheinung eintritt, müssen durch Variation der begleitenden Umstände ermittelt und . . . in den verschiedenen zusammengehörigen Versuchen planmäßig verändert werden, indem man sie teils in einzelnen Versuchen ganz ausschaltet, teils in ihrer Stärke oder Qualität abstuft“ (Wundt 1907, 308).

Erscheinen hier die formalen Kriterien des Experiments in der Funktion, wissenschaftliche von unwissenschaftlicher (Selbst-) Beobachtung zu scheiden, so gilt, was theoretisch bedeutsam ist, umgekehrt der Einsatz der Selbstbeobachtung als Legitimation des Experiments in der Psychologie. Vergegenwärtigt man sich die Fragestellungen der „Physiologischen Psychologie“, wurde tatsächlich jedes in diesem Kontext entworfene Experiment, das auf Selbstbeobachtung verzichtete, zum physiologischen Experiment (vgl. Metge, a.a.O., 186).

Es dürfte aus den vier zitierten Regeln, die im einzelnen als Kann-Bestimmungen galten, erkennbar sein, daß die Einsatzmöglichkeiten wissenschaftlicher Selbstbeobachtung begrenzt waren auf relativ einfache, d.h. „überschaubare“,

Vorgänge, die zum Zwecke der objektiven Kontrolle möglichst auf physische Reize bezogen sein sollten, auch wenn reine Erlebnisbeobachtung zulässig war, sofern einige der Regeln eingehalten wurden. Das aber heißt, daß schon die Selbstbeobachtungsmethodik der meisten Wundt-Schüler als nicht mehr regelgerecht galt.

2.3 Die systematische experimentelle Selbstbeobachtung

2.3.1 Die konkrete Vorgehensweise

Im Gegensatz zur heute ausgefeilten Methodik etwa der Testkonstruktion oder der Versuchsplanung gibt es keine „Einführung in die Methodik der Selbstbeobachtung“. Vielmehr scheint es so, als ob die sie anwendenden Psychologen und ihre Versuchspartner durchweg glauben, man beherrsche Selbstbeobachtung natürlicherweise, als ob es einer Methodenlehre der Selbstbeobachtung nicht bedürfe, allenfalls einer Grenzbestimmung ihrer Möglichkeiten. Jedoch erfordert auch eine solche abgrenzende Kritik eine systematische methodische Auseinandersetzung mit dieser empirischen Vorgehensweise, und so haben es auch die Psychologen um die Jahrhundertwende gesehen. Aus ihrem konkreten Vorgehen läßt sich ihre implizite Methodenlehre entwickeln und dann diskutieren. Wir schildern daher als erstes ihr methodisches Vorgehen, wobei wir eine der sorgfältigsten Anwendungen zugrunde legen, die durch N. Ach (1905).

Ach (S. 8) bestimmt als *Ziel* der experimentellen Selbstbeobachtung, „ein vollständiges, zuverlässiges und unbefangenes Bild der wirklich vorhandenen Bewußtseinsinhalte“ zu geben. Dieses Ziel erreicht die „systematische experimentelle Selbstbeobachtung“, allgemein beschrieben, auf folgendem Weg (Ach, 1905, S.8f.):

„Die Methode der systematischen experimentellen Selbstbeobachtung geht, wie bereits bemerkt, darauf aus, das durch äußere experimentelle Hilfsmittel veranlaßte Erlebnis der Versuchsperson jedesmal in der dem Versuche unmittelbar folgenden Zeit einer vollständigen Beschreibung und Analyse zu unterwerfen. Hierbei findet ein fortwährender enger Gedankenaustausch zwischen der beobachtenden Versuchsperson und dem protokollierenden Versuchsleiter statt. Da jede Versuchsanordnung im allgemeinen durch ein vorbereitendes Signal, welches die notwendige Einstellung der Aufmerksamkeit bezweckt, eingeleitet wird, so lassen sich beim psychologischen Einzelversuch drei Zeitabschnitte unterscheiden:

- 1) die *Vorperiode*, welche die Zeit zwischen Signal und Eintritt des Reizes umfaßt,
- 2) die *Hauptperiode*, welche das eigentliche experimentell zu untersuchende Erlebnis in sich schließt,
- 3) die *Nachperiode*, welche die sich unmittelbar an den Abschluß des Experimentes anschließende Zeit umfaßt.

Der gesamten jeweiligen Versuchsreihe hat außerdem die Angabe der Instruktion vorauszugehen. Die Instruktion der Versuchsperson hinsichtlich der Selbstbeobachtung lautet dahin, die in der Vorperiode und Hauptperiode erlebten Vorgänge in der Nachperiode eingehend zu schildern. Selbstverständlich hatte die Versuchsperson auch die Pflicht, bemerkenswerte Erlebnisse in den Zwischenpausen zwischen den einzelnen Versuchen, so eine stattfindende associative Einübung u. dergl. dem Versuchsleiter anzugeben.“

Es wird deutlich, daß Ach hier Selbstbeobachtung als Aufgabe, zeitlich unmittelbar vorausgegangenes Erleben zu beschreiben, auffaßt, so wie Stern (1911, S. 38) als Aufgabe der Selbstbeobachtung vorgibt „. . . die Feststellung eines *akuten Merkmals* (Hervorhebung von Stern), d.h. eines bestimmten, in einem gegebenen Augenblick vorhandenen Phänomens oder Aktes in der sich beobachtenden Persönlichkeit.“ Davon unterschieden wird Selbstbeurteilung, die als Aufgabe des Forschers nach Bedingungen für psychische Prozesse und Strukturen sucht.

2.3.2 Maßnahmen zur Sicherung der Ergebnisse

Der Versuchsleiter hat im Rahmen dieses allgemeinen Schemas bestimmte Maßnahmen zu ergreifen, die gewährleisten sollen, daß die Vp ihre Erlebnisse vollständig, zuverlässig und unverfälscht schildert. Dazu gehört als erstes, die Nachperiode von den beiden anderen zu trennen, und nur in der Nachperiode das Erlebte zu schildern, und diese Sequenz öfters zu wiederholen, damit die *Beobachtung des Erlebens oder schon die Absicht zu beobachten den Erlebnisablauf oder seinen Inhalt* nicht stören kann. Da dieses Argument bis heute zu den zentralen in der Auseinandersetzung um die Selbstbeobachtung gehört, geben wir es vollständig wieder (Ach 1905, S.9):

„Es fällt hiermit jener Einwand weg, der schon von Kant (Ach verweist auf: *Metaphysische Anfangsgründe der Naturwissenschaft*. Vorrede S. XI, 1786) angedeutet und seitdem häufig wiederholt wurde, daß eine direkte Beobachtung der psychologischen Phänomene während ihres Erlebens oder die Absicht, während des Vorganges zu beobachten, den zu untersuchenden Prozeß unmöglich macht. Denn hier findet während des Erlebnisses für gewöhnlich keine Beobachtung statt, ebensowenig besteht die Absicht, während des Erlebens das zu untersuchende Geschehen zu beobachten. Daß die Selbstbeobachtung auf das Erlebnis, solange dasselbe sich nicht öfters wiederholt hat, einen störenden Einfluß ausübt, davon konnte ich mich bei meinen Untersuchungen vielfach überzeugen.“

Um eine *verfälschende Auswahl* durch die Vp aus ihrem Erleben zu vermeiden und um die Vp in *Unwissenheit* darüber zu lassen, woran genau der VI interessiert ist, fordert Ach, jedesmal das gesamte Erlebnis so vollständig wie nur möglich zu beobachten und zu berichten, insbesondere nicht nur das gerade wichtig erscheinende oder lebhaft hervortretende (S. 14). Da die gewünschte Vollständigkeit „bei den überaus reichhaltigen psychischen Erlebnissen“ sich

von keiner Vp erreichen lasse, seien Wiederholungen nötig, bei denen man nach und nach der Vollständigkeit näher komme (S. 16f.).

Drittens fordert Ach eine Kontrolle darüber, ob die von der Vp gewählte „sprachliche Bezeichnung wirklich den adäquaten Ausdruck des zugehörigen geistigen Inhaltes darstellt“. Ach sieht das Problem in erster Linie darin, daß der *Sprachgebrauch* von Vp und VI nicht übereinstimmen, weniger darin, daß die Sprache grundsätzliche Mängel bei der Abbildung von Erlebten aufweisen könnte (s. Linschoten 1959, 1961). Die fehlende Übereinstimmung läßt sich durch *Nachfragen* klären und vielleicht beseitigen: „Der Versuchsleiter hat deshalb die Pflicht, die gegebene Schilderung durch Fragestellungen zu ergänzen.“ (S. 14). Als Beispiele finden wir:

„Die Fragestellungen bezogen sich auf die zeitliche Aufeinanderfolge, so daß z.B. nach der Schilderung der Versuchsperson gefragt wurde: was ging diesem Zustande vorher? was war zwischen diesen beiden Vorgängen? schlossen sie sich unmittelbar aneinander? standen sie in irgendeiner bewußten Beziehung? Auch der simultane Inhalt wurde in ähnlicher Weise besprochen, z.B. waren die Vorgänge gleichzeitig im Bewußtsein? welchem war die Aufmerksamkeit zugewendet? wie war der Vorgang im Bewußtsein? was für Merkmale hat dieser Vorgang? waren Gefühle dabei u.s.w.? Ist der Vorgang gleich einem vorhergehenden Vorgang? .“ (S. 17).

Viertens, gewissermaßen als zusätzliche Sicherung gegen „Täuschung“, verwendet Ach fast nur solche Beobachtungen, „. . . welche bei verschiedenen Versuchspersonen übereinstimmend gefunden wurden“ (S.20), und die *interindividuelle Replikation* geschieht nicht über beliebige Vpn, sondern möglichst an *Psychologen*, wegen der „Schwierigkeit der Durchführung“, und auch bei den Psychologen hat „Schulung und stetige Kontrolle“ stattzufinden (S.23). Das Thema „Übung in der Selbstbeobachtung“ ist bis heute nicht systematisch empirisch untersucht worden.

Unabdingbar ist fünftens, daß es sich um Selbstbeobachtung unter *experimentell variierten Bedingungen* handelt, und:

„Die systematische experimentelle Selbstbeobachtung hat jedoch keinen Wert, wenn es nicht gelingt, durch Änderung der äußeren Versuchsanordnung und der Instruktion auch eine dem jeweiligen Zwecke entsprechende Änderung des inneren Erlebnisses herbeizuführen, und so durch Variierung der äußeren Umstände auch eine Kontrolle der in der Selbstbeobachtung gemachten Angaben durchzuführen.“

Der Forscher muß also Annahmen darüber formulieren, wie äußere Bedingungen und Erleben zusammenhängen, und ihre Bestätigung stärkt das Vertrauen in die verwendete Methode. Der konsequent weitergehende Schluß, ausbleibende Bestätigungen könnten nicht nur zu einer Revision der Theorie, sondern auch der Methodik führen, wurde im Detail in der vorbehavioristischen Zeit nicht gründlich genug, und mit dem Behaviorismus allzu vehement gezogen.

Es gibt, bei Ach selbst wie bei seinen Zeitgenossen, zahlreiche Varianten der Methode, z.B. mit und ohne Fragen des VI, mit und ohne Vorphase, häufig ohne experimentelle Variation der situativen Bedingungen usw.; diese Varianten müssen jedoch nicht diskutiert werden, wenn die grundsätzliche methodische Problematik aufgezeigt werden soll.

2.3.3 Begründung der Möglichkeit von Selbstbeobachtung

Wie jede andere Methode der Datenerhebung bedarf auch die der Selbstbeobachtung des Nachweises, daß die Beobachtungen, die man mit ihrer Hilfe gewinnen will, prinzipiell auch gewonnen werden können. Das Kriterium für einen solchen Nachweis besteht darin zu zeigen, daß (empirisch gesicherte) theoretische Annahmen über den zu erfassenden Gegenstandsbereich, hier psychische Strukturen und Prozesse des Erlebens, und implizite Voraussetzungen der Methode über die Natur dieses Gegenstandsbereiches miteinander vereinbar sind.

Zu den notwendigen psychologischen Annahmen gehört: Die experimentell erzeugte Situation führt an der gleichen Person - von gesetzmäßigen oder kontrollierbaren Reihenfolgewirkungen einmal abgesehen - zu einem Erlebnis, das auch über verschiedene Replikationen soweit identisch ist, daß sich seine Behandlung als „gleiches Phänomen“ theoretisch rechtfertigen läßt. Für die Untersuchung von Erlebnissen - etwa im Gegensatz zu offen beobachtbaren Verhaltensweisen - ist diese Forderung deshalb kritisch, weil über sein Erleben nur der eine Beobachter Auskunft geben kann. Ferner muß man annehmen (Ach, S. 10), Selbstbeobachtung beziehe sich auf Bewußtseinsinhalte, welche die Tendenz haben, „im Bewußtsein weiter zu verharren“. Man muß also ein Minimum an Gedächtnisleistung voraussetzen, und weiter annehmen, das Erinnernte sei mit dem Erlebten identisch oder in bekannter Weise verschieden, wie Ach (S. 15) formuliert, man müsse „die Identität des perseverierenden Erlebnisses mit dem wirklich vorhandenen“ voraussetzen.

Schließlich ist als notwendige Voraussetzung zu erwähnen, daß die psychischen Gegebenheiten überhaupt bewußt und „bewußtseinsfähig“ sind. Die frühen Anwender der Introspektion gehen keinesfalls davon aus, alles Psychische sei bewußt oder könne bewußt gemacht werden. Dieser Ausgangspunkt hat jedoch nicht zu einer Theorie oder Kontroverse geführt, was unter welchen Umständen prinzipiell erlebt werden kann; die Grenzziehung wurde der Vp, dem VI und ihrer gemeinsamen alltäglichen und wissenschaftlichen Sprache überlassen.

2.3.4 Anmerkungen zu typischen Ergebnissen

Ergebnisse vorbehavioristischer Studien mit einer Methode der Selbstbeobachtung liegen oft auf drei verschiedenen Abstraktionsstufen vor: 1) als umfangreiche wörtliche Mitschrift, in der Regel wegen des Umfangs nicht oder nicht vollständig veröffentlicht, auch wegen ihrer Unvollständigkeit wohl nicht gerne aus der Hand gegeben, wobei sich Ach einiges von der Einführung des Phonographen verspricht 2) als kondensiertes Protokoll, deren Wiedergabe oft die Hälfte einer Publikation ausmacht und den Leser im interpretativen Nachvollzug zu 3) den theoretischen Schlußfolgerungen führen soll. Der Übergang von der einen zur anderen Ebene ist noch nicht systematisch kontrolliert; Prinzipien und Methoden etwa der Inhaltsanalyse waren damals noch nicht bekannt.

„Was freilich in der Frühzeit des experimentellen Arbeitens mit Selbstbeobachtungen ebenso wie bei manchen philosophischen Autoren vergessen wurde, ist der mittelbare Aussagewert von sprachlichen Äußerungen über innerseelische Vorgänge. Jede Aussage über ein Phänomen stellt ja nicht ohne weiteres die Abbildung einer Tatsache dar. Sie ist zunächst einmal eine verbale Reaktion auf eine Situation, die interpretiert werden muß wie irgendeine andere Reaktion.“ (Thomae 1960, S.32).

Thomae hat konsequent für Selbstbeobachtungen bei intraindividuellen Konflikten Verfahren der systematischen Inhaltsanalyse auf verbale Schilderungen von Erleben und Verhalten angewendet; Feger und Feger (1969 a, b) haben die Anwendung der Inhaltsanalyse auf erlebnisdeskriptives Material der Entscheidungsforschung weiterentwickelt. Um die nachfolgende Kritik verständlich zu machen, sei hier ein Teil eines typischen kondensierten Protokolls (Ach, S. 38 f.) wiedergegeben:

„Bei Jetzt wurde der Finger auf den Taster niedergedrückt mit dem Wissen, daß er niedergedrückt werden soll. Dann wurde die Blechplatte (Verschlußplatte des Kartenwechslers) fixiert und innerlich gesprochen „wird gleich kommen“ oder „jetzt kommts“, „jetzt kommts“ mit der Bedeutung, daß dort, wo fixiert wird, etwas (i.e. weiße Karte) eintreten wird. Dabei bestanden Spannungsempfindungen als sinnliche Begleiterscheinungen der Aufmerksamkeitskonzentration in den Augen, Stirn- und Schläfen, zuweilen auch in den Gesichtsmuskeln und in den Schultern, sowie ein Anhalten des Atems. Spannungsempfindungen in der Hand oder im Finger waren nur ausnahmsweise vorhanden. Trotzdem war in dem gesamten Spannungszustand das Wissen enthalten, daß sofort reagiert werden soll, ohne daß dies innerlich gesprochen wurde, oder sonst phänomenologisch repräsentiert war. Außerdem bestand die Bewußtheit, daß in kurzer Zeit das Erwartete eintreten, d.h. die Karte kommen wird, also neben der sonstigen Bestimmtheit des Erwartens auch eine zeitliche Komponente.

Die Erwartung selbst konzentrierte sich auf die kommende Karte, so daß diese im Mittelpunkt des gesamten Erlebnisses stand. Aber nur ausnahmsweise war dieselbe visuell gegeben (1. Versuch des 5. Tages), und auch hier war es nur „wie die Andeutung eines visuellen Streifens an der oberen Kante der Verschlußplatte, dessen Helligkeits-

qualität nicht als weiß zu bezeichnen war.“ Sonst war die weiße Karte nur als Bewußtheit im Erwartungsinhalt gegenwärtig d.h. Versuchsperson wußte, daß dort, wo sie fixierte, die weiße Karte erscheinen wird, ohne daß dieser auf die weiße Karte sich beziehende Vorstellungsinhalt anschaulich repräsentiert war.“

In diesem Protokoll erscheinen schon wesentliche theoretische Termini wie „Spannungsempfindung“, „Aufmerksamkeitskonzentration“, „Bewußtheit“, „Erwartungsinhalt“ und „Vorstellungsinhalt“. Es ist nicht durchgängig klar, für welche verbale Äußerungen der Vp sie jeweils stehen, und es ist nicht nachprüfbar, wie stark in der sich bisweilen über Wochen erstreckenden Interaktion zwischen VI und Vp die Anregungen des VI waren, entsprechende sprachliche Beschreibungen zu erzeugen. Pointiert formuliert wird die Vp zur Projektionsfläche für die theoretischen Vorstellungen des VI, und zwar so wenig kontrolliert, wie wir es uns heute kaum noch vorstellen können und gestatten würden.

2.4 Die behavioristische Kritik der „Introspektion“

Der Behaviorismus, den Watson (1913; 1968) von Anfang an selbst so taufte, verstand sich, ebenso von Anfang an, als Gegenwendung gegen eine mit Introspektion arbeitende Bewußtseinspsychologie. Das bekannte behavioristische Manifest von 1913 beginnt mit den Sätzen:

„Psychologie, wie sie der Behaviorist sieht, ist ein vollkommen objektiver, experimenteller Zweig der Naturwissenschaft. Ihr theoretisches Ziel ist die Vorhersage und Kontrolle von Verhalten. Introspektion spielt keine wesentliche Rolle in ihren Methoden . . .“ (1968, 13).

Watsons Vorbehalte waren insofern doppelter Art, als sie sich ineins gegen das Bewußtsein als Forschungsgegenstand der Psychologie und gegen die Introspektion als Forschungsmethode wandten, wobei die beiden Seiten dieser Kritik wie - in Watsons Sicht - auch ihre beiden Objekte einander bedingten. Es wäre heute nur mehr von historischem Interesse, Watsons oft durch Polemik vergrößerte Kritik der Introspektion in einen methodologischen Beitrag aufzunehmen. Die behavioristische Kritik als solche findet sich ohnehin differenzierter bei dem Protagonisten eines Radikalen Behaviorismus, B. F. Skinner (1953; 1963; 1974). Seine Kritik richtet sich nur gegen die Methode der Introspektion, nicht gegen das Bewußtsein. Als „private world within the skin“ gilt das Innere als ein kleiner Teil des gleichen Universums, das wir außen besser beobachten können. Aber für den inneren Teil gilt:

„Wir fühlen ihn und beobachten ihn auch in einem gewissen Sinn, und es wäre unsinnig, diese Informationsquelle bloß deswegen zu vernachlässigen, weil nie mehr als eine Person Kontakt mit einer inneren Welt aufnehmen kann. Nur

bedarf unser Verhalten bei dieser Kontaktnahme der Überprüfung“ (Skinner 1974, 24).

Berücksichtigt man den sozialen Ursprung unserer Selbstkenntnis und die Tatsache, daß etwas Fühlen und darüber Berichten zweierlei ist, dann ist auch die behutsame Verwendung von Berichten über „private Ereignisse innerhalb der Haut“ zu vertreten. über den „praktischen Nutzen von Berichten über die innere Welt, die gefühlt und introspektiv beobachtet wird“, heißt es, daß sie Anhaltspunkte ergeben (1) für vergangenes Verhalten und die Bedingungen, unter denen es stand, (2) für laufendes Verhalten und dessen Bedingungen und (3) für Bedingungen, die sich auf künftiges Verhalten beziehen“ (Skinner, a.a.O., 35).

Die Nutzenanwendung dieser Empfehlung findet sich in den entsprechenden mit Selbst-Beobachtung und Selbstkontrolle operierenden Techniken der (klinischen) Verhaltensmodifikation (vgl. hierzu Kanfer 1975; Braun 1978).

Für die Grundlagenforschung hingegen dominieren in Skinners eigenen Arbeiten die Bedenken gegenüber dem Einsatz introspektiver Methodik. Sie lassen sich dahingehend zusammenfassen, daß (1) introspektive Berichte nie genau genug sein können, weil die Entsprechung zwischen den privat bleibenden inneren Ereignissen und den sie (wenn überhaupt) begleitenden Umweltkontingenzen und dem sie kommentierenden (erklärenden oder beschreibenden?) verbalen Verhalten nie ganz verlässlich überprüft werden kann und zwar prinzipiell nicht; (2) die Berücksichtigung mentaler Ereignisse für eine *funktionale* Analyse des Verhaltens entbehrlich sei. In der funktionalistisch verstandenen Kausalkette ist es wichtig, das erste Glied (genetische und/oder Umweltbedingungen) zu kennen bzw. zu beherrschen, um - unter Außerachtlassen des mittleren mentalen Gliedes - das dritte Glied des manifesten Verhaltens vorhersagen bzw. modifizieren zu können. Die Kenntnis des zweiten Gliedes vermag ein gewisses Licht auf die Gesamtbeziehung zu werfen, sie aber nicht zu ändern (Skinner 1953, 35). Außerdem - und man mag dies als einen eigenen Einwand betrachten - fördert (3) die Berücksichtigung mentaler Ereignisse die Neigung und die Gefahr, allzusehnell Verhaltensweisen und -änderungen auf hypothetische innere Variablen zurückzuführen, statt weiter nach äußeren Ursachen zu forschen.

Wenn man ebenso abwägend wie Skinner die Vor- und Nachteile der Selbstbeobachtung darstellt, die Skinnersche und mit ihr überhaupt die behavioristische Kritik an introspektiven Verfahren beurteilt, kann man, wie es D.A. Lieberman (1979) getan hat, durchaus für eine (limitierte) Wiederzulassung der Selbstbeobachtung plädieren. Zur spät- bzw. nachbehavioristischen Rehabilitation der introspektiven Methodik vgl. auch Bakan (1959); Natsoulas (1970; 1978); Dilkington & Glasgow (1967); Radford (1974) und White (1980); spe-

ziell zur historischen Rehabilitation ihrer Darstellung die - vor allem Boring (1953) korrigierende - Arbeit von Kurt Danziger (1980).

2.5 Die Technik des lauten Denkens

Was wir bei Kindern ohne Schwierigkeit beobachten und wobei wir auch uns selbst, wenn wir ganz allein sind, noch gelegentlich ertappen, ist ein mehr oder weniger fragmentarisches lautes Denken. Beim Lösen eines Problems hören wir den Betreffenden einzelne Lösungsschritte ankündigen, andere bereits getane rekapitulieren, gegenwärtige kommentieren usw.

Es war der Schweizer Denkpsychologe Edouard Claparède, der, unzufrieden mit der systematischen experimentellen Selbstbeobachtung der bisherigen Denkpsychologie, die von ihm „gesprochenes Denken“ (reflexion parlée) genannte Methode 1917 einführte (Claparède 1965). Karl Duncker übernahm sie dann 1926 (Duncker 1926; 1966).

„(Die Methode) besteht darin, jemanden die Lösung eines mehr oder minder schwierigen Problems aufzutragen... und diese Versuchsperson zu bitten, laut zu denken. Es handelt sich hier nicht um Introspektion, denn diese besteht aus der Analyse von Bewußtseinsprozessen, aus der Beschreibung ihrer Eigenart und ihrer Struktur. Hier handelt es sich einfach darum, die Denkschritte zu erzählen; es gilt zu beobachten, was das Denken tut, nicht, was es *ist*. Wenn Sie wollen, ist es eine behavioristische Methode, die den Ablauf des inneren Verhaltens zu bestimmen versucht. Vor der Introspektion hat sie den Vorteil, daß sie keine Spaltung der Versuchsperson verlangt: diese muß nicht zugleich denken und sich denken sehen“ (Claparède, a.a.O., 110).

Ähnlich argumentiert und operiert Duncker (1926, 664; 1966, 2), der die Methode unter Bezeichnungen des „thinking aloud“ und des „lauten Denkens“ bekannt gemacht hat. Sicher gibt es, wie eigene Empirie belegt, immer wieder mal Versuchspersonen, denen das Verbalisieren schwerfällt, die der Zwang zum Sprechen beim Denken irritiert. Und ebenso sicher gibt es keine Möglichkeit zu überprüfen, wieviele Gedanken ungeäußert bleiben. Trotzdem hat diese Methode, vor allem, wenn sie beim Problemlösen mit Verhaltensbeobachtung gepaart werden kann, in der modernen Denkforschung ihren festen Platz gefunden; man vergleiche etwa die Untersuchungen von Lüer (1973) und Dörner (1974) und die Behandlung der Methode als „weitverbreitete“ Datenquelle bei Ericsson & Simon (1980).

2.6 Phänomendeskription

Schon in den späteren Auflagen der Arbeiten von Wilhelm Wundt setzte sich allmählich der Begriff des Erlebens durch, zuerst noch gebunden als „Bewußt-

seinerlebnis“. Dieser neue bald und bis heute eingebürgerte Themenbegriff der Psychologie stand schließlich für alle Modalitäten unmittelbarer Erfahrung, ob Wahrnehmen, Vorstellen, Denken, Urteilen, Fühlen oder Wollen. Parallel zu dieser Entwicklung, sie wohl mitbedingend, hatte sich in der Phänomenologie ab der Jahrhundertwende eine neue Bewußtseinskonzeption durchgesetzt, deren wesentliches Bestimmungsstück Intentionalität war (s.O.; vgl. hierzu auch Gurwitsch 1966). An die Stelle eines „Behältnismodells“ des Bewußtseins, in dessen „Innerem“ man Inhalte und Akte (Prozesse) ansetzte, die man durch „Introspektion“ im engeren Sinne zu beobachten trachtete, trat das Modell eines „intentional“ auf die „Dinge selbst“ gerichteten Bewußtseins (bzw. Erlebens) (vgl. hierzu Graumann 1966; Nuttin 1955).

Das, was sich diesem Bewußtsein zeigte („Phänomen“), konnte der Deskription und weiteren Analysen unterzogen werden, so *wie es sich zeigte und rein in den Grenzen, in denen es sich zeigte*. Das methodische Postulat, das die Phänomendeskription betreibende Psychologie sich - unter dem Eindruck der Husserlschen Phänomenologie - verordnete, lautet in der erkenntniskritisch vielleicht anspruchsvollsten Fassung bzw. Forderung:

„Das Vorgefundene zunächst einfach hinzunehmen, wie es ist; auch wenn es ungewohnt, unerwartet, unlogisch, widersinnig erscheint und unbezweifelten Annahmen oder vertrauten Gedankengängen widerspricht. Die Dinge selbst sprechen zu lassen, ohne Seitenblicke auf Bekanntes, früher Gelerntes, ‚Selbstverständliches‘, auf inhaltliches Wissen, Forderungen der Logik, Voreingenommenheiten des Sprachgebrauchs und Lücken des Wortschatzes. Der Sache mit Ehrfurcht und Liebe gegenüberzutreten, Zweifel und Mißtrauen aber gegebenenfalls zunächst vor allem gegen die Voraussetzungen und Begriffe zu richten, mit denen man das Gegebene bis dahin zu fassen suchte“ (Metzger 1954, 12).

Nur um einem gegenüber einer deskriptiv vorgehenden oder gar phänomenologisch orientierten Psychologie häufig aktualisierten Vorurteil entgegenzuwirken, sei der obigen Forderung noch mit Wolfgang Metzger die Anmerkung beigelegt, daß dieses Deskriptionsgebot keinen Verzicht auf weiterführende Hypothesenbildung und deren experimentelle Prüfung bedeutet, wohl aber, „daß es sinnlos ist und zu Fehlansätzen führen muß, wenn man zu Annahmen und Untersuchungen über Ursachen und über Wirkungen des zunächst Gegebenen übergeht, ohne dieses überhaupt recht zu kennen“ (Metzger a.a.O., 13).

Modellbeispiele für eine derartige Phänomendeskription, die hier aus Raumgründen schlecht wiedergegeben werden können (sie sind, weil differenziert, entsprechend umfangreich), finden sich in der phänomenologischen Literatur etwa bei W. Schapp (1976), in der psychologischen Literatur etwa bei David Katz (1911; 1929); im übrigen sind viele der heute klassisch genannten Arbei-

ten der Gestaltpsychologen seit Wertheimer (1912) und Köhler (1921) mit Hilfe solcher Phänomendeskription zu ihren Ergebnissen gekommen. Demgegenüber ist die neuere Psychologie arm an derartigen Dokumenten reiner Erlebnisdeskription (vgl. aber Gibson 1950; deutsch 1973).

2.7 Behavioristische Selbstwahrnehmung

Eine der neueren Darstellungen der behavioristischen Auffassungen gibt Bern (1972, S. 2), der die folgende Zusammenfassung als die zentralen Annahmen seiner Theorie der Selbstwahrnehmung bezeichnet:

„Individuals come to „know“ their own attitudes, emotions, and other internal states partially by inferring them from observations of their own overt behavior and/or the circumstances in which this behavior occurs. Thus, to the extent that internal cues are weak, ambiguous, or uninterpretable, the individual is functionally in the same position as an outside observer, an observer who must necessarily rely upon those same external cues to infer the individual's inner states.“

Diese Position geht im wesentlichen zurück auf die radikal-behavioristische Analyse sog. „private events“ durch Skinner (1945, 1953, 1957). Bern behandelt das Thema weiter unter der aufschlußreichen Kapitelüberschrift „The ontogeny of self-attributions“. Um innere Zustände richtig beschreiben zu können, muß ein Kind dies lernen, und zwar von jemandem, der es Namen für diese Zustände lehrt und ihm beim Unterscheiden ähnlicher Zustände hilft. Dabei ergibt sich das Problem, daß dieser Lehrer, der „außenstehende“ Beobachter, den Zeitpunkt und die Gelegenheit feststellen muß, wo beim Kind der „kritische innere Reiz“ auftritt, z.B. Schmerz, wenn es sich den Kopf stößt. Die Beschreibung „das tut weh“ mag dann in das Reaktionsrepertoire des Kindes aufgenommen werden, wenn beispielsweise die Mutter sagt: „Weine nicht, ich weiß ja, daß es dir weh tut.“ Das Kind kann diese Beschreibungsreaktion generalisieren auf die Wirkung einer größeren Zahl schmerzzeugender Reize. Am Anfang stand - nach Skinner wie Bern - ein beobachtbarer äußerer Reiz (Kopf stoßen) und vielleicht eine beobachtbare Reaktion (Weinen). Es ist dann rigoros und konsequent anzunehmen, „. . . that we have virtually no knowledge at all until we have been explicitly trained. Internal identifications that we have not been taught remain internal identifications that we cannot make.“

Die Überprüfung der Theorie, insbesondere die Ableitung spezifischer, empirisch prüfbarer Hypothesen, ist noch nicht weit vorangekommen. Berns (1965; 1966) hierzu angestellte Experimente sind schwierig zu bewerten und setzen sich im wesentlichen mit der Dissonanztheorie auseinander. Wir wenden uns daher einigen Unterschieden zu, die Bern (1972) zwischen Selbstwahrnehmung und interpersonaler Wahrnehmung herausstellt. Die erste Unterscheidung

wird als die zwischen *insider versus Outsider* getroffen. Dem Selbstbeobachter des eigenen Erlebens stünden Reize zur Verfügung, die dem Fremdbeobachter nicht verfügbar seien, und auch zwischen diesen Reizen könne diskriminiert werden, wenn auch deshalb relativ schlecht, weil die Gemeinschaft derer, die die gleiche Sprache sprechen, nur begrenzt imstande ist, solche Differenzierungen einzuüben. Die zweite Unterscheidung ist die zwischen *intimate versus stranger*. Für den Selbstbeobachter steht die Kenntnis der eigenen Vergangenheit zur Verfügung, die für den Außenstehenden nicht als Bezugssystem dienen kann. Ihm fehlen damit Anhaltspunkte, um die inneren und äußeren Reize und Reaktionen zu bewerten. Weiter wird abgehoben auf die besonderen und möglicherweise verschiedenen Interessenlagen von *self versus other*, die zu Verzerrungstendenzen führen könnten. Schließlich wird die Perspektive des Selbstbeobachtenden als Handelnden gegen die nur observierende des Fremdbeobachters gestellt - actor versus *observer*, und aus den unterschiedlichen Perspektiven können unterschiedliche Aspekte der Situation relevant werden.

2.8 Neuere Untersuchungen über bildhafte Vorstellungen

Das in den letzten Jahrzehnten verstärkte Interesse an kognitiver Psychologie und an Phänomenen wie den bildhaften Vorstellungen (mental images oder imagery), die in der Tradition der älteren Psychologie als Bewußtseinsstatsachen zu bezeichnen wären, könnte vermuten lassen, man hätte Verfahren zur Erfassung von Erleben in diesem Zusammenhang diskutiert, oder gar gezielt untersucht und verbessert. Uns interessiert hier nicht die Rolle, die bildhafte Vorstellungen in verschiedenen kognitiven Theorien spielen, sondern die Methoden zu ihrer Analyse. Richardson (1980) berichtet in seiner Monographie drei Vorgehensweisen, die fast die gesamte Forschung beschreiben:

- (1) Vpn werden instruiert, sich etwas bildhaft vorzustellen, wenn sie bestimmte Aufgaben - meist des Lernens und Erinnerns - im Experiment ausführen.
- (2) Den Vpn wird unterschiedliches Reizmaterial vorgelegt und man erhebt, in welchem Ausmaß das Material bildhafte Vorstellungen erzeugt, und ob dieses berichtete Ausmaß mit Leistungsvariablen zusammenhängt. Meist wird geprüft, ob Material, das mehr bildhafte Vorstellungen hervorruft, besser behalten wird.
- (3) Man vergleicht Vpn, die sich in der Lebhaftigkeit der von ihnen als erlebt berichteten bildhaften Vorstellungen unterscheiden, nach ihren Leistungen in Tests, welche die Fähigkeit bei räumlichen Manipulationen erfassen, z.B. bei der vorgestellten Rotation vorgegebener oder vorgestellter Figuren. Reaktionszeit ist eine hier typische abhängige Variable.

Um die in diesem Forschungsbereich verwendeten Methoden zu diskutieren, schildern wir das Vorgehen anhand der Arbeiten von Shepard (Übersichtsarti-

kel 1978). Shepard untersucht bildhafte Vorstellungen von realen Objekten, und zwar vorgestellten gegenüber als Vorlage im Versuch vorgegebenen, die nicht ambivalent sind, also z.B. nicht mehrdeutige optische Figuren oder uneindeutige, TAT-ähnliche Reizvorlagen. Auf diese Weise versucht er, die Probleme zu umgehen, die sich daraus ergeben, daß subjektive Deutungen - man sieht etwas als etwas - bei einigen Reizvorlagen die entscheidende Rolle für interindividuelle Varianz spielen. Die Grundannahme dieses Versuches (Shepard & Chipman, 1970) wie zahlreicher anderer besteht darin, eine Äquivalenz zwischen Wahrnehmung und Vorstellen zu vermuten, und zwar „. . . a more abstract or ‚second order‘ isomorphism in which the functional relations among objects as imagined must to some degree mirror the functional relations among those same objects as actually perceived.“ (Shepard 1978, S. 131). Dann müßten - eine weitere Annahme - Vpn Fragen nach der Ähnlichkeit von Komponenten des vorgestellten Bildes in etwa so beantworten wie Fragen nach der Ähnlichkeit von Komponenten des visuell vorhandenen Bildes. Als typisches Ergebnis werden statistisch nicht unterscheidbare Matrizen von Ähnlichkeitskoeffizienten berichtet, und man schließt, „. . . the subjects were performing very similar mental processes in the perceptual and imaginal conditions“ (S. 132). Da bei den Vpn ohne visuelle Vorgabe eine bildhafte Vorstellung nicht experimentell hergestellt wird, ist es denkbar, daß sie ohne solche Vorstellungen zu ihrem Verhalten kamen, daß sie etwa den demand characteristics des Experimentes entnahmen, wie sie reagieren sollten (vgl. Mitchell & Richman 1980). Forscher wie Finke & Kosslyn (1980) reagieren auf diesen Einwand mit dem Bemühen, experimentelle Situationen zu erfinden, in denen Vpn nicht wissen können, was die „richtigen“ Reaktionen wären, weil diese zu kompliziert sind.

Andere Versuchspläne sind gegenüber dieser Kritik noch anfälliger, wenn beispielsweise eine Gruppe von Vpn aufgefordert wird, sich vor der Reizdarbietung eine Vorstellung dieses Reizes zu machen, und diese Gruppe mit einer anderen ohne diese Instruktion verglichen wird (z.B. Shepard & Metzler 1971). Während in der zuvor erwähnten Untersuchung Ähnlichkeitseinstufungen als abhängige Variable verwendet wurde, also verbale Berichte über Ergebnisse kognitiver Prozesse, werden im zweiten Experiment, wie auch sonst häufig, Reaktionszeiten verwendet, also Verhaltensbeobachtungen als abhängige Variablen. Es darf als charakteristisch für die heutige kognitive Psychologie, und als Folge der behavioristischen Kritik am Introspektionismus gelten, daß Annahmen über Erleben, über bewußte Prozesse, an ihren vermuteten Auswirkungen auf Verhalten überprüft werden; nur selten greift man auf Selbstbeschreibungen zurück. Kognitive Vorgänge werden eher als intervenierende Variablen denn als hypothetische Konstrukte aufgefaßt.

In einer dritten Art von Versuchsplänen bemüht man sich, die Vp zu direkten Operationen an ihren bildlichen Vorstellungen zu veranlassen, und diese Ope-

rationen sollen die gleichen sein wie an vorgegebenem Material. Podgorny & Shepard (1978) beispielsweise legten ihren Vpn in der visuellen Versuchsbedingung den in Abb. 1 A gezeigten Blockbuchstaben F vor, in der Vorstellungsbedingung das leere Raster B in Abb. 1, zugleich mit der Instruktion, sich den mit Hilfe eines zuvor gelernten Codes beschriebenen Buchstaben F in diesem Raster vorzustellen. Wenn die Vp ihre volle Bereitschaft signalisierte, wurde ihr als Probereiz ein Punkt im Raster, Teil C in Abb. 1, z.B. mit der Frage dargeboten, ob der Probereiz auf der Figur läge oder nicht. Außer der Antwort wurde auch die Reaktionszeit erfaßt. Auch bei komplizierten Vorgaben stimmten die Ergebnisse der visuellen und der Vorstellungs-Bedingung überein, so waren beispielsweise die Reaktionen kürzer, wenn mehrere Punkte eines zusammengesetzten Probereizes auf die Figur fielen als wenn sie auf unterschiedliche Teile des Rasters fielen. Anscheinend wurden die Vorstellungen so erzeugt, daß gleichartige Operationen, jedenfalls solche mit vergleichbarem Zeitbedarf, bei den gleichen Aufgaben durchgeführt wurden.

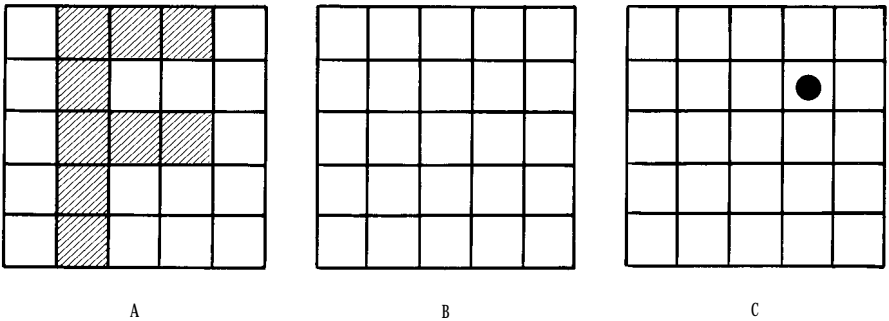


Abb. 1: Vorlagen im Versuch von Podgorny & Shepard, 1978

Beim nächsten Typus von Versuchsplan sollten laut Instruktion Operationen mit den Vorstellungen selbst vorgenommen werden, beispielsweise bei Shepard & Metzler (1971) eine Rotation dreidimensionaler Objekte im „Vorstellungsraum“. Gestützt auf mehrere solcher Experimente kommt Shepard (1978, S. 134) zu dem Schluß: „. . . we have established, more directly, that the intermediate states of the internal process do indeed have a one-to-one correspondence to intermediate orientations of the external Object. Our results, in fact, show that there is actually something rotating during the course of a mental rotation - namely, the orientation in which the corresponding external stimulus, if it were to be presented, would be most rapidly discriminated from other possible stimuli.“

Zweifellos stellen Ergebnisse wie diese und ähnliche jemanden, der sie ohne das Auftreten *bildhafter* Vorstellungen anzunehmen erklären will, vor eine

schwierige Aufgabe. Dennoch ist in allen diesen Untersuchungen der theoretisch interessierende Zustand, methodisch gesprochen: die unabhängige Variable, nicht experimentell hergestellt worden; der Erfolg der experimentellen Manipulation, die oft lediglich aus einer Instruktion an die Vpn besteht, ist kaum oder meistens gar nicht kontrolliert worden. Wenn auch im letzten Jahrzehnt die Annahmen über die Natur der untersuchten kognitiven Prozesse und Strukturen deutlicher geworden sind, so gilt dies nicht für die den Methoden impliziten Annahmen, die eine Methode bei ihrer Anwendung zu einer geheimen Theorie des Gegenstandsbereiches machen, zu dessen Erforschung sie gerade eingesetzt werden.

2.9 Methoden der Metakognitionsforschung

Die „kognitive Wende“ oder gar „kognitive Revolution“, die seit einigen Jahren selbst aus den Texten der ernsthaftesten psychologischen Grundlagenforscher und Theoretiker in die Augen springt (und an die „behavioristische Wende“ nach 1913 erinnert), besagt vorerst nur, daß tatsächlich der kognitive Wortschatz eine unwahrscheinliche Blüte erlebt. Weniger klar ist, inwieweit der neue Diskurs eine „Wende“, d.h. eine Abkehr vom „Behaviorismus“, symbolisiert. Zumindest steht das, was unter dem Zeichen eines Kognitivismus, zumeist auf dem Gebiet der „Artificial Intelligence“, allgemein der Informationsverarbeitungsmodelle, geleistet wird, nicht unbedingt in Widerspruch zu elaborierten Mediationsmodellen des Neobehaviorismus. Auch von einer nennenswerten Wiederbelebung der bewußtseinstheoretischen Diskussion kann keine Rede sein. Insofern steht auch die Weiterentwicklung der erlebnisdeskriptiven Methodik nicht unbedingt auf dem Programm der „Kognitiven Psychologie“.

Eine Ausnahme läßt sich jedoch erkennen: War die (iterative) Reflexivität des *cogito me cogitare* . . ., des ‚Ich weiß, daß ich weiß‘, auch ein jahrhundertealtes Thema philosophischer Reflexion, die Psychologie erreichte das Thema des Wissens über Wissen nur da, wo sich eine phänomenologische Einstellung, in der Bewußtsein thematisiert wird, realisierte. Das hat sich nun insofern geändert, als jetzt mehr und mehr Psychologen die alte Erkenntnis, daß das Tun eines Subjekts vom Bewußtsein dieses Tuns begleitet sein kann, sozusagen wiederentdecken und vor allem für Entwicklungs- und Pädagogische Psychologie fruchtbar zu machen suchen.

Um nicht in die nach wie vor noch tabuisierte Bewußtseinsterminologie der älteren Psychologie verstrickt zu werden, ist schlicht die Rede von *Metakognitionen* (metamemory, metalearning, metaattention), auch wenn mit (schlecht definierbaren) „Kognitionen über Kognitionen“ (Meichenbaum et al. 1979) doch nur „knowing about knowing“ (Brown 1978) gemeint ist. Immerhin

dürfte dieser Bereich der menschlichen Reflexivität, des Wissens, daß wir wissen und, günstigenfalls, wie wir lernen, denken, uns erinnern, am ehesten von allen kognitiven Forschungsgebieten auf den Einsatz erlebnisdeskriptiver Verfahren angewiesen sein.

Wiederum hängen aber der Einsatz und die Verwendung der Methoden ab von der Verwendung (Bedeutung) der das Feld bezeichnenden Konzepte.

Wenngleich der Begriff bzw. die Vokabel *Metagedächtnis* am Anfang dieser neueren Entwicklung stand, muß man ihn wohl heute dem Konzept der *Metakognition* zuordnen:

„Metakognition bezieht sich auf das Wissen eines Menschen über seine eigenen kognitiven Prozesse und Produkte oder was damit in Beziehung steht, wie lernrelevante Eigenschaften von Informationen oder Daten. So betreibe ich z.B. Metakognition . . . , wenn ich bemerke, daß ich mehr Schwierigkeiten habe, A als B zu lernen; wenn mir aufgeht, daß ich C nochmal prüfen sollte, ehe ich es als Tatsache akzeptiere; wenn mir einfallt, daß ich besser jede einzelne Alternative in den Multiple-choice-Aufgaben durchprüfe, ehe ich entscheide, welche die richtige ist; wenn ich das Gefühl habe, ich sollte mir D lieber notieren, damit ich es nicht vergesse . . .“ (Flavell 1976, 232).

Rein von der Funktion her wird Metakognition primär als Selbstkontrolle und -Steuerung verstanden; als „*monitoring*“, was nach Flavell (1979, 2) soviel heißt wie „keeping track of how it is going and taking appropriate measures whenever it needs to go differently“. Unabhängig davon kann man das „metakognitive“ Wissen primär als *Inhalt* (stored contents) oder primär als Prozeß bzw. *Aktivität* auffassen (Cavanaugh & Perlmutter 1980).

Je nach Akzentsetzung fallen denn auch die methodischen Präferenzen aus: beim inhaltsorientierten Vorgehen Interviews und Fragebogen, bei Aktivitätsorientierung verbale Protokolle, Verhaltensbeobachtung, „feeling-of-knowing“-Technik.

Als jüngster Versuch in der Geschichte der Psychologie, wenigstens des (reflexiven) Wissens unseres Bewußtseins habhaft zu werden, verdient seine Methodik besondere Aufmerksamkeit.

Doch wird derjenige, der die Entwicklung oder den Einsatz einer neuartigen Methode erwartet, vorerst enttäuscht. Wie schon gelegentlich in der Diagnostik praktiziert, wird dieser Mangel durch eine „Batterie“ von Verfahren kompensiert. Dabei stellen Interview und Fragebogen die bisher verbreitetsten Verfahren dar, deren Schwächen, im Prinzip bekannt, neu diskutiert werden (Adair & Spinner 1979; Meichenbaum & Butler 1980). Bei dem von Hart (1965; 1966; 1967) entwickelten ‚Feeling-of-knowing‘-Verfahren werden Vpn, die sicher sind, die Lösung eines Problems bzw. die Antwort auf eine Frage

schon „auf der Zunge“ zu haben, sie nur nicht aussprechen zu können („tip-of-the-tongue-Phänomen“), zu Aussagen über die entsprechenden Items veranlaßt, die dann mit späteren Aussagen darüber verglichen werden (kritisch hierzu: Cavanaugh & Perlmutter 1980). Problematisch scheint auch die Verwendung der Reaktionszeit für eine Frage als Indikator der Gewißheit der Vp bezüglich ihres Wissens zu sein (Lachman & Lachman 1980). Nahe liegt, daß zur Identifikation von Metakognitionen auch die Methode des lauten Denkens wieder aufgegriffen wird. Sie wird ergänzt durch eine Art Nachfaßtechnik (probe technique), bei der die Vpn nach jeder Aufgabe nach den Strategien und Hypothesen gefragt werden, die sie bei der Bearbeitung benutzt haben. Ähnlich hatte schon Giorgi (1967), um die Gleichgewichtigkeit von „experimental data“ und „experiential data“ zu demonstrieren, nach einem traditionellen Experiment zum seriellen Lernen seine Vpn in einer postexperimentellen Befragung zur subjektiven Schwierigkeit u.ä. vernommen und dadurch erst Zugang zu den Lösungsansätzen (und zur sinnvollen Interpretation seiner Daten) gefunden. Es besteht die Hoffnung, daß durch das jüngste Interesse an Metakognitionen die methodisch noch entwicklungsfähige Kombination von Verhaltens- bzw. Leistungs- und Erlebnisdaten neue Impulse erfährt.

3. Aktuelle Probleme der Verhaltensbeobachtung

Anders als im Artikel über die wissenschaftliche Beobachtung (Feger, in diesem Band), in dem allgemeine methodische Probleme der wissenschaftlichen Beobachtung besprochen wurden, wenden wir uns hier Ansätzen zu, die versuchen, Beobachtungsverhalten mit Hilfe psychologischer Begriffe und Theorien zu beschreiben und zu erklären. Schließlich ist Beobachten menschliches Verhalten, und somit Bestandteil des Gegenstandsbereiches unseres Faches. Wir befassen uns zunächst mit der Frage, wie der Gegenstand einer psychologischen Verhaltensbeobachtung zu bestimmen sei, dann mit der teilnehmenden Beobachtung, mit den Regeln, nach denen Beobachter ihre Beobachtungseinheiten festlegen, und schließlich mit den Zusammenhängen zwischen Beobachtung, Gedächtnis und verbalem Bericht über das Beobachtete.

3.1 Der Gegenstand psychologischer Verhaltensbeobachtung

Wir stellen zunächst die Frage, was gegenwärtig typischerweise *Untersuchungsgegenstand psychologischer Beobachtungsstudien* ist und wie er genau anzugeben sei, wobei wir der Übersicht von Longabaugh (1980) weitgehend folgen. Longabaugh unterscheidet zwei Brennpunkte der Analyse: Was zeigt sich als Verhalten, und *wie* zeigt es sich:

„The study of ‚what‘ is the study of what it is actors communicate to one another: both what is intended and encoded and what is received or decoded. The study of ‚how‘ is the study of how the content is expressed.“ (p. 64).

Die Untersuchung der Erscheinungsformen hat sich in den letzten Jahrzehnten stark ausgebreitet und spezialisiert, wobei sich zwei Schwerpunkte gebildet haben: Die Analyse des Sprechens und der nichtverbalen Kommunikation (z.B. Duncan 1969) und der Bewegungen des Körpers (z.B. Birdwhistell 1970, Schefflen 1975).

Neben Erscheinungsform und Inhalt werden, um den Untersuchungsgegenstand genau zu bestimmen, „actors, targets, and settings“ anzugeben sein. Man wird also diejenigen bestimmen müssen, von denen das Verhalten ausgeht, diejenigen, auf die es gerichtet ist, und die Umstände schildern müssen, unter denen das Verhalten verwirklicht wird. Longabaugh führt ein Klassifikationssystem von Lambert (1960) weiter, in dem für Verhaltensträger, Verhaltensziele und Verhaltensumwelt unterschieden wird, ob sie als einmalig und spezifisch festgelegt oder als Klasse umschrieben oder unspezifiziert gelassen werden. Für Verhaltensträger beispielsweise bedeutet die erste Kategorie, daß eine Einzelfallstudie durchgeführt wird, die zweite Kategorie, daß beispielsweise zwölfjährige Mädchen in Bayern untersucht werden, die dritte Kategorie, daß Ergebnisse ohne Angabe des Bereichs der Verallgemeinerbarkeit mitgeteilt werden. Als wichtig unterstreicht nun Longabaugh, daß auch Verhaltensziele und settings nach dem Grad der Spezifität charakterisiert werden können und müssen: „Because of the variable and sometimes overwhelming effects of targets and settings on behavior, researchers of behavior in naturalistic settings cannot let any of these elements go unidentified.“ Dabei sollte die Charakterisierung nach jenen Merkmalen erfolgen, die als für das beobachtete Verhalten als wesentliche Ursachen, Bedingungen, Voraussetzungen etc. angesehen, oder - besser noch - aus einer Theorie des Beobachteten *abgeleitet* werden. Dann ergibt sich als weiterer Vorteil neben der eindeutigen Abgrenzung des Aussagenbereichs die Möglichkeit, das Beobachtete umzuklassifizieren, insbesondere dann, wenn in einer Bedingungskombination mehr Heterogenität als erwartet beobachtet wurde.

Die Beschreibung der Umwelt kann nach Longabaugh aus drei Perspektiven erfolgen: erstens als die physikalische Beschreibung der Objekte und des Terrains (z.B. Barker 1968, Whiting & Whiting 1975), zweitens als Beziehung zwischen der beobachteten Person und ihrer Umwelt, wobei die *relevanten* Aspekte der Umwelt - zusammengefaßt als Situation des Beobachteten - aus der Kovariation von Verhalten und Umweltvariablen erschlossen wird (z.B. Skinner 1953), und schließlich drittens als die Bedeutung, die einer Gelegenheit, einem Ort, einem Anlaß von den typischen Benutzern, Anwesenden und Teilnehmern gemeinsam und übereinstimmend zugeschrieben wird (z.B. Barker & Schoggen 1973).

3.2 Analyse des Beobachters als Meßinstrument

3.2.1 Die Ermittlung von „Fehlern“

Beobachter werden häufig dann eingesetzt, wenn es keine Meßinstrumente gibt, um das zu erfassen, woran der Wissenschaftler interessiert ist. Der Beobachter übernimmt dann die Funktion eines Meßinstruments:

„When a person makes a perceptual judgment, he acts as a kind of measuring instrument, and the theory of perceptual judgments is an attempt to understand how these measurements are produced.“ (Greeno 1968, S. 1).

Der Gedanke liegt nahe, man müsse die Eigentümlichkeiten, Unvollkommenheiten, „Fehler“ dieses menschlichen Meßinstrumentes kennen, um sie bei Schlüssen aus den Beobachtungsergebnissen berücksichtigen zu können. Aus anderer Perspektive geben diese Fehler Aufschluß über den Menschen als kognitives System, das zu analysieren Aufgabe der Allgemeinen Psychologie ist. Im ersten Fall interessiert nach wie vor primär das Beobachtete, und die Fehleranalyse ist in Gefahr, ein Inventar von Kuriosa zu bleiben; im zweiten Fall interessiert der Beobachter als solcher. Zwischen diesen beiden Polen schwankten die Auffassungen in der Psychologiegeschichte, wobei der Fehlerauffassung historisch früher nachgegangen wurde, und zwar besonders dann, wenn in einem bestimmten, relativ umgrenzten Forschungsbereich - etwa bei der Personenbeschreibung in der Diagnostik und Sozialpsychologie, als Erwartungsfehler in der Psychophysik oder Kontexteffekt in der Wahrnehmungsforschung - sich solche Unregelmäßigkeiten zeigten, die mit dem umgrenzten Bereich anscheinend nichts zu tun hatten. Die Fehler scheinen relativ unabhängig von Inhalt des jeweils Wahrgenommenen und deshalb Merkmale von Beobachtern allgemein zu sein.

Guilford (1954, S. 278ff.) gibt eine Übersicht der bis dahin häufiger untersuchten Fehlerarten: der Fehler der zentralen Tendenz (error of central tendency), der Fehler der Milde (leniency) oder der Großzügigkeit (generosity error), logische Fehler, verzerrende Effekte der zeitlichen Reihenfolge: stärkerer Einfluß der zuerst (primacy-) oder zuletzt (recency effect) aufgenommenen Elemente einer Serie, und besonders der Halo-Effekt. Man könnte Eigentümlichkeiten im Umgang mit bestimmten Urteilsprachen hinzunehmen, etwa die bevorzugte Verwendung bestimmter Kategorien einer Rating-Skala, oder Einflüsse, die von anderen Zielen und Absichten des Beobachters ausgehen als gerade zu beobachten, beispielsweise die Wirkungen der sozialen Erwünschtheit von Reaktionen.

Von „Fehlern“ kann man nur sprechen, wenn es Diskrepanzen zwischen den Beobachtungen und Resultaten anderer, unabhängiger methodischer Zugänge zum gleichen Gegenstand gibt. Und die Gültigkeit und Verlässlichkeit dieser

Ergebnisse anderer Methoden muß hinreichend gesichert sein. Betrachtet man die erwähnten Fehlerarten in dieser Hinsicht näher, so erscheint die Problemlage nicht einheitlich und für die Annahme, es handle sich um Fehler, nicht günstig. Beispielsweise wurden Fehler deshalb angenommen, weil Urteile verschiedener Beobachter untereinander nicht übereinstimmten und man systematische Einflußquellen ermitteln konnte - z.B. beim Mildefehler um so größeres Wohlwollen, je bekannter der Beurteilte dem Beurteiler war. Oder Beobachtungsleistungen wurden als Fehler bezeichnet, weil sie mit Selbstaussagen der Beobachteten nicht übereinstimmten; Selbstaussagen sind sicher nicht garantiert valide. Beim Vergleich mit physikalischen Meßwerten, die an den beobachteten Phänomenen gewonnen wurden, ist das Fehlerkonzept in dem Augenblick überholt, wo man Wahrnehmung als aktive Herstellung einer psychischen Wirklichkeit durch den Organismus begreift. Wir wählen daher im folgenden exemplarisch einige Ansätze aus, die primär am Beobachter, und dann an seinen Urteilen interessiert sind.

3.2.2 *Der Einfluß von semantischen Gedächtnisstrukturen auf Verhaltensbeschreibungen*

D'Andrade (1974) untersucht die Auswirkungen von langfristigem (länger als 10 Min.) Gedächtnis auf Urteile über Verhalten. In einer früheren Arbeit (1965) zeigte er, daß Urteile über die semantische Ähnlichkeit von Eigenschaftsbezeichnungen einerseits, Verhaltenseinschätzungen andererseits nahezu die gleiche Struktur der Interkorrelationen aufwiesen. Die Isomorphismus-Hypothese deutet diesen Befund so, die Ähnlichkeit der Eigenschaftsnamen bestehe, weil sie ein Niederschlag der Erfahrungen über den tatsächlichen, beobachtbaren Zusammenhang von Verhaltenseigenschaften sei. Hingegen interpretiert die Verzerrungshypothese: Semantische Zusammenhänge verfälschen Urteile über Verhalten so, daß sie in der Weise interkorrelieren, wie es der semantischen Ähnlichkeit entspricht.

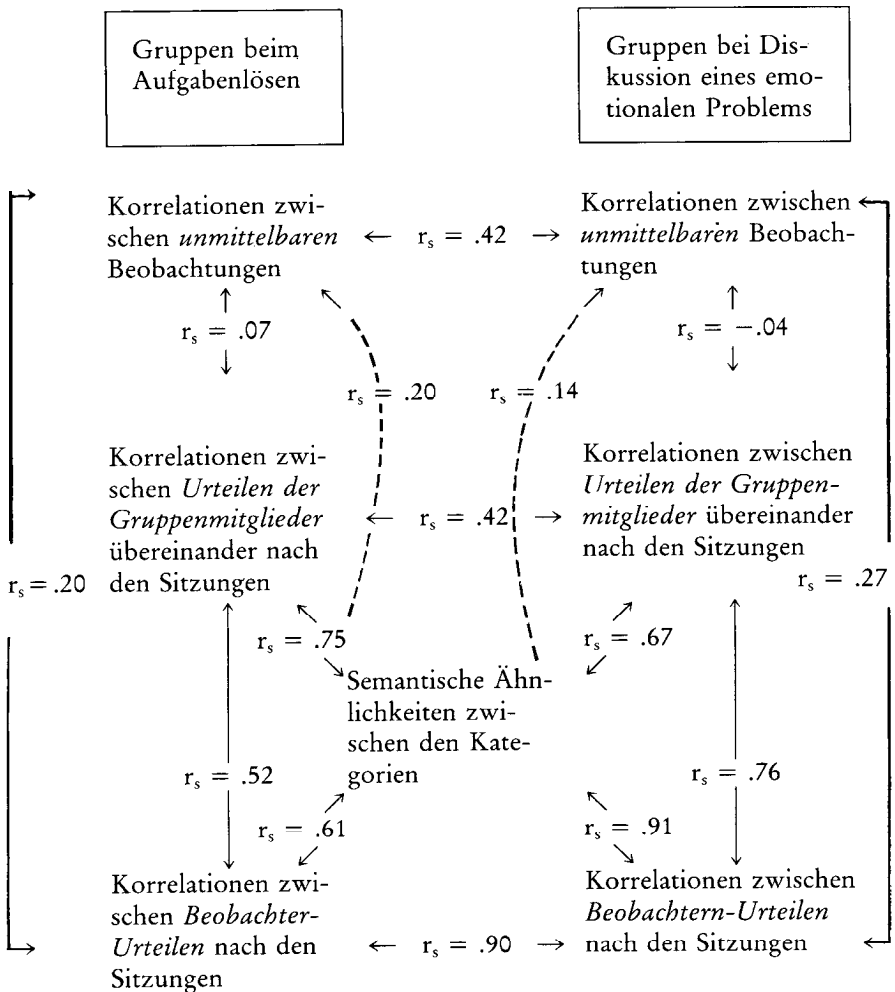
Zwischen beiden Hypothesen kann man nach D'Andrade entscheiden, wenn man über zweierlei Daten verfügt: unmittelbare, detaillierte Verhaltensbeobachtungen und Beschreibungen nach Ablauf eines längeren Zeitintervalls.

„If the observer's memory-based ratings showed a very different pattern of correlations from that found for the data based on the actual behavior of the subjects (but a pattern similar to judgments of semantic similarity), it would be reasonable to reject the isomorphism hypothesis and to consider the systematic-distortion hypothesis supported.“ (D'Andrade 1974, S. 162).

D'Andrade untersucht diese Frage anhand der Studien von Borgatta et al. (1958) und Mann (1959), deren Daten er Einschätzungen seiner eigenen Vpn hinzufügte, welche die semantische Ähnlichkeit zwischen den (in Anlehnung

an Bales formulierten) verhaltensdeskriptiven Kategorien eingestuft hatten. Bestimmt wurde dann die Ähnlichkeit zwischen den Korrelationsmatrizen, indem Spearmans rho zwischen den entsprechenden Korrelationen in je zwei Matrizen berechnet wurden. Die Ergebnisse, die sich auf die Mann (1959)-Studie beziehen, seien in Tab. 1 wiedergegeben. Sie stützen nach D'Andrade eindeutig die Verzerrungs-Hypothese. Im Gedächtnis würden demnach Beobachtungen bereits bestehenden, semantischen Strukturen angeglichen; allen-

Tabelle 1: Vergleich von Korrelationsmatrizen mit Hilfe von Spearmans rho (= D'Andrade 1974, S. 176).



falls könnte man Beschreibungen trauen, die unmittelbar nach der Beobachtung entstehen. (Borg & Bergermaier, 1979, haben in anderem Zusammenhang die Problematik solcher Matrizenvergleiche herausgestellt.)

Mit der Schilderung der Arbeiten von D'Andrade sind wir in eine Kontroverse eingestiegen, die seit Beginn der wissenschaftlichen Psychologie zunächst aus der Perspektive der Beurteilungsfehler geführt wurde: Wenn Beobachter Verhalten beschreiben und (angeblich) verhaltensbezogene Schätzurteile abgeben, interessiert oft nicht nur die berichtete Ausprägung eines Merkmals, sondern der Zusammenhang zwischen mehreren Merkmalen. Und man vermutet, Schätzungen des Zusammenhanges von Merkmalen könnten verfälscht sein, weil die Beurteiler schon ihre Vorstellungen über den Zusammenhang der Merkmale hätten, und diese Vorstellungen gingen in noch näher zu bestimmender Weise in die Schätzungen ein.

Die Thematik wurde erstmals in den Forschungen zum *Halo-Effekt* angesprochen (Wells 1907, Thorndike 1920, Rugg 1922; s. auch Rudinger & Feger, 1970). Unter Halo-Effekt versteht man einen überhöhten Zusammenhang zwischen (beurteilten) Merkmalen, verursacht durch den Gesamteindruck über den Merkmalsträgern und beim Beurteiler bestehende Annahmen über die Beziehungsstruktur zwischen Merkmalen im allgemeinen. Hat ein Beurteiler einen insgesamt positiven Eindruck von einer Person gewonnen, so werde er dazu neigen, sie relativ günstig bei von ihm als positiv bewerteten, miteinander zusammenhängenden Eigenschaften einzuschätzen. Korreliert man nun die Urteile zu zwei Merkmalen (in der Literatur: der gleichen oder verschiedener Personen, von einem oder mehreren Beurteilern), so fallen die Korrelationen verfälscht aus. Die Evidenz in den frühen Arbeiten für die Existenz eines Halo-Effekts war zunächst nur das Auftreten unplausibel hoher Korrelationen zwischen Eigenschaftsratings.

Einen Schritt weiter ging Newcomb (1931) und neuerdings D'Andrade (s.o.), der den Einfluß eines Halo-Effektes aus dem Unterschied zwischen Korrelationen schloß. Die eine Korrelation bezog sich auf Einstufungen von zwei Merkmalen unmittelbar nach der Verhaltensbeobachtung, die andere auf Einstufungen aus dem Gedächtnis, und diese fiel durchgängig höher aus. Was auch immer die Verfälschung bewirkte, es brauchte eine gewisse Zeit zu seiner Wirkung oder es wurde durch den unmittelbaren Bezug zum beobachteten Verhalten an seiner Wirkung gehindert.

Daß Beurteiler, Personen überhaupt, relativ stabile und interindividuell vergleichbare Vorstellungen darüber haben, wie (Persönlichkeit-)Eigenschaften kovariieren, ist spätestens seit Bruner & Tagiuri (1954) und Cronbach (1955) Ergebnis und Gegenstand der Forschung zur „impliziten Persönlichkeitstheorie“. Relevant sind solche Befunde auch für Arbeiten im Bereich der Persönlichkeitsforschung, die ihre Aussagen stark auf Korrelationen zwischen Ver-

haltenseinschätzungen stützen (z.B. Cattell 1957, Lorr & McNair, 1965). Ihr Anspruch, relativ stabile Persönlichkeitsdimensionen gefunden zu haben, wurde durch andere Forscher (z.B. Mulaik 1964, Passini & Norman 1966) in Frage gestellt, die diese Dimensionen als „Produkt stereotyper und semantischer Annahmen der Beurteiler“ deuteten (so Berman & Kenny 1976). Eine vergleichbare Problematik ergibt sich bei Korrelationen zwischen klinischen, diagnostischen Urteilen (z.B. Chapman & Chapman 1967, 1969; Starr & Katkin 1969).

Berman & Kenny (1976) gehen wieder einen Schritt weiter: Während in den früheren Untersuchungen die „wahre“ Korrelation zwischen den Merkmalen unbekannt war, variieren die Autoren diese Korrelation experimentell, indem sie ihren Beurteilern Bilder von Personen zeigten, und mit jedem Bild eine Aussage darboten, die die Ausprägung des Merkmals angab (z.B.: John ist sehr freundlich). Da John auch andere Merkmale zugeschrieben wurden, konnten Berman & Kenny den „wahren“, d.h. von ihnen hergestellten Zusammenhang zwischen den Eigenschaften variieren. Sie wählten nun nicht beliebige Merkmale, sondern solche, die nach Einstufungen anderer Vpn in einer Vorstudie paarweise stark positiv, stark negativ oder gar nicht korrelierten. Die Vpn der Hauptstudie stuften aus ihrer Erinnerung die ohne Aussagen erneut dargebotenen Personenbilder ein. Verglichen wurden die aus diesen Einstufungen berechneten Korrelationen mit den von den Autoren hergestellten.

Gleichgültig, wie hoch die experimentell hergestellte Korrelation war: Jene Merkmalspaare, die nach Meinung der Vpn der Vorstudie hoch korrelieren, zeigten auch die aus den Einstufungen in der Hauptstudie berechneten höchsten Korrelationen. Analoges galt für die negativen und Null-Korrelationen, und dies galt auch für über alle Beurteiler gemittelte Einstufungen. Als paralleler Befund wird auf die Arbeiten von Lay & Jackson (1969) und Stricker et al. (1974) verwiesen, in denen große Ähnlichkeit zwischen den Korrelationsmustern der Antworten in Persönlichkeits-Inventaren wie dem MMPI einerseits und Annahmen von Vpn über den Zusammenhang der durch diese Inventare erfaßten Merkmale gezeigt wurde.

Die Arbeit von Berman & Kenny hat eine Diskussion ausgelöst, in die Block (1977) folgende Argumente einbringt:

- (1) Die experimentelle Manipulation der Korrelation zwischen den Merkmalen sei „künstlich“, denn die paarweisen Merkmale seien bei hoher vermuteter Korrelation Homonyme, bei negativen Korrelationen aber Gegensätze gewesen. Wie man aus Arbeiten wie denen von Cohen & Schümer (1968), Cohen (1971), Schümer (1971) entnehmen könne, bemerken Vpn meistens, wenn und daß sie widersprüchliche Information über eine zu

beurteilende Person angeboten bekommen, und das führe für zumindest einige Vpn zu einem „different context of understanding“.

- (2) Die Situation sei für die Vpn von Berman & Kenny nicht die einer Verhaltensbeurteilung gewesen, da ihnen kein Verhalten gezeigt wurde; es sei eher paired associate Lernen verlangt worden, was andere Prozesse auslöse.
- (3) Die Informationsbasis der Vpn von Berman & Kenny entspreche wegen ihrer kurzen zeitlichen Dauer allenfalls dem Bilden eines ersten Eindrucks.
- (4) Block listet Anforderungen an Einstufungen von Verhalten auf, die seines Erachtens allgemein zu erfüllen seien, bei Berman & Kenny jedoch teilweise vernachlässigt worden wären, und zwar:
 - (a) mehrere Beobachter schätzen unabhängig voneinander ein;
 - (b) Beobachter sollten ihre Urteile auf extensive, und - wegen der Wirkung von Kontexteffekten - auf verschiedenartige Gelegenheiten stützen können;
 - (c) die Beobachter sollten eine gemeinsame „evaluative metric“ benutzen, so daß es erlaubt ist, ein aggregiertes Urteil zu bilden;
 - (d) ipsative Urteile seien außernormorientierten vorzuziehen;
 - (e) „Prior calibration of the observers“, also Schulung, und konstante Kontrolle der Leistung der Beobachter sollte gewährleistet sein.

Die Erwiderung findet sich in Berman & Kenny (1977). Sie kommentieren das Argument, ihren Vpn sei kein beobachtbares Verhalten als Urteilsbasis angeboten worden, mit dem Hinweis, seit langem - so z.B. schon Guilford (1954) - hätten Psychometriker als wichtige Möglichkeit, Urteilsfehler zu reduzieren, vorgeschlagen, die Anforderungen an das Urteilsverhalten der Beobachter zu minimieren. Das sei in ihrer Studie geschehen, und trotzdem seien die ausgeprägten Verzerrungen festgestellt worden. - Zum Gedanken, Urteilsfehler durch Wiederholungen „auszumitteln“, argumentieren Berman & Kenny (1977): Um Urteile über eine Mittelwertsbildung gültiger zu machen, um also das Gesetz der großen Zahl anwenden zu können, benötige man Messungen, deren Fehler unkorreliert seien. Gerade dies sei bei Verhaltenseinstufungen in der Regel nicht der Fall. Es sei - noch anspruchsvoller - sogar zu fordern, daß die Fehler bei der Beurteilung des einen Merkmals durch den einen Beobachter unabhängig seien von den Fehlern bei den Einschätzungen eines anderen Merkmals durch einen anderen Beobachter. Analysen von Berman & Kenny legten den Schluß nahe, daß auch diese Bedingung nicht immer erfüllt sei.

Damit haben wir die Diskussion der Abhilfeschläge eingeleitet, die meist mit der Mahnung beginnen, die Urteilsskalen, ihre Etikettierung und die zugehörige Instruktion sorgfältig zu wählen. Jeder Beurteiler solle alle Personen, zunächst oder ausschließlich, nur hinsichtlich eines Merkmals einstufen. Johnson (1965) konnte jedoch auch bei diesem Vorgehen die Verzerrung nicht

vermeiden. Statistische Adjustierungen wurden von Chi (1937) und Stanley (1961) vorgeschlagen, wozu Berman & Kenny (1976) bemerken, die Korrektur geschehe bei diesen Verfahren aufgrund von Unterschieden zwischen den Beobachtern, deren „common stereotypes“ würden also nicht berücksichtigt.

Die in diesem Abschnitt behandelten Arbeiten beziehen sich auf Beurteilungen, nicht auf Wahrnehmungen, aber jede Beobachtung muß in irgendeiner Form - und in der Psychologie ist es oft die sprachliche - ihr Ergebnis mitteilen. Die Studien zeigen: Wird das Verhalten nicht sogleich (vom Beobachter) aufgezeichnet, so können systematische Verzerrungen eintreten. Erklärt wurden diese „Fehler“ zunächst als bedingt durch einen Gesamteindruck, den der Beobachter vom Beobachtungsgegenstand, einer Person, gewonnen hat, dann als bedingt durch eine implizite Theorie über das Zusammen-vorkommen der Merkmale, wobei diese die Beurteilungen steuernde „Theorie“ weder spezifisch auf einen bestimmten Beobachtungsgegenstand bezogen ist noch als charakteristisch für einen bestimmten Beobachter angesehen wird. Der Halo-Effekt wurde vom Beobachtungsfehler zu einem von vielen Indizien dafür, wie konkrete einzelne Beobachtungen mit früher gesammelten Erfahrungen integriert werden.

3.2.3 Die Theorie der Signalentdeckung: Der Beobachter als Sensorium und als Entscheidungsinstanz

Wir geben hier keine Darstellung auch nur der Grundzüge der theory of signal detection (erstmalig Peterson et al. 1954, van Meter & Middleton 1954), in die es mehrere gute Einführungen, z.B. Coombs et al. (1970), McNicol (1972), systematische Darstellungen (Egan 1975), und übersichten ihrer Anwendung in der Psychologie gibt (Swets 1964, Price 1966, Pastore & Scheirer 1974). Wir werden vielmehr an einem fiktiven und in einigen Punkten absichtlich unzulänglichen Beispiel die Leistungen und Möglichkeiten der Theorie der Signalentdeckung skizzieren.

Ausgangspunkt der Theorie, und gegenüber der bisherigen Psychophysik das wesentlich Neue ist die analytische Trennung zwischen dem Beobachter als einem *Sensorium*, einerseits, d.h. als Registriersystem, das mit einer gewissen, empirisch festzustellenden Sensitivität auf die externe Reizvariation reagiert, und andererseits dem Beobachter als *Entscheidungsinstanz*, d.h. als Bewertungssystem, das für seine Urteile über eine Reizgegebenheit nicht nur den sensorischen input berücksichtigt, sondern auch seine subjektiven Erwartungen und Werte. Die Erwartungen können sich auf die Wahrscheinlichkeiten beziehen, daß ein Reiz, Signal, oder nur noise, kein Signal auftritt. Die Bewertungen beziehen sich auf die belohnenden Folgen eines richtigen, und die bestrafenden Folgen eines falschen Urteils. Die Theorie wird angewendet,

wenn das Signal schwach ist relativ zu den (störenden) Hintergrundreizen. Für starke Reize, die stets eindeutig vom Hintergrund oder im Vergleich zu anderen Reizen unterschieden werden, ist die Theorie nicht erforderlich. Die Theorie legt fest, welches Urteil über den Reiz man in einer unsicheren Situation fällen sollte, indem sie Kriterien und Eingangsgrößen spezifiziert, nach denen die Wahl zwischen Reaktionsmöglichkeiten zu treffen ist.

Das Beispiel: Ein Beobachter soll immer dann eine bestimmte Kodierungsmarke benutzen, wenn eine Vp einer anderen „freundlich zulächelt“. Das freundliche Zulächeln ist in dem Sinne ein „schwaches Signal“, als es zum einen durch rein äußerliche Umstände - wenn etwa sich die Vp vom Beobachter mehr oder weniger abwendet - nur teilweise oder verzerrt als Reiz angeboten wird, zum anderen nicht präzise definiert ist, was „freundlich zulächeln“ bedeutet - es gibt keinen eindeutigen Vergleichsstandard. Andererseits ist „noise“ stets vorhanden, auch für den Beobachter zeigt die Vp ununterbrochen Ausdrucksverhalten, das bisweilen dem Zulächeln ähnlich ist und damit verwechselt werden kann. Um die Theorie der Signalentdeckung anwenden zu können, muß nun eindeutig geklärt sein, wann in der Umwelt des Beobachters der Reiz auftritt, und wann nicht. Wir stoßen hier auf ein Problem, das die Übertragung der Signalerkennungs-Theorie auf Beobachtungssituationen außerhalb des psychologischen Laboratoriums erschwert hat, und das - in anderem Zusammenhang - D'Andrade (1974, S. 159) so formuliert:

„Attempting to specify what aggressive means by defining the term with reference to more specific acts, such as hitting or insulting, still fails to make the decision process explicit, since the rules for coding behavior remain dependent on a set of undefined terms. Unless the measurement process is based solely on physical properties, at some point such verbal coding rules always fall back on undefined terms.“

Die Theorie nimmt nun weiter an, die sensorische Evidenz, auf die sich der Beobachter für sein Urteil stützt, ließe sich als ein Kontinuum zunehmender oder abnehmender Stärke darstellen, als eine „evidence variable“ (Ingleby 1974) oder „decision axis“, deren inhaltliche Zusammensetzung zunächst nicht geklärt werden muß. Über diesem Kontinuum existieren zwei Verteilungen, die eine für die Fälle, in denen das Signal vorhanden ist, die andere für die Fälle mit ausschließlich Störhintergrund. Diese beiden Wahrscheinlichkeits-Verteilungen, daß bei einer bestimmten Evidenzstärke das Signal vorhanden ist oder nicht, hat der Beobachter aus seiner Alltagserfahrung gelernt. Die Beobachtungssituation legt auch fest, sei es indirekt etwa über eine Instruktion, frühere Erfahrungen etc., sei es direkt durch experimentelle Manipulation einer Auszahlungsmatrix, wie großen Wert der Beobachter darauf legt, zutreffend „Lächeln“ und „Nicht-Lächeln“ zu identifizieren, und wie gravierend für ihn die Fehler des false alarm (er registriert ein Lächeln, das nicht auftrat) und des Übersehens sind.

Die Theorie definiert nun zwei statistisch voneinander unabhängige Kennwerte, von denen d' die Empfindlichkeit des Beobachters (hier: seine Sensitivität für zutreffende Identifikation von Lächeln aus dem übrigen Ausdrucksgeschehen) und β einen Cut-Off-Punkt auf der Entscheidungsachse darstellt, der vorschreibt, bis zu welcher Evidenzausprägung das Urteil „ich habe Lächeln beobachtet“ abzugeben sei: Die analytische Trennung zwischen dem Beobachter als Sensorium und als Entscheidungsinstanz geschieht so, daß Erwartungen und Werte des Beobachters ausschließlich in die Bestimmung von β , nicht aber von d' eingehen. Um β bestimmen zu können, muß man noch festlegen, welcher Entscheidungsregel der Beobachter als *bester* folgen soll, und dafür gibt es mehrere Möglichkeiten, etwa die, den maximal erwarteten Nutzen zu erstreben, oder den minimal zu erwartenden Verlust. Ein Beobachter, der in seinem Beurteilungsverhalten einer solchen Regel folgt, heißt rationaler *Beobachter*. Wenn er nun noch Leistungen zeigt, die nur durch die Struktur des Reizangebotes in seiner Umwelt, nicht aber durch seine Schwächen als Sensorium begrenzt werden, bezeichnet man ihn als *idealen Beobachter* (oder Empfänger). Ändert sich die Umwelt, so ändert sich die Struktur des idealen Beobachters in solcher Weise, daß auch unter den neuen Bedingungen die bestmöglichen, nicht unbedingt 100% zutreffenden Urteile über die Reizgegebenheiten gefällt werden (Egan 1975).

Empirische Untersuchungen (z.B. Coombs et al. 1970) zeigen, daß reale Beobachter deutlich schlechtere Leistungen zeigen als ideale, und daß dies u.a. wahrscheinlich an unzulänglichen Speicherungen des Signals im Gedächtnis liegt. Wir können hier nicht die ständig wachsende Zahl empirischer Arbeiten wie etwa die von Ingleby (1974) referieren; uns kam es lediglich darauf an, die Perspektive der Theorie der Signalentdeckung aufzuzeigen und ihre Anwendung auch auf nicht streng oder ausschließlich physikalisch definierte Reize zu diskutieren und anzuregen.

3.2.4 *Verhaltenseinschätzungen als Testscores*

Man kann, wie dies van der Kamp & Mellenbergh (1976) tun, Verhaltenseinstufungen als „a special class of test scores“ auffassen, und dann konsequent annehmen „. . . that each rater may be regarded as a test instrument“ (S. 311). Mit der Anwendung der Testtheorie (s. Kristof und Fischer in diesem Band) betrachtet man dann die Frage, in welcher Hinsicht Beobachter übereinstimmen. Jeder Beobachter wird dabei zu einer Methode, und gefragt ist nach der Äquivalenz verschiedener Methoden, eine „zugrundeliegende“ oder „hypothetische“ Variable zu messen.

Die strengste Form der Äquivalenz ist die Austauschbarkeit (ratings are interchangeable), die sich mit einem Kriterium nach Wilk-Votaw (s. Gulliksen 1950, 1968) prüfen läßt. Für die Urteile von Beobachtern bedeutet dies, sie

seien in der gleichen Weise - am gleichen Skalenuersprung - verankert (gleiche Mittelwerte), benutzen die gleiche Skaleneinheit (gleiche Varianzen), erfassen die Beziehungsstruktur zwischen verschiedenen Variablen in der gleichen Weise (gleiche Kovarianzen), und ihre Einstufungen seien deshalb auch für alle Beobachter gleich verlässlich und gleichermaßen gültig. Diese unrealistischen Anforderungen werden nun von den Autoren Schritt für Schritt abgeschwächt, unterschiedliche Reliabilität der Beobachter wird eingeräumt, etc.

Werts et al. (1976) gehen ebenfalls von der klassischen Testtheorie aus und betrachten den Fall, daß mehrere Beobachter die gleichen Objekte in mehr als einer Hinsicht einschätzen. Unter diesen Umständen sei es für einen Beobachter kaum möglich, seine Einschätzungen auf der einen Dimension nicht von seinen Kenntnissen seiner Urteile auf den anderen Dimensionen beeinflusst sein zu lassen. „This kind of contamination means that the errors of measurement for one dimension may be correlated with the errors on other dimensions, i.e., the *intrajudge* measurement errors are correlated.“ (S. 319). Dann sind die Kovarianzen zwischen Einstufungen auf verschiedenen Dimensionen durch den gleichen Beobachter nicht gleich den Kovarianzen zwischen den zugrunde liegenden wahren Werten, wie normalerweise in der klassischen Testtheorie angenommen werde. Auch die übliche Minderungskorrektur um die Korrelation zwischen wahren Werten verschiedener Dimensionen zu bestimmen, sei nicht anwendbar. Die Autoren entwickeln ein Analyseverfahren, in dem - unter der Annahme, die Fehler eines einzelnen Beobachters seien korreliert, nicht jedoch die zwischen verschiedenen Beobachtern - sich die Korrelationen zwischen den wahren Werten auf verschiedenen Dimensionen, die Reliabilitäten jedes Beobachters auf jeder Dimension und die Korrelationen zwischen den Urteilsfehlern des einzelnen Beobachters berechnen lassen. Es interessiert in diesem Ansatz also nicht die psychologische Ursache für meist wohl überhöhte Korrelationen zwischen verschiedenen Variablen, die am gleichen Beobachtungsgegenstand erfaßt werden; vielmehr wird gezeigt, unter welchen Annahmen über den Beobachter man dennoch die „wahren“ Zusammenhänge bestimmen könne.

3.2.5 Brunswiks probabilistischer Funktionalismus: Beobachtung als Leistung

Als ein Beispiel für die Ansätze der kognitiven Psychologie, die das Zustandekommen von Wahrnehmungsurteilen detailliert beschreiben, erwähnen wir kurz Brunswiks Modell (Brunswik 1952, 1955, 1956; als gute Einführung: Postman & Tolman 1959; wir folgen weitgehend dieser Darstellung). Brunswik unterscheidet die einer zentralen und motorischen Reaktion vorhergehenden Bedingungen als Reize mit unterschiedlicher Entfernung vom Organismus. Er betrachtet zum einen distale Reize, Gegebenheiten der Umwelt, mit denen der Organismus keinen unmittelbaren Kontakt hat, zum anderen proximi-

male Reize, die an den Grenzen des Organismus, beispielsweise als Netzhautbild gegeben sind.

Brunswik untersucht nun drei funktionale Beziehungen. Als erste sei die zwischen distalen und proximalen Reizen betrachtet, die dann als cues und als proximale Wirkungen erscheinen. Einige Hinweisreize variieren mit größerer Wahrscheinlichkeit systematisch als Funktion von distalen Reizänderungen. Die Stärke der Kovariation zwischen proximalem Hinweisreiz und distalem Merkmal definiert die ökologische Validität dieses Hinweisreizes hinsichtlich jenes Merkmals. Die ökologischen Validitäten beschreiben die ererbte und erlernte - u. U. auch im Beobachtungstraining - Einbettung des Beobachters in seine Umwelt. Sie zu kennen ist für Schlüsse aus Beobachtungen auf das Beobachtete wesentlich. Die zweite Funktion beschreibt als Beziehung zwischen proximalen Wirkungen und zentralen, perzeptiven Reaktionen den Gebrauch, den der Beobachter von Hinweisreizen macht. Um sich an distale Gegebenheiten anzupassen, gebraucht der Organismus die ihm verfügbaren proximalen Hinweisreize. Da diese nur begrenzte ökologische Validität haben, muß der Organismus - und, vielleicht bewußter, der Beobachter - eine Regel finden, wie er verschiedene Hinweisreize gewichten und kombinieren soll, um die distalen Gegebenheiten möglichst richtig abzuschätzen. Die dritte Beziehung, zwischen distalen Merkmalen und perzeptiver Reaktion, wird als funktionale Validität bezeichnet. Sie quantifiziert das Ausmaß, in dem der Organismus die Wahrnehmung der Umwelt geleistet hat.

Bei der Auseinandersetzung mit einer Umwelt kann man demnach Beobachtung als Leistung in zweifacher Hinsicht untersuchen: als Transformation von distalen in proximale Reize und als Integration verschiedener Hinweisreize zu einer kognitiven Reaktion. Um die Leistung eines Organismus einschätzen zu können, muß man ihn in seinem natürlichen Habitat untersuchen, denn dort treten die Hinweisreize in jener Kovariation auf, an die der Organismus seine Anpassung vollzogen hat. Werden die Hinweisreize, wie oft bei Wahrnehmungsexperimenten im Labor, isoliert und voneinander unabhängig variiert, dann müssen die Erkenntnisse über Teile des Systems kein zutreffendes Bild vom Gesamtsystem ergeben. Ein im Sinne Brunswiks repräsentativer Versuchs- und Beobachtungsplan sollte daher eine unverfälschte Stichprobe von Situationen aus der Ökologie des Organismus ziehen (Anwendungen der Brunswikschen Gedanken in der Lerntheorie unter dem Stichwort *multiple-cue probability learning task* oder *probabilistic concept identification task*, in der Diagnostik s. Lür & Kluck in diesem Band; formale Weiterentwicklungen z.B. bei Castellan 1973, Steward 1976).

3.3 Die Wahl von Beobachtungseinheiten durch Beobachter

Im Artikel über die wissenschaftliche Beobachtung wurde bereits die Frage angesprochen, wie die *Beobachtungseinheit* vom Forscher zu definieren und dem Beobachter vorzugeben sei. Wir wiesen darauf hin, daß der Forscher keine vollständige Kontrolle darüber haben kann, wie der Beobachter seine Beobachtungseinheiten festlegt. Dann ist es also eine empirische Aufgabe, die tatsächlich gewählten Einheiten zu bestimmen und die Regeln zu finden, wie Einheiten zustande kommen. Mit dieser Frage, wie - in welche Einheiten - Beobachter das Beobachtete gliedern haben sich in letzter Zeit Newton und seine Mitarbeiter befaßt (Newton 1973, 1976; Newton & Engquist 1976; Newton & Rindner 1979; Newton et al. 1977). Nach Newton gliedern Beobachter den Verhaltensstrom, indem sie aufeinanderfolgende Definitionspunkte wählen. Diese points of definition, auch breakpoints genannt, sind Stellen, die der Wahrnehmende als jene identifiziert, an denen sich Handlungen ereignet haben.

Beobachter scheinen diese Punkte so auszuwählen, daß sie eine Zusammenfassung der Information darstellen, die sie aus der Beobachtung der Handlungsabfolgen gewonnen haben. Newton (1973) geht davon aus, daß Wahrnehmende aktiv die Informationsaufnahme bei der Beobachtung kontrollieren, indem sie eine größere oder kleinere Anzahl von Definitionspunkten wählen, also ihre Einheit der Segmentierung größer oder kleiner festlegen. Wie kann man nun nachweisen, daß „observational units“ existieren, und wie kann man einzelne konkrete Beobachtungseinheiten identifizieren? Wenn der Strom des beobachteten Verhaltens ununterscheidbar kontinuierlich erschiene, könnte weder der gleiche Beobachter bei wiederholten Gelegenheiten noch verschiedene Beobachter am gleichen Material übereinstimmend Markierungspunkte feststellen. Die Untersuchungen von Newton und Mitarbeitern zeigen jedoch hohe intraindividuelle Konsistenz und hohes interrater-agreement, wenn Verhalten in Einheiten unterteilt werden sollte. Die Vpn mußten Grenzen bestimmen, durch die eine „behavior unit“ von einer anderen getrennt wird (Instruktion: „... press the button whenever, in your judgment, one meaningful action ends and a different one begins“). Ferner konnte diese Forschergruppe zeigen, daß diese Definitionspunkte andere Merkmale aufwiesen als willkürlich aus der Verhaltenssequenz herausgegriffene andere Punkte. Wurden kurze Segmente aus Filmen herausgeschnitten, so konnten die Auslassungen richtiger entdeckt werden, wenn Definitionspunkte entfernt worden waren als bei anderen Punkten. Wurden Bilder verwendet, die Definitionspunkten entsprachen, so ließ sich aus ihnen das Geschehen leichter und zutreffender interpretieren als aus anderen Bildern, und die Reihenfolge der Bilder ließ sich für Definitionspunkte eher richtig bestimmen. Definitionspunkt-Bilder werden auch besser im Gedächtnistest wiedererkannt. Diese Punkte, so schließen die Autoren aus diesen Befunden, stellen kein Artefakt dar, sondern stützen die

Annahme, Verhalten werde als in Einheiten gegliedert wahrgenommen und geschildert. Definitionspunkte werden als Reize mit besonders hoher Information für den Beobachter aufgefaßt. Mit dem Signal, daß eine neue Einheit beginnt, ist auch Information verbunden, wie es weiter geht. Für die Antizipation oder Vorhersage des neuen Geschehens wird durch die Information im Definitions-Reiz eine Basis geschaffen, die im Verhalten der Beobachteten vor dem Auftreten dieses Punktes nicht vorhanden war. Für das Verstehen von Verhalten wäre demnach notwendig, die Information zu kennen, die durch die Definitionspunkte geboten wird. Für vertraute Verhaltenseinheiten, die in sich relativ abgeschlossen waren, weil sie z.B. die Erledigung einer einfachen Aufgabe mit ersichtlichem Zweck erforderten, konnten Newton und Mitarbeiter ihre soeben geschilderte Auffassung gut belegen. Schwierigkeiten ergaben sich für Verhaltenssequenzen mit rhythmischen Wiederholungen, z.B. Tanz zu Rockmusik.

Inhaltlich lassen sich Definitionspunkte allgemein kennzeichnen als die Stellen, an denen sich gegenüber zuvor diskriminierten Zuständen Veränderungen ergeben. Vpn, die spontan oder auf Instruktion hin kleinere Einheiten kodieren, wählen nicht andere, sondern feiner aufgelöste Einheiten. D.h., es gibt eine Hierarchie der Einheiten; bei der größten Einheitenbildung (unitization) werden kleinere zusammengefaßt, und zwar nach übergeordneten Zielen, die dem Beobachteten zugeschrieben werden, Unterziele bestimmen die kleineren Einheiten. Kleine Einheiten werden spontan gewählt, wenn Verhalten irregulär, lose organisiert erscheint, was mit der Annahme übereinstimmt, Definitionspunkte hätten eine wichtige Funktion für die Vorhersagbarkeit. Bleibt das Verhalten gleichförmig, so werden allmählich größere, zeitlich längere Einheiten gewählt. Treten unerwartete Ereignisse ein, so kehrt der Beobachter zur Feingliederung zurück.

Wie auch Wilder (1978) stellten Newton und seine Mitarbeiter einen Trend fest, mit fortschreitender Beobachtungsdauer zunehmend größere Einheiten zu wählen. Anscheinend fühle sich der Beobachter nach einiger Zeit für seine Zwecke hinreichend informiert, und weitere Information werde dann nicht mehr beachtet, oder der Aufwand, sie zu verarbeiten, lasse nach. Die vom Beobachter gewählte Größe der Beobachtungseinheit bestimmt dann die obere Grenze, wieviel Information gewonnen und wie rasch Zufriedenheit mit dem eigenen Informationsstand erreicht werden kann.

In diesem Zusammenhang seien einige Befunde über die Beziehungen zwischen skaliertem Informationsmenge und klinischen Urteilen berichtet, die auf Beobachtungen (meist Filmen) basierten (Feger, 1972, Kap. 7). 1. Je länger die Darbietungsdauer eines Filmes über eine Person in Testsituationen, je ausgedehnter also die Möglichkeit zur Beobachtung war, als desto höher wurde die vorhanden geglaubte Informationsmenge eingestuft. Die Zunahme der Informationsmenge erwies sich als kurvilinear, negativ beschleunigt. 2. Es gibt über

verschiedene Zeitpunkte der Informationsaufnahme und über verschiedene zu beurteilende Merkmale konstante interindividuelle Unterschiede in der subjektiven Informationsmenge, über die ein Beobachter zu verfügen glaubt. 3. Beobachter, die sich besser informiert glauben, stufen Verhalten extremer ein; ein Zusammenhang mit Konfidenz - der subjektiven Sicherheit, richtig beurteilt zu haben - besteht nicht. 4. Beobachtertraining erhöht die Konfidenz.

3.4 Der Entstehungsprozeß von Beschreibungen

Alle Beobachtungen, alle Wahrnehmungen schlechthin, müssen wenigstens kurzzeitig im Gedächtnis gespeichert werden, damit sie berichtet werden können. Die typische Art, Beobachtungen zu berichten, ist die verbale Beschreibung. Ericsson & Simon (1980) fassen Beschreibungen von Beobachtungen (verbal reports) als Daten auf, d.h. sie fordern - um die Aussagefähigkeit der Beschreibungen abschätzen zu können - eine Erklärung der Prozesse, wie verbale Berichte zustande kommen, und von welchen Bedingungen sie abhängen. Wie man bei diesen Autoren erwarten darf, legen sie eine kognitive *Rahmentheorie des Entstehungsprozesses von Beschreibungen* vor, deren Grundannahmen sich folgendermaßen skizzieren lassen: Man muß wissen, was ein Beobachter im Kurzzeitgedächtnis gespeichert hat, denn das kann er *verlässlich* verbalisieren. Bezieht sich die Frage eines Forschers (oder eines Beobachters an sich selbst) jedoch auf Material, das nicht gespeichert ist, so muß er schlußfolgern, und Schlüsse dieser Art können falsch sein. Aufforderungen wie „Denken Sie laut“ ändern kognitive Prozesse nur dann, wenn durch sie eine Vp gezwungen wird, ihre Aufmerksamkeit auf Gegebenheiten zu richten, denen sie sich ohne eine solche Instruktion nicht zugewandt hätte.

Ericsson & Simon beginnen ihre Argumentation im einzelnen mit einem Hinweis auf die „schizophrene“ Behandlung verbaler Berichte durch Behavioristen. In der Forschung über Begriffsbildung beispielsweise werde als hartes Datum akzeptiert, wenn eine Vp mit ja oder nein die Frage beantwortet, ob die Vorlage eine Instanz des Konzeptes sei, aber nicht akzeptiert, wenn sie berichtet, sie prüfe die Hypothese, das gesuchte Konzept sei „ein kleiner, gelber Kreis“. Es fehle ein klares *Kriterium*, was denn nun ein akzeptabler Report sei und was nicht. Ein solches einfaches Kriterium geben nun auch Ericsson & Simon nicht. Sie wenden das Problem vielmehr so, daß ein Verwender von Berichten aus deren Entstehungsprozeß beurteilen muß, ob sie für seine Zwecke brauchbar sind. Die Frage ist dann also nicht mehr, ob eine Vp „wahrheitsgetreu“ berichtet, sondern vielmehr, wie sie dazu kommt, etwas Bestimmtes auszusagen.

Unterschiede im Entstehen von Berichten ergeben sich aus den Umständen, unter denen sie gewonnen wurden. Die Autoren klassifizieren diese Umstände

zunächst danach, ob es die hauptsächliche Aufgabe der Vp war, einen verbalen Bericht zu geben, oder ob die Beschreibung nur ein Nebenprodukt war, das bei der Bewältigung einer anderen, der eigentlichen Aufgabe anfiel. Die Verbalisierung kann weiter gleichzeitig mit der Aufgabenbewältigung geschehen oder retrospektiv. Dabei kann man der Vp gezielte Fragen stellen oder solche, die von ihr Verallgemeinerungen verlangen. Einige Einschränkungen für die Benutzung von Berichten ergeben sich sogleich aus dieser Klassifikation: Ist die Verbalisation nur Nebensache, können die Berichte unvollständig und für die Fragestellung des Forschers irrelevant sein. Steht die Verbalisation im Vordergrund, so besteht die Gefahr der Interferenz mit der Bewältigung der Versuchsaufgabe. Bittet man die Vp um generelle Aussagen, so wird sie um so mehr schlußfolgern statt berichten, je allgemeiner und umfassender ihre Aussage sein soll. Wird der Bericht retrospektiv verlangt, dann werden Vollständigkeit und Gültigkeit davon abhängen, in welchem Umfang bei der früheren Aufgabenbewältigung zufällig das den Forscher Interessierende gespeichert wurde.

Wie in der neueren Psychophysik wird unterschieden zwischen der inneren Repräsentation der gespeicherten Information und dem Bericht, und zwischen beiden findet der Prozeß der Verbalisation statt. Je nach der einer Vp gestellten Berichtaufgabe kann dieser Prozeß bedeuten, (1) daß noch nicht verbal gespeicherte Information in verbale Form recodiert werden muß, (2) daß Such- und Filterprozesse ablaufen müssen, wenn der Bericht bestimmte Anforderungen erfüllen soll, z.B. nur dieses, nicht aber jenes, oder in einer bestimmten Form zu schildern, und (3) daß Abstraktionen und Verallgemeinerungen vom Beobachter verlangt werden.

Ericsson & Simon klassifizieren gängige Methoden, wie Forscher von Personen Berichte erfragen (s. Tab. 2). Für die Unterscheidungen und Beispiele im einzelnen muß auf die Veröffentlichung verwiesen werden. Wesentlich für die Bewertung der Verfahren - es ergeben sich um so mehr Vorbehalte, je mehr man in der Tabelle nach unten und nach rechts geht - ist das Ausmaß, in dem Vpn statt sich zu erinnern, um Urteile gebeten werden, und je weniger Vpn sich tatsächlich erinnern können, weil die erforderliche Information nicht mehr im Kurzzeitgedächtnis ist, nur unvollständig dorthin aus dem Langzeitgedächtnis zurückgeführt werden kann, oder nie im Gedächtnis war.

Das methodische Vorgehen beim Gewinnen der Berichte ist nur ein Gesichtspunkt, nach dem deren Qualität beurteilt werden kann. Ein anderer ist, die psychologischen Bedingungen zu berücksichtigen, unter denen Wahrnehmungen zustande kommen und zu Berichten werden können. Die (schon bestehenden) kognitiven Strukturen und die bei einer Beobachtungsmöglichkeit ablaufenden Prozesse - wie Schwankungen der Aufmerksamkeit, Störungen durch die Umwelt, Automation der Vollzüge - bestimmen, was bewußt werden kann, und somit auch, worauf sich die Aufmerksamkeit richten kann. Das

Tabelle 2: Eine Klassifikation verschiedener Arten von Verbalisierungsaufgaben als Funktion des Zeitpunktes der Verbalisierung (Zeilen) und der Abbildung von beachteter auf verbalisierte Information (Spalten); Tabelle 1 bei Ericsson & Simon (1980, S. 244)

| | | Beziehung zwischen beachteter und verbalisierter Information | | |
|--|------------------------------|--|--|-------------------------------------|
| | | Mittelbare Verarbeitung | | |
| Zeitpunkt der Verbalisierung | Direkt eins zu eins | viele zu eins | unklar | keine Beziehung |
| Aufmerksamkeit der Information zugewendet | Laut Sprechen Laut Denken | | | |
| | | mittelbare schlußfolgernde und generative Prozesse | | |
| Information befindet sich noch im Kurzzeitgedächtnis | gleichzeitiges Erfragen | | | |
| Nach Abschluß der aufgabenbezogenen Prozesse | Retrospektives Erfragen | Erbitten allgemeiner Berichte | Erfragen hypothetischer Zustände | Erfragen allgemeiner Zustände |

wiederum legt fest, was wie gespeichert wird und abgerufen werden kann, wenn ein Bericht gewünscht wird. Alle Faktoren, die fördernd oder hemmend auf Speichern und Abrufen wirken, beeinflussen deshalb auch positiv oder negativ eine zutreffende Wiedergabe (kognitiver) Gedächtnisinhalte. So wird die These begründet, es gebe Bereiche, in denen im allgemeinen und besonders nach einer Einübungsphase die Prozesse so schnell und ohne Aufmerksamkeitszuwendung ablaufen, daß ein zutreffender Bericht nicht möglich sei. Das wird beispielsweise für Enkodierungsprozesse bei der Wahrnehmung, beim Retrieval aus dem Gedächtnis und bei motorischen Prozessen angenommen.

3.5 Verhaltenseinschätzung (behavioral assessment)

Im folgenden schildern wir aus methodischer Sicht einige Ergebnisse der Studien über Verhaltensbeobachtung, die im Zusammenhang mit verhaltensthera-

peutischen Zielen durchgeführt wurden (Überblick Ciminero 1977). Wir möchten aus dieser Forschung jene Arbeiten zusammenfassen, die Einflüsse auf Beobachtungen berichten, welche sich aus der Beobachtungssituation, aus dem spezifischen beobachteten Verhalten und aus Merkmalen der Beobachter ergeben. Verhaltensbeobachtung ist für Auswahl, Steuerung und Bewertung der Therapie erforderlich. Nach dem Selbstverständnis der Forscher in diesem Bereich unterscheidet sich ihr Vorgehen von der traditionellen Persönlichkeitsforschung und der dort üblichen Verhaltensbeobachtung u.a. dadurch, daß sie nicht versuchen, zugrundeliegende Einheiten, Faktoren, der Persönlichkeitsstruktur aus dem Verhalten zu erschließen, vielmehr versuchen, die Bedingungen zu erfassen, die in spezifischen Situationen ganz bestimmte Verhaltensweisen auslösen (ausführlich in Gottfried & Kent, 1972). Peterson (1968, S. 114) gibt eine konzise Zusammenfassung:

„The central features of the method are (1) systematic observation of the problem behavior to obtain a response frequency baseline, (2) systematic observation of the stimulus conditions following and/or preceding the behavior, with special concern for antecedent discriminative cues and consequent reinforcers, (3) experimental manipulation of a condition which seems functionally, hence causally, related to the problem behavior, and (4) further observation to record any changes in behavior which may occur.“

Charakteristisch ist also, daß dem Anspruch nach der Bereich dessen, was beobachtet wird, hier wesentlich erweitert ist: die Ereignisse in der Umwelt, die dem gezeigten Verhaltenssegment vorausgehen und nachfolgen, werden ausdrücklich (und nicht nur in der Deutung durch den Beobachter implizit) hinzugenommen, und man greift in diese Umwelt gezielt ein, um die Gegebenheiten herauszufinden, die das Verhalten beeinflussen.

Zu den Methoden der Verhaltenseinschätzung werden außer den hier nicht besprochenen physiologischen Messungen Berichte der untersuchten Person (self report, self recording, hier Selbstberichte genannt) sowohl des eigenen Erlebens als auch insbesondere des eigenen Verhaltens und die Fremdbeobachtungen, gewöhnlich durch den Therapeuten, gerechnet. Zu den *Selbstberichten* zählen Interviews des Patienten durch den Therapeuten über das Verhalten des Patienten (behavioral interviews), die Anwendung von Verhalten-Fragebögen (Surveys and inventories) und die Anwendung von Registrierverfahren durch den Patienten (self-monitoring procedures), womit aus ethischen oder praktischen Gründen nicht direkt beobachtbare Verhaltensweisen, aber auch Erleben - wie der Wunsch, sich eine Zigarette anzuzünden - erfaßt werden sollen. *Fremdbeobachtungen* werden danach unterschieden, ob sie in der natürlichen, alltäglichen Umgebung (naturalistic setting) der untersuchten Person angestellt werden oder im Labor, und zwar dort in Situationen, die dem Alltag des Patienten nachgebildet sind (contrived analogue settings). Wir behandeln drei methodische Themenbereiche, die für alle Beobachtungsarten

relevant sind und vergleichsweise ausführlich untersucht wurden, und zwar Fragen der Reliabilität, der Reaktivität und der Auswirkung von Beobachtererwartungen.

3.5.1 Die Verlässlichkeit von Selbstberichten und Fremdbeobachtungen

Da Beobachter meist für die Erfassung einer *bestimmten Art von* Verhaltensweisen geschult werden, und ein häufiges Erfolgskriterium der Schulung die Reliabilität ist, stellt sich die Frage, ob eine einmal erreichte Verlässlichkeit erhalten bleibt, wenn sich - wie etwa bei Schizophrenie und manisch-depressiven Störungen - das Verhalten der Beobachteten deutlich ändert. Redfield & Paul (1976) berichten gleichbleibende Reliabilität. Kent & Foster (1977) berichten, in einer neuen Umgebung fiel die Übereinstimmung zwischen Beobachtern zunächst ab und stieg dann wieder an. Verschiedene Verhaltensarten, so wird in der Literatur häufig vermutet, jedoch selten geprüft, sollten sich unterschiedlich reliabel beobachten lassen. Jones et al. (1975) nehmen an, die Verlässlichkeit sei eine Funktion der Komplexität des Verhaltens, wobei sie Komplexität durch das vorgegebene Kategoriensystem als Anzahl der Unterscheidungen definieren, die ein Beobachter bei der Datenaufnahme treffen muß. Jones und Mitarbeiter fanden höhere Komplexitätswerte in Sitzungen kurz vor und nach Sitzungen mit Reliabilitätsprüfung. Die Auswirkung der „Komplexität“ des Kategoriensystems untersuchten Mash & McElwee (1974), indem sie verbales Material nach zwei jeweils erschöpfenden Kategoriensystemen beurteilen ließen. Das erste enthielt nur vier Kategorien, beim zweiten waren diese Kategorien noch jeweils unterteilt in zwei weitere. Höhere Übereinstimmung fand sich für das System mit weniger Kategorien. Schulung an einem kurzen Videotape, das gut vorhersagbares Interaktionsverhalten zeigte, führte zu einer geringeren Übereinstimmung bei der Beurteilung eines später gezeigten Bandes als Training an einem Tape mit wenig vorhersagbarer Interaktion (bestätigt in Mash & Makohoniuk 1975).

Gut belegt ist der positive Einfluß der Erwartung der Beobachter, hinsichtlich der Reliabilität kontrolliert zu werden, auf eben diese Reliabilität (Reid 1970, Romanczyk et al. 1973). Taplin & Reid (1973) ließen 18 undergraduate Studenten die Interaktion zwischen Mutter und Kindern auf Videotape beobachten. Nach einer Vorschulung wurden sie in drei Gruppen eingeteilt, die sich durch folgende Informationen unterschieden: I Die Versuchsleiter würden die Reliabilität nicht prüfen. II Jeder beliebige Teil des Protokolls könne später vom VI geprüft werden. III Es gäbe Stichproben (spot checks), die zuvor als solche angekündigt würden. Unter Bedingung II zeigte sich die höchste Übereinstimmung, ein mittleres Maß unter I und geringstes unter III, außer für jene Protokollteile, für die eine Kontrolle zuvor angekündigt wurde. Taplin & Reid (1973, S. 553) schlußfolgern:

„The major thrust of the findings . . . is to point to the danger of assuming that unmonitored observational data are reliable. Instead of being established a priori, the reliability of observational data ought to be empirically demonstrated. Without such demonstrations, the accuracy of observational data can only be a matter of faith.“

Beobachter können die Interpretation eines Kodiersystems zum einen während der Schulung auch im direkten Vergleich ihres Verhaltens mit dem anderen Beobachter lernen. Zum anderen können sie aber auch während der Anwendungsphase indirekt lernen, etwa über mitgeteilte Reliabilitätskoeffizienten oder inter rater agreement. Offensichtlich passen sich länger zusammenarbeitende Beobachter untereinander im Gebrauch von Kodiersystemen an. Dafür sprechen sowohl Befunde, nach denen im Verlauf längerer Beobachtungen die Übereinstimmung in einem Beobacherteam anwächst als auch der Befund, daß die Übereinstimmung in einem Team meistens größer ist als zwischen Beobachtern aus verschiedenen Teams, selbst wenn alle Beobachter zuvor durch die gleiche Schulung gegangen sind (vgl. Kent & Foster 1977).

Wir wenden uns nun der *Reliabilität von Selbstberichten* zu. Eine typische Aufgabe beim self recording besteht beispielsweise für einen Patienten darin zu registrieren, wie oft er sein eigenes Gesicht berührt. Die Reliabilität, definiert als Übereinstimmung mit einem externen Beobachter, ist in der Regel dann relativ hoch ($r = .8$), wenn die Patienten wußten, daß ihre Zuverlässigkeit geprüft wurde, und wenn ihre Zuverlässigkeit immer wieder belohnt wurde (Fixen et al. 1972, Lipinski & Nelson 1974, Lipinski et al. 1975, Nelson et al. 1975, Taplin & Reid 1973). Vergleicht man die Reliabilität von Selbstberichten mit der von Daten externer Beobachtungen, so zeigen sich Selbstberichte oft als weniger zuverlässig. Simkins (1971) erwägt dafür folgende Gründe: Externe Beobachter erhalten ein besseres oder anderes Training als Selbstberichter, sie stehen unter verschiedenartigen Belohnungskontingenzen und benutzen verschiedene Kriterien für Beurteilung und Registrierung. Bei Selbstberichtern könnten Reaktionen eine Registrierung behindern, die es für externe Beobachter nicht gibt, beispielsweise könnte das zu registrierende Interaktionsverhalten die Teilnehmer, nicht aber externe Beobachter ablenken (s. Cavior & Marabotto 1976).

Für eine *Zusammenfassung* zahlreicher, auch hier nicht berichteter Befunde erscheint es dem gegenwärtigen Methodenstand in diesem Bereich angemessen, wenn man bei der Diskussion der Reliabilitätsfrage davon ausgeht, daß ein Training die Beobachter schon bis zu einem Minimum an Verlässlichkeit führt, und dann nach Bedingungen gesucht wird, welche diesen einmal erreichten Stand beeinflussen können. Die Befunde sprechen dafür, daß Faktoren, die zwischen Schulung und Einsatz liegen, und solche, hinsichtlich derer sich Situationen der Schulung von denen des Einsatzes unterscheiden, deutliche Unterschiede in der Reliabilität bewirken. Zu diesen Faktoren gehören insbe-

sondere Erwartungen der Beobachter, die deshalb später in einem eigenen Abschnitt besprochen werden.

3.5.2 *Reaktivität*

Wir definieren Reaktivität (bewußt quantitativ) als das Ausmaß, in dem ein Verfahren das Phänomen ändert, das mit ihm untersucht wird. Wir schildern hier nicht allgemein die umfangreiche Literatur zu Versuchsleiterereffekten (z.B. Rosenthal 1966, 1976; Rosenthal & Rosnow 1969) oder zu nichtreaktiven Meßverfahren (Webb et al. 1966), obwohl vermutlich vieles, das sich in dieser Literatur auf strikt experimentelles Vorgehen bezieht, auch für Beobachtungsstudien relevant sein kann. Vielmehr beschränken wir uns hier auf Arbeiten, die Reaktivität speziell bei Beobachtungen zum Zweck der Verhaltenseinschätzung untersucht.

Zunächst befassen wir uns mit der möglichen Reaktivität von Beobachtungen, die durch *externe Beobachter* angestellt werden. Nach Kent & Foster (1977) können folgende Faktoren der Reaktivität entgegenwirken: (1) Gewöhnung an die Beobachter, dies wurde allerdings in einer Studie gezeigt, in der die Gewöhnung über mehrere Wochen hinweg möglich war, bevor die Beobachtung selbst begann; (2) ein Beobachtungsplan mit häufigen und nicht vorhersagbaren Beobachtungsphasen; (3) Verzögerung des feedbacks über Beobachtungsergebnisse an die Beobachteten. Auch unter sonst gleichen Beobachtungsbedingungen und an den gleichen beobachteten Personen kann man offensichtlich nicht bei allen *Variablen* in gleichem Ausmaß Reaktivität erwarten (Roberts & Renzagha 1965).

Beim typisch experimentellen Vorgehen interagieren in der Regel nur Versuchsperson und Experimentator, der oft mit dem Forscher identisch ist. Wenn jedoch Beobachter und Forscher nicht identisch sind, etwa wenn Eltern über ihre Problemkinder berichten sollen, kann sich Reaktivität nicht nur im Verhalten der Beobachteten zeigen und i. w. S. auch beim Forscher, sondern auch beim Beobachter. Johnson & Lobitz (1974) analysieren diese Beobachtungssituation mit Hilfe des Konzeptes der *demand characteristic*: Welche spezifischen Situationsmerkmale ergeben sich für alle Personen in einer Situation mit einem Beobachter aus dessen Anwesenheit? Johnson & Lobitz nehmen beispielsweise für Eltern an, die ihr Kind als therapiebedürftig beschrieben haben, für sie ergäbe sich bei der Beobachtung in der Familie vielleicht unbewußt der Wunsch, ihr Kind tatsächlich als problematisch erscheinen zu lassen. Die Autoren baten Eltern in Familien ohne Problemkinder, ihre Kinder entweder als „brav“ oder als „ungezogen“ erscheinen zu lassen. Johnson & Lobitz konnten entsprechende Unterschiede im Verhalten von Eltern und Kindern nachweisen und argumentieren, wenn es unterschiedliche demand

characteristics gäbe, seien Personen in der Lage, darauf unterschiedlich zu reagieren und somit die Validität von Beobachtungen zu gefährden.

Im letzten Jahrzehnt hat man sich besonders um Techniken, aber auch um Versuchspläne bemüht, die eindeutige Schlüsse über Reaktivität zulassen. Die Verwendung von Kleinst(radio)sendern wurde mehrfach untersucht, z.B. von Moos (1968) an psychiatrischen Patienten. Moos verglich das Verhalten bei eingeschaltetem und bei ausgeschaltetem Sender. Insgesamt zeigte sich wenig Reaktivität, ihr Ausmaß war jedoch abhängig von der jeweiligen Situation und zeigte sich stärker bei den stärker gestörten Patienten. Johnson & Bolstad (1975) schnitten die Interaktion von Familienmitgliedern auf Tonband mit und fanden keine Unterschiede bei Anwesenheit oder Abwesenheit eines externen Beobachters. Goldfried & Linehan (1977) wenden in ihrem hier zugrundegelegten Überblick gegen beide Studien das Fehlen einer Kontrollgruppe ein, die eindeutig nicht beobachtet wurde, zumindest nach Meinung der betroffenen Vpn. In der Studie von Moos wußten die Vpn, daß sie auch dann beobachtet wurden, wenn sie den Sender nicht trugen, und in der Untersuchung von Johnson & Bolstad lief das Tonband auch dann, wenn kein Beobachter anwesend war. Diesen Einwand berücksichtigen - vielleicht ethisch bedenklich - Hagen et al. (1975). Sie arbeiteten mit einem verborgenen Mikrophon, das ständig die Interaktion von Personal und Patienten aufnahm, und verglichen das Verhalten bei An- und Abwesenheit eines Beobachters, ohne Unterschiede zu finden. Diesen Versuchsplan entwickelten Johnson et al. (1976) weiter, die Kinder mit einem Sender versahen, der das verbale Verhalten auf Tonband übertrug. Das Tonband wurde nach einem time sampling Schema ein- und ausgeschaltet, das weder dem Kind noch den übrigen Familienmitgliedern bekannt war.

Während die Daten externer Beobachter beim behavior assessment kaum durch Reaktivität beeinflusst zu werden scheinen, sind Reaktivitätseffekte bei *Selbstberichten* so stark - sie reduzieren meistens die problematische Verhaltensweise - daß self monitoring inzwischen regelmäßig, wie Goldfried & Linehan berichten, als therapeutische Technik eingesetzt wird. Erklärt wird dies damit, daß der Selbstbericht zu einem Spezialfall von feedback wird, das das Verhalten ändert. Statt auf feedback legen Ciminero et al. (1977, p. 208) mehr Wert auf die Tatsache, daß „. . . the presence of an observer alters the usual stimulus situation, thereby producing behavior changes“ und „. . . when an individual begins to self-observe his own behavior, there is also a change in the usual stimulus situation.“ Es wäre interessant zu erfahren, worin genau die Veränderung besteht, und warum sie in beiden Situationstypen so verschieden ist, daß so starke Unterschiede in der Reaktivität auftreten.

Ciminero et al. geben einen Überblick über verschiedene *Versuchspläne*, die zur Analyse von Reaktivität bei Selbstberichten verwendet wurden. Als Beispiel für eine *Einzelfallstudie* sei die Arbeit von Maletzky (1974) erwähnt:

Fünf Patienten beobachteten ihr störendes Verhalten, das während der Selbstbeobachtung zurückging. Eine Unterbrechung des Selbstberichts ließ die Frequenz wieder steigen; erneute Selbstbeobachtung führte zu weiterem Abfall der Häufigkeit. Wie in anderen Untersuchungen dieser Art fehlt auch bei Maletzky eine Angabe zur Verlässlichkeit der Selbstbeobachtung. Diese lassen sich in einer Studie von Herbert & Baer (1972) aus den Angaben externer Beobachter in den Familien der untersuchten Mütter und Kinder gewinnen, die dadurch zu einem *within-subject experimental design* wurde, daß eine Umkehrphase (reversal) eingeführt wurde, in der die Registrierung des eigenen Verhaltens durch die Mütter unterbrochen wurde. Einen Vergleich zwischen verschiedenen Vpn, also ein *between-subjects experimental design* finden wir in der Arbeit von Johnson & White (1971). Sie untersuchten drei Gruppen von College-Studenten. Die erste Gruppe beobachtete selbst ihre Studienaktivitäten, die zweite ihre Zeit, die sie mit Freundinnen verbrachten, die dritte Gruppe wurde nicht zu Selbstberichten angehalten. Abhängige Variable war die wöchentlich erhobene Benotung der Studienleistung. Darin zeigten sich signifikante Unterschiede zwischen der ersten und dritten Gruppe zugunsten der ersten, während sich die zweite Gruppe nicht überzufällig von den beiden anderen unterschied. Das Erstellen von Selbstberichten über eine bestimmte Verhaltensweise (dating) kann demnach auch zu Effekten auf eine andere Verhaltensweise (Studiengewohnheiten) führen, also generalisieren, und dann die abhängige Variable indirekt beeinflussen.

Die Stärke der Reaktivität hängt von der Bewertung des beobachteten Verhaltens und der Motivation, es zu ändern ab: Wird das Verhalten positiv eingeschätzt, steigt seine Frequenz, bei negativer Bewertung sinkt sie (Brodin et al. 1971, Kazdin 1974, Cavior & Marabotto 1976). Kanfer (1970) erklärt die Wirkung der Reaktivität als einen Prozeß mit drei Stadien: Im ersten Schritt erfolgt die Selbstbeobachtung, im zweiten bewertet die Person selbst ihr Verhalten nach ihren eigenen Normen, danach führt im dritten Schritt eine positive Bewertung zum Anstieg, die negative zum Abfall der Auftretenshäufigkeit. Diese von Ciminero et al. als „mediational explanation“ - wegen der Vermittlung über die Bewertung - bezeichnete Erklärung wird von ihnen mit der „operant explanation“ von Rachlin (1974) kontrastiert. Nach Rachlin haben die Folgen, die man sich selbst verschafft, nicht die Funktion von Verstärkern, sondern von Hinweisreizen. Diese Hinweisreize wirken als Signale für das, was schließlich an externen Konsequenzen zu erwarten sei, wenn man sich auf ein bestimmtes Verhalten einläßt. Selbstberichte von eigenem Verhalten schliesse dann tendenziell zumindest auch Beobachten der Folgen dieses eigenen Verhaltens mit ein.

3.5.3 Einflüsse bestehender Erwartungen der Beobachter

Einige der in den vorausgegangenen Abschnitten geschilderten Ergebnisse lassen sich integrieren, wenn man annimmt, daß Beobachter *Erwartungen entwickeln*, denen sie sich in ihrem Beurteilungsverhalten anpassen. Kent & Foster (1977) bezeichnen die Kovariation von geäußelter Erwartung über vermutlich auftretendes Verhalten und vom gleichen Beobachter berichtetem Verhalten als *expectation bias* (auch als Rosenthal-Effekt bekannt). Die Erwartungen der Beobachter beziehen sich unter anderem auf Art und Ausprägung des vermutlich auftretenden Verhaltens der zu beobachtenden Personen, auf Beurteilungsverhalten etwaiger Mitbeobachter und auf Kontrollverhalten der Forscher. Tritt höhere Verhaltensvariation des Beobachtungsobjektes auf als die Beobachter erwarten, oder zeigt sich das zu Beobachtende in einer für den Beobachter neuen Umgebung, so könnten Kategoriensystem und Beobachtungsinstruktion neu interpretiert werden müssen (was die hierbei gefundenen, vorübergehenden Reliabilitätsverminderungen erklären kann). Lernt der Beobachter, was er von Mitbeobachtern zu erwarten hat, so kann er sich anpassen, was zu der meist erwünschten Steigerung der Übereinstimmung führt. Auf Motivation, Aufmerksamkeit und Konzentration schließlich wirkt eine Kontrollerwartung zugunsten verbesserter Verlässlichkeit, während die oft erhöhten Anforderungen an Gedächtnis, Aufmerksamkeit und ähnlichen kognitiven Leistungen, welche die Einsatzsituation oft kennzeichnen, auch die Verlässlichkeit der Beobachtungen beeinträchtigen können.

Schon 1961, vor Rosenthal also, zeigten Azrin et al., daß Beurteilungen, die Beobachter über Meinungsäußerungen abgaben, mit ihren Erwartungen systematisch variierten, welche Meinung von den Vpn geäußert werden würde. Auf welche Schwierigkeiten Studien zum Rosenthal-Effekt stoßen, wird aus der Dissertation von Kent (1972, hier nach Kent & Foster 1977) deutlich: Zunächst wurden alle 40 Vpn gemeinsam geschult, und zwar 40 Stunden lang, verteilt über eine Periode von fünf Wochen anhand von Aufzeichnungen mit Videotape. Wie sich zeigt, ist gemeinsame Schulung wichtig, um nicht zu Unterschieden zwischen den Beobachtern zu kommen, die schon vor ihrer Zuteilung zu Versuchsbedingungen bestehen und Vergleiche zwischen den Versuchsbedingungen konfundieren. Gegen Ende der Schulung stellte sich heraus, daß sich die Übereinstimmung von $r = 0,60$ zwischen den Beobachtern nicht mehr verbessern ließ - zur Berechnung wurde jeder Beobachter mit einem aus den übrigen 39 zufällig herausgegriffenen verglichen. Danach wurden die Beobachter in zufälligen Gruppen zu je 5 Personen den 8 Versuchsbedingungen mit unterschiedlichen Ergebniserwartungen zugeteilt. Schon nach drei Tagen stieg die mittlere Verlässlichkeit innerhalb dieser Fünfergruppe auf 0,7. Zugleich entwickelten sich, und zwar vor jeder experimentellen Variation Unterschiede zwischen den Gruppen. Dieser Effekt wurde als „consensual observer drift“ (s. Johnson & Bolstad 1975) bezeichnet und führt Kent & Foster (1977, S. 283) zu der Warnung:

„... any study in which groups of observers collect data only within a particular class-room or treatment condition may badly confound differences in use of the behavioral code with the variables under investigation.“

Um Schwierigkeiten mit Vortreatment-Unterschieden auszuräumen, gingen Kent et al. (1974) so vor: Zehn Paare von Beobachtern (undergraduates) wurden so lange trainiert, bis jeder mit seinem Partner hinreichend hohe Übereinstimmung zeigte. Dann wurde at random je ein Mitglied eines Paares einer der beiden Versuchsbedingungen zugeteilt und beobachtete getrennt Videotapes, die angeblich teils das störende Verhalten eines Kindes vor, teils während der Behandlung zeigten. Der Versuchsleiter in der einen Experimentalbedingung informierte die Beobachter dahingehend, das Störverhalten werde während der Behandlung abnehmen, während in der anderen Bedingung - wahrheitsgemäß - informiert wurde, es werde sich keine Änderung zeigen. In der Auswertung ergab eine globale Frage, ob Veränderung eingetreten sei, den Rosenthal-Effekt; wertete man jedoch die detaillierten Verhaltensprotokolle zu neun operational definierten Kategorien aus, so zeigten sich keine erwartungsbedingten Unterschiede. Kent & Foster halten es für möglich, daß Erwartungseffekte insbesondere bei globalen Bewertungen, nicht jedoch bei spezifischen Verhaltensbeschreibungen auftreten. Allerdings zeigten sich in einer Studie von O'Leary et al. (1975) auch bei relativ explizit definierten Kategorien Effekte von Äußerungen von Versuchsleitern, die eine Erwartung bewirken sollten.

Literatur

- Ach, N. 1905. über die Willenstätigkeit und das Denken. Göttingen: Vandenhoeck & Ruprecht.
- Adair, J. & Spinner, B. 1979. Subjects' access to cognitive processes: Demand characteristics and verbal report. Unpublished manuscript, University Manitoba.
- Azrin, N. H., Holz, W., Ulrich, R. & Goldiamond, I. 1961. The control of the content of conversation through reinforcement. *Journal of experimental analysis of behavior*, **4**, 25-30.
- Bacon, F. 1974. *Neues Organ der Wissenschaften* (1620). (Dtsch. von A. T.Brück). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Bakan, D. 1959. A reconsideration of the problem of introspection. *Psychological Bulletin*, **51**, 105-118.
- Barker, R. 1968. *Ecological psychology: concepts and methods for studying the environment of human behavior*. Stanford: Stanford Univ.Press.
- Barker, R. & Schoggen, P. 1973. *Qualities of community life*. San Francisco: Jossey-Bass.

- Bern, D. J. 1965. An experimental analysis of self-persuasion. *Journal of Experimental Social Psychology*, 1, 199-218.
- Bern, D. J. 1966. Inducing belief in false confessions. *Journal of Personality and Social Psychology*, 3, 707-710.
- Bern, D. J. 1972. Self-perception theory. In: Berkowitz, L. (ed.): *Advances in experimental social psychology*. Vol.6, New York: Academic Press, 1-62.
- Berman, J. S. & Kenny, D. A. 1976. Correlational bias in observer ratings. *Journal of Personality and Social Psychology*, 34, 263-273.
- Berman, J. S. & Kenny, D. A. 1977. Correlational bias: Not gone and not to be forgotten. *Journal of Personality and Social Psychology*, 35, 882-887.
- Birdwhistell, R. L. 1970. *Kinesics and context*. Philadelphia: Univ. of Philadelphia Press.
- Block, J. 1977. Correlational bias in observer ratings: Another perspective on the Berman and Kenny study. *Journal of Personality and Social Psychology*, 35, 873-880.
- Blumenthal, A. L. 1975. A reappraisal of Wilhelm Wundt: *American Psychologist*, 30, 1081-1088.
- Borgatta, E. F., Cottrell, L. S. & Mann, J. H. 1958. The spectrum of individual interaction characteristics: An inter-dimensional analysis. *Psychological Reports*, 4, 279-319.
- Braun, P. 1978. Verhaltenstherapeutische Diagnostik. In: L. Pongratz (ed.): *Klinische Psychologie (Handbuch der Psychologie. Bd. 8/2)*. Göttingen: Hogrefe, 1649-1725.
- Brodén, M., Hall, R. V. & Mitts, B. 1971. The effect of self-recording on the classroom behavior of two eighth-grade students. *Journal of Applied Behavior Analysis*, 4, 191-199.
- Brown, A. L. 1978. Knowing when, where, and how to remember: A problem of metacognition. In: R. Clark (ed.): *Advances in instructional psychology*, Vol. I. Hillsdale, N.J.: Erlbaum.
- Bruner, J. S. & Tagiuri, R. 1954. The perception of people. In: Lindzey, G. (ed.): *Handbook of social psychology*. Vol.2, Cambridge, Mass.: Addison-Wesley, 634-654.
- Brunswik, E. 1952. The conceptual framework of psychology. *International Encyclopedia of Unified Science*, 1, No.10.
- Brunswik, E. 1955. Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62, 193-217.
- Brunswik, E. 1956. *Perception and the representative design of experiments*. Berkeley: Univ. of California Press.
- Castellan, N. J., Jr. 1973. Comments on the „lens model“ equation and the analysis of multiple-tue judgment tasks. *Psychometrika*, 38, 87-100.
- Cattell, R. B. 1957. *Personality and motivation: Structure and measurement*. Yonkers, N. Y.: World Book.

- Cavanaugh, J. C. & Perlmutter, M. 1980. Metamemory - a critical examination. Unpublished paper (draft). Institute of Child Development. University of Minnesota.
- Cavior, N. & Marabotto, C. 1976. Monitoring verbal behaviors in dyadic interaction. *Journal of Consulting and Clinical Psychology*, 44, 68-76.
- Chapman, L. J. & Chapman, J. P. 1967. The genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 72, 193-204.
- Chapman, L.J. & Chapman, J. P. 1969. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 271-280.
- Chi, P. L. 1937. Statistical analysis of personality. *Journal of Experimental Education*, 5, 229-245.
- Ciminero, A. R. 1977. Behavioral assessment: An overview. In: Ciminero, A. R., Calhoun, K. S. & Adams, H. E. (eds): *Handbook of behavioral assessment*. New York: Wiley, 3-13.
- Ciminero, A. R., Nelson, R. O. & Lipinski, D. P. 1977. Self-monitoring procedures. In: Ciminero, A. R., Calhoun, K. S. & Adams, H. E. (eds): *Handbook of behavioral assessment*. New York: Wiley, 195-232.
- Claparède, E. 1965. Die Entdeckung der Hypothese. In: C. F. Graumann (ed.): *Denken*. Köln: Kiepenheuer & Witsch, 109-115.
- Cohen, R. 1971. An investigation of the diagnostic processing of contradictory information. *European Journal of Social Psychology*, 1, 475-492.
- Cohen, R. & Schümer, R. 1968. Eine Untersuchung zur sozialen Urteilsbildung. I. Die Verarbeitung von Informationen unterschiedlicher Konsonanz. *Archiv für die gesamte Psychologie*, 120, 151-179.
- Comte, A. 1949. *Cours de philosophie positive (1830-1842)*. 2 Bände. Paris: Garnier.
- Coombs, C. H., Dawes, R. M. & Tversky, A. 1970. *Mathematical Psychology*. Englewood Cliffs, N. J.: Prentice-Hall.
- Cronbach, L. J. 1955. Processes affecting scores of „understanding of others“ and „assumed similarity“ *Psychological Bulletin*, 52, 177—193.
- D'Andrade, R. G. 1965. Trait psychology and componential analysis. *American Anthropologist*, 67, 215-228.
- D'Andrade, R. G. 1974. Memory and the assessment of behavior. In: Blalock, H. M. Jr. (ed.): *Measurement in the social sciences*. Chicago: Aldine, 159-186.
- Danziger, K. 1980. The history of introspection reconsidered. *Journal of the History of the Behavioral Sciences*, 16, 241-262.
- Dörner, D. 1974. *Die kognitive Organisation beim Problemlösen*. Bern: Huber.
- Duncan, S. 1969. Nonverbal communication. *Psychological Bulletin*, 72, 118-137.
- Duncker, K. 1926. A qualitative (experimental and theoretical) study of productive thinking (solving of comprehensible problems). *The Pedagogical Seminary and Journal of Genetic Psychology*, 33, 642-708.
- Duncker, K. 1966. *Zur Psychologie des produktiven Denkens (1935)*. Berlin: Springer.

- Egan, J. P. 1975. Signal detection theory and ROC analysis. New York: Academic Press.
- Ericsson, K. A. & Simon, H. A. 1980. Verbal reports as data. *Psychological Review*, 37, 215-251.
- Feger, H. 1972. Skalierte Informationsmenge und Eindrucksurteil. Bern: Huber.
- Feger, H. & Feger, B. 1969 (a). Beiträge zur inhaltsanalytischen Untersuchung von Entscheidungen. Teil I: Methode und Vergleich der Materialstichproben. *Archiv für die gesamte Psychologie*, 121, 205-232.
- Feger, H. & Feger, B. 1969 (b). Beiträge zur inhaltsanalytischen Untersuchung von Entscheidungen. Teil II: Kontingenzanalyse und Paralleluntersuchung. *Archiv für die gesamte Psychologie*, 121, 233-254.
- Finke, R. A. & Kosslyn, S. M. 1980. Mental imagery acuity in the peripheral visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 126-139.
- Fixen, D. L., Phillips, E. L. & Wolf, M.M. 1972. Achievement place: The reliability of self-reporting and peer-reporting and their effects on behavior. *Journal of Applied Behavior Analysis*, 5, 19-30.
- Flavell, J. H. 1976. Metacognitive aspects of problem-solving. In: L. B. Resnick (ed.): *The nature of intelligence*. Hillsdale, N. J.: Erlbaum.
- Flavell, J. H. 1979. Monitoring social-cognitive enterprises: Something else that may develop in the area of social cognition. Paper presented for the Social Science Research Council Committee on Social and Affective Development During Childhood, January.
- Gibson, J. J. 1950. *The perception of the visual world*. Boston: Houghton Mifflin, (Deutsch: *Die Wahrnehmung der visuellen Welt*. Weinheim: Beltz, 1973).
- Giorgi, A. 1967. A phenomenological approach to the problem of meaning and social learning. *Review of Existential Psychology and Psychiatry*, 7, 106-118.
- Goldfried, M. R. & Kent, R. N. 1972. Traditional versus behavioral assessment: A comparison of methodological and theoretical assumptions. *Psychological Bulletin*, 77, 409-420.
- Goldfried, M. R. & Linehan, M. M. 1977. Basic issues in behavioral assessment. In: Ciminero, A. R., Calhoun, K. S. & Adams, H. E. (eds): *Handbook of behavioral assessment*. New York: Wiley, 15-46.
- Graumann, C. F. 1966. Bewußtsein und Bewußtheit - Probleme und Befunde der psychologischen Bewußtseinsforschung. In: W. Metzger (ed.): *Wahrnehmung und Bewußtsein* (Handbuch der Psychologie, I, 1). Göttingen: Hogrefe, 79-127.
- Graumann, C. F. 1978. Wahrnehmung und Beurteilung der anderen und der eigenen Person. In: A. Heigl-Evers (ed.): *Lewin und die Folgen* (Die Psychologie des 20. Jahrhunderts, Bd. VIII). Zürich: Kindler, 154-183.
- Graumann, C. F. 1980. Experiment, Statistik, Geschichte - Wundts erstes Heidelberger Programm einer Psychologie. *Psychologische Rundschau*, 31, 73-83

(Auch in: W. G. Bringmann & R. D. Tweney (eds): *Wundt Studies*. Toronto: Hogrefe, 1980, 33-41).

- Greeno, J. G. 1968. *Elementary theoretical psychology*. Reading, Mass.: Addison-Wesley.
- Grüner, K.-W. 1974. *Beobachtung*. Stuttgart: Teubner (Studienskripten).
- Guilford, J. P. 1954. *Psychometric methods*. New York: McGraw-Hill, (2. Aufl.).
- Gulliksen, H. 1950. *Theory of mental tests*. New York: Wiley.
- Gulliksen, H. 1968. Methods for determining equivalence of measures. *Psychological Bulletin*, 70, 534-544.
- Gurwitsch, A. 1966. *Studies in phenomenology and psychology*. Evanston, Ill.: Northwestern.
- Hagen, R. L., Craighead, W. E. & Paul, G. L. 1975. Staff reactivity to evaluative behavioral observations. *Behavior Therapy*, 6, 201-205.
- Hart, J. T. 1965. Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 56, 208-216.
- Hart, J. T. 1966. Methodological note on feeling-of-knowing experiments. *Journal of Educational Psychology*, 57, 347-349.
- Hart, J. T. 1967. Second-try recall, recognition, and the memory-monitoring process. *Journal of Educational Psychology*, 58, 193-197.
- Herbert, E. W. & Baer, D. M. 1972. Training parents as behavior modifiers: Self-recording of contingent attention. *Journal of Applied Behavioral Analysis*, 5, 139-149.
- Holzkamp, K. 1980. Zu Wundts Kritik an der experimentellen Erforschung des Denkens. In: *Wilhelm Wundt - Progressives Erbe, Wissenschaftsentwicklung und Gegenwart*. Leipzig: Karl-Marx-Universität, 141-153.
- Ingleby, J. D. 1974. Further studies of the human observer as a statistical decision maker. *Organizational Behavior and Human Performance*, 12, 299-314.
- James, W. 1950. *The principles of psychology*. 2 vols. (1890). New York: Dover.
- Johnson, D. M. 1963. Reanalysis of experimental halo effects. *Journal of Applied Psychology*, 47, 46-47.
- Johnson, S. M. & Bolstad, O. D. 1975. Reactivity to home observation: A comparison of audio recorded behavior with observers present or absent. *Journal of Applied Behavioral Analysis*, 8, 181-185.
- Johnson, S. M. & Lobitz, G. K. 1974. Parental manipulations of child behavior in home observations. *Journal of Applied Behavior Analysis*, 7, 23-32.
- Johnson, S. M. & White, G. 1971. Self-observation as an agent of behavioral change. *Behavior Therapy*, 2, 488-497.
- Johnson, S. M., Christensen, A. & Bellamy, G. T. 1976. Evaluation of family intervention through unobtrusive audio recordings: Experiences in bugging children. *Journal of Applied Behavioral Analysis*, 9, 213-219.

- Jones, R. R., Reid, J. B. & Patterson, G. R. 1975. Naturalistic observation in clinical assessment. In: McReynolds, P. (ed.): *Advances in psychological assessment*. Vol. 3. San Francisco: Jossey-Bass, 42-95.
- Kanfer, F. H. 1970. Self-monitoring: Methodological limitations and clinical applications. *Journal of Consulting and Clinical Psychology*, 35, 148-152.
- Kanfer, F. H. 1975. Self-management methods. In: F. H. Kanfer & A. P. Goldstein, *Helping People Change*. New York: Pergamon.
- Kant, I. 1800. *Anthropologie in pragmatischer Hinsicht*. 2. Aufl. Königsberg: Nicolovius.
- Kant, I. 1903. *Metaphysische Anfangsgründe der Naturwissenschaft (1786)*. In: I. Kant: *Gesammelte Schriften*, Bd. 4. Berlin: Reimer.
- Katz, D. 1911. Die Erscheinungsweisen der Farben und ihre Beeinflussung durch die individuelle Erfahrung. *Zeitschrift für Psychologie, Erg.band 7*.
- Katz, D. 1929. Der Aufbau der Tastwelt. *Zeitschrift für Psychologie, Erg.band 11*.
- Kazdin, A. E. 1974. Reactive self-monitoring: The effects of response desirability, goal setting, and feedback. *Journal of Consulting and Clinical Psychology*, 42, 704-716.
- Kent, R. N. 1972. Expectancy bias in behavioral observation. Unpubl. doc.diss., State Univ. of New York, Stony Brook, New York.
- Kent, R. N. & Foster, S. L. 1977. Direct observational procedures: Methodological issues in naturalistic settings. In: Ciminero, A. R., Calhoun, K. S. & Adams, H. E. (eds): *Handbook of behavioral assessment*. New York: Wiley, 279-328.
- Kent, R. N., O'Leary, K. D., Diament, C. & Dietz, A. 1974. Expectation biases in observational evaluation of therapy change. *Journal of Consulting and Clinical Psychology*, 42, 774-780.
- Köhler, W. 1921. *Intelligenzprüfungen an Menschenaffen*. Berlin: Springer.
- Lachman, J. L. & Lachman, R. 1980. Age and the actualization of world knowledge. In: L. W. Poon et al. (eds): *New directions in memory and aging*. Hillsdale, N.J.: Erlbaum, 285-313.
- Lambert, W. W. 1960. Interpersonal behavior. In: Mussen, P. H. (ed.): *Methods in child development*. New York: Wiley.
- Lay, C. H. & Jackson, D. N. 1969. Analysis of the generality of traitinferential relationship. *Journal of Personality and Social Psychology*, 12, 12-21.
- Lieberman, D. A. 1979. A (limited) call for a return to introspection. *American Psychologist*, 34, 319-333.
- Linschoten, J. 1959. *Op weg naar een fenomenologische psychologie*. Utrecht: Bijleveld, (übers. von F. Mönks: *Auf dem Wege zu einer phänomenologischen Psychologie*. Berlin: de Gruyter, 1961).
- Lipinski, D. P. & Nelson, R. O. 1974. The reactivity and unreliability of self-recording. *Journal of Consulting and Clinical Psychology*, 42, 118-123.
- Lipinski, D. P., Black, J. L., Nelson, R. O. & Ciminero, A. R. 1975. The influence of

- motivational variables on the reactivity and reliability of self-recording. *Journal of Consulting and Clinical Psychology*, 43, 637-646.
- Longabaugh, R. 1980. The systematic observation of behavior in naturalistic settings. In: Triandis, H. C. & Berry, J. W. (eds): *Handbook of cross-cultural psychology*. Vol. 2: Methodology. Boston: Allyn & Bacon, 57-126.
- Lorr, M. & McNair, D. M. 1965. Expansion of the interpersonal behavior circle. *Journal of Personality and Social Psychology*, 2, 823-830.
- Lüer, G. 1973. *Gesetzmäßige Denkabläufe beim Problemlösen*. Weinheim: Beltz.
- Maletzky, B. M. 1974. Behavior recording as treatment: A brief note. *Behavior Therapy*, 5, 107-111.
- Mann, R. D. 1959. The relation between personality characteristics and individual performance in small groups. Ph. D. dissertation, Univ. of Michigan.
- Mash, E. J. & McElwee, J. D. 1974. Situational effects on observer accuracy: Behavior predictability, prior experience, and complexity of coding categorie. *Child Development*, 45, 367-377.
- Mash, L. J. & Makohoniuk, G. 1975. The effects of prior information and behavioral predictability on observer accuracy. *Child Development*, 46, 513-519.
- McNicol, D. 1972. *A primer of signal detection theory*. London: Allen & Unwin.
- Meichenbaum, D., Burland, S., Gruson, L. & Cameron, R. 1979. Metacognitive assessment. Paper presented at the Conference on the Growth of Insight. Wisconsin Research and Development Center, October.
- Meichenbaum, D. & Butler, L. 1980. Cognitive ethology: Assessing the streams of cognition and emotion. In: K. Blankstein, P. Pliner & J. Polivy (eds): *Advances in the study of communication and affect: Assessment and modification of emotional behavior*. Vol. 6. New York: Plenum.
- Metge, Anneros. 1980. Zum Problem der Selbstbeobachtung bei Wundt. In: *Wilhelm Wundt - Progressives Erbe. Wissenschaftsentwicklung und Gegenwart*. Leipzig: Karl-Marx-Universität, 183-190.
- Metzger, W. 1954. *Psychologie*. 2. Aufl. Darmstadt: Steinkopf.
- Mitchell, D. B. & Richman, C. L. 1980. Confirmed reservations: Mental travel. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 58-66.
- Moos, R. H. 1968. Behavioral effects of being observed: Reactions to a wireless radio transmitter. *Journal of Consulting and Clinical Psychology*, 32, 383-388.
- Mulaik, A. 1964. Are personality factors rater's conceptual factors? *Journal of Consulting Psychology*, 28, 506-511.
- Natsoulas, T. 1970. Concerning introspective ‚knowledge‘. *Psychological Bulletin*, 73, 89-111.
- Natsoulas, T. 1978. Residual subjectivity. *American Psychologist*, 33, 269-283.
- Nelson, R. O., Lipinski, D. P. & Black, J. L. 1975. The effects of expectancy on the reactivity of self-recording. *Behavior Therapy*, 6, 337-349.

- Newcomb, T. M. 1931. An experimental design to test the validity of a rating technique. *Journal of Educational Psychology*, 22, 279-289.
- Newtonson, D. 1973. Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28, 28-38.
- Newtonson, D. 1976. Foundations of attribution: The unit of perception of ongoing behavior. In: Harvey, J., Ickes, W. & Kidd, R. (eds): *New directions in attribution research*. Hillsdale, N. J.: Erlbaum, 223-247.
- Newtonson, D. & Engquist, G. 1976. The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12, 436-450.
- Newtonson, D. & Rindner, R. J. 1979. Variation in behavior perception and ability attribution. *Journal of Personality and Social Psychology*, 37, 1847-1858.
- Newtonson, D., Engquist, G. & Bois, J. 1977. The objective basis of behavior units. *Journal of Personality and Social Psychology*, 35, 847-862.
- Nisbett, R. E. & Wilson, T. D. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Nuttin, J. 1955. Consciousness, behavior, and personality. *Psychological Review*, 62, 349-355.
- O'Leary, K. D., Kent, R. N. & Kanowitz, J. 1975. Shaping data collection congruent with experimental hypotheses. *Journal of Applied Behavior Analysis*, 8, 43-51.
- Passini, F. T. & Norman, W. T. 1966. A universal conception of personality structure? *Journal of Personality and Social Psychology*, 4, 44-49.
- Pastore, R. E. & Scheirer, C. J. 1974. Signal detection theory: Considerations for general application. *Psychological Bulletin*, 81, 945-958.
- Peterson, D. R. 1968. *The clinical study of social behavior*. New York: Appleton-Century-Crofts.
- Peterson, W. W., Birdsall, T. G. & Fox, W. C. 1954. The theory of Signal detectability. *Institute of Radio Engineers Transactions*, PGIT-4, 171-212.
- Pilkington, C. W. & Glasgow, W. D. 1967. Towards a rehabilitation of introspection as a method in psychology. *Journal of Existentialism*, 7, 329-350.
- Podgorny, P. & Shepard, R. N. 1978. Functional representations common to visual perception and imagination. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 21-35.
- Postman, L. & Tolman, E. C. 1959. Brunswik's probabilistic functionalism. In: Koch, S. (ed.): *Psychology: A study of a science*. Vol. I. New York: McGraw-Hill, 502-564.
- Price, R. H. 1966. Signal detection methods in personality and perception. *Psychological Bulletin*, 66, 55-62.
- Rachlin, H. 1974. Self-control. *Behaviorism*, 2, 94-107.
- Radford, J. 1974. Reflections on introspection. *American Psychologist*, 29, 245-250.
- Redfield, J. & Paul, G. L. 1976. Bias in behavioral observation as a function of observer

- familiarity with subjects and typicality of behavior. *Journal of Consulting and Clinical Psychology*, 44, 156.
- Reid, J. B. 1970. Reliability assessment of observational data: A possible methodological problem. *Child Development*, 41, 1143-1150.
- Richardson, J. T. E. 1980. *Mental imagery and human memory*. London: Macmillan.
- Roberts, R. R. Jr. & Renzaglia, G. A. 1905. The influence of tape recording on counseling. *Journal of Counseling Psychology*, 12, 10-16.
- Rohracher, H. 1963. *Einführung in die Psychologie*. 8. Aufl. Wien: Urban & Schwarzenberg.
- Romanczyk, R. G., Kent, R. N., Diament, C. & O'Leary, K. D. 1973. Measuring the reliability of observational data: A reactive process. *Journal of Applied Behavior Analysis*, 6, 175-184.
- Rosenthal, R. 1976. *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts, 1966; enlarged edition: New York: Irvington Publ.
- Rosenthal, R. & Rosnow, R. L. (eds) 1969. *Artifact in behavioral research*. New York: Academic Press.
- Rudinger, G. & Feger, H. 1970. Die Beurteilung formaler Persönlichkeitsmerkmale durch Rating-Skalen: Eine Generalisierbarkeitsstudie. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 2, 96-112.
- Rugg, H. 1922. Is the rating of human Character practicable? *Journal of Educational Psychology*, 13, 30-42.
- Schapp, W. 1976. *Beiträge zur Phänomenologie der Wahrnehmung*. Wiesbaden: Heymann.
- Schefflen, A. E. 1975. *How behavior means*. New York: Aronson.
- Schümer, R. 1971. Eine experimentelle Untersuchung zur sozialen Eindrucksbildung. *Zeitschrift für Sozialpsychologie*, 2, 92-108.
- Shepard, R. N. 1978. The mental image. *American Psychologist*, 33, 125-137.
- Shepard, R. N. & Chipman, S. 1970. Second order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1, 1-17.
- Shepard, R. N. & Metzler, J. 1971. Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Simkins, L. 1971. The reliability of self-recorded behaviors. *Behavior Therapy*, 2, 83-87.
- Skinner, B. F. 1945. The operational analysis of psychological terms. *Psychological Review*, 52, 270-277.
- Skinner, B. F. 1953. *Science and human behavior*. New York: Macmillan, (Deutsch: *Wissenschaft und menschliches Verhalten*. München: Kindler, 1973).
- Skinner, B. F. 1957. *Verbal behavior*. New York: Appleton.
- Skinner, B. F. 1963. Behaviorism at fifty. *Science*, 140, 951-958.
- Skinner, B. F. 1974. *About behaviorism*. New York: Knopf.

- Stanley, J. C. 1961. Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. *Psychometrika*, 26, 205-219.
- Starr, D. J. & Katkin, E. S. 1969. The clinician as aberrant actuary: Illusory correlation and the Incomplete Sentences Blank. *Journal of Abnormal Psychology*, 74, 670-675.
- Stern, W. 1911. *Die Differentielle Psychologie in ihren methodischen Grundlagen*. Leipzig: Barth.
- Steward, T. R. 1976. Components of correlation and extensions of the lens model equation. *Psychometrika*, 41, 101-120.
- Stricker, L. J., Jacobs, P. I. & Kogan, N. 1974. Trait interrelations in implicit personality theory and questionnaire data. *Journal of Personality and Social Psychology*, 30, 198-207.
- Swets, J. A. (ed.) 1964. *Signal detection and recognition by human observers*. New York: Wiley.
- Taplin, P. S. & Reid, J. B. 1973. Effects of instructional set and experimenter influence on observer reliability. *Child Development*, 44, 547-554.
- Thorndike, E. L. 1920. A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25-29.
- Thomae, H. 1960. *Der Mensch in der Entscheidung*. München: Barth.
- Titchener, E. B. 1914. *A primer in psychology*. New York: Macmillan.
- van der Kamp, L. J. T. & Mellenbergh, G. J. 1976. Agreement between raters. *Educational and Psychological Measurement*, 36, 311-317.
- van Meter, D. & Middleton, D. 1954. Modern statistical approaches to reception in communication theory. *Institute of Radio Engineers Transactions*, PGIT-4, 119-145.
- Webb, E. J., Campbell, D. T., Schwartz, R. D. & Sechrest, L. 1966. Unobtrusive measures. *Nonreactive research in the social sciences*. Chicago: Rand McNally, (7. Aufl. 1971; deutsch als: *Nichtreaktive Meßverfahren*. Weinheim: Beltz, 1975).
- Wells, F. L. 1907. A statistical study of literary merit. *Archives of Psychology*, 1 (7).
- Wertheimer, M. 1912. Experimentelle Studien über das Sehen von Bewegungen. *Zeitschrift für Psychologie*, 61, 161-265.
- Werts, C. E., Jöreskog, K. G. & Linn, R. L. 1976. Analyzing ratings with correlated intrajudge measurement errors. *Educational and Psychological Measurement*, 36, 319-328.
- White, P. 1980. Limitations on verbal reports of internal events: A refutation of Nisbett and Wilson and of Bern. *Psychological Review*, 87, 105-112.
- Whiting, B. B. & Whiting, J. W. 1975. *Children in six cultures: a psycho-cultural analysis*. Cambridge, Mass.: Harvard Univ. Press.
- Wilder, D. 1978. Effects of predictability on units of perception and attribution. *Personality and Social Psychology Bulletin*, 4, 281-284.

- Wundt, W. 1862. Beiträge zur Theorie der Sinneswahrnehmung. Leipzig/Heidelberg: Winter.
- Wundt, W. 1863. Vorlesungen über die Menschen- und Thierseele. 2 Bde. Leipzig: Voß.
- Wundt, W. 1885. Die Aufgaben der experimentellen Psychologie (1882). In: W. Wundt: Essays. Leipzig: Engelmann.
- Wundt, W. 1888. Selbstbeobachtung und innere Wahrnehmung. Philosophische Studien, 4, 292-309.
- Wundt, W. 1907. über Ausfrageexperimente und über Methoden zur Psychologie des Denkens. Psychologische Studien, 3, 301-360.