11. Kapitel

# Hypothesen

***Walter Hussy und Holger Möller***

## *1. Zur Definition, Klassifikation, Generierung und Funktion von Hypothesen*

### **1.1** Wissenschaftliche und nicht-wissenschaftliche Hypothesen

„Morgen wird es regnen." Dieser Satz stellt eine Aussage, eine Vorhersage bestimmter Ereignisse dar: Er drückt offenkundig eine **Vermutung** über das Wetter des nächsten Tages aus. Man könnte auch sagen, daß es sich um die **Hypothese** handelt, daß es am nächsten Tag regnen wird. Ist diese Hypothese nun wissenschaftlich, kann man sie als **wissenschaftliche Hypothesen** [WH bzw. WHn] bezeichnen? Diese vermeintlich leicht zu beantwortende Frage läßt sich jedoch losgelöst aus dem Kontext, in dem die jeweilige Hypothese aufgestellt wurde, nicht beantworten, wie die beiden folgenden Beispiele illustrieren:

Eventuell ist die Person, die diese Hypothese aufstellt, Teilnehmer einer Gartenparty und äußert die Vermutung zu Beginn eines Gesprächs, vielleicht um Kontakt zu einer anderen Person zu erhalten. Ganz offensichtlich sind es jedoch nicht diese Art von Hypothesen, die in der psychologischen Forschung eine noch näher zu bestimmende Rolle spielen.

Eventuell ist unsere Person aber auch ein Meteorologe, der zu dieser Vermutung im Rahmen seiner Forschungsaktivitäten kommt, aus denen er Vorhersagen für das Wetter des nächsten Tages ableiten kann. In diesem Falle wurde es sich um eine WH handeln. Worin mögen die Unterschiede beider Hypothesentypen, die doch in ihrer Aussage allein nicht differieren, begründet liegen? Was kennzeichnet eine WH und hebt sie von einfachen Aussagen ab? Wie läßt sich der Begriff der (wissenschaftlichen) Hypothese definieren?

## 1.2 Allgemeine Begriffsbestimmungen

Man geht häufig davon aus, daß wissenschaftliche Hypothesen theoretische Vermutungen bzw. Annahmen über die Zusammenhänge von interessierenden Sachverhalten sind oder auch vorläufige (provisorische) Antworten auf ein sich stellendes wissenschaftliches Problem beinhalten. Den WHn wird für gewöhnlich unterstellt, daß sie allgemeine Aussagen darstellen, die das Eintreten bestimmter Ereignisse, Erscheinungen oder Zusammenhänge vorhersagen und damit gleichzeitig das Eintreten anderer Ereignisse etc. ausschließen. Um von einer **wissenschaftlichen** Hypothese sprechen zu können, sollten zwei Merkmale erfüllt sein:

(a) sie muß im Sinne einer synthetischen Aussage formuliert sein, also sich als falsch oder korrekt erweisen können („Wenn der Luftdruck fällt, regnet es!"; diese ausformulierte Hypothese liegt der Aussage des Meteorologen zugrunde, wenn er vermutet „Morgen wird es regnen"), und
(b) sie muß überprüfbar sein (man muß prinzipiell feststellen können, ob sich der Luftdruck verändert und man muß beobachten können, ob es regnet oder nicht regnet).

Im Abschnitt 2.1 werden wir uns ausführlicher mit diesen Merkmalen beschäftigen.

WHn wird ein **vorläufiger** Charakter unterstellt, da sie im Rahmen der wissenschaftlichen Forschung sowohl verworfen als auch zunächst beibehalten bzw. weiterentwickelt werden können. Von den Konzepten, die in WHn auftreten, wird größtenteils angenommen, daß sie über die direkt beobachtbaren Sachverhalte hinausgehen, daß sie im Falle des Beibehaltens einer WH als Erklärungen dienen können, warum spezifische Sachverhalte aufgetreten sind bzw. nicht aufgetreten sind. Wir werden hierauf zurückkommen. Einschränkendere Definitionen betonen die „Beziehungsherstellung zwischen Determinanten und Resultanten", die „in Form von Je-desto' oder ‚Wenn-dann'-Aussagen" Vorfindbar sind (z.B. Krapp, Hofer & Prell, 1982).

## 1.3 Arten von Hypothesen

Wissenschaftliche Hypothesen lassen sich nach zahlreichen Gesichtspunkten unterscheiden. Zur Illustration zunächst einige Beispiele:

(1) „Kurt hat heute die dritte Aufgabe der vorliegenden Klausur intelligent gelöst."
(2) „In diesem Hörsaal gibt es männliche Personen, die einen HAWIE-IQ von mehr als 130 Punkten erreichen."

(3) „Alle Teilnehmer dieser Vorlesung Varianzanalyse haben einen HAWIE-IQ von mehr als 150 Punkten."
(4) „Frustrierte Personen reagieren zumeist aggressiv."
(5) „Die Mittelwerte der beiden Gruppen A und B unterscheiden sich."
(6) „Frustration erzeugt Aggression."

Ein möglicherweise nützliches und sinnvolles Kriterium zur Differenzierung ist die Art der Einschränkungen, die den in Hypothesen angeführten Personenkreisen, Situationen, Aktivitäten, Zeiten etc. auferlegt werden. So wird z.B. in der o.g. Hypothese (3) für alle Personen einer spezifischen Lehrveranstaltung zu einer spezifischen Zeit eine Aussage getroffen (man könnte auch sagen, daß der Anwendungsbereich der Hypothese eine raum-zeitliche Beschränkung aufweist: beschränkt universelle Hypothese), während in Hypothese (6) weder Personen noch Orte oder Zeiten eingeschränkt werden: Hypothese (6) soll immer und für alle Personen in allen Situationen gelten (Universalhypothese oder unbeschränkte, universelle Hypothese). Groeben und Westmeyer (1981) haben so acht verschiedene Hypothesenarten differenziert, auf die wir kurz eingehen wollen. Die sehr unterschiedlichen Überprüfungsmöglichkeiten dieser Hypothesentypen werden uns erst später beschäftigen.

**Singuläre Hypothesen** haben sowohl hinsichtlich des Personenkreises, auf den sie sich beziehen, wie auch hinsichtlich der möglichen Situationen und den dort ausgeführten Aktivitäten Einschränkungen (siehe Beispiel (1) oben).
**Pseudo-singuläre (idiographische) Hypothesen** haben entgegen den singulären Hypothesen keine Einschränkungen in Raum und/oder Zeit: „Kurt ist intelligent."
**Unbestimmte Existenzhypothesen** enthalten mindestens einen Existenzquantor ohne weitere Einschränkungen: „Es gibt ein psycho-physiologisches Korrelat der menschlichen Emotion."
**Lokalisierende (bestimmte) Existenzhypothesen** beschränken ihre Existenzaussagen auf bestimmte Räume und/oder Zeiten (vgl. Beispiel (2) oben).
**Quasi-universelle Hypothesen** sind ähnlich wie (echt) universelle Hypothesen formuliert, beinhalten jedoch Einschränkungen, durch die die generellen Aussagen eher vage werden. Die quasi-universelle Hypothese soll nur mit einer (zumeist unbestimmten) Wahrscheinlichkeit gültig sein, Ausnahmen von den in der Hypothese aufgestellten Regeln sprechen nicht unbedingt gegen die Hypothese (siehe Beispiel (4) oben). Dieser Hypothesentyp ist in der Psychologie sehr häufig anzutreffen. Groeben und Westmeyer vermuten, daß Psychologen zwar einerseits generelle Aussagen von unbedingter Gültigkeit aufstellen möchten, andererseits jedoch die Generalität nicht einlösen können.
**Statistische Hypothesen** sind Hypothesen über Verteilungen von Werten bzw. über z.B. Mittelwerte, Varianzen, Korrelationen von Datenreihen (evtl. auch ohne Verteilungsannahmen; vgl. Beispiel (5) oben).

**Beschränkte universelle Hypothesen** beziehen sich einerseits auf alle Fälle einer Population von Individuen, führen aber Beschränkungen in Raum und Zeit ein (siehe Beispiel (3) oben). Groeben und Westmeyer versuchen die Generalisierungen der Sozialwissenschaften durch diese Hypothesenart zu rekonstruieren, die offenkundig eine universelle Gültigkeit beanspruchen, sich jedoch auf bestimmte ,,historisch-gesellschaftliche Rahmenbedingungen" beschränken.

**Unbeschränkte universelle Hypothesen** beziehen sich auf alle Fälle einer bestimmten Art, haben keinerlei weitere Einschränkungen (vgl. Beispiel (6) oben).

Dazu noch einige Erläuterungen anhand des Meteorologenbeispiels, dem eine raumzeitliche Beschränkung anhaftet. Die vom Meteorologen implizit mitgedachte Ausformulierung ,,Wenn der Luftdruck fällt, regnet es!", läßt verschiedene Einordnungen in das besprochene Klassifikationsschema zu: es handelt sich um eine (a) ***quasi-universelle Hypothese,*** wenn sie nur Wahrscheinlichkeitscharakter aufweist (wenn der Meteorologe unsicher ist); (b) ***beschränkt universelle Hypothese,*** wenn es räumliche und/oder zeitliche Einschränkungen gibt; ***(c) unbeschränkt universelle Hypothese,*** wenn keinerlei Einschränkungen in Ort, Zeit und Wahrscheinlichkeit mitgedacht sind.

Abschließend noch einige Bemerkungen zu den inhaltlichen Hypothesen aus unseren o. g. Beispielen. Vergleichen wir o. g. Hypothese (5) etwa mit Hypothese (4), so fällt auf, daß die erste eine **statistische** Hypothese ist (also Aussagen über statistische Parameter wie Mittelwerte, Korrelationen, Kovarianzen, Varianzen etc. beinhaltet), während die zweite eine **inhaltliche** Hypothese darstellt. Vergleichen wir weiterhin Hypothese (4) mit Hypothese (2), so fällt auf, daß sich Hypothese (2) auf konkret beobachtbare Dinge bezieht, Sachverhalte also, die empirisch beobachtbar sind, während Hypothese (4) sich auf Dinge bezieht, die eigentlich nicht empirisch beobachtbar sind. Die Begriffe, die in dieser Hypothese verwendet werden, sind eher Konstruktionen (theoretische Konstrukte), die als Erklärung für bestimmte Phänomene herangezogen werden.

> Die aus Voruntersuchungen, eigenen Beobachtungen, Überlegungen bzw. aus Theorien abgeleiteten Vermutungen bezüglich des in Frage stehenden Untersuchungsgegenstandes bezeichnen wir als ***Forschungshypothese. . . . .*** Der Forschungshypothese nachgeordnet ist die ***operationale Hypothese.*** Mit der operationalen Hypothese prognostiziert der Forscher den Ausgang einer konkreten Untersuchung (der natürlich im Einklang mit der allgemeinen Forschungshypothese stehen muß). (Bortz, 1984, S. 366)

Es erscheint uns sinnvoll, eine Trennung von theoretischen und empirischen Begriffen auch mit Blick auf die WHn durchzuführen. Daher wollen wir mit Hager (1984) eine solche Hypothese, die sich auf theoretische Begriffe und

Konstrukte bezieht, als **theoretisch-inhaltliche Hypothese** [TIH] bezeichnen. Hypothesen wie o.g. Hypothese (2) beziehen sich auf empirisch-beobachtbare Begriffe. Diese Art der Hypothese soll ebenfalls in Anlehnung an Hager (1984) als **empirisch-inhaltliche** Hypothese [EIH] bezeichnet werden. Die von den inhaltlichen Hypothesen abgegrenzten statistischen Hypothesen werden an dieser Stelle noch nicht weiter untergliedert. Von den inhaltlichen Hypothesen wollen wir annehmen, daß in ihnen im Regelfall psychologisch-inhaltliche Konzepte enthalten sind, daß statistische Konzepte wie Verteilungen, Mittelwerte o.ä. nicht in ihnen enthalten sind. Ausnahmen wären dann diejenigen psychologisch-inhaltlichen Hypothesen, die mathematische und/oder statistische Terme enthalten (z. B. Mathematische Modelle menschlichen Lernens etc.). Diese inhaltlichen Hypothesen lassen sich zwar ebenso unter o.g. Kategorien fassen, die dazu notwendigen Erläuterungen wurden uns jedoch zu weit vom eigentlichen Thema entfernen.

Die Unterscheidung in die zwei o. g. inhaltlichen Hypothesenebenen - je nach Art der verwendeten Begriffe - macht die eingangs genannten Äußerungen überdenkenswert: TIHn können das Eintreten bestimmter Ereignisse weder vorhersagen noch ausschließen, wenn diese Ereignisse beobachtbar sein sollen. Streng genommen haben TIHn keine direkte Verbindung mit beobachtbaren Sachverhalten, mit dem, was Forscher auch als Empirie bezeichnen. Dieses Problem wird uns später noch detaillierter beschäftigen, wenn wir die Frage beantworten wollen, wie WHn überprüft und gegebenenfalls als „falsch“ oder ,,richtig“ bezeichnet werden können.

## 1.4 Generierung von Hypothesen

Wie kommt ein Wissenschaftler eigentlich zu seinen Hypothesen? Eine allgemeingültige Anweisung dafür existiert nicht. Vielmehr muß man die wissenschaftliche Tätigkeit als Problemlösevorgang verstehen (vgl. z.B. Dörner, 1979; Hussy, 1983). Der wissenschaftlich tätige Psychologe interessiert sich für das ***wie*** und *warum* menschlichen Verhaltens und Erlebens. Er stellt Fragen und sucht nach Antworten. Hypothesen sind - wie gesehen - vorläufige Antworten auf solche Fragen, Sie sind Ergebnisse des Problemlösevorgangs (Lösungsmöglichkeiten), die überprüft werden müssen.

So kann man sich fragen, wie es kommt, daß kurzfristig nur eine begrenzte Zahl an Informationen behalten werden kann. Legt man einer Person eine siebenstellige Telefonnummer vor, so kann sie diese in der Regel für eine kurze Zeit behalten, etwa so lange, bis sie die Ziffern gewählt hat. Ist die Nummer deutlich länger, wird sie Schwierigkeiten bekommen. Ähnliche Schwierigkeiten wird sie haben, wenn die Telefonnummer nicht unmittelbar nach der Präsentation, sondern erst nach einer viertel Minute gewählt wird. Ein Wissen-

schaftler, der eine Vielzahl solcher und weiterer Beobachtungen angestellt hat, wird zu allgemeineren Vermutungen über das kurzfristige Behalten kommen, z. B. in Form der Hypothese: Wenn mehr als ca. sieben Informationseinheiten in unmittelbarer Abfolge präsentiert werden, dann kommt es bei der sich unmittelbar anschließenden Wiedergabe des Materials zu Fehlern. Der eigentliche Problemlösevorgang besteht hierbei in der Abstraktion der Gemeinsamkeiten der beobachteten Ereignisse. Es muß an dieser Stelle betont werden, daß Abstraktion hier im Sinne eines kognitiven Prozesses verstanden wird und keineswegs Bezug genommen wird auf die induktive Logik im Sinne von Carnap (1936; vgl. auch Westermann und Gerjets, Kapitel 10 in diesem Band). Dennoch bezeichnet man diese Art der Hypothesenfindung als ***induktives Vorgehen,*** weil von speziellen Einzelereignissen zu allgemeineren Vermutungen übergegangen wird.

Umgekehrt verhält es sich beim ***deduktiven Vorgeben.*** Hierbei werden aus einer vorliegenden allgemeinen Theorie spezielle (neue) Vermutungen abgeleitet. Ein gutes Beispiel dafür ist die Untersuchung von Lepper, Greene und Nisbett (1973), die die Selbstwahrnehmung (kognitive Dissonanztheorie) betrifft. Gemäß der Theorie wird zwischen zwei (oder mehreren) Gedanken, Erlebnissen etc. (sogenannten „kognitiven Elementen“) dann eine Dissonanz entstehen, wenn diese zueinander im Widerspruch stehen. Zentraler Gedanke der Theorie ist dann, daß der Mensch in solchen Fällen eine Tendenz verspürt, diese Dissonanz zu reduzieren. Er wird Maßnahmen ergreifen, die ihm eine solche Dissonanzreduktion ermöglichen, er wird Handlungen oder möglicherweise Umbewertungen seiner kognitiven Elemente vollführen.

In einer typischen Untersuchung zu dieser Theorie nehmen Personen an einem Experiment teil, in welchem recht stupide Aufgaben zu erledigen sind. Ein Teil von ihnen wird dafür gut, der andere Teil schlecht bezahlt. Die schlechtbezahlten Personen berichten hinterher, daß ihnen das Experiment gefallen hat, die gutbezahlten dagegen finden es langweilig. Gemäß der Theorie wurde man vermuten, daß bei den schlechtbezahlten Personen eine kognitive Dissonanz besteht - so wenig Geld für eine so langweilige Tätigkeit -, die dadurch beseitigt wird, daß man die Situation im Nachhinein als doch ganz interessant erlebt. Lepper et al. fragten sich nun, ob auch das Gegenteil zutrifft. Sie leiteten aus der Theorie die Hypothese ab, daß das (gute) Bezahlen einer kurzweiligen Tätigkeit dazu führt, daß diese Tätigkeit im Nachhinein als weniger kurzweilig erlebt wird. Tatsächlich bestätigten die Ergebnisse der Untersuchung diese Vermutung.

Gleichgültig ob man den induktiven oder deduktiven Weg beschreitet (in den meisten Fällen wird es ohnehin eine Mischform sein), profitiert der Vorgang der intentionalen Hypothesengenerierung - wie jeder Problemlöseprozeß - von einem umfangreichen, wohlstrukturierten, problembezogenen Faktenwis-

sen auf seiten des Wissenschaftlers. Allerdings kann auch der Zufall eine entscheidende Rolle spielen, wenngleich in der Literatur solche Fälle selten berichtet werden. So geht die Theorie des klassischen Konditionierens von Ivan Pawlow auf eine Zufallsentdeckung zurück. Er war ursprünglich - als Physiologe - an der Verdauung, speziell der Sekretion der Speicheldrüsen interessiert, etwa an der Frage, wie lange es bis zur Speichelsekretion dauert, wenn ein Hund gefüttert wird. Im Zuge seiner Untersuchungen fand er, daß die Tiere mit zunehmender Vertrautheit mit der Fütterungssituation (das gleiche Fressen, der gleiche Napf, der gleiche Pfleger) sogar schon Speichel produzierten, bevor sie das Fressen im Maul hatten. Das ging so weit, daß die Speichelproduktion schon beim Sehen des Pflegers begann. Er interessierte sich weiter für dieses überraschende Phänomen, das er psychische Sekretion nannte, da er vermutete, daß es auf die mentalen Aktivitäten der Tiere zurückzuführen sei und leistete mit seiner daraus entwickelten Theorie der klassischen Konditionierung einen wesentlichen, ursprünglich nicht geplanten Beitrag zu den psychologischen Lerntheorien.

## 1.5 Die Hypothese im Forschungsprozeß

Nicht nur das Generieren von Hypothesen ist als ein Vorgang des Problemlösens verstehbar. Auch die Entwicklung von Theorien (vgl. Kapitel 8) und die Aufstellung und Abarbeitung von Forschungsprogrammen (vgl. Kapitel 6) stellen Problemlöseprozesse dar. Sie repräsentieren die übergeordneten Ziele, zu deren Erreichung wissenschaftliche Hypothesen einen zentralen Beitrag leisten. Letztere bestimmen die Richtung der Forschungsarbeit und leiten das wissenschaftliche Arbeiten eines jeden Forschers (vgl. Bredenkamp, 1980; Gadenne, 1976, 1984 oder Hager & Westermann, 1983).

Wie bereits besprochen und in Abbildung 1 veranschaulicht, können WHn aus mehr oder minder elaborierten Theorien oder gar Theorienetzen abgeleitet sein (deduktives Vorgehen). In diesem Falle (den man wohl als Idealfall psychologischer Forschung bezeichnen müßte) wird die Bewertung der WHn auch von Bedeutung für die Bewertung der Theorien sein, aus denen sie abgeleitet wurden. Wenn WHn in größere Konzeptionen eingebettet sind, wie dieses bei Theorien und - verstärkt - bei Theorienetzen der Fall ist, können die Einzelerkenntnisse, die sich im Laufe des Überprüfungsprozesses der WHn ergeben, zu einem umfassenderen Erkenntnisfortschritt auf Theorieebene führen. Die besprochene Arbeit von Lepper et al. (1973) verdeutlicht diese Überlegung.

Gemäß Abbildung 1 führt der zweite Weg direkt von den Beobachtungen im Interessen- bzw. Problembereich zu den Hypothesen, etwa deshalb, weil noch keine theoretischen Vorstellungen zur Fragestellung vorliegen (induktives

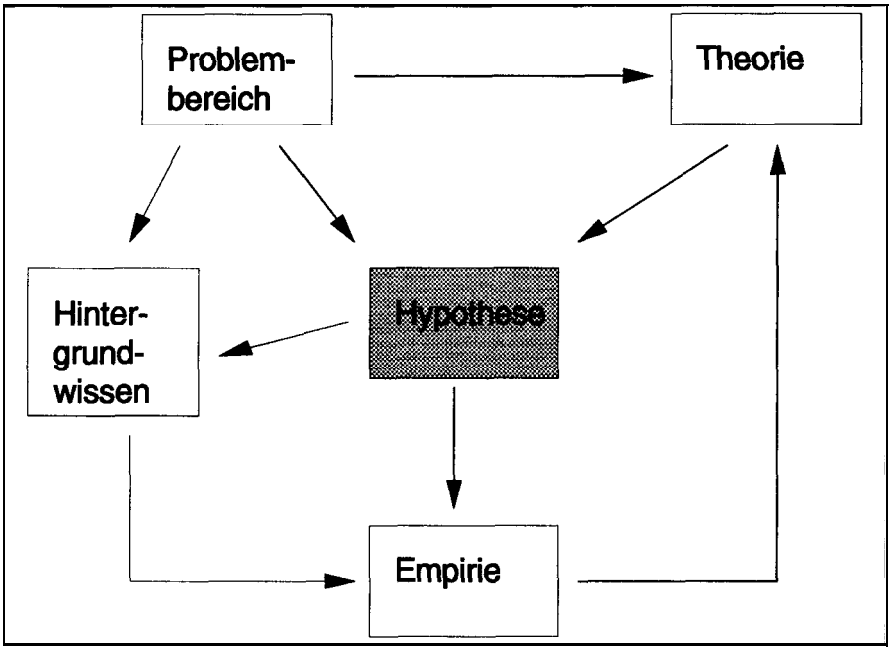


Abb. 1: Zusammenhänge von Theorie, Hypothese und Empirie

Vorgehen). In diesem Fall trägt die Hypothesenprüfung mit ihren Ergebnissen zur Schaffung einer soliden empirischen Datenbasis bei und bildet somit die Grundlage für die Entwicklung von Theorien. Vergegenwärtigen wir uns das obengenannte Beispiel zum kurzfristigen Behalten. Aus einer Vielzahl von Einzelbeobachtungen zum fraglichen Gegenstandsbereich wurde eine WH gebildet und überprüft. Weitere diesbezügliche Untersuchungen (etwa zur Hypothese: „Wenn der Abruf von einmal präsentierten Informationen mehr als 15 Sekunden verzögert wird, dann kommt es zu Fehlern bei der Wiedergabe") verbreitern die Ergebnisbasis und helfen beim Formulieren von Theorien. In diesem Beispiel führen Untersuchungen, die die Beibehaltung der Hypothesen stützen, zur Entwicklung des Konzepts des Kurzzeitgedächtnisses mit begrenzter Kapazität, sowohl bezüglich des Informationsumfangs als auch der Behaltensdauer.

Hypothesen haben eine weitere wichtige Funktion im Wissenschaftsprozeß: Sie stellen die „theoretische Brille" dar, durch die der Forscher seine empirische Arbeit sieht. Man kann die Auffassung vertreten, daß jede Beobachtung und Datenerhebung im Grunde genommen hypothesengeleitet erfolgt (wenngleich mit unterschiedlichem Explikationsgrad). Die Entscheidung des Forschers, an einer bestimmten Stichprobe von Versuchspersonen (VP bzw. VPn) ein spezifisches Treatment zu applizieren und genau definierte Maße empirisch

zu erheben, gründet sich immer auf mehr oder weniger klare Vorstellungen und Erwartungen, auf Vermutungen über Zusammenhänge der interessierenden Phänomene, die einerseits nichts anderes als Hypothesen sind (Hempel, 1974, S. 27), andererseits aber auch das Hintergrundwissen repräsentieren, welches durch die Fragestellung und die Hypothese (gegebenfalls auch durch die Theorie) aktiviert wird und auf die Gestaltung der empirischen Untersuchung zur Hypothesenprüfung - wie wir noch sehen werden - erheblichen Einfluß nimmt (vgl. Abbildung 1).

Generell kann man feststellen, daß Hypothesen im Rahmen des hypothetico-deduktiven Vorgehens das Bindeglied zwischen den Theorien und den empirischen Daten darstellen. Sie bilden die Grundlage für die Präzisierung der Fragestellung und führen zur Planung der Anlage der Untersuchung ebenso wie sie den Rückschluß von den statistischen Untersuchungsergebnissen auf die Ausgangstheorie (oder zu erstellende Theorie) bestimmen. Auf die dabei auftretenden Probleme werden wir im weiteren Verlauf des Textes eingehen. Sie gilt es in jedem Fall zu berücksichtigen, gleichgültig auf welcher Ebene eine Hypothese generiert wird (TIH oder EIH bzw. deduktiv oder induktiv), oder mit Hilfe welcher quantitativen Forschungsmethode (Experiment, Korrelationsstudie) sie überprüft wird.

## *2. Zur Logik der Überprüfung wissenschaftlicher Hypothesen*

Da eine Vermutung über etwas „richtig" oder „falsch" sein kann, stellen sich an dieser Stelle mindestens zwei Probleme:

(a) Läßt sich einer WH „Richtigkeit" bzw. „Falschheit" attestieren, „trifft die WH zu oder trifft sie nicht zu"?
(b) Wenn ja, wie könnte der Überprüfungsprozeß aussehen bzw. wie lassen sich WHn überprüfen?

Unter (a) lassen sich die Vorbedingungen für die Überprüfbarkeit von Hypothesen zusammenfassen. Wir gehen im folgenden Abschnitt kurz darauf ein. Im Anschluß daran beschäftigen wir uns ausführlicher mit dem in (b) angesprochenen Überprüfungsprozeß, der eine Reihe von Ableitungsschritten umfaßt.

### *2.1* Widerspruchsfreiheit und Operationalisierbarkeit von Hypothesen

Bevor man sich dem in der Regel nicht unerheblichen Aufwand unterzieht, eine WH empirisch zu überprüfen, sollte man sich vergewissern, daß einige Vorbedingungen eingehalten sind, deren Verletzung die Ergebnisse einer Un-

tersuchung wertlos machen bzw. eine sinnvolle Untersuchung erst überhaupt nicht ermöglichen. Es zählen dazu Überlegungen zur **Formulierung** von Hypothesen sowie zur prinzipiellen **Widerlegbarkeit** einer Hypothese und der **Operationalisierbarkeit** der darin enthaltenen Konzepte.

Der Art der Formulierung von Hypothesen wird generell ein großer Freiraum eingeräumt. Einem gewissen Standardformat entsprechen die ,,wenn,dann"-Hypothesen oder ,,je,desto"-Hypothesen (McGuigan, 1968, S. 313f.). Zum einen offenbaren diese Formulierungen in besonderer Weise den vorläufigen Charakter von Hypothesen und zum anderen differenzieren sie (bezogen auf die empirische Ebene) bereits zwischen unabhängigen und abhängigen Variablen (Kausalhypothesen) bzw. Prädiktor und Kriterium (Zusammenhangshypothesen). Wichtiger als diese sprachlichen Aspekte ist die Gefahr der widersprüchlichen Formulierung, wenngleich diesem Problem selten viel Platz eingeräumt wird. Dieses mag u.a. damit zusammenhängen, daß bei einfachen Hypothesen ein Widerspruch in sich sofort auffallt. Allerdings muß bereits in solchen einfachen Fällen darauf geachtet werden, daß Widerspruchsfreiheit auch bzgl. der Zusatzannahmen (Hintergrundwissen, vgl. Abschnitt 2.2) besteht, welche bei der Operationalisierung der WH benötigt werden. Einen schwierigen Schritt stellt die Überprüfung der Widerspruchsfreiheit vor allem bei komplexen Hypothesen oder Hypothesengeflechten dar (vgl. z.B. Hussy und von Eye, 1988). Die in jüngerer Zeit erfolgte Entwicklung von Strukturgleichungsmodellen (z. B. EQS, Bentler, 1986; LISREL, Jöreskog und Sörbom, 1988) ermöglicht eine simultane Überprüfung mehrerer komplexer statistischer Hypothesen abgeleitet aus sogenannten Kausalmodellen. Neue und weiterführende Einsichten zur Lösung der Problematik der Überprüfung von Hypothesengeflechten scheinen daraus jedoch nicht zu resultieren (z. B. Möller, 1991).

Einige Überlegungen sollten auch zur prinzipiellen Widerlegbarkeit von Hypothesen angestellt werden. Daß die These ,,Wenn der Hahn kräht auf dem Mist, ändert sich das Wetter, oder es bleibt wie es ist" nicht widerlegbar ist, liegt auf der Hand (vgl. Huber, 1987, S.53). Trifft eine solche Hypothese durch die Art ihrer Formulierung praktisch immer zu, so gibt es keinerlei empirische Sachverhalte, die zu den in der Hypothese behaupteten Sachverhalten in Widerspruch stehen können. Man sagt auch, daß die betreffende Hypothese keinen ***empirischen Gehalt*** hat. Der empirische Gehalt einer Hypothese wird häufig über die Anzahl von Falsifikatoren einer Hypothese erfaßt (je mehr mögliche Falsikatoren eine Hypothese hat, desto höher ist ihr empirischer Gehalt; vgl. Opp, 1976).

Die prinzipielle Widerlegbarkeit ist auch dann gefährdet, wenn die in der Hypothese enthaltenen Begriffe nicht bzw. nicht eindeutig operationalisierbar sind. Die Hypothese ,,Erweitertes Bewußtsein erhöht die Fähigkeit zur De-

fokussierung der Aufmerksamkeit und mindert die Fähigkeit zur Fokussierung der Aufmerksamkeit“ ist solange prinzipiell nicht widerlegbar (und bedarf von daher keiner weiteren empirischen Überprüfung) als nicht eindeutig angegeben werden kann, was unter Bewußtsein verstanden und wie es erfaßt wird. Man spricht in diesem Fall auch davon, daß die Hypothese nicht mit empirischen Daten konfrontiert werden kann (vgl. Schulz, Muthig & Koeppler, 1981, S. 30).

Abschließend sei darauf hingewiesen, daß natürlich auch die in der Hypothese gemachten Einschränkungen (bzw. das Fehlen solcher) und die logische (Re-)Konstruktion der Hypothese einen erheblichen Einfluß auf die prinzipielle Widerlegbarkeit (bzw. Belegbarkeit, Verifikation) der Hypothese haben: Je nach Art (vgl. Abschnitt 1.3) wird eine Hypothese z.B. eventuell überhaupt nicht widerlegbar (dieses gilt z.B. für die o. g. unbestimmten Existenzhypothesen) bzw. belegbar sein.

## 2.2 Die Überprüfung wissenschaftlicher Hypothesen

Nach diesen kurzen Vorbemerkungen zu den Vorbedingungen für die überprüfbarkeit wissenschaftlicher Hypothesen kommen wir nun zur Frage (b), wie nämlich der Überprüfungsprozeß aussieht. Diese Frage wollen wir zunächst unter Rückgriff auf Abschnitt 1.3 und die dort dargelegten Konzepte der verschiedenen Hypothesenarten nach Groeben und Westmeyer (1981) behandeln. Daß die Überprüfungen von Hypothesen immer nur unter Rückgriff auf ein spezifisches Hintergrundwissen möglich sind, soll zunächst weiter ausgeklammert bleiben.

Obgleich sowohl singuläre als auch pseudo-singuläre Hypothesen als wissenschaftliche Hypothesen zu gelten haben, wollen wir Überprüfungsmöglichkeiten beider Arten hier nicht behandeln. Sie werden in der Psychologie zumeist als Beschreibungen (adverbialer bzw. adjektivischer Art) angewendet, es wird z.B. die Testleistung eines Probanden in einer bestimmten Aufgabe ausgedruckt.

Die unbestimmten Existenzhypothesen („Es gibt ein psycho-physiologisches Korrelat der menschlichen Emotion“) lassen sich prinzipiell bestätigen: Es reicht der Nachweis, daß die Hypothese bei einem einzigen Individuum bzw. einer einzigen Instanz gilt. Andererseits lassen sich unbestimmte Existenzhypothesen **nicht** widerlegen, da es niemals auszuschließen ist, daß irgendwann einmal eine positive Instanz gefunden wird.

Bestimmte Existenzhypothesen sind demgegenüber durchaus (zumindest prinzipiell) beleg- sowie widerlegbar. Durch die Einschränkungen in Raum und/oder Zeit ist eine Untersuchung aller Einheiten, aller Fälle möglich.

Anders verhält es sich mit den o. g. quasi-universellen Hypothesen, die in der Psychologie häufig vorliegen: Durch die Vagheit der eigentlich generellen Hypothese ist eine strenge Prüfung **nicht** möglich. Man wird auf Hilfskonstruktionen und Zusatzannahmen angewiesen sein, die uns im folgenden Abschnitt näher beschäftigen werden.

Beschränkte universelle Hypothesen lassen sich ähnlich den bestimmten Existenzhypothesen prinzipiell untersuchen, da Einschränkungen im Raum-Zeit-Bereich gemacht werden. Beispiele für solche Überprüfungen sind Meinungsumfragen, Einstellungsuntersuchungen etc., in denen die gesamte (endliche und abgeschlossene) Menge von Objekten untersucht wird.

Die unbestimmten universellen Hypothesen (die Allaussagen formuliert haben) sind (streng genommen) nicht zu belegen, sondern nur zu widerlegen. Da sie keinerlei Einschränkungen kennen, ist der Überprüfungsprozeß eigentlich nie abgeschlossen. Selbst wenn bislang bei allen Untersuchungen sich die Hypothese bewähren konnte, so ist prinzipiell nicht auszuschließen, daß nicht doch irgendwann eine negative Instanz auftritt. Das Auftreten einer solchen negativen Instanz reicht dann zur Widerlegung der Hypothese aus. Hier liegt insofern eine Umkehrung der unbestimmten Existenzhypothese vor.

Bei vielen WHn ist es (zumindest in der Psychologie) nicht immer leicht, sie als eine solche Art von Hypothese zu rekonstruieren, wie sie oben beschrieben werden. Daher läßt sich auch das Problem der Überprüfung einer WH immer nur unter dem Blickwinkel der spezifischen Rekonstruktion der betreffenden WH sehen („Diese Hypothese scheint die Struktur einer unbestimmten Existenzhypothese zu haben, und daher folgt für ihre Überprüfung, daß ..."). In der Psychologie finden wir sehr häufig die bereits erwähnten quasi-universellen Hypothesen (Wahrscheinlichkeitshypothesen). Da eine streng-logische Überprüfung nicht möglich ist, werden andere Wege beschritten werden müssen, über die wir nun berichten.

### *2.2.1 Der Weg von den Inhalten zur Statistik*

Wir wollen zunächst die o.g. Hypothese (4) als Beispiel heranziehen: „Frustrierte Personen reagieren zumeist aggressiv." Diese Hypothese ist einerseits eine **inhaltliche** Hypothese, da sie auf inhaltliche Konzepte (Frustration, Aggression) abzielt. Andererseits ist sie eine **theoretisch-inhaltliche Hypothese,** da sich bei genauerer Betrachtung die o.g. inhaltlichen Konzepte als nicht beobachtbar (zumindest im wissenschaftlichen Sinne) herausstellen. Wollen wir das potentielle Zutreffen oder Nicht-Zutreffen dieser Hypothese möglichst stringent wissenschaftlich überprüfen, so werden wir die Hypothese mit der Realität, der Empirie, konfrontieren müssen. Dazu müssen wir vorher

darlegen, **was *wir im Experiment*** unter Frustration und Aggression verstehen wollen. Erst daran anschließend können wir die Hypothese überprüfen (z.B. durch ein entsprechend geplantes Experiment).

Zunächst sind die unabhängigen und abhängigen Variablen, die als theoretische Begriffe in der Hypothese enthalten sind, auf der Basis des vorliegenden Hintergrundwissens so festzulegen, daß sie einer Beobachtung, Messung und Erfassung zugänglich werden. Dieses Vorgehen wird als **Operationalisierung** bezeichnet (vgl. Hager, 1987, S. 53) und stellt einen der wichtigsten, aber auch schwierigsten Schritte in der psychologischen Forschung dar. Da die theoretischen Konzepte sich einer Beobachtung entziehen, müssen wir ihnen empirische Konzepte zuordnen.

Die Zuordnung von empirischen zu theoretischen Konzepten ist nicht möglich ohne (teils erhebliche) Reduktion des semantischen Gehaltes der theoretischen Konzepte, in einigen Fällen werden die vorgeschlagenen oder durchgeführten Operationalisierungen bei keinem zweiten Forscher Zustimmung finden. Dieses kann u. a. schon daraus resultieren, daß viele theoretische Konzepte einen nur sehr schwer zu definierenden semantischen Gehalt haben. Es scheint zwar in den meisten Fällen eine Art stiller Übereinkunft zwischen verschiedenen Forschern zu existieren, was sich theoretisch z.B. unter „Aggression“ verstehen läßt, wenn es dann jedoch daran geht, für ein konkretes Experiment Maßnahmen zu planen, die zu einer „Aggression“ der VPn führen sollen, findet man keinen Konsens.

Die Zuordnungsbeziehung sollte idealerweise wissenschaftlich exakt definierbar sein, jedoch fanden bislang vorgebrachte Lösungsvorschläge keine allgemeine Zustimmung: Als ein Beispiel dafür sei an die Vorschläge des logischen Empirismus via Carnap erinnert, die auf eine Definition der theoretischen Konzepte über beobachtbare Variablen hinausliefen, wobei die Zuordnung über sogenannte Reduktionssätze erfolgen sollte (vgl. u.a. Westermann, 1987a, S.11-12). Inwieweit aktuelle Ansätze aus dem Bereich der mathematischen Statistik (z.B. Einbezug von meßtheoretischen Modellen, damit möglicherweise präzisere Definition der Zuordnungsbeziehung von sogenannten latenten und manifesten Variablen über Strukturgleichungsmodelle wie z.B. LISREL etc.) hier eventuell Abhilfen schaffen können, muß an dieser Stelle offen bleiben (vgl. aber den Beitrag von Steyer in diesem Band, Kapitel 15). Bei letztgenannten Ansätzen scheint u.a. die Beziehung zwischen den theoretischen Konzepten und den latenten Variablen der Strukturmodelle unklar.

Die Operationalisierung erfolgt unter Bezugnahme auf das sogenannte Hintergrundwissen (Abbildung 1, vgl. u.a. Schulz, Muthig & Koeppler, 1981), einem Wissensfundus, der dem Forscher zur Verfügung steht (oder zumindest stehen sollte). Dieses Hintergrundwissen enthält die zur Zeit in einem spezifischen Wissenschaftsbereich bekannten Gesetzmäßigkeiten, die zur Verfü-

gung stehenden technischen Voraussetzungen, weitere Vorannahmen, experimentelle Techniken, empirische Resultate etc., die in den Prozeß der Operationalisierung der theoretischen Variablen einmünden. Dieses Wissen ist selbst nicht Gegenstand der Überprüfung der WHn, es wird praktisch (ungeprüft) als zutreffend vorausgesetzt.

Damit ist einerseits ein praktischer Vorteil für jeden Forscher verbunden: Er muß nicht jedesmal erneut nachweisen, daß z.B. die Technik des ,,Freien Reproduzierens" zur Operationalisierung der Behaltensleistung einer VP tauglich ist, sondern kann dieses Wissen, welches er aus anderen Studien, Untersuchungen etc. hat, ungeprüft auf seine Forschung anwenden. Andererseits ist es durchaus möglich, daß sich in dem Hintergrundwissen fehlerhafte Annahmen etc. befinden, die von dem Forscher jedoch ungeprüft übernommen werden und so potentiell zu falschen Bewertungen seiner WHn führen können.

Für die gewählten Operationalisierungen der theoretischen Begriffe werden (aus der TIH) empirisch-inhaltliche Hypothesen abgeleitet. Im einfachsten Falle jeweils nur einer Operationalisierung pro theoretischer Variablen resultiert auch nur eine inhaltliche Hypothese, die auf diese Operationalisierungen Bezug nimmt. Fehler, die in diesem frühen Stadium der Überprüfung der WH gemacht werden, lassen sich im folgenden kaum noch korrigieren. Daher muß dieser Schritt sorgfältig überdacht und durchgeführt werden.

Die dem Experiment und den dort realisierten empirischen Sachverhalten zugeordneten Hypothesen haben wir oben als empirisch-inhaltliche Hypothesen gekennzeichnet. Es sei hier darauf hingewiesen, daß für ein theoretisches Konzept durchaus verschiedene Operationalisierungen möglich sind und in der Praxis auch angewendet werden (vgl. z.B. für die „Behaltensleistung" die experimentellen Erhebungsmethoden des ,,Freien Reproduzierens", des ,,Wiedererkennens" oder des ,,Wiedererlernens"). Daher ist es durchaus möglich, daß einer TIH in verschiedenen Experimenten unterschiedliche inhaltliche Hypothesen zugeordnet werden, die sich auf die unterschiedlichen Operationalisierungen beziehen. Die TIH kann so über verschiedene EIHn durchaus unterschiedliche Bewertungen erfahren. Es folgt hieraus unmittelbar die Aufforderung an den Forscher, sowohl seine TIHn wie auch die gewählten EIHn im Rahmen des Forschungsberichtes explizit zu machen, damit dem Rezipienten des Berichtes ein kritisches Nachvollziehen ermöglicht wird.

In der TIH ,,Frustrierte Personen reagieren ***zumeist*** aggressiv" taucht der Begriff ***„zumeist"*** auf. Diesen müssen wir bei der Ableitung der EIH ebenso berücksichtigen, wie die Begriffe ,,Frustration" und ,,Aggression". Es sind nicht nur die Konzepte oder Variablen der TIH zu operationalisieren, um eine EIH abzuleiten. Die zwischen den theoretischen Begriffen bestehenden Relationen müssen ebenso auf die EIH übertragen werden. Das bedeutet an

dieser Stelle, daß eine Umschreibung und Klärung des Begriffes ,,zumeist“ erforderlich ist: Wir müssen darlegen, **was *wir* unter „zumeist” verstehen wollen.**

So wäre ein Experiment mit zwei Gruppen denkbar, wobei in Gruppe 1 die VPn z.B. durch unausweichbaren Mißerfolg frustriert werden[1], während dieses in der Gruppe 2 nicht der Fall ist. Es müßte ein Maß für Aggression erhoben werden[2] (Sprechgeschwindigkeit, Anzahl von Verbalinjurien pro Zeiteinheit ...). welches in Gruppe 1 höher liegen sollte als in Gruppe 2. Um den Begriff ,,zumeist“ in dem Experiment zu definieren, muß an dieser Stelle festgelegt werden, was „höher liegen” bedeuten soll (alle VPn der Gruppe 1 haben höhere Werte als die der anderen Gruppe; von den 10 höchsten Werten kommen 75 % aus Gruppe 1 ...?).

Die EIH ist eine **inhaltliche** Hypothese, die sich auf die gewählten Operationalisierungen bezieht. In unserem Beispiel wurde die EIH bspw. lauten können: ,,Die Personen mit Mißerfolgserlebnissen zeigen in der sich anschließenden Diskussion durchschnittlich mehr Verbalinjurien als die Personen ohne induzierte Mißerfolgserlebnisse (wobei die Anzahl von Verbalinjurien innerhalb der Zeit von x Minuten von y unabhängigen und geschulten Beurteilern auf der Kategorienskala Z erhoben wird).“ Auf der Ebene dieser Hypothese werden empirische Daten gewonnen, die zur Beurteilung der Hypothese herangezogen werden sollen. Die Beurteilung der EIH anhand der Daten könnte nach vorher festgelegten, **inhaltlichen** Kriterien erfolgen. Leider liegen solche Kriterien (zumindest derzeit) kaum vor.

Die empirischen Daten psychologischer Experimente sind im allgemeinen fehlerbehaftet, d.h. es wird in aller Regel eine Diskrepanz zwischen den empirisch erhobenen Werteausprägungen und den zugrunde liegenden (wahren) Werten auftreten, die viele Gründe hat (z.B. könnte eine VP müde und unkonzentriert sein und wurde so ihre wirkliche Leistungsfähigkeit nicht zeigen können, sie wurde im psychologischen Test bzw. im psychologischen Experiment einen schlechteren Wert erhalten als unter normalen Bedingungen). Auf derartige Grunde wollen wir hier nicht näher eingehen, wohl aber auf eine Konsequenz, die daraus in den meisten psychologischen Untersuchungen gezogen wird.

Inhaltliche Hypothesen, die sich als quasi-universeile Hypothesen rekonstruieren lassen, werden in der Psychologie zumeist **statistisch** überprüft. Haagen und Seifert sprechen demzufolge von einer „Operationalisierung der psychologischen Hypothese in statistische Konzepte“ (1979, S. 167). Ihnen werden **statistische Hypothesen** zugeordnet, die wir mit Hager (1987) als **statistische**

1 Operationalisierung des Begriffes ,,Frustration“
2 Operationalisierung des Begriffes ,,Aggression“

**Vorhersagen** [SV bzw. SVn] bezeichnen wollen, um in diesem Zusammenhang ihre Funktion im Rahmen des Überprüfungsprozesses einer inhaltlichen Hypothese anzudeuten. Über die SV wird dann in aller Regel mit Hilfe eines oder mehrerer Signifikanztests entschieden. Die in diesem Zusammenhang somit nachgeordneten Signifikanztests testen sehr spezifische statistische Hypothesen, die wir als **Testhypothesen** [$H_0$, $H_1$] bezeichnen wollen. Die Entscheidung über die SV erfolgt im günstigen Fall über einen Signifikanztest, dessen Null (H,) oder Alternativhypothese (H,) der interessierenden SV entspricht. Andernfalls sind mehrere Tests durchzuführen, die einzelnen Entscheidungen über die Null und/oder Alternativhypothesen der Tests sind über logische Verknüpfungen zur Entscheidung über die SV heranzuziehen.

Obgleich eine Vielzahl von SVn als Zuordnung zu den EIHn denkbar sind, werden in der Psychologie nahezu ausschließlich sogenannte Mittelwertshypothesen (unter der sogenannten Normalverteilungsannahme) statistisch überprüft (vgl. auch nächster Abschnitt).

### *2.2.2 Die Beziehung von inhaltlichen Hypothesen und statistischen Vorhersagen*

Wenn statistische Vorhersagen zur Überprüfung inhaltlicher Hypothesen herangezogen werden, wird man in der Regel die Beziehung beider Hypothesenarten näher zu betrachten haben, um die Art und Aussagekraft der Überprüfung beurteilen zu können. Blicken wir nochmals auf o.g. Beispiele ***(4)*** und (5), so wird wohl unmittelbar einsichtig, daß sich beide Hypothesen auf unterschiedliche Dinge und Sachverhalte beziehen. Inhaltliche Hypothesen [Beispiel (4)] haben zumeist (offene) Populationen von **Individuen** als Grundlage, die Gültigkeit der Hypothese wird für alle Individuen der Population behauptet. Diese Individuen sind die **Merkmalsträger,** eine Menge von Personen oder Tieren (VPn, Versuchstiere), denen durch die Hypothese eine spezifische Eigenschaft zugesprochen wird (z.B. aggressiv zu sein).

Statistische Vorhersagen [Beispiel (5)] beziehen sich auf statistische Konzepte, sie haben in diesem gewählten Kontext **Daten** zum Gegenstand, die (isoliert betrachtet) ohne inhaltliche Bedeutung sind und nur bestimmten Auflagen (z.B. normalverteilt zu sein) zu folgen haben. Diese Auflagen stehen in der Regel ohne Verbindung zur WHn und sind auch nur dann notwendig, wenn z.B. (theoretische) Modellparameter im Rahmen statistischer Modelle geschätzt werden sollen (sogenannte Reparametrisierungsbedingungen) oder um Testverteilungen mathematisch zu definieren, die einen (parametrischen) Signifikanztest überhaupt erst ermöglichen (sogenannte Verteilungs- und Varianzhomogenitätsannahmen etc.).

Aus den inhaltlich-psychologischen Hypothesen der hier beschriebenen Art werden normalerweise keine Kriterien hervorgehen, die eine spezifische Zuordnung statistischer Vorhersagen zu den inhaltlichen Hypothesen möglich machen wurde. Solche inhaltlich-psychologische Aussagen lassen sich ,,unmöglich auf eine Aussage über Populationsmittelwerte“ reduzieren (Gadenne, 1984, S. 107), sie lassen sich allenfalls ,,auf dem Umweg über eine Populationsaussage“ testen (Gadenne, ebd.). Wir haben weiter oben davon gesprochen, daß die **Relationen,** die zwischen den Variablen der TIH behauptet werden, möglichst direkt auf die EIH übertragen werden sollten. Dieses ist zumeist relativ problemlos möglich (,,ist größer“, „ist gleich wie“, ,,erniedrigt die Werte“...). Die Relationen lassen sich ebenso auf die SV übertragen (>, =, A ≺ B). Gelingt eine solche Übertragung der Relationen, so spricht Hager (1984) von einer **adäquaten** Zuordnung der SV zur EIH.

Wie wir jedoch gesehen haben, ist eine direkte Übertragung der theoretischen oder empirischen **Variablen** der inhaltlichen Hypothesen auf statistische Variablen der SV nahezu unmöglich. Daraus folgt, daß wir die Überprüfung inhaltlicher Hypothesen durch eine Entscheidung über statistische Vorhersagen als sehr **indirekt** kennzeichnen wollen (z.B. Gadenne, 1984; Schulz, Muthig & Koeppler, 1981). Nicht nur die Beurteilung der TIH kann in verschiedenen Experimenten unterschiedlich sein in Abhängigkeit von der gewählten EIH, sondern auch die Bewertung einer konkreten EIH in Abhängigkeit von der gewählten Zuordnung der SV zu der EIH. In der Forschungspraxis wird diese Zuordnung zudem häufig durch die Verwendung spezifischer Signifikanztests determiniert (Testhypothese und statistische Vorhersage werden nicht unterschieden, die statistische Vorhersage wird nicht mit Bezugnahme auf die EIH kritisch reflektiert und dieser zugeordnet). Wir werden im Rahmen der Validitätsproblematik auf dieses Problem zurückkommen.

Diese nur sehr indirekte Beziehung beider Hypothesenarten wurde in der Vergangenheit des öfteren kritisiert, die Funktion der Statistik in der Psychologie in Frage gestellt. Auch Gadenne (1984) berührt diesen Punkt, wenn er schreibt:

> Ungeklärt ist bereits, wie die in statistischen Tests geprüften Hypothesen mit den psychologischen Theorien, um deren Beurteilung es geht, logisch zusammenhängen. Die Kenntnis dieses Zusammenhanges ist Voraussetzung für die Beantwortung der Frage, welche Konsequenzen die Annahme oder Ablehnung einer statistischen Prüfhypothese für die Beurteilung einer psychologischen Theorie haben sollte. (S. 104)

Zur Lösung dieses Problems wurde u.a. gefordert, daß zwischen EIH und SV eine (logische) **Implikationsbeziehung** [Wenn A, dann B; Aus A folgt B] bestehen (z.B. Bredenkamp, 1980), oder daß die SV folgerichtig aus der EIH abgeleitet sein sollte (Meehl, 1967).

Die EIH sollte über das Prinzip des modus tollens falsifizierbar werden, wenn sich die SV als falsch herausstellen sollte. Diese Falsifikation wurde sich jedoch logisch nur dann rechtfertigen lassen, wenn eine Implikationsbeziehung zwischen EIH und SV bestünde. Eine Bestätigung der EIH wäre danach allerdings nicht möglich bzw. allein mit logischen Argumenten nicht zu rechtfertigen. Leider scheint zumindest aus der Sicht der Forschungspraxis die Annahme einer solchen Beziehung zwischen EIH und SV kaum realistisch zu sein. So erfolgt bspw. die Ableitung und Überprüfung inhaltlicher Hypothesen immer auch unter Rückgriff auf das Hintergrundwissen (vgl. Abschnitt 2.2). Wenn überhaupt, so wäre nur die Verbindung aus inhaltlicher Hypothese und Hintergrundwissen (die logische Konjunktion) falsifizierbar, nicht die EIH allein. Überdies läßt sich die SV mit logischen Mitteln gar nicht falsifizieren, sondern es kann immer nur eine Entscheidung des Forschers über die Gültigkeit oder Ungültigkeit der SV getroffen werden. Da sich jedoch EIH und SV u.E. auf unterschiedlichen und nicht vergleichbaren Ebenen befinden, läßt sich selbst diese Entscheidung nicht direkt auf die EIH übertragen.

Eine Implikationsbeziehung wäre wohl nur dann möglich, wenn die psychologischen (inhaltlichen) Hypothesen anders aussähen als die empirisch-inhaltlichen Hypothesen (EIH), über die wir hier berichten. Die psychologischen Hypothesen [PH] sollten dann neben den inhaltlichen Aspekten auch Hinweise enthalten, die eine zweifelsfreie Übertragung auf die statistische Ebene ermöglichen wurden. In diese Form von PH werden also idealerweise statistische Konzepte einbezogen, sie stellen einen Bestandteil der PH dar (diese Auffassung der PH wird in diesem Band im Kapitel 14 von Erdfelder und Bredenkamp sowie im Kapitel 15 von Steyer vertreten). So gibt es in der Psychologie denn auch inhaltliche Hypothesen, die statistische Konzepte enthalten: Man denke nur an einige der Gesetzmäßigkeiten, die in der Psychophysik aufgestellt wurden. In den allermeisten Fällen sind jedoch Theorien und Hypothesen bei weitem nicht so exakt formuliert: Frustration schafft Aggression.

Aus Sicht der heutigen Forschungspraxis muß man u.E. befürchten, daß eine Forderung, PH ,,statistischer zu formulieren", eher dazu fuhren wird, die PH dann gleich als SVn zu formulieren (Gleichsetzung von Inhalt und Statistik), um damit einigen Interpretationsproblemen aus dem Wege zu gehen. Dieses ließe sich fürwahr bei einigen Hypothesen der psychologischen Forschung bei oberflächlicher Betrachtung vermuten. Ein aufmerksames Lesen der als ,,Diskussion" bzw. ,,Zusammenfassung" ausgewiesenen Textstellen wissenschaftlicher Arbeiten offenbart jedoch fast ausnahmslos, daß die Autoren sich gerade nicht für die statistischen Parameter (und deren Schätzung) interessieren. Sie wollen in der Regel Aussagen über Verhaltensweisen, Wahrnehmungsleistungen, Denkvorgänge, Gedächtnisleistungen und dergl. ihrer VPn treffen.

Da wir die Beziehung zwischen inhaltlichen Hypothesen und statistischen Vorhersagen somit weiterhin als sehr indirekt kennzeichnen wollen und sich für diese Art psychologisch-inhaltlicher Hypothesen eine Implikationsbeziehung zur SV [wenn EIH, dann SV] in nahezu allen Fällen nicht ergeben wird, bliebe es zu erwähnen, daß der Signifikanztests **als Institution** im Rahmen psychologischer Forschung durchaus in Frage gestellt wird (z.B. Harnatt, 1975 oder Morrison & Henkel, 1970).

Wenn über SVn Entscheidungen getroffen werden sollen, ohne den Signifikanztest als (Hilfs)Kriterium zu nutzen, muß zunächst die Fehlerbehaftetheit psychologischer Daten ins Kalkül gezogen werden. Diese Entscheidungen sind wiederum mit spezifischen (zunächst aber unbekannten) **Fehlerrisiken** verbunden, welche wir jedoch kennen oder zumindest vorher abwägen sollten, bevor wir zu Entscheidungen kommen. Beide Problempunkte lassen sich jedoch im allgemeinen nicht aufgrund inhaltlicher Vorannahmen und Theorien lösen.

Die statistischen und mathematischen Theorien, die den bekannten Signifikanztests zugrunde liegen, ermöglichen dadurch, daß sie den Daten bestimmte Verteilungsfunktionen, Lokationsparameter, Streuungen usw. unterstellen, die Abschätzung der Fehlerrisiken, mit denen die Entscheidungen verbunden sind. Damit bietet der Signifikanztest Kriterien an, an denen wir uns orientieren können. Dadurch erhält er eine Rechtfertigung als Entscheidungskriterium in der psychologischen Forschung (vgl. auch Gadenne, 1984), wenngleich an dieser Stelle keineswegs eine Institutionalisierung des Signifikanztests im Sinne einer ,,kochbuchartigen Standarddurchführung“ favorisiert werden soll.

Um die Abschätzung der Fehlerstreuungen und die Kalkulation der Fehlerrisiken zu erlangen, muß der Forscher allerdings bereit sein, den Daten ein spezifisches (mathematisch-statistisches) Datenmodell zu unterstellen und übernimmt damit ferner die Verpflichtung, für das Einhalten der Forderungen des adaptierten Datenmodelles Sorge zu tragen, es gegebenenfalls gegen ein eventuell geeigneter erscheinendes auszuwechseln. Statistische Entscheidungstheorien ermöglichen nur unter diesen Voraussetzungen die valide Abschätzung der Fehlerrisiken (der Entscheidungen).

Das Einhalten o. g. Verpflichtungen ist für den Forscher in einigen Fällen relativ leicht, in anderen Fällen kaum möglich. So ist es für den Forscher sehr schwierig, z.B. auf eine **Normalverteilung** der Daten (und noch dazu in einer meistens nicht bekannten Population) zu achten. Entspricht jedoch die Verteilung seiner Daten nicht der Verteilung, die das Modell des Signifikanztests voraussetzt, so werden die für den Signifikanztest als maximal tolerierbar geplanten Fehlerwahrscheinlichkeiten (die nominellen Werte) von den „realen“ Fehlerwahrscheinlichkeiten, die mit seinen Daten verbunden sind, abweichen.

Die Entscheidungsrationale unseres Forschers fußt in diesem Falle auf falschen oder nicht passenden Voraussetzungen, zumindest was die Verteilung der Daten anbetrifft. Die Konsequenzen dieser Nichtbeachtung solcher Voraussetzungen des Signifikanztests wird allerdings je nach wissenschaftstheoretischer Einbettung anders bewertet, wobei diese Bewertungen mit einer durchaus unterschiedlichen Interpretation des Signifikanztests zusammenhängen (Westermann, 1987b; zusammenfassend Hager, 1987, S. 119-121).

Wenn wir nun an die unterschiedlichen Hypothesenebenen denken, können wir für jede Ebene (getrennt) mögliche Wahrscheinlichkeiten des Zutreffens der Bewertungen (TIH und EIH werden wir nicht weiter trennen) bzw. der Entscheidungen (SV und Signifikanztest) definieren, deren Zusammenhänge weitere Aufschlüsse über die Beziehungen von inhaltlichen und statistischen Hypothesen geben. Zur sprachlichen Vereinfachung wollen wir diese Wahrscheinlichkeiten im folgenden etwas ungenau als „Fehlerrisiken“ bezeichnen. Um die Beziehungen dieser Fehlerrisiken weiter zu untersuchen, übernehmen wir zunächst in Tabelle 2.1 bis Tabelle 2.3 die Definitionen von Westermann (1987a, S.38) und Hager (1987, S.77, 116 und 132). Im nächsten Abschnitt werden wir dann die Zusammenhänge der definierten Fehlerrisiken näher beleuchten.

Auf der (untersten) Ebene der Signifikanztests definieren sich mögliche Fehlerrisiken der Entscheidung zugunsten der $H_0$ bzw. $H_1$-Hypothese wie in Tabelle 2.1 angegeben. Mit $\alpha$ und $\beta$ werden hier die Fehlerrisiken eines Signifikanztests bezeichnet. Es sind dieses die Wahrscheinlichkeiten von Fehlentscheidungen über die Testhypothesen des Signifikanztests.

Tabelle 2.1: Wahrscheinlichkeiten für richtige und falsche Entscheidungen in Signifikanztests

| | Wahrer, aber unbekannter Sachverhalt | |
|---|---|---|
| Entscheidung für | $H_0$ gilt | $H_1$ gilt |
| $H_0$ | $1-\alpha$ | ß |
| $H_1$ | a | 1-b |

Auf der (nächsthöheren) Ebene der statistischen Vorhersage werden die Fehlerrisiken einer Entscheidung über die SV gemäß Tabelle 2.2 definiert. Mit $\varepsilon$ wird die Wahrscheinlichkeit einer fälschlichen Entscheidung zugunsten der SV, mit $\varphi$ die Wahrscheinlichkeit einer fälschlichen Entscheidung zuungunsten der SV bezeichnet.

Auf der (obersten) Ebene der EIH (und TIH) werden die Wahrscheinlichkeiten des Zutreffens der Bewertungen der inhaltlichen Hypothesen gemäß Tabelle 2.3 definiert. Mit f wird die Wahrscheinlichkeit bezeichnet, die EIH aufgrund der durchgeführten Untersuchung fälschlicherweise als nicht zutreffend zu charakterisieren, mit e die Wahrscheinlichkeit, die EIH fälschlicherweise als zutreffend zu bezeichnen.

Tabelle 2.2: Wahrscheinlichkeiten für richtige und falsche Entscheidungen über die statistische Vorhersage aufgrund eines oder mehrerer Signifikanztests

| | Wahrer, aber unbekannter Sachverhalt | |
|---|---|---|
| Entscheidung für | SV gilt | SV gilt nicht |
| SV | $1-\varphi$ | $\varepsilon$ |
| $\neg$SV | $\varphi$ | 1 - e |

Tabelle 2.3: Wahrscheinlichkeiten für richtige und falsche Beurteilungen der empirisch-inhaltlichen Hypothese aufgrund der Entscheidungen über die SV sowie weiterer, inhaltlicher Kriterien

| | Wahrer, aber unbekannter Sachverhalt | |
|---|---|---|
| Beurteilung der EIH als | EIH trifft zu | EIH trifft nicht zu |
| zutreffend | 1 - f | e |
| nicht zutreffend | f | 1 - e |

Obgleich die letztgenannten Wahrscheinlichkeiten e und f numerisch nicht bestimmbar sind, da ihre Ausprägungen zusätzlich noch von anderen Faktoren als nur den statistischen Entscheidungen abhängen (z. B. der internen Validität der Untersuchung), hängen sie doch mit den Fehlerrisiken $\varepsilon$ und $\varphi$ und letztendlich mit $\alpha$ und ß zusammen (vgl. auch nächster Abschnitt). Eine Erhöhung der $\varepsilon$-Fehlerwahrscheinlichkeit wird (andere Faktoren als konstant angenommen) in einer Erhöhung der e-Fehlerwahrscheinlichkeit resultieren und eine Erhöhung der $\varphi$-Fehlerwahrscheinlichkeit wird eine Erhöhung der f-Fehlerwahrscheinlichkeit nach sich ziehen. Die vier Wahrscheinlichkeiten der statistischen Ebene ($\varepsilon$, $\varphi$, $\alpha$ und $\beta$) lassen sich (unter Vorbehalt, vgl. 3.3) numerisch bestimmen (vgl. dazu Hager, 1987). Der Signifikanztest macht die genauere Abschätzung von $\alpha$ und ß möglich (bei unexakten Hypothesen wie z.B. „$\mu_1 > \mu_2$“ eventuell als Maximalwahrscheinlichkeit; vgl. u. a. Hays, 1978). Damit kann es ebenso ermöglicht werden, $\varepsilon$ und $\varphi$ zu bestimmen. Schließlich wird darüber eine gewisse Kontrolle der Fehlerwahrscheinlichkeiten e und f realisierbar.

### *2.2.3 Die Fehlerkontrolle von α und β*

Wenn wir im folgenden einmal annehmen, wir hätten zwei Gruppen von VPn jeweils unterschiedlich frustriert (vgl. oben), die zweite nicht und die erste über irgendeine spezifische Technik im Rahmen der Untersuchung, so wurde gemäß der übergeordneten TIH „Frustrierte VPn reagieren aggressiver" zu erwarten sein, daß die VPn der ersten Gruppe mit erhöhter Aggressivität reagieren. Wenn wir z.B. die „Anzahl von Verbalinjurien gegen den Versuchsleiter innerhalb einer bestimmten Zeiteinheit" [AVV] als Operationalisierung des Begriffes „Aggression" definieren (vgl. Bredenkamp, 1980), sollte dieser Wert in der ersten Gruppe höher sein als in der zweiten (abgekürzt: $AVV_1 > AVV_2$). Wir ordnen dieser EIH die statistische Vorhersage zu, daß der Mittelwert der abhängigen Variablen in der ersten Gruppe **größer** ist als in der zweiten:

$$\text{EIH: } AVV_1 > AVV_2 \Rightarrow \text{SV: } \mu_{AVV_1} > \mu_{AVV_2} \tag{2.1}$$

Das Zeichen „⇒" stellt **keine logische Implikation** dar, sondern soll nur die o. g. Zuordnungsbeziehung verdeutlichen. über das mögliche Zutreffen oder Nicht-Zutreffen der SV läßt sich mit einem einfachen, gerichteten t-Test entscheiden, da diese SV der Alternativhypothese eines gerichteten t-Tests entspricht:

$$\text{SV: } \mu_{AVV_1} > \mu_{AVV_2} \equiv H_1\text{: } \mu_1 > \mu_2 \quad \text{(t-Test)} \tag{2.2}$$

Das Zeichen „≡" deutet eine **Äquivalenzbeziehung** von SV und der Alternativhypothese des t-Tests an. Damit wird ebenso verdeutlicht, daß die SV zwar explizit nur eine Hypothese über zwei Mittelwerte darstellt, jedoch implizit auch spezifische weitere Annahmen (streng genommen damit weitere Hypothesen) macht, die nicht genannt werden (z.B. die Annahme der Normalverteilung etc.). Erst durch diese zusätzlichen Annahmen wird sie zu einer **testbaren** Hypothese. Diese zusätzlichen Annahmen folgen in den allermeisten Fällen nicht aus den inhaltlichen Theorien oder Hypothesen, sondern sind, wie bereits oben erwähnt, mehr technischer Natur.

Es läßt sich nun zeigen (vgl. Hager, 1987), daß die Fehlerwahrscheinlichkeit φ (fälschliche Ablehnung der SV) hier identisch ist mit der Fehlerwahrscheinlichkeit β, die sich aus dem gerichteten t-Test ergibt. Ferner ist die Fehlerwahrscheinlichkeit ε (fälschliche Annahme der SV) numerisch gleich der Fehlerwahrscheinlichkeit α. Wenn wir davon ausgehen, daß alle anderen Faktoren als konstant vorausgesetzt werden können, so resultieren daraus folgende Konsequenzen:

(1) Ein **nicht-signifikantes** Ergebnis des t-Tests wurde die Ablehnung der $H_1$-Hypothese zur Folge haben. Daraus wurde die Entscheidung resultieren, die SV als nicht zutreffend zu bezeichnen. Eine solche Entscheidung kann (und wird im allgemeinen) die Beurteilung zur Folge haben, daß sich die EIH (zumindest in der vorliegenden Untersuchung) nicht bewähren konnte. Je größer also die Fehlerwahrscheinlichkeit $\beta$ wird, desto größer wird damit auch die Wahrscheinlichkeit, daß wir die SV fälschlicherweise ablehnen (9). Dadurch wächst notgedrungen auch die Wahrscheinlichkeit f einer fälschlichen Beurteilung der EIH als „unzutreffend" bzw. „nicht bewährt".

(2) Ein **signifikantes** Ergebnis des t-Tests wurde die Ablehnung der $H_0$-Hypothese zur Folge haben. Daraus würde die Entscheidung resultieren, die SV als zutreffend zu bezeichnen. Diese Entscheidung wiederum kann (und wird im allgemeinen) die Beurteilung zur Folge haben, daß die EIH sich (zumindest in der vorliegenden Untersuchung) bewähren konnte. Je größer also die Fehlerwahrscheinlichkeit $\alpha$ wird, desto größer wird damit auch die Wahrscheinlichkeit, daß wir die SV fälschlicherweise annehmen (E). Dadurch wächst notgedrungen auch die Wahrscheinlichkeit e einer fälschlichen Beurteilung der EIH als „zutreffend" bzw. „bewährt".

Setzen wir nunmehr voraus, daß die Überprüfung der interessierenden inhaltlichen Hypothesen mit möglichst geringen Fehlerrisiken erfolgen sollte ***(stren***ge und faire Prüfungen; vgl. Hager & Westermann, **1983),** so folgt daraus, daß sowohl eine Kontrolle des $\alpha$ **wie auch** des $\beta$-Fehlers des statistischen Tests erfolgen muß. Auf das eher technische Problem, **wie** diese Kontrolle praktisch zu erfolgen hat, gehen wir in diesem Zusammenhang nicht weiter ein (vgl. dazu Hager, 1987). Nötig wird eine **Simultankontrolle** der Determinanten des Signifikanztests, die Hager **(1987)** als **Testplanung** beschrieben hat. Dabei kann sich der „Testplaner" die funktionalen Zusammenhänge der Determinanten des Signifikanztests (die sogenannten Teststärkefunktionen; vgl. Cohen, 1977; Hager & Möller, 1986) zunutze machen, um eine Kontrolle der Einflußfaktoren des Signifikanztests zu erreichen.

Da die Verbindung der hier dargestellten inhaltlichen zu den statistischen Hypothesen als nur sehr „lose" (Meehl, 1967) charakterisiert wurde, wollen wir im nächsten Abschnitt beschreiben, warum es in Anlehnung an die Forschungspraxis günstiger ist, wenn wir von einer **Bewertung** bzw. **Beurteilung** der inhaltlichen Hypothesen (u. a. aufgrund der signifikanztheoretischen Entscheidungen über statistische Hypothesen) sprechen, die zudem nicht ohne subjektive Anteile ist. Der Überprüfungsprozeß führt somit nicht zu einer Verifikation oder Falsifikation der inhaltlichen Hypothesen.

## 2.3 Der Weg von der Statistik zurück zu den Inhalten

Wir haben den Weg der Überprüfung inhaltlicher Hypothesen dadurch umschrieben, daß wir von TIHn ausgegangen sind, aus denen EIHn abgeleitet werden. Diesen EIHn werden SVn adäquat (und ***suffizient;*** vgl. Hager, 1987) zugeordnet, die im Anschluß via Signifikanztests überprüft werden, um zu einer Entscheidung über die **Testhypothesen** einerseits und die übergeordneten SVn andererseits zu gelangen. Der Begriff der Überprüfung der inhaltlichen Hypothesen (d.h. TIH und EIH) wurde zugunsten des Begriffes der Bewertung der inhaltlichen Hypothesen aufgegeben.

Wie läßt sich nun verfahren, wenn wir auf der untersten Ebene des Signifikanztests zu Ergebnissen gekommen sind? Können wir von signifikanten oder nicht-signifikanten Resultaten auf die Gültigkeit oder Ungültigkeit der inhaltlichen Hypothesen schließen? Wie läßt sich (aus dem Blickwinkel der Testhypothesen) der „Weg zurück" beschreiben?

Die Frage nach dem Weg zurück ist gleichbedeutend mit der Frage nach einer Bewertung der wissenschaftlich-inhaltlichen Hypothesen (TIH und EIH). Die Antwort auf diese Frage läßt sich ohne Einbezug unterschiedlicher Güte- oder Validitätskriterien nicht geben. Wir haben die Betrachtung der unterschiedlichen Validitätskriterien an das Ende dieses Beitrages plaziert. In diesem Abschnitt sollen zuvor verschiedene Entscheidungs- und Bewertungsaspekte jeweils getrennt für die einzelnen Ebenen kurz skizziert werden.

### *2.3.1 Die Ebene des Signifikanztests*

Die Entscheidung zugunsten der Alternativ bzw. Nullhypothese eines Signifikanztests ist mit den numerisch bestimmbaren Fehlerrisiken $\alpha$ und $\beta$ verbunden und ist als eine ***konventionalistische*** Entscheidung zu kennzeichnen: Unter jeder Testhypothese sind zunächst alle empirisch-statistischen Ergebnisse möglich, wenn auch unterschiedlich wahrscheinlich. Ein spezifisches empirisches Datum kann insofern nicht gegen oder für eine statistische Hypothese sprechen. Der Forscher wird sich dazu entschließen, eine der beiden sich widersprechenden Hypothesen anzunehmen. Er fallt damit eine Entscheidung.

### *2.3.2 Die Ebene der statistischen Vorhersage*

Wenn die SV direkt in eine testbare statistische Hypothese eines Signifikanztests mündet, so ist die Entscheidung aus der untersten Ebene des Signifi-

kanztests mit der der nächsthöheren Ebene der SV identisch. Entspricht die SV der $H_1$-Hypothese des Signifikanztests, so führt ein signifikantes Resultat im Signifikanztest zur Annahme der SV, ein nicht-signifikantes Resultat wurde zur Ablehnung der SV führen. Wenn die SV mit einer $H_0$-Hypothese eines Signifikanztests identisch ist, wurde ein signifikantes Resultat im Test gerade zur Ablehnung der SV führen und umgekehrt. Die Entscheidung zugunsten bzw. zuungunsten der SV ist mit den Risikowahrscheinlichkeiten $\varepsilon$ bzw. $\varphi$ verbunden. Je nach Beziehung zwischen SV und Testhypothesen kann z.B. $\varepsilon$ der Fehlerwahrscheinlichkeit $\alpha$ oder $\beta$ entsprechen (Genaueres siehe Hager, 1987, S. 131-133).

Etwas komplizierter werden die Beziehungen, wenn **mehrere Signifikanztests** zur Entscheidung über eine **SV** herangezogen werden sollen. Die einzelnen Resultate der Signifikanztests lassen sich **logisch** miteinander verknüpfen. So könnte bspw. aus einer SV folgen, daß jeweils die Alternativhypothesen ($H_{1k}$) von k = 3 gerichteten t-Tests gelten sollten:

$$SV: \mu_1 > \mu_2 > \mu_3 > \mu_4 \equiv (H_{11}: \mu_1 > \mu_2) \wedge (H_{12}: \mu_2 > \mu_3) \wedge (H_{13}: \mu_3 > \mu_4) \quad \textbf{(2.3)}$$

In dem genannten Beispiel werden die drei t-Tests mit einer **Konjunktion** verbunden, d.h. es müßten sich in allen drei Tests signifikante Resultate einstellen, um den Forscher zu einer Entscheidung zugunsten der SV zu bewegen. Die Beziehungen der Fehlerrisiken $\varepsilon$ und $\varphi$ auf der einen und $\alpha$ und $\beta$ auf der anderen Seite werden etwas komplexer, einfache numerische Entsprechungen wie in dem Falle identischer SVn und Testhypothesen, sind nicht mehr ohne weiteres denkbar. Im übrigen ist es zudem möglich, daß die einzelnen Tests mit unterschiedlichen Fehlerrisiken $a_i$ und $\beta_i$ durchgeführt werden. Bei Hager (1987, S. 171-177) finden sich weitergehende Betrachtungen zur sogenannten **Kumulation** der Fehlerrisiken der statistischen Tests. In die Entscheidungen über die SV werden außerdem spezifische Validitätsaspekte einfließen, über die wir jedoch erst im Abschnitt 3 berichten.

### *2.3.3 Die Ebene der inhaltlichen Hypothesen*

Während wir auf den statistischen Ebenen (der Testhypothesen und der SV) von Entscheidungen gesprochen haben, werden wir bzgl. der Ebene der **EIH** nunmehr **von Beurteilungen** sprechen. Die Beurteilung der EIH „im Lichte der Daten“ (basierend auf den Entscheidungen über die SV) wird in Abhängigkeit von verschiedenen Gütekriterien der Untersuchung durchaus unterschiedlich ausfallen und nicht mehr relativ statisch bzw. automatisch wie noch z.B. im Signifikanztest ablaufen. Hier ist bspw. die Frage relevant, inwiefern die **Ableitungsvalidität der statistischen Hypothese** einerseits und die **statistische Validität** andererseits als hinreichend erfüllt angesehen werden kön-

nen. Wir werden auf diese Aspekte im nächsten Abschnitt zu sprechen kommen.

Für die Bewertung der inhaltlichen Hypothesen basierend auf den statistischen Testergebnissen wurden in der Vergangenheit unterschiedliche Vorschläge unterbreitet. Wir werden uns hier auf eine mögliche Konvention stützen, die von Hager (1984) als Erweiterung der Betrachtungen von Westermann und Hager (1982) eingeführt worden ist. Voraussetzung dieser Beurteilungen ist eine hinreichende Erfüllung der Validitätskriterien einerseits und eine entsprechend gelungene Kontrolle der Fehlerwahrscheinlichkeiten $\alpha$, $\beta$, $\varepsilon$ und $\varphi$ andererseits.

Diese Beurteilung stützt sich nicht nur auf die Information, ob die Tests signifikant wurden, sondern gewichtet gleichzeitig Informationen über die eingetretenen statistischen Effekte (Effektstärken). Gestützt auf eine **Testplanung,** die vor dem Experiment stattfand und der Simultankontrolle der Determinanten des Signifikanztests dient (vgl. Hager, **1987),** läßt sich der **Beurteilungsraum der EIH** in vier bzw. drei Bereiche aufteilen (andere Kriterien werden u.a. bei Westermann & Hager, 1982, dargestellt). Danach wird die EIH als **„bewährt“** oder **„bedingt bewährt“** oder **„bedingt nicht bewährt“** oder gar **„nicht bewährt“** betrachtet, je nachdem, ob einerseits die relevante SV angenommen wurde, und andererseits ein vor dem Experiment geplanter statistischer Effekt über bzw. unterschritten worden ist. Die Beurteilung richtet sich im einfachsten Falle der Äquivalenz der SV mit einer Testhypothese danach, ob die SV der $H_0$ oder der $H_1$-Hypothese entspricht und demzufolge der geplante statistische Effekt einen **Mindest-** oder **Maximaleffekt** darstellt. Für vertiefende Betrachtungen sei auf die Ausführungen von Hager (1987) verwiesen.

Der direkte Einbezug der von der Stichprobengröße bekanntermaßen unabhängigen Effektstärke läßt in engen Bahnen einen Einbezug inhaltlicher Überlegungen in den Signifikanztest zu (vgl. auch Hays, 1978). Der Forscher kann bspw. ein Testergebnis trotz signifikantem Ausgang auch als „nur bedingt für die EIH sprechend“ **interpretieren,** wenn die mit dem signifikanten Resultat verbundene Effektstärke ihm als zu niedrig erscheint, um eine **praktische Bedeutsamkeit** zu erhalten.

Die Bewertung der **TIH** aufgrund der bisherigen Überlegungen wird sich anschließend nahezu ausschließlich auf nicht-statistische Überlegungen stützen. Hier werden Bewertungen der Validitäten der Untersuchungen in hohem Ausmaß relevant: Ist die interne Validität der Untersuchung als hinreichend hoch anzusehen, wie sieht es mit der Sicherung der Populationsvalidität, der Situationsvalidität aus? Wurden die theoretischen Variablen annähernd valide operationalisiert (Variablenvalidität; vgl. Hager, 1984, S. 70), wurden die inhaltlichen Hypothesen folgerichtig abgeleitet etc.?

Die Fehlerwahrscheinlichkeiten e und f, die wir weiter oben dargestellt haben, werden in der Regel mit nicht-statistischen Überlegungen verbunden, so daß eine numerische Bestimmung von e und f ausgeschlossen sein wird. Des weiteren werden immer mehr subjektive Elemente Eingang in den Forschungsprozeß finden, je mehr wir uns wiederum den oberen, inhaltlichen Ebenen nähern. Exakt logisch begrundbare Beurteilungskriterien lassen sich auf diesen Ebenen nicht mehr anführen. Muß bspw. ein Forscher davon ausgehen, daß aufgrund vielfältiger Störungen seines Experimentes die interne Validität als nicht gesichert angesehen werden muß, so kann er ebenfalls davon ausgehen, daß seine statistischen Ergebnisse keine Relevanz (zumindest für die TIH) haben werden. Dieses gilt selbst dann, wenn die statistische Validität der Untersuchung als sehr hoch einzuschätzen ist. Wir erkennen an diesem kleinen Beispiel, daß die Betrachtung der unterschiedlichen **Untersuchungsvaliditäten** den eigentlich entscheidenden Anteil an der Beurteilung der inhaltlichen Hypothesen hat.

## *3. Zur Validität der Überprüfung wissenschaftlicher Hypothesen*

Für eine wissenschaftliche Untersuchung werden unterschiedlichste Gütekriterien angegeben, nach denen die Gültigkeit der Untersuchung und deren Aussagen und Bewertungen beurteilt werden können. Je nach der spezifischen inhaltlichen Hypothese und je nach wissenschaftstheoretischer Orientierung des Forschers wird dabei dem einen Kriterium mehr, dem anderen weniger Bedeutung beizumessen sein. Die Gewichtung wird im Einzelfall zu erfolgen haben und kann nicht Gegenstand dieses Textes sein. Insoweit werden wir uns im folgenden nicht so sehr auf die Populationsvalidität, die ökologische Validität, die Situationsvalidität sowie die interne Validität (Ceteris-Paribus-Validität; vgl. Westermann, 1987a) konzentrieren (vgl. dazu z.B. die Abhandlungen von Bredenkamp, 1980 oder Hager & Westermann, 1983), sondern auf Gütekriterien eingehen, die sich direkt auf die Untersuchung und deren Hypothesenbewertungen beziehen. Wir werden dabei die Ausführungen von Hager (1984) aufnehmen bzw. ergänzen. Damit sollen jedoch andere Validitätsaspekte keineswegs als weniger bedeutsam eingeschätzt werden. So sind für eine intern nicht valide Untersuchung alle hier gemachten Überlegungen ohne Bedeutung. Es wäre ohne Relevanz, ob diese Untersuchung bspw. eine hohe Populationsvalidität besitzt etc., denn ohne Sicherung der internen Validität läßt sich nicht begründet darauf schließen, daß die Veränderung in den abhängigen Variablen, die sich in der Untersuchung eingestellt haben, überhaupt etwas mit den unabhängigen Variablen der Untersuchung zu tun haben. Dieses gilt selbst dann, wenn die unabhängigen und abhängigen Variablen der Untersuchung optimal operationalisiert worden sind, dieses gilt ebenfalls selbst dann, wenn die durchgeführten statistischen Tests die bestmöglichen Tests dar-

stellen und die erhobenen Daten eine vollkommene Reliabilität aufweisen wurden.

## 3.1 Die Zuordnung der empirisch-inhaltlichen zu den theoretisch-inhaltlichen Hypothesen

Wir haben erwähnt, daß die theoretischen Variablen der TIH operationalisiert werden, um empirische Variablen konkret zu untersuchen. Die Zuordnung der empirischen Variablen zu den theoretischen Begriffen kann im Experiment durchaus unterschiedlich erfolgen, so daß wir die Güte dieser Operationalisierung durch den Begriff der Variablenvalidität erfassen wollen. Diese kann sowohl für die unabhängigen als auch für die abhängigen Variablen der Untersuchung beeinträchtigt sein. So stellt sich bspw. das Problem, ob das „Auszählen von Verbalinjurien innerhalb einer bestimmten Zeit" inhaltlich betrachtet tatsächlich etwas von dem intendierten Konstrukt der Aggression abzubilden vermag. Weiter kann es durchaus fraglich erscheinen, ob die experimentellen Maßnahmen, die die VPn frustrieren sollen, tatsächlich etwas ausgelöst haben, was sich im Sinne des Konstruktes interpretieren ließe.

Wichtig ist, daß allein die Bedeutungsreduktion des theoretischen Begriffes auf der experimentellen Ebene noch nicht ein Indiz für mangelnde Variablen-Validität darstellt, denn wir werden uns im Experiment immer auf bestimmte Facetten des allgemeineren Begriffes einschränken müssen. Da wir jedoch die Möglichkeit besitzen, mehrere Untersuchungen durchzufuhren, können wir in diesen dann jeweils andere Operationalisierungen i.S. der konzeptuellen Replikation (Bredenkamp, 1980) durchführen. So können wir aus der Feststellung, daß für bestimmte empirische Interpretationen (Operationalisierungen) des theoretischen Begriffes die Hypothesen nicht anwendbar sind, wichtige Erkenntnisse gewinnen. Eine zu geringe oder gar fehlende Variablenvalidität wird jedoch die Validität der gesamten Untersuchung infrage stellen. Leider läßt sich in der Forschungspraxis zumeist nicht eindeutig feststellen, in welchem Umfang die Variablenvalidität eingeschränkt worden ist, es lassen sich keine Werte, keine Validitätskoeffizienten o.ä. angeben, wie man dieses z.B. im Rahmen der Testkonstruktion psychologischer Tests oft versucht. Die Variablenvalidität hängt direkt mit dem Operationalisierungsproblem zusammen, über das wir weiter oben bereits berichtet haben.

Neben dem Problem der Sicherung der Variablenvalidität bleibt zu klären, ob wir bei der Übertragung der **Relationen** zwischen den theoretischen Begriffen auf die empirische Ebene Fehler gemacht haben. Die **Ableitungsvalidität der empirisch-inhaltlichen Hypothese** [AblVal(EIH)] kann beeinträchtigt sein

(vgl. Hager, 1984, S. 63). Einen solchen Fehler wurden wir bspw. dann begehen, wenn in der TIH ausgesagt wurde, daß „A *höher* ***ist*** *als* B“, wir als EIH jedoch „A* ***ist ungleich*** B*“ schließen wiirden[3].

In Abhebung zu der bislang benutzten Definition wollen wir an dieser Stelle den Begriff der Operationalisierung in einer erweiterten Fassung vorstellen: Da die Güte der Operationalisierung von der Variablenvalidität **und** der AblVal(EIH) abhängt, wurde u. E. „Operationalisieren“ demzufolge bedeuten, sowohl die Konzepte wie auch die Relationen zwischen diesen Konzepten einer Beobachtung, Messung und Erfassung zugänglich zu machen.

## 3.2 Die Zuordnung der statistischen Vorhersagen zu den empirisch-inhaltlichen Hypothesen

Der EIH wird eine SV zugeordnet. Die ***Übersetzung*** der empirischen Variablen in statistische scheint (nach unseren vorangegangenen Ausführungen) nicht direkt möglich, nur eine ***Zuordnung*** statistischer Konzepte zu empirischen Variablen erscheint realistisch. Die Güte dieser Zuordnungsbeziehung hängt u. a. vom Skalenniveau der erhobenen Daten ab. Es stellt sich die Frage, ob alle ***relevanten*** Informationen, die die EIH betreffen, sich auf der statistischen Ebene (in einer entsprechend veränderten Form) wiederfinden, Ist dieses der Fall, so können wir mit Hager (1987) von einer **suffizienten** Zuordnung sprechen.

> So könnte eine Einschränkung einer suffizienten Zuordnung vorliegen, wenn in der SV eine Ranghypothese vertreten wird, obgleich sich aus den Daten Mittelwertsinformationen gewinnen lassen. Die Reduktion auf eine Ranghypothese bringt einen Informationsverlust mit sich, da die Abstandsinformationen der Rohdaten verloren gehen.

Werden die Relationen zwischen den Variablen der EIH korrekt auf die statistische Ebene übertragen, haben wir die Zuordnung als **adäquat** bezeichnet. Faßt man beide Kriterien zusammen, so lassen sich Beeinträchtigungen als Störungen der **Ableitungsvalidität der statistischen Vorhersagen** [AblVal(SV)] bezeichnen. So wurde bspw. eine Beeinträchtigung der AblVal(SV) vorliegen, wenn einer EIH „A* ist größer als B*“ die SV zugeordnet werden wurde, daß „$\mu_A \neq \mu_B$“.

3 Mit dem hochgestellten * soll verdeutlicht werden, daß es sich um die Operationalisierung der theoretischen Variablen handelt.

## 3.3 Die Zuordnung der Signifikanztests zu den statistischen Vorhersagen

Eine Beeinträchtigung der **Ableitungsvalidität der Testhypothesen [Abl-Val($H_0$,$H_1$)]** liegt vor, wenn Signifikanztests herangezogen werden, die zur Prüfung der SV nur eingeschränkt geeignet sind. Gehen wir vereinfachend davon aus, daß Mittelwerte eine suffiziente Zuordnung statistischer zu empirischen Konzepten darstellen, dann wurde im folgenden Beispiel eine Beeinträchtigung der AblVal($H_0$,$H_1$) vorliegen:

Ein Forscher bildet vier Gruppen von VPn und äußert die Vermutung, daß die unterschiedlichen Treatments, die den Gruppen appliziert werden, dazu führen, daß die Arbeitszufriedenheit, die über einen spezifischen Fragebogen erfaßt wurde [AZ],über die vier Gruppen ansteigt und in der vierten Gruppe die höchste AZ vorliegen sollte. Das Ansteigen sollte von Gruppe zu Gruppe vorliegen. Er ordnet dieser EIH folgende SV zu: $\mu_1 < \mu_2 < \mu_3 < \mu_4$. Diese SV kann man als adäquate Zuordnung bezeichnen. Der Forscher testet diese Hypothese im folgenden mit einer einfaktoriellen Varianzanalyse und bewertet ein signifikantes Ergebnis als für seine EIH sprechend. Da die einfaktorielle Varianzanalyse als Testhypothese jedoch die $H_0$: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ gegen die unspezifische $H_1$: $\mu_i \neq \mu_j$ für mindestens ein Paar mit $i \neq j$ testet, liegt eine Störung der AblVal($H_0$,$H_1$) vor.

Die Kriterien einer adäquaten und suffizienten Zuordnung verschiedener Hypothesenarten zueinander lassen sich ebenso auf die Beziehung von Testhypothesen und SV anwenden. In dem o. g. Beispiel kann man davon sprechen, daß weder eine adäquate, noch eine **suffiziente Ableitung** der Testhypothesen vorliegt: Die Signifikanz des varianzanalytischen Tests spricht allein noch nicht für die vorangestellte SV. Hätte der Forscher im Anschluß an die Varianzanalyse sogenannte (gerichtete) Post hoc-Vergleiche durchgeführt, so wäre zwar die Beeinträchtigung der AblVal($H_0$,$H_1$) zumindest geringer, ließe sich jedoch aufgrund der wahrscheinlich vergrößerten Fehlerrisiken für $\varepsilon$ und/oder $\varphi$ (die mit den unterschiedlichen Teststärken der verwendeten Testverfahren zusammenhängen) immer noch nicht ausschließen.

> Da es u. E. in Einzelfällen denkbar erscheint, daß ein Forscher irgendeinen leider vollkommen ungeeigneten Signifikanztest als Überprüfung seiner SV heranzieht, wird man in solchen Fällen kaum noch von einer Ableitungsvalidität sprechen können. In diesem sicherlich extremen Fall wurde zwischen der SV und der/den Testhypothese(n) keine Beziehung bestehen und die Fehlerwahrscheinlichkeiten $\varepsilon$ und $\varphi$ wären demzufolge numerisch ***nicht*** bestimmbar.

## 3.4 Die Validität des statistischen Schlusses

Schließlich bleibt noch anzumerken, daß die Verwendung von Signifikanztests auf spezifischen Voraussetzungen beruht, deren Erfüllung bzw. Nichterfüllung zu einer Beeinträchtigung der **statistischen Validität**[4] per se führen können. Während die Fehlerwahrscheinlichkeiten $\varepsilon$ und $\varphi$ mit der $AblVal(H_0,H_1)$ zusammenhängen, bewirkt eine Einschränkung der statistischen Validität (meistens) unkontrollierte Veränderungen der Fehlerwahrscheinlichkeiten für $\alpha$ und ß. Die tatsächlichen Stichprobenkennwerteverteilungen der Tests weichen von den tabellierten ab, die tatsächlichen Fehlerraten erster und zweiter Art können (teils erheblich) von den nominellen, die ein Forscher sich vorgibt, differieren. Die statistische Validität hängt zusätzlich u. a. mit dem Skalenniveau der Daten und einigen anderen Faktoren zusammen, auf die wir in diesem Zusammenhang nicht näher eingehen können (vgl. u.a. Bredenkamp, 1980; Hager & Westermann, 1983 oder Hager, 1987).

# *4. Abschließende Bemerkungen*

Der Prozeß der Beurteilung wissenschaftlicher Hypothesen, ihre Überprüfung, hängt von zahlreichen Bedingungen ab und enthält an vielen Stellen subjektive Komponenten. Die Einteilung der Hypothesen in vier mögliche Ebenen mag für einige Zwecke zu undifferenziert sein, kann jedoch spezifische Probleme der Forschungspraxis verdeutlichen. Sobald ein Forscher zum Zwecke der Bewertung inhaltlicher (und theoretischer) Hypothesen Untersuchungen anstellt, wird er Entscheidungen fallen müssen, die richtig, aber auch falsch, optimal, aber auch suboptimal sein können.

Die unterschiedlichen Validitätsaspekte wissenschaftlicher Untersuchungen behandeln letztendlich diese potentiellen Fehlermöglichkeiten und versuchen sie an zahlreichen Stellen sogar zu quantifizieren. Ein reines „Schielen“ nach signifikanten Ergebnissen vermag wohl den Anforderungen der heutigen psychologischen Forschung kaum noch gerecht zu werden, zumal ein Rücktransfer der (signifikanztheoretischen) „Wahrheit“ auf die inhaltlichen Hypothesen u.E. nicht möglich ist. Die Beziehung zwischen den hier beschriebenen inhaltlichen und den statistischen Hypothesen ist dafür bei weitem zu indirekt, als daß eine solche Übertragung gelingen könnte.

Andererseits wäre wohl auch ein Umformen der inhaltlichen Hypothesen gemäß den Erfordernissen der Statistik und des Signifikanztests nicht das, was u. E. eine Lösung erbringen könnte. Schließlich stellt sich die Psychologie als inhaltliche und naturwissenschaftliche Disziplin dar, die den Methodenkanon

4 Validität des statistischen Schlusses

der Statistik, der Mathematik, eigentlich nur als Hilfswissenschaft nutzt. Die wissenschaftlich sicherlich wichtigen Fragen eines Mathematikers sind eben nicht die inhaltlich relevanten Fragen eines Psychologen. Daraus folgt u.E. konsequent die Unmöglichkeit, valide psychologische Untersuchungen dadurch durchführen zu wollen, indem man nur die statistischen Entscheidungsprozesse optimiert. Sicherlich ist letzteres eine wichtige Perspektive psychologischer Forschung, die endgültigen Bewertungen der inhaltlichen Hypothesen können jedoch immer nur unter Bezugnahme auf nicht-statistische Kriterien gemacht werden, die sich (leider?) auch nicht (noch nicht?) quantifizieren lassen.

## *Literatur*


Bentler, ***P.*** M. (1986). ***EQS - Ein Ansatz zur Analyse von Strukturgleichungsmodellen für normal- bzw. nichtnormalverteilte quantitative Variablen.*** In C.Möbus & W. Schneider (Hrsg.), Strukturmodelle für Längsschnittdaten und Zeitreihen (S. 27-56). Bern: Huber.

Bortz, J. (1984). ***Lehrbuch der empirischen Forschung.*** Berlin: Springer.

Bredenkamp, J. (1980). ***Theorie und Planung psychologischer Experimente.*** Darmstadt: Steinkopff.

Carnap, R. (1936). Testability and meaning. ***Philosophy of Science, 3,*** 419-471.

Cohen, J. (1977). ***Statistical power analysis for the behavioral sciences*** (2nd ed.). Hillsdale, NJ: Erlbaum

Dörner, D. (1979). ***Problemlösen als Informationsverarbeitung (2.*** Auflage). Stuttgart: Kohlhammer.

Gadenne, V. (1976). ***Die Gültigkeit psychologischer Untersuchungen.*** Stuttgart: Kohlhammer.

Gadenne, V. (1984). ***Theorie und Erfahrung in der psychologischen Forschung.*** Tübingen: Mohr.

Groeben, N. & Westmeyer, H. (1981). ***Kriterien psychologischer Forschung (2.*** Aufl.). München: Juventa.

Haagen, K. & Seifert, H. G. ***(1979). Methoden der Statistik für Psychologen (Band 2).*** Stuttgart: Kohlhammer.

Hager, W. (1984). Aspekte eines deduktiven Forschungsansatzes in der empirischen Pädagogik: Fragen der Hypothesenvalidität, der Untersuchungsplanung und der Hypothesenbeurteilung. *Zeitschrift für Empirische Pädagogik und Pädagogische Psychologie, 8,* 55--75.

Hager, W. (1987). ***Grundlagen einer*** Versuchsplanung ***zur Prüfung empirischer Hypothesen in der Psychologie.*** In G. Lüer (Hrsg.), Allgemeine experimentelle Psychologie (Kap. 3, S. 43-264). Stuttgart: Fischer.

Hager, W. & Möller, H. (1986). Tables and procedures for the determination of power and sample sizes in univariate and multivariate analyses of variance and regression. ***Biometrical*** Journal, 28, 643-667.

Hager, W. & Westermann, R. (1983). Planung und Auswertung von Experimenten. In J. Bredenkamp & H. Feger (Hrsg.), ***Hypothesenprüfung. Enzyklopädie der Psychologie, Forschungsmethoden der Psychologie,*** Bd. 5 (S. 24-238). Göttingen: Hogrefe.

Harnatt, J. (1975). Der Signifikanztest in kritischer Betrachtung. ***Psychologische Beiträge,*** 17, 595-612.

Hays, W. L. (1978). ***Statistics for the social sciences*** (2nd ed.). New York: Holt, Rinehart & Winston.

Hempel, C. G. (1974). ***Philosophie der Naturwissenschaften.*** München: DTV.

Huber, O. (1987). ***Das psychologische Experiment.*** Stuttgart: Huber.

Hussy, W. (1983). Komplexe menschliche Informationsverarbeitung: das SPIV-Modell. ***Sprache & Kognition, 2, 47-62.***

Hussy, W. & Eye, A. von (1988). ***On cognitive operators in information processing and their effects on short-term memory performance in different age groups.*** In F.E. Weinert & M. Perlmutter (Eds.), Memory development: Universal changes and their individual differences (pp.275-291). Hillsdale, NJ: Erlbaum.

Jöreskog, K. G. & Sörbom, D. (1988). ***LISREL 7 - a guide to the program and applications.*** Chicago: Scientific Software Inc.

Krapp, A., Hofer, M. & Prell, S. (1982). Forschungs-Wörterbuch. München: Urban & Schwarzenberg.

Lepper, M. R., Greene, D. & Nisbett, R.E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the 'overjustification' hypothesis. **Journal *of Personality and Social Psychology, 28,*** 129-137.

McGuigan, F. J. ***(1968). Experimental psychology*** (2nd Ed.). Englewood Cliffs, Ca.: Prentice-Hall.

Meehl, P.E. (1967). Theory testing in psychology and physics: A methodological paradox. ***Philosophy of Science,*** 34, 103-115.

Möller, H. (1991). ***Möglichkeiten und Grenzen Linearer Strukturmodelle zur Parametrisierung ereigniskorrelierter Potentiale.*** Eine Untersuchung am Beispiel von emotional bedeutsamem Reizmaterial. Trier: WVT.

Morrison, D.E. & Henkel, R.E. (Eds.). (1970). ***The significance test controversy.*** Chicago: Aldine.

Neyman, J. & Pearson, E.S. (1933). The testing of statistical hypotheses in relation to probabilities a priori. ***Proceedings of the Cambridge Philosophical Society, 29,*** 492-510.

Opp, K. D. ***(1976). Methodologie der Sozialwissenschaften*** (3. Aufl.). Reinbek bei Hamburg: Rowohlt.

Schulz, T., Muthig, I? & Koeppler, K. (1980). ***Theorie, Experiment und Versuchsplanung in der Psychologie.*** Stuttgart: Kohlhammer.

Westermann, R. (1987a). ***Wissenschaftstheoretische Grundlagen der experimentellen Psychologie.*** In G.Lüer (Hrsg.), Allgemeine experimentelle Psychologie (Kap. 2, S. 5-42). Stuttgart: Fischer.

Westermann, R. (1987b). ***Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie.*** Berlin: Springer.