erschienen in: Th.Herrmann & W.Tack (Hrsg.) (1993). *Methodische Grundlagen der Psychologie. Enzyklopädie der Psychologie, Bereich B, Serie I, Bd.1*, (S.558-603). Göttingen: Hogrefe.

13. Kapitel

Von Zahlzeichen zu Skalen

Rainer Mausfeld

1. Der Zahlgebrauch in der Psychologie

Testen, Skalieren, Messen: drei Bezeichnungen in der Psychologie für einen Sachverhalt, den der Physiker Messung nennt oder den er als ‚meßbar machen' umschreibt, falls er erst die Grundlagen der Messung bereitstellt. Die begriffliche Vielfalt in der Psychologie ist nicht Resultat einer logisch-systematischen Einteilung unterschiedlicher Meßoperationen im Bereich des Psychischen, sondern vorrangig historische Konsequenz unterschiedlicher Perspektiven, aus denen heraus in der Psychologie das Meßproblem betrachtet wurde.

In der diagnostischen oder psychometrischen Tradition stehen der unterschiedliche Ausprägungsgrad von Eigenschaften verschiedener Individuen, wie beispielsweise Intelligenz, Konzentrationsfähigkeit und Neurotizismus, im Mittelpunkt des Interesses. Diese vorrangig sozialtechnologisch orientierte Tradition versteht unter dem Begriff des Messens die Erstellung geeigneter Indizes innerhalb statistischer Vorhersagetechniken; die Bewertung dieser Indizes ergibt sich dabei eher aus der pragmatischen Brauchbarkeit denn aus einem Theoriegefüge.

In der psychophysikalischen Tradition werden allgemeine Prinzipien der Abhängigkeit ‚subjektiver Dimensionen', wie Lautstärke, Helligkeit und Farbton, von den ihnen zugrunde liegenden physikalischen Größen untersucht; diese Abhängigkeiten sucht man durch Skalen zu modellieren und die Beziehungen zwischen verschiedenen psychologischen ‚Dimensionen' zu charakterisieren. Mit dem Ziel, die Natur und Struktur perzeptueller Codes zu erfassen, orientiert sich hier der Meßbegriff an dem Vorbild der Physik.

In der meßtheoretischen Tradition schließlich werden abstrakt die Bedingungen behandelt, unter denen eine physikalische oder psychologische Eigenschaft durch Zahlen im Sinne der Mathematik charakterisiert werden kann. Hier sucht man für psychologische Größen den Größenbegriff in ähnlich präziser Weise zu fassen, wie die Physik dies für ihren Größenbegriff tut, und zugleich zu zeigen, daß psychologische Relationen strukturell so reichhaltig sein können, daß sie sich durch numerische Relationen symbolisch repräsentieren lassen.

In der Geschichte des Gebrauchs von Zahlen lassen bis heute – stark vereinfachend – zwei Pole bestimmen, die in gewisser Weise die historische Entwicklung der Arithmetik umspannen. An dem einen Pol finden wir den konkreten Umgang mit individuellen Zahlzeichen; dieser steht in engem Zusammenhang mit Alltagsproblemen, etwa des Handwerks, des Warentausches oder konkreter Maßtechniken. Der operative Umgang des Zählens und Rechnens differenzierte sich allmählich in vielfältiger Weise aus und führte, wie bei den Babyloniern, zu hochentwickelten Rechentechniken und Regelsystemen für das Operieren mit Zahlzeichen. Dieser konkret-operativen Umgangsweise steht als anderer Pol eine abstrakt-theoretische Auffassung von Zahlen gegenüber. Historisch findet sich diese Haltung zuerst bei den Griechen, die eine abstrakte Definition der Zahl suchten und erkannten, daß der Begriff der Zahl dem der Größe untergeordnet ist. Zahlen stellen hier eigenständige theoretische Entitäten dar, die ihre Bedeutung durch eine umfassende abstrakte Theorie erhalten. Verschiedene Zahlzeichen können dabei dieselbe Zahl darstellen. Dieser hochgradig theoretische Gebrauch des Konzeptes der Zahl ist kennzeichnend für die Mathematik.

Beide Pole des Zahlengebrauchs spielten bei der Entwicklung der neuzeitlichen Naturwissenschaft eine wichtige Rolle; die moderne Naturwissenschaft verdankt ihre theoretischen Erfolge jedoch vorrangig der Anbindung ihrer Theorieentwicklung an die Mathematik. Erst mit dem theoretischen Konzept der Zahl, wie es die Mathematik entwickelte, konnte die Idee einer umfassenden Quantifizierung ihre Faszination als Leitbild der Theorieentwicklung entfalten. Ob man, wie Cusanus, das Wort ‚mens' mit dem Wort ‚mensura' in Beziehung setzt oder wie Kepler der Überzeugung ist, wie sich das Auge auf das Erkennen von Farben und das Ohren auf das Erkennen von Tönen richte, so richte sich der menschliche Geist auf die Erkenntnis des Quantitativen: in dem Maße, wie sich in der neuzeitlichen Wissenschaft das Streben nach Naturerkenntnis mit dem Bedürfnis nach Gewißheit dieser Erkenntnis verbindet, läßt sich ein theoretischer Erkenntnisfortschritt nur in Anbindung an die Mathematik erreichen.

Die Erfolge, welche das quantitativ-theoretische Denken für die Theorieentwicklung der Physik verzeichnen konnte, trugen dazu bei, daß auch die Psychologie ihre Entwicklung als eigenständige, empirisch orientierte Wissenschaft an die Bedeutung band, die sie dem Quantitativen in ihrer Theorieentwicklung zu geben vermochte. In jenen Bereichen der Psychologie, deren Entwicklung durch theoretische Intentionen geleitet war, wie die Fechnersche Psychophysik, wird man eine solche Haltung, in geeigneter Präzisierung, durchaus als berechtigt ansehen können. Doch folgte die Psychologie auch in jenen Bereichen diesem Leitbild, wo sie sich für die Bewältigung von Alltagsproblemen mit einer operativen Verwendung von Zahlzeichen hätte zufrieden geben können. Ohne Beachtung der schon in ihrer Geschichte angelegten Heterogenität der Psychologie wird man die Rolle des Quantitativen in der Psychologie nicht diskutieren können. Theoretische Intentionen auf der einen Seite und konkrete, mit Alltagsproblemen verbundene Zielsetzungen auf der anderen charakterisieren in der Psychologie zwei sich parallel entwickelnde Wissenschaftsbereiche, die trotz mannigfacher Querbezüge von gänzlich unterschiedlichen Prinzipien und Erkenntniszielen geprägt sind. Als Stellvertreter für diese beiden Wurzeln der Psychologie seien Fechner und Binet genannt. Dem Quantitativen kommt nun in beiden – hier schematisch gegenübergestellten – Bereichen eine jeweils andere Bedeutung zu: Sucht man eine Theorie über die Natur und Funktionsweisen des menschlichen Geistes zu erstellen, so kann man – wie Fechner – fragen, inwieweit der Aufbau einer solchen Theorie auf der Entwicklung eines eigenständigen Konzeptes psychologischer Größen beruht und ob in dieser Theorie Konzepte, die sich in Analogie zur Physik als subjektive Dimensionen beschreiben lassen, eine Rolle spielen. Zielt man jedoch aus einer sozialtechnologischen Perspektive auf die Erfassung und Beherrschung von Zusammenhängen in komplexen Systemen, so möchte man quantitative Indizes gewinnen, die sich zur Lösung des jeweiligen Problems als besonders geeignet erweisen. Nicht nur der Erklärungsbegriff ist hier ein anderer, sondern auch die im Vordergrund stehenden Aspekte des Konzeptes der Zahl. Die sich in einer naturwissenschaftlich orientierten Psychologie möglicherweise stellende Frage, ob ihre theoretischen Strukturen ähnlich denen der Physik zu eigenständigen Größen, Dimensionen und Skalen führen, ob also die Zahlzeichen durch den strukturellen Gehalt der Theorie als Zahlen aufgefaßt werden können, ist für eine sozialtechnologisch orientierte Psychologie unerheblich (vgl. Herrmann, in diesem Band, Kap. 6.5.3). Die funktionale Beherrschung eines Systems setzt weder eine substantielle, an einem kausalen Erklärungsbegriff orientierte Theoriebildung noch einen eigenständigen Größenbegriff voraus. Die empirische Struktur der Indizes, wie man sie durch Tests, Rating-Skalen u.ä. gewinnt, kann so reichhaltig sein, daß sie die jeweils erforderlichen Einsichten in die Funktionszusammenhänge eines komplexen Systems bereitstellen kann. Daß derartige Indizes, d. h. Zahlzeichen, innerhalb komplexer Rechenprozeduren, etwa mathematisch-statistischer Schätz- und Prüfverfahren, wieder als Zahlen im theoretischen Sinne (in der Regel als reelle Zahlen) aufgefaßt werden, rechtfertigt nicht, diese Indizes als Messung einer psychologischen Größe, d. h. als Skala, zu betrachten – es sei denn man liberalisiert die Begriffe der Größe, Dimension und Skala in entsprechender Weise; mit dieser Liberalisierung verliert jedoch das Problem der Quantifizierbarkeit psychologischer Größen sein theoretisches Interesse. (Die Interpretation von Parametern in einem wahrscheinlichkeitstheoretischen Modell als Skalenniveau ist Folge einer solcher Bedeutungsliberalisierung.) Das naturwissenschaftliche Ideal einer Quantifizierung bindet das Quantitative an den strukturellen Gehalt einer substantiellen Theorie. Statistische Theorien können aber kein Ersatz für inhaltliche Theorien sein; losgelöst von diesen können sie zu einer Quantifizierung nicht beitragen.

Dieses Kapitel befaßt sich mit der Frage, ob es eigenständige Größen in der Psychologie gibt, die sich durch Skalen numerisch darstellen lassen; es diskutiert einige der mit ihr verbundenen Konzepte und Probleme. Nicht zuletzt aus den genannten Gründen soll dies exemplarisch am Beispiel der Psychophysik erfolgen: Die Psychophysik ist zur Untersuchung einer solchen Frage in einer vergleichsweise glücklichen Lage, da sie über Bereiche mit reichhaltigen theoretischen Strukturen verfügt; auch lassen hier die Möglichkeiten einer präzisen physikalischen Beschreibung des Reizes sowie seine hohe experimentelle Kontrollierbarkeit es am ehesten als sinnvoll erscheinen, den Versuch einer Definition psychologischer Größen zu unternehmen und die Beziehungen dieser Größen zueinander zu studieren. Die Frage der Meßbarkeit und Quantifizierung, wie sie sich hier stellt, läßt zudem die Bezüge zum allgemeinen Problem der Rolle des Quantitativen in den Naturwissenschaften deutlich erkennen.

Innerhalb der Psychologie wird man beim Konzept der Skala überwiegend an Skalen aus dem Bereich der Sozialpsychologie und Diagnostik denken: Aggressions- und Vorurteilsskalen sind ebenso geläufig wie Intelligenz- und Neurotizismusskalen. Der Skalenbegriff spielt also gerade in jenen Bereichen der Psychologie eine wichtige Rolle, die sich mit der Entwicklung einer Methodologie zur Erfassung komplexer Funktionszusammenhängen beschäftigen. Doch hängt dieser Skalenbegriff nur noch metaphorisch mit dem Skalenbegriff zusammen, aus dem die Frage der Meßbarkeit psychologischer Größen ihre theoretische Bedeutung bezieht. Für die Einstellungsmessung und Diagnostik ist dies freilich keine Einschränkung, da die Erreichbarkeit ihrer jeweiligen Ziele in keiner Weise davon abhängt,-daß es einen eigenständigen psychologischen Größenbegriff gibt; daher ist es schwer verständlich, warum sich hier der Blick oftmals auf Probleme einengt, die mit einer Rechtfertigung der Meßbarkeit verbunden sind. Bei der Behandlung gesellschaftlicher Probleme oder als interessant erachteter Alltagsfragen haben numerische Indizes ganz unterschiedlicher Art seit jeher eine große Rolle gespielt (s. Gould, 1981). Die Psychologie hat für die Behandlung dieser Probleme ein umfangreiches Instrumentarium und eine eigene Methodologie hervorgebracht, deren pragmatischer Nutzen jenseits der metatheoretischen Frage einer Meßbarkeit des Psychischen außer Zweifel steht. Doch wirft die Diskrepanz zwischen Methodologie und theoretischem Substrat, zu der Suppes (1962, S.260) bemerkt, „It is a paradox of scientific method that the branches of empirical science that have the least theoretical developments often have the most sophisticated methods of evaluating evidence", wiederum Fragen auf. Insbesondere birgt sie einerseits die Gefahr, daß in den Teilen der Psychologie, die ihre psychologisch-theoretischen Ansprüche reduzieren und ihren Gegenstandsbereich unter dem Postulat der unmittelbaren Anwendbarkeit verkürzen, eine Funktionalisierung der Theorieentwicklung für externe Zwecke (Böhme et al., 1974) überhand nimmt und daß andererseits bei einer Behandlung des Meßbarkeitsproblems Fragen nach der Inkongruenz von theoretischem Verständnis und sozialtechnischer Brauchbarkeit als ungerechtfertigt erscheinen und die Methodologien selbst als Lösung des Meßbarkeitsproblems betrachtet werden.

Im folgenden untersuchen wir zunächst den Begriff der psychologischen Größe und beschäftigen uns mit der Frage, ob es bereits a priori Einwände gegen eine theoriebezogene Quantifizierung des Psychischen gibt (die weiterführende Frage, ob die Theorieentwicklung in der Psychologie tatsächlich zur Entwicklung eigenständiger fundamentaler Skalen geführt hat, werden wir hier nicht behandeln können; vgl. jedoch Mausfeld, in diesem Band, Kap. 4). Sodann werden wir einige jüngere Entwicklungen, die mit dem Ziel der Skalenkonstruktion verbunden sind, diskutieren. Die verschiedenen Arten der Skalenkonstruktion können nach vielerlei Gesichtspunkten klassifiziert werden, beispielsweise nach der Art der sich aus unterschiedlichen psychologischen Relationen ergebenden Datenstruktur (Coombs, 1964; Roskam, 1983). Da in diesem Kapitel metatheoretische Betrachtungen im Vordergrund stehen sollen, scheint eine abstraktere und zugleich gröbere Klassifikation geeigneter zu sein, in der sich grundsätzliche theoretische Perspektiven der Skalenkonstruktion ausdrücken. Schließlich werden wir uns der Frage zuwenden, inwieweit man sicherstellen kann, daß numerische Relationen und Aussagen in sinnvoller Weise an psychologische angebunden sind.

2. *Qualität und Quantität: Ist Psychisches meßbar?*

‚Quantität' und ‚Messung', auf diesen Begriffen, die im Erkenntnisprozeß eine besondere Wirkungskraft zu entfalten scheinen, gründet sich wesentlich die epistemologische Dignität der Naturwissenschaften, voran der Physik. Ihren Erkenntnisidealen folgend bemühte sich auch eine empirisch orientierte Psychologie mit der Eliminierung alles philosophisch Spekulativen um den Nachweis der prinzipiellen Möglichkeit einer Quantifizierung des Psychischen.

Die klassische Grenzziehung von Qualität und Quantität war in der Physik schon früh durch eine fortlaufende Neubestimmung dieser Begriffe aufgehoben worden. Die Physik löste sich von der unmittelbaren Erfahrung, qualitative Unterschiede suchte man auf Differenzen des Quantitativen zurückzuführen, sei es der geometrischen Struktur oder der Zahl. Mit der pythagoräischen Idee, die Einheit der Welt in einer abstrakten Harmonie zu suchen, bildet sich ein neues Erkenntnisideal aus. Es kehrt wieder im neuzeitlichen Gedanken einer Mathesis Universalis, einer allgemeinen Theorie der Größen und Größenverhältnisse, in der die sich ausdifferenzierenden und zunehmend von der Welt unmittelbarer Erfahrungen sich lösenden Wissenschaften eine methodische Einheit finden sollen. Seitdem prägte es die Entwicklung der neuzeitlichen Naturwissenschaften, voran der Physik, auf deren wichtigstes Charakteristikum es hinweist: ihre Anbindung an die Mathematik.

Zu einer Mathesis *Universalis* zählen all jene Unterfangen, wo, so Descartcs, „nach Ordnung und Maß geforscht wird, und daß es hierbei gar nicht darauf ankommt, ob man dieses Maß nun in den Zahlen, oder in den Figuren oder den Gestirnen oder den Tönen oder in irgendeinem anderen Gegenstand zu suchen hat" (zit. nach Mittelstraß & Schroeder-Heister, 1986). Der für diese formale Einheit zu entrichtende Preis war freilich ein Wachsen der Kluft zwischen alltäglicher Erfahrungswelt und Welt der Physik. Eben dieses Ideal einer theoretischen Quantifizierung, das sich in der Physikgeschichte allmählich entfaltete, tritt mit G. Th. Fechner weitgehend unvermittelt in die Psychologiegeschichte ein (daß Fechner den Neologismus Psychophysik für sein Forschungsprogramm bildet, ist ein Indiz für diese Diskontinuität) und stellt fortan die Richtschnur eines sich naturwissenschaftlich begreifenden psychologischen Dcnkens dar. Hingegen fand das ursprüngliche Warum, Fechners philosophische Perspektive nämlich, der zählende und messende Gott der Weisheit Salomons habe auch die psycho-physische Einheit nach Zahl und Maß geordnet, ihres metaphysischen Gehaltes wegen in der sich rein empirisch verstehenden Psychologie keinen Platz mehr. Fechner ging es nämlich, vor dem Hintergrund des Leib-Seele-Problems, noch im eigentlichen Sinne um ‚Empfindungsmessung'.

Hat nun das Fechnersche Forschungsprogramm in seinen modernen Varianten den Anspruch eingelöst, psychologische Größen in einem theoretisch fruchtbaren Sinne zu quantifizieren und der Psychologie durch so erstellte Skalen ein theoretisches Fundament bereitzustellen? Allein schon die Auseinandersetzung mit dieser Fragen läßt hoffen, daß dadurch der Gegenstandsbereich der Psychologie selbst erhellt wird.

Zwei Problemkreise, zwei Begriffskomplexe berühren sich in der Frage nach der Metrisierbarkeit psychologischer Größen. Da ist zunächst der Begriff der psychologischen Größe oder Eigenschaft. Zweitens der Begriff der Metrisierbarkeit und mit ihm verbunden die Begriffe ‚Zahl', ‚Größe' und ‚Quantität'. Welches Verhältnis haben Größen, seien es nun Länge, Gewicht oder Fechners Konzept der Stärke der Empfindung, und Zahlen zueinander? Warum lassen sich die aus der Wirklichkeit abstrahierten Größen durch mathematische Entitäten so erfolgreich beschreiben?

Werfen wir einen kurzen Blick auf das Verhältnis von Zahlen und empirisch interpretierten Größen. Diese Beziehung ist für unser Verständnis von Wissenschaft grundlegend, steht doch „am Anfang der Mathematik und am Anfang der Philosophie und am Anfang des abendländischen Denkens diese Begeisterung für die Zahl" (Martin, 1956, S. 16). Im griechischen Sinne waren Zahlen nur das, was in heutiger Terminologie positive ganze Zahlen sind. Allgemeiner als der Begriff der Zahl war der der Größe. Als Größe wurde alles aufgefaßt, was in Bestandteile zerlegbar ist, wobei die Teile einer Größe für sich bestehen können und selbst wieder Größen sind, wie etwa die Teile einer Länge wieder Längen sind. Diese beiden Konzepte, Zahl und Größe, welche entscheidende Wurzeln des naturwissenschaftlich-mathematischen Denkens bildeten, aus denen heraus Mathematik und Physik gleichermaßen erwuchsen, ließen sich durch das Konzept der Proportion (als einem Verhältnis zweier Größen) in eleganter Weise miteinander in Beziehung setzen, freilich nur unter der stillschweigenden Voraussetzung, daß je zwei Größen kommensurabel sind, d. h. ein gemeinsames Maß haben. Mit der Erschütterung, welche die Entdeckung der Inkommensurabilität auslöste, mit dem Nachweis von Größen nämlich, die keine durch ganze Zahlen ausdrückbare Proportion haben, zerbrach auch die ursprünglich angenommene Beziehung von Zahl und Größe. So finden wir bei Euklid Zahlen und Größen streng geschieden. Es war der von Eudoxos entwickelte Kalkül mit Größenverhältnissen und keineswegs das griechische Konzept der Zahl, das zur Entwicklung der so mächtigen Theorie der reellen Zahlen führte (zwischen der klassischen Proportionenlehre und der Theorie der Automorphismengruppe extensiver Strukturen besteht eine enge Beziehung, wie Krull, 1960, gezeigt hat). Eine der heute üblichen Präzisierungen des Konzepts der reellen Zahlen, nämlich die von Dedekind, knüpft unmittelbar an das 5. Buch Euklids an, das die Darstellung der Eudoxischen Größenlehre zum Inhalt hat. Angesichts diese Entwicklung sieht man, daß unser heutiges Konzept der Zahl, das ja allem Quantifizieren und Messen unterliegt, historisch selbst aus der Behandlung geometrischer und physikalischer Größen erwachsen ist. Die reellen Zahlen sind nicht etwas in der Mathematik Vorgefundenes, sondern eine Abstraktionsleistung des menschlichen Geistes, konstruiert zu dem Zwecke, Größen gedanklich handhabbarer zu machen. Die reellen Zahlen stellen gleichsam eine universelle Größenstruktur dar, durch welche die formale Struktur aller eindimensionalen Größen erfaßt wird (vgl. Bourbaki, 1966, S. 12). Interessanterweise knüpft auch in der abstrakten Meßtheorie Niederée (1987, 1992a,b) an die Proportionenlehre an; er entwickelt eine modelltheoretische Verallgemeinerung von ihr, die er zur Grundlegung und Verallgemeinerung meßtheoretischer Konzepte, wie sie im folgenden beschrieben werden, heranzieht. Anders als der herkömmliche Repräsentationsansatz, doch diesen gleichsam fundierend, geht Niederée nicht von einem vorgegebenen Zahlbereich (wie dem der reellen Zahlen) als Grundmengen der repräsentierenden numerischen Struktur aus.

Statt dessen geht sein Typenansatz von einem abstrakten Konzept des iterativen *Meßverfahrens* aus, und ‚Zahlen'[6] im Sinne von ‚Meßwerten' werden sodann im Kontext der jeweiligen qualitativen Struktur auf theoretischer Ebene aus dem betreffenden Meßverfahren heraus gleichsam konstruiert. Dieser Ansatz ermöglicht zum einen eine befriedigendere Behandlung einer Reihe grundsätzlicher Fragen (z. B. zur ‚Konstruierbarkeit' von Skalen und zur Rolle Archimedischer Axiome) als bisher möglich. Darüber hinaus erweitert sich der ‚meßtheoretische Horizont', da im Prinzip auch Zahlbereiche denkbar erscheinen, welche wesentlich anderer Natur sind als der Bereich der reellen Zahlen (der ja seinem Ursprung nach primär auf elementare physikalische und geometrische Probleme zugeschnitten ist).

Meßtheoretische Betrachtungen im eigentlichen Sinne wurden durch erkenntnistheoretische Fragen der Physik veranlaßt. Man wandte sich in abstrakter Weise der Größenmessung selbst zu, um die erkenntnistheoretischen und formalen Grundlagen zu legen, durch die jede Art der Größenmessung ihre Rechtfertigung finden soll. Diesem Ziel widmete 1887 Helmholtz seine Schrift *Zählen und Messen erkenntnistheoretisch betrachtet.* „Was", so Helmholtz, „ist der objective Sinn davon, dass wir Verhältnisse reeller Objecte durch benannte Zahlen als Größen ausdrücken, und unter welchen Bedingungen können wir dies thun?" Auf der Suche nach einer Antwort stellte Helmholtz – hierbei zugleich eine Präzisierung des traditionellen Begriffs der extensiven Größe anstrebend – der arithmetischen Addition eine physikalische Verknüpfung, der arithmetischen Gleichheit ein Verfahren der „physikalischen Vergleichung" gegenüber. Eine Größenmessung für die betrachteten Objekte ist nun, Helmholtz zufolge, genau dann gegeben, wenn die den Objekten zugeordneten Zahlen die physikalische Gleichheit respektieren und wenn die arithmetische Addition mit der physikalischen Addition verträglich ist. Das durch den Zahlgebrauch als Quantitativ Aufscheinende ist nichts anderes als das Spiegelbild der qualitativen Relationen zwischen den Objekten: eine folgenreiche Relativierung (oder genauer eine Neufassung) der traditionellen Dichotomie von Qualität und Quantität. Diese Behandlung von Größen wurde in einer einflußreichen Arbeit von Hölder (1901) mathematisch weitergeführt und erweitert; damit waren die Grundlagen der abstrakten Meßtheorie gelegt, die später im Scott-Suppes-Paradigma zur Standardtheorie der *fundamentalen Messung* in Psychologie und Okonomie wurden.

Helmholtz zufolge scheint jedes fundamentale Messen extensives Messen zu sein und erfordert folglich eine additionsanaloge Verknüpfung zwischen den empirischen Entitäten. Unter *fundamentalem Messen* wird, nach einer von dem Physiker N. R. Campbell in den 20er Jahren getroffenen Unterscheidung, ein Messen verstanden, daß im Gegensatz zur sog. *abgeleiteten Messung* grundlegenden Charakter hat und nicht selbst wiederum auf einer Messung beruht. In modernen Worten: Fundamentale Größenmessung liegt dann vor, wenn die Zahlen den qualitativ-strukturellen Gehalt der physikalischen Relationen und Operationen treu widerspiegeln. Mit der Anbindung des fundamentalen an das extensive Messen schien a priori eine fundamentale Messung psychologischer ‚Größen' ausgeschlossen zu sein, denn was ließe sich unter einer Addition im Bereiche des Psychischen vorstellen? Fechners Lösungsvorschlag, eine Empfindung mit Hilfe gerade bemerkbarer Unterschiede als eine „Summierung eines Soundsovielmal des Gleichen" zu bestimmen, hatte mit der extensiven Messung über direkte, assoziative und monotone Verknüpfungsoperationen wenig gemein und ließ sich nicht als ein fundamentales Messen betrachtet. v. Kries (1882) hatte schon eingewandt, das Fechnersche Vorgehen beruhe nicht auf eigentlicher Messung, sondern lediglich auf einer willkürlichen Konvention. Bis zur Mitte dieses Jahrhunderts blieb die Auffassung, fundamentales Messen in der Psychologie sei a priori nicht möglich, vorherrschend. Obwohl der Psychophysiker S. S. Stevens im Zusammenhang mit neuartigen Konstruktionsprinzipien psychophysikalischer Skalen wesentliche Anstöße gab, aus operationalistischer Perspektive über diese Auffassung hinaus zu denken, fand sie ihren deutlichsten Ausdruck 1940 in einem entsprechenden Urteil einer eigens zur Untersuchung dieser Frage eingesetzten Kommission der British Royal Society (vgl. Narens & Luce, 1986). Doch zeigte sich alsbald, daß der tatsächliche Umfang einer allgemeinen Theorie der in einem fundamentalen Sinne meßbaren Größen sehr viel weiter war. Hatte sich schon Fechner durch Bernoullis Überlegungen zur Nutzentheorie leiten lassen, so stellten auch hier wieder Betrachtungen zur Nutzentheorie entscheidende Impulse für eine geeignete Verallgemeinerung des Konzepts fundamentaler Messung bereit (Ramsey, 1931; von Neumann & Morgenstern, 1947; Debreu, 1959) und führten schließlich zum Konzept *additiv-verbundener Messung.* Hinzu kam die modelltheoretische Ausarbeitung des Gedankens ‚Messen als numerische Repräsentation qualitativer empirischer Strukturen' durch den mathematischen Logiker Dana Scott und den Philosophen, Physiker und mathematischen Psychologen Patrick Suppes im Jahre 1958. Diese löste eine Fülle weiterer Untersuchungen aus, durch die sich erneut und in neuem Lichte die Frage psychologischer Messung stellte (Krantz, Luce, Suppes & Tversky, 1971). Mit der Abstrakten Meßtheorie konstituierte sich ein sich rasch verselbständiges Forschungsfeld – von Cliff (1992) als „one of the most important developments of scientific psychology“ bezeichnet –, dessen weitreichende Bedeutung für eine naturwissenschaftlich orientierte Psychologie erst zukünftige Entwicklungen erschließen werden.

Die Ausarbeitung des Konzepts der additiv-verbundenen Messung (s. Krantz et al., 1971, Kap.5) bildete den wichtigsten Ertrag dieser Untersuchungen. Statt sich wie im extensiven Fall auf eine explizite und empirisch interpretierte Verknüpfungsoperation zu beziehen, geht man bei *additive-verbundenen Strukturen* von einer Produktstruktur aus, auf der lediglich die Relation einer schwachen Ordnung definiert ist. Empirisch entsprechen Produktstrukturen Attributen, die durch mindestens zwei unabhängige Komponenten bestimmt werden. Die entscheidenden strukturellen Annahmen, die in Form sog. Axiome formuliert werden, beziehen sich auf eine Zerlegbarkeit der Produktstruktur in geordnete Teilstrukturen sowie auf einen *trade off* bzw. auf eine Kompensierbarkeit zwischen den Teilstrukturen. Ein psychophysikalisches Beispiel für eine solche additiv-verbundene Struktur ist die binaurale Lautstärke, bei der eine Versuchsperson (Vp) Töne, die sich jeweils aus einer Reizkomponente des linken und rechten Ohres zusammensetzen, der Lautstärke nach ordnet (Levelt, Riemersma & Bunt, 1972). Für additiv-verbundene Strukturen läßt sich dann die Repräsentierbarkeit in einer über der geordneten additiven Gruppe der reellen Zahlen definierbaren numerischen Struktur beweisen (wobei sich dies unter geeigneten Verallgemeinerungen des Scott-Suppes-Ansatzes ebenfalls als Homomorphismus formulieren läßt). Daß dabei ‚lediglich' aus Ordnungsinformationen eine additive Repräsentation mit Intervallskaleninvarianz erhalten wird, ist weniger erstaunlich, wenn man sich verdeutlich, daß in den Anforderungen an diese Ordnung gerade die Strukturrestriktionen verborgen sind, die einer abstrakten additiven Operation entsprechen. Additiv-verbundene Strukturen sind gleichsam „krypto-extensive" Strukturen (Niederée, 1992a), die ohne explizite Angabe einer empirisch interpretierbaren Verknüpfung eine fundamentale Messung in gleich strengem Sinne erlauben wie die extensiven Strukturen.

Es lassen sich weitere Arten fundamentaler Messung formulieren, die mit den additiv-verbundenen Strukturen verwandt sind und ebenfalls keine extensive Verknüpfung voraussetzen. Die wichtigsten sind Differenzenstrukturen und Bisymmetriestrukturen. Erstere beziehen sich auf eine Ordnung auf Intervallen, durch die sich implizit eine additionsanaloge Verknüpfung von Intervallen formulieren läßt. Für eine Anwendung in der auditiven Psychophysik siehe Schneider, Parker & Stein (1974). Bisymmetriestrukturen, die zuerst von Pfanzagl (vgl. 1971) formal studiert wurden, beziehen sich auf eine Operation $\bigcirc$, die hinsichtlich einer Ordnung monoton ist, doch weder positiv ist (d. h. $a \bigcirc b \succ a$) noch assoziativ oder kommutativ, jedoch eine die Axiome der Assoziativität und Kommutativität abschwächende Bedingung erfüllt, nämlich die Bisymmetrie: $(a \bigcirc b) \bigcirc (c \bigcirc d) \sim (a \bigcirc c) \bigcirc (b \bigcirc d)$. Das Studium solcher Strukturen leitet sich aus der Untersuchung algebraischer Eigenschaften von Mittelwerten von Paaren reeller Zahlen ab. Empirisch lassen sich derartige Strukturen mit den Bisektionsmethoden der Psychophysik in Beziehung setzen: Interpretiert man $a \bigcirc b$ als die Operation der Mittenbildung zwischen den Reizen a und b, so ist infolge vor. Hysterese-Effekten i.d.R. die Kommutativität nicht erfüllt, d. h. $(a \bigcirc b) \not\sim (b \bigcirc a)$, wohl aber die Bisymmetrie. Das Bisektionsverfahren wurde erstmals von Plateau (1872) in der Psychophysik angewandt, um eine Skala von Grautönen zu erstellen.

Für die Psychologie bedeuten die formalen Resultate zu neuartigen Strukturen fundamentaler Messung dies: Der aprioristische Einwand, fundamentales Messen in der Psychologie sei nicht möglich, da ihre Größen nicht-extensiver Natur seien, ist beseitigt. Oftmals ist der qualitative Gehalt einer empirischen Struktur so reichhaltig, daß er auch ohne additive Verknüpfung eine Repräsentation durch die Struktur der reellen Zahlen, d.h. ein fundamentales Messen gestattet. Der Größenbegriff erweist sich somit als sehr viel komplexer und umfassender, als es eine Theorie physikalischer Größen zunächst nahelegt. Konzeptuell scheint folglich der grundsätzlichen Möglichkeit einer quantitativen Erfassung psychologischer Größen nichts im Wege zu stehen. Damit ist freilich nur ein einzelner Aspekt psychologischer Größenmessung geklärt, der sich auf ein begriffslogisches Problem bezieht. Alle weiterführenden konzeptuellen Fragen, etwa welchen Status die psychologischen Relationen haben und was die eigentlichen Objekte einer psychologischen Messung sind, werden dadurch in keiner Weise berührt. Fechner wollte Empfindungen messen und sah folglich diese als eigentliche Objekte der Messung an. Die ‚Empfindung' ist verlorengegangen und an ihre Stelle sind die ‚Reize' zusammen mit bestimmten Urteilsaufgaben getreten; doch was ist dann der Gegenstand der Messung? Zudem bleibt zu zeigen, daß numerischen Skalen bei der Entwicklung psychologischer und psychophysikalischer Theorien tatsächliche eine Bedeutung zukommt (vgl. Mausfeld, in diesem Band, Kap.4). Die Theorieentwicklung in Naturwissenschaften wie der Chemie und Biologie hat nämlich nicht zu einer eigenständigen Konstruktion fundamentaler Skalen über die physikalischen hinaus geführt; eine naturwissenschaftliche Theoriebildung muß weder notwendig auf einer eigenen Art fundamentaler Messung beruhen, noch setzt sie numerische Skalen voraus.

Wie bedeutend oder unbedeutend sich eigene Skalen auch für die Theoriebildung in der Psychologie erweisen mögen, es bleibt Verdienst der Psychophysik, die Beschäftigung mit diesen abstrakten Fragen der fundamentalen Meßbarkeit jenseits des Konzepts der extensiven Messung ausgelöst zu haben. Dadurch wurde in der Folge ein sehr viel tieferes Verständnis und zugleich eine Neubestimmung der klassischen Dichotomie von Qualität und Quantität erzielt, in der prinzipiell einer fundamentalen Messung psychologischer Größen nichts entgegensteht.

3. *Arten der Skalenkonstruktion*

Im folgenden Abschnitt werden typische Vorgehensweisen bei der Konstruktion von Skalen betrachtet; dadurch wollen wir Aufschlüsse darüber gewinnen, was die Objekte einer solchen Messung sein können und welche Rolle die erhaltene Skala für die Theoriebildung haben könnte. Für derartige Betrachtungen bietet die Psychophysik einen besonders geeigneten Rahmen: Beispielhaft lassen sich hier sowohl die formalen Techniken wie auch die Anbindung formaler Konzepte an solche der substantiellen Theoriebildung in präziser Weise behandeln; zudem bildet die Psychophysik den ideengeschichtlichen Ursprung, von dem andere Bereiche der Psychologie, wie die Entscheidungstheorie, die Diagnostik und die Sozialpsychologie, entscheidende Anstöße und Intuitionen der Skalenkonstruktion bezogen. Die Möglichkeit einer präzisen physikalischen Beschreibung des zu skalierenden Attributes in der Psychophysik läßt Grundzüge und Probleme der Skalenkonstruktion deutlicher als in anderen Bereichen hervortreten. Was sich bei einer solchen Behandlung an Einsichten gewinnen läßt, kann zudem in wesentlichen Teilen in andere mit einer Skalenkonstruktion befaßte Bereiche der Psychologie übertragen werden. Insbesondere gilt dies für die unterschiedlichen Weisen, wie sich der Begriff eines Fehlers konzipieren Iäßt. Der nun folgende Abschnitt stellt weder eine Einführung noch eine Übersicht über Methoden der Skalierung darstellt; hierzu sei auf Luce & Galanter (1963a,b), Roskam (1983), Tack (1983), Falmagne (1985), Luce & Krumhansl (1988) und Drösler (1989) verwiesen.

Verschiedene Arten der Skalenkonsrruktion können nach unterschiedlichen Gesichtspunkten klassifiziert werden, etwa danach, ob sie auf ,direkten' oder ,indirekten' Schätzrnethoden beruhen, nach der Art der Urteilsaufgabe, danach, ob sie unter dem Aspekt der Identifikation, Klassifikation, Entdeckbarkeit oder Diskriminierbarkeit von Reizen betrachtet werden, nach der Art der Vorstellungen über die interne Reizrepräsentation (*random* vs. constant *utility* model) oder nach der Dimensionalität der Repräsentation. Diese und andere Klassifikationen kreuzen sich in vielfältiger Weise, und eine einheitliche Systematik psychophysikalischer Verfahren gibt es nicht. Für unsere Zwecke scheint folgende abstrakte Einteilung zweckmäßig zu sein:

a) lokale vs. globale Verfahren,
b) probabilistische vs. deterministische Verfahren,
c) Verfahren, die rneßtheoretisch formuliert sind vs. solche, für die dies nicht gilt.

Die Einteilung ,lokal vs. global' bezieht sich darauf, ob sich die Gesamtskala aus Diskriminationsurteilen in den jeweiligen Umgebungen ihrer Elemente zusammengesetzt oder ob die strukturellen Annahmen, auf denen die Skala beruht, sich auf den Gesamtbereich beziehen. Diese traditionelle Einteilung erscheint bei genauerer begrifflicher Bestimmung der beiden Unterscheidungsmerkmale als nur schwer präzisierbar; tatsächlich wird sich, wie wir noch sehen werden, bei probabilistischen Verfahren diese Grenzziehung verwischen.

Die Einteilung ‚probabilistisch vs. deterministisch' unterscheidet Verfahren, deren Grundkonzepte eine Behandlung von Zufallsfluktuationen erlauben von solchen, die dies nicht zulassen.

Die dritte Unterscheidung schließlich bezieht sich darauf, ob sich ein Verfahren so darstellen läßt, daß sich die Skala aus der Gültigkeit qualitativer Bedingungen als numerische Repräsentation im Sinne der Meßtheorie ergibt.

Verdeutlichen wir die Klassifikation anhand einiger Beipiele. Zu den deterministischen, lokalen Verfahren gehören sowohl die Fechnerschen Verfahren, sofern die Schwelle deterministisch bestimmt wird, als auch die Semiordnungen. Zu den deterministischen, globalen Verfahren zählen die auf dem algebraischen additiv-verbundenen Messen beruhenden Vorgehensweisen, ferner das genannte Bisektionsverfahren, Stevens' indirekte Verfahren sowie die nichtmetrische multidimensionale Skalierung. Zu den probabilistischen, lokalen Verfahren sind alle Verfahren zu rechnen, die zur Bestimmung von Diskriminationswahrscheinlichkeiten das Konzept der psychometrischen Funktion (oder ein analoges Konzept) heranziehen, zu den probabilistischen, globalen Verfahren hingegen jene, bei denen eine additiv-verbundene Struktur auf Wahrscheinlichkeiten oder Zufallsvariablen formalisiert wird, sowie probabilistische Varianten der Stevensschen Verfahren und der multidimensionalen Skalierung. Die dritte Kategorie der Klassifikation kreuzt sich mit den andern, da es in jeder der Teilkategorien meßtheoretisch begründete Verfahren gibt. Als Beispiele seien die Krantz-Shepard-Reformulierung der Stevensschen Verfahren genannt, die meßtheoretische Darstellung der multidimensionalen Skalierung und die Schwellenrepräsentationen. Wir werden hierauf noch zurückkommen. Das Attribut ‚meßtheoretisch' deutet hauptsächlich eine bestimmte metatheoretische Haltung an; der formale Kern eines Verfahren kann nämlich in unterschiedliche Sichtweisen eingebettet sein, wie die Kontroverse über *conjoint* measurement vs. functional measurement (s. Anderson, 1982) illustriert.

3.1 Globales, deterministisches Vorgehen I: Die Stevensschen Verfahren

Zu globalen, deterministischen Verfahren sind Ratingskalen-Methoden sowie die ‚direkten' Skalierverfahren von Stevens zu rechnen. Von diesen sind die wichtigsten die Größenschätzung (magnitude estimation) und der Intermodalvergleich (*cross* modality matching). Da auch die Größenschätzung als ein Vergleich zweier Bereiche betrachtet werden kann, nämlich dem eines physikalischen mit einem Zahlenkontinuum, kann sie als Spezialfall des *cross* modality matching betrachtet werden. Statt der zuvor betrachteten Ordnungsrelation liegt diesen Methoden die Relation einer Korrespondenz oder Gleichheit zugrunde.

Bei der Methode der Größenschätzung ist es Aufgabe der Vp, jedem dargebotenen Reiz ein Zahlzeichen zuzuordnen, so daß die Zahlen den durch die Reize hervorgerufenen Empfindungen proportional sind; dabei wird zuvor ein Bezugsreiz durch eine willkürliche numerische Zuordnung ausgezeichnet, um die Response-Skala zu verankern. Bei den Herstellungsmethoden soll zu einem dargebotenen Reiz ein anderer von der Vp eingestellt werden, der eine ‚halbe oder doppelt so starke Empfindung' hervorruft wie der Ausgangsreiz. Bei dem Intermodalvergleich schließlich werden zwei Modalitäten zunächst willkürlich miteinander verankert, indem ein Reiz aus einer Modalität vorgegeben wird, der als ‚gleich' zu einem bestimmten Reiz einer anderen Modalität definiert wird. Mit dieser Verankerung hat die Vp dann zu beliebigen anderen Reizen der ersten Modalität solche der zweiten anzugeben oder zu adjustieren, die dem Reiz der ersten Modalität wiederum ‚gleich' sind. Bereits hier stellen sich eine Reihe konzeptueller Probleme im Zusammenhang mit der Frage, welche psychologischen Prozesse sich in den durch diese Methoden erhobenen Daten widerspiegeln könnten: Verwenden die Vpn Halbierungsinstruktionen so, als würden sie numerische Meßwerte halbieren? Bezeichnen die Zahlen-Äußerungen der Vpn tatsächlich Zahlen, die der Forscher als numerische Skalen im Sinne der Meßtheorie interpretieren kann? Wie läßt sich Stevens' Behauptung, durch Größenschätzung gewonnene Skalen seien Verhältnisskalen, rechtfertigen? Welche interne Empfindungsarithmetik wird bei diesen Verfahren postuliert? Diese Beispiele konzeptueller Fragen deuten bereits an, daß sich hinter der scheinbaren Einfachheit dieser Verfahren zahlreiche metatheoretische Probleme verbergen. Ausführliche theoretische Behandlungen der Stevensschen Methoden finden sich in Luce & Galanter (1963b), Krantz (1972a, b) sowie Baird & Noma (1978), der Kategorienmethoden bei Restle & Greeno (1970, S. 149ff.).

Stevens hatte vor den Hintergrund operationalistischer und behavioristischer Denkweisen nach psychophysikalischen Methoden gesucht, die lediglich auf Reiz-Reaktions-Beziehungen basieren. (Dennoch redet er weiterhin von ‚Empfindungen' und ‚Empfindungsskalen'.) Stevens hielt zudem Skalierverfahren für ungeeignet, die, wie die Fechnerschen Methoden, allein im Bereich der Unsicherheit zwischen Reizen anwendbar sind. Denn diese klassischen Skalierverfahren, die Stevens irreführend „indirekt" nannte und denen er seine „direkten" Methoden gegenüberstellte, erlauben eine Skalenkonstruktion nur dort, wo eine genaue Unterscheidung zwischen den Reizen nicht möglich ist; aus der Perspektive deterministischer Modelle macht sich dies als Inkonsistenz der Urteile bemerkbar. Der Fechnersche (wie auch der Thurstonesche) Ansatz beruhe darauf, „to transform variabilities, discriminal dispersions, or confusions into units of measure", und behaupte, „that all we can know about magnitude is what confusion tells us" (Stevens, 1957, S. 154). Die Möglichkeit, innerhalb probabilistischer Modelle eine psychophysikalische Skala erstellen zu können, wurde von Stevens also ebenso in Abrede gestellt wie die Fechnersche Annahme, daß jeder ebenmerkliche Zuwachs subjektiv gleich groß ist, daß also gleichen Reizverhältnissen gleiche Empfindungsdifferenzen entsprechen.

Diese Kritik des Fechnerschen Ansatzes findet sich bereits in klassischen Arbeiten des vergangenen Jahrhunderts. Techniken einer ‚globalen' Skalenkonstruktion, die nicht auf Urteilen über ‚lokale' Diskriminierbarkeit beruhen und im Stevensschen Sinne „direkt" sind, wurden von Merkel (1888) entwikkelt, etwa die „Methode der doppelten Reize". Brentano (1874/1924, S. 99f.) vertrat die Auffassung, daß empirische Daten eher nahelegen anzunehmen, daß gleichen Reizverhältnissen gleiche Empfindungsverhältnisse entsprechen. Plateau (1872, S. 383) hatte diese Annahme als eine sinnvolle a priori Annahme diskutiert, die er dadurch gestützt sah, daß Helligkeitsverhältnisse unter Beleuchtungsänderung konstant bleiben. Diese Annahme führt, worauf Fechner (1877, S. 24ff.) hinweist, auf das von Plateau (1872) postulierte Potenzgesetz zwischen Reiz x und Empfindung $\Psi(x)$:

$$\Psi(x) = c \cdot x^r \quad (c, r > 0). \tag{3.1}$$

Empirisch konnte für viele eindimensionale Reizkontinua, die sog. prothetischen Kontinua (Stevens, 1957) gezeigt werden, daß sich die mittlere Größenschätzung $Y(x)$ durch eine Potenzfunktion der physikalischen Reizintensität gut approximieren läßt und daß sich zudem im Intermodalvergleich der Exponent aus dem Verhältnis der unabhängig gewonnenen Exponenten für die beiden einzelnen Dimensionen vorhersagen läßt, d. h. mit $\Psi_1(x) = c \cdot x^r$ und $\Psi_2(y) = c' \cdot y^{r'}$ ergibt sich aus dem Vergleich $\Psi_1(x) = \Psi_2(y)$, daß $y = (c/c')^{1/r'} \cdot x^{r/r'}$. (Prothetische Kontinua sind solche, die sich, wie etwa Lautstärke und Helligkeit, nur auf die Intensität beziehen, während die sog. metathetischen Kontinua auf Qualitäten wie Tonhöhe, Farbton oder Geruch Bezug nehmen.)

Die Stevensschen Verfahren können als deterministisch betrachtet werden, da die Urteile der Vp unmittelbar als Skalenwerte der psychophysikalischen Skala angesehen werden und üblicherweise das Fehlerproblem nicht explizit gestellt wird; vielmehr wird hier als ‚Fehler' indirekt das definiert, was durch Mittelung über Vpn und durch eine Kleinste-Quadrate-Anpassung einer Geraden in log-log-Koordinaten eliminiert wird. Die psychologische Skala ergibt sich als eine Transformation der physikalischen Skala, zu der sie gleichsam ein internes Analogon ist. Die Fehlerbehandlung folgt daher einer physikalischen Fehlerbehandlung, ohne daß freilich wie in der Physik Kriterien darüber vorliegen, was als akzeptabel anzusehen ist und was nicht. Auf diesen Punkt werden wir später zurückkommen.

Die Methoden der Größenschätzung und des Intermodalvergleichs lassen sich meßtheoretisch reformulieren, so daß sich die Skala als Konsequenz der Gültigkeit qualitativer Bedingungen ergibt: In der Krantz-Shepard-Theorie (Krantz, 1972a; s. a. Tack, 1983), die davon ausgeht, daß alle Urteile implizit oder explizit relational sind, wurden die idealisierten empirischen Regularitäten, die sich in empirischen Studien mit diesen Methoden zeigten, herangezogen, um aus ihnen die Existenz und die Art der numerischen Repräsentation abzuleiten. Luce (1990) erhält die Potenzfunktion als Repräsentation in nicht-relationaler Weise durch die Forderung, daß die Korrespondenz- bzw. Gleichheitsrelation invariant unter bestimmten Paaren von Translationen der beiden beteiligten physikalischen Strukturen ist. Qualitative Bedingungen für die Größenschätzung und den Intermodalvergleich, deren Gültigkeit Voraussetzung für den Erhalt einer Verhältnisskala ist, formuliert Narens (1993).

Die Theorien von Krantz-Shepard und Luce sind der Natur ihrer Grundkonzepte nach deterministisch und liefern keine Kriterien zur Bewertung von Zufallsfluktuationen der Gleichheitsurteile. Wahrend bei den klassischen Stevensschen Verfahren der ‚Fehler' in geometrisch-numerischer Weise als Abweichung von der postulierten numerischen Funktion definiert wird, werden beim meßtheoretischen Vorgehen als ‚Fehler' alle auftretenden Arten von Verletzungen der qualitativen Bedingungen angesehen.

Ein wichtiger Unterschied in der Forschungsperspektive zwischen den Stevenschen Verfahren und ihren meßtheoretischen Reformulierungen sei abschließend noch einmal hervorgehoben: Es charakterisiert die meßtheoretische Perspektive, daß sie das qualitative Testen eines Modells in den Mittelpunkt stellt und eine konkrete numerische Skala nur als Nebenprodukt der Modellgültigkeit betrachtet. Nicht die numerische Skala als solche, sondern der strukturelle Gehalt des Modelles steht hier im Vordergrund.

3.2 Globales, deterministisches Vorgehen II: Die additiv-verbundene Messung

Das Vorgehen bei der additiv-verbundenen Messung wollen wir schematisch an einem idealisierten Beispiel der Untersuchung binauraler Lautstärke betrachten. Dabei werden dem linken und rechten Ohr simultan Töne bestimmter Intensität dargeboten. Die Menge der Reize für das linke Ohr sei mit $A = \{a,b,c...\}$ bezeichnet, die für das rechte Ohr mit $P = \{p,q,r,...\}$, wobei die Elemente jeweils 1000 Hz-Sinustönen darstellen, die sich nur in der Intensität unterscheiden. Aufgabe der Vp ist es, (etwa in einem Paarvergleichsdesign) anzugeben, welche von zwei dargebotenen Kombinationen (a,p), (b,q) lauter ist. Die Urteile haben also die Form $(a,p) \succ (b,q)$. Der Vp werden alle Paare an Kombinationen aus A x P dargeboten, und aus diesen paarweisen Urteilen wird eine verbundene und transitive Ordnungsrelation auf A x P erstellt. Für die erhaltene Ordnung auf A x P kann sodann überprüft werden, ob sie die qualitativen Bedingungen einer additiv-verbundenen Struktur erfüllt. Wenn diese Bedingungen, die oft als Axiome bezeichnet werden, erfüllt sind, so folgt mathematisch, daß es eine additive Repräsentation der Daten gibt, d. h. reellwertige Funktionen f auf A und g auf P mit

$$(a,p) \succcurlyeq (b,q) \Leftrightarrow f(a) + g(p) \geq f(b) + g(q). \tag{3.2}$$

Unter den Axiomen einer additiv-verbundenen Struktur spielt die Bedingung der Doppelkürzbarkeit eine wichtige Rolle; diese Bedingung lautet:

> Wenn für beliebige $a,b,c \in A$ und $p,q,r \in P$ gilt $(c,p) \succcurlyeq (b,r)$ und $(a,r) \succcurlyeq (c,q)$, so muß auch $(a,p) \succcurlyeq (b,q)$ gelten.

Die Wahrscheinlichkeit, daß eine Ordnung, welche die Unabhängigkeit und die Doppelkürzbarkeit erfüllt, keine additive Repräsentation hat, ist nämlich sehr gering (Arbuckle & Larimer, 1976; McClelland, 1977), so daß diese beiden Bedingungen bereits eine starke Modellrestriktion darstellen.

Zumeist will man sich nicht mit dem Nachweis der Existenz einer additiven Repräsentation begnügen, sondern eine solche konkret erstellen. Erfüllen die Daten das Modell einer additiv-verbundenen Struktur fehlerfrei, so kann die Bestimmung einer Skala über das Lösen eines entsprechenden Systems linearer Ungleichungen erfolgen, wodurch sich der gesamte Lösungsraum, d. h. die Menge der möglichen additiven Repräsentationen, bestimmen läßt (McClelland & Coombs, 1975). Die Größe des Lösungsraumes hängt von der Größe der Grundmengen ab und konvergiert mit steigender Anzahl von Ungleichungen gegen die durch Intervallskalentransformationen beschriebene Lösungsmenge, d. h. alle Skalen der Lösungsmenge gehen durch Intervallskalen-Transformationen auseinander hervor. Bei einem größeren Lösungsraum kann man nach gewissen Kriterien eine repräsentative Lösung auswählen.

Erweisen sich die Daten hinsichtlich des obigen Modells als fehlerhaft, so sind zwei Betrachtungsweisen möglich: Entweder man geht zu einem probabilistischen Modell über, worauf wir noch zu sprechen kommen werden, oder man behandelt den Fehler in geometrisch-statistischer Weise direkt auf der Ebene der numerischen Skala. Dies kann durch Techniken der isotonen Regression (s. Robertson et al., 1988) geschehen, durch die sich additive Repräsentationen f und g bestimmen lassen, die in einem bestimmten Sinne, nämlich der Minimierung des sog. Stress-Index, möglichst nahe an den erhaltenen Daten liegen. Hat man durch ein geeignetes Verfahren Skalen f und g bestimmt, so betrachtet man sie als psychophysikalische Skalen für das Attribut ‚binaurale Lautstärke'.

In einem weiteren Schritt kann man die Frage nach der Beziehung dieser Skalen zu der zugrunde liegenden physikalischen Skala, hier etwa der des Schalldrucks oder der Schallintensität der verwendeten Reize, stellen. So untersuchten Levelt, Riemersma & Bunt (1972), ob sich die Skalen f und g als Potenzfunktionen der physikalischen Skala ergeben.

Das obige Modell einer additiv-verbundenen Struktur ist in rein algebraischer Weise formuliert und läßt die Zufallsschwankungen empirischer Daten außer acht. Will man diese unmittelbar in die Formulierung des Modells mit einbeziehen, so kann man zu analogen probabilistischen Modellen übergehen, die wir als globale, probabilistische Verfahren klassifizieren.

3.3 Lokales, deterministisches Vorgehen

Als Beispiel wollen wir Verfahren und Repräsentationen anführen, denen ein nicht-probabilistisches Schwellenkonzept zugrunde liegt.

Betrachten wir zunächst die vierstellige Relation $\succcurlyeq$ in einer additiv-verbundenen Struktur $(A \times P, \succcurlyeq)$; diese wird ebenso als transitiv angenommen wie die entsprechenden zweistelligen Relationen $\succcurlyeq$ einer einfachen Ordnungsstruktur $(A, \succcurlyeq)$ oder einer extensiven Struktur $(A, \circ, \succcurlyeq)$. In empirischen Daten können aber Intransivititäten der folgenden Arten auftreten: (i) $a \sim b$ und $b \sim c$, jedoch $a \succ c$, (ii) $a \succ b$ und $b \succ c$, jedoch $c \succ a$. Will man diese Intransitivitäten nicht lediglich auf Seiten der numerischen Repräsentation durch einen numerisch-geometrisch inspirierten Fehlerbegriff behandeln, sondern zu einem Bestandteil der qualitativen Charakterisierung des Modells machen, lassen sich verschiedene Wege beschreiten. Man kann die relationale Struktur in Termini von Wahrscheinlichkeiten oder von geeigneten Statistiken von Zufallsvariablen beschreiben oder man kann die Intransitivitäten, insbesondere des Typs (ii), als Hinweise auf eine ‚Mehrdimensionalität' der Struktur betrachten und zu entsprechenden Formulierungen der Struktur übergehen; drittens schließlich kann man die Anforderungen an die Ordnungsrelation selbst modifizieren, indem man diese, wie im Fall der Semiordnungen, in eine transitive strenge Ordnung und eine intransitive und symmetrische Indifferenzrelation zerlegt, wobei ein Indifferenzintervall niemals größer als ein Ordnungsintervall sein kann. Im letzten Fall erhält man die übliche numerische Repräsentation, nur daß nun die Indifferenzrelation einem Schwellenbereich entspricht. Eine formale meßtheoretische Behandlung aller drei Vorgehensweise findet sich bei Suppes et al. (1989, Kap. 16 & 17). Ersetzt man etwa bei der additiv-verbundenen Messung die Relation $\succcurlyeq$ durch eine Semiordnung $\succ_m$, so erhält man folgende Repräsentation:

$$(a,p) \succ_m (b,q) \Leftrightarrow f(a) + g(p) \geqslant f(b) + g(q) + \delta \; (\delta > 0).$$

Ein besonders interessanter Fall ergibt sich, wenn auf einer extensiven Struktur $\langle A, \bigcirc, \succcurlyeq \rangle$ zusätzlich eine psychologische Ordnungsrelation $\succ_m$ definiert wird durch:

$a \succ_m b \Leftrightarrow$ „a ist merklich größer (z. B. heller, lauter, länger, schwerer) als b"

Diese Relation $\succ_m$, die als asymmetrisch, transitiv und irreflexiv angenommen wird, erlaubt für jede additive Repräsentation f der extensiven Struktur eine Schwellenrepräsentation der Art

$$a \succ_m b \Leftrightarrow f(a) > f(b) + \Delta_f(b). \tag{3.3}$$

Die Gültigkeit des Weberschen Gesetzes läßt sich nun durch bestimmte qualitative Bedingungen charakterisieren (s. Narens, 1985, S. 89), so insbesondere durch die Additivität des qualitativen Schwellenoperators J, wobei

$$J(a) = \sup\{b \mid b \bigcirc a \not\succ_m a).$$

$J(a)$ bezeichnet das größte physikalische ‚Inkrement', daß zu a ‚hinzuaddiert' von a nicht merklich unterscheidbar ist. Additivität bedeutet dann gerade:

$$J(a \bigcirc b) = J(a) + J(b).$$

Hieraus folgt die Additivität des entsprechenden numerischen Terms 4; es gilt dann nämlich $\Delta_f(a) = k \cdot f(a)$ $(k > 0)$. Die Schwellenrepräsentation (3) geht in diesem Fall über in

$$a \succ_m b \Leftrightarrow f(a) > f(b) + k \cdot f(b) \text{ bzw. } f(b)(1 + k).$$

Mit $u = \log f$ und $d = \log(1 + k)$ erhält man dann die Fechnerschc Rcpräsentation

$$a \succ_m b \Leftrightarrow u(a) > u(b) + d,$$

in der alle ebenmerklichen Unterschiede denselben numerischen Wert erhalten. Diese Repräsentation mit konstantem Schwellenwert ist ein formales Analogon zu Fechners Annahme, daß gleichen Reizverhältnissen gleiche Empfindungsdifferenzen entsprechen.

Ausführliche Behandlungen von Schwellenrepräsentationen finden sich bei Luce & Galanter (1963a, S. 201ff.), Pfanzagl (1971, S. 176ff.) und Suppes et al. (1989, S. 353ff.).

3.4 Lokales, probabilistisches Vorgehen

Die bisher betrachteten Modelle gehen davon aus, daß die Beziehungen zwischen Reiz, ‚Empfindung' und Urteil in den betrachteten Ausschnitten idealiter deterministisch sind. Inkonsistenzen zwischen Daten und Theorie, insbesondere Urteilsfluktuationen, werden also im Sinne externer Fehlerkonzepte behandelt. Probabilistische Modelle nehmen demgegenüber gerade Charakteristiken der Variabilität zum Ausgangspunkt. ‚Fehler' im Sinne dieser Variabilität wird hier als wesentlicher Theoriebestandteil, d. h. als theorieintern, angesehen. Bei einem solchen Vorgehen gibt es konzeptuell wie formal sehr unterschiedliche Wege, den Begriff der Variabilität in einem Modell zu fassen und Skalen als Parameter probabilistischer Modells zu erhalten. Beispielsweise kann man die interne Repräsentation eines physikalischen Reizes selbst als eine Zufallsvariable auffassen und erhält dann die sog. random utility-Modelle, die ihren Namen analogen nutzentheoretischen Modellen verdanken und von denen das Thurstone-Modell das bekannteste ist. Die ihnen zugrunde liegenden abstrakten Prinzipien lassen sich zur Formulierung einer modelltheoretischen ‚Probabilisierung' von Konzepten der abstrakten Meßtheorie heranziehen (Heyer & Niederée, 1992). Den constant utility-Modellen in der Psychophysik liegt demgegenüber das Bild zugrunde, daß Reizen in deterministischer Weise jeweils eine Empfindung zugeordnet ist, während das Urteil von Personen Fluktuationen unterworfen ist. Beide Klassen von Modellen können jedoch auch unabhängig von diesen Bildern rein formal betrachtet und verglichen werden (siehe die ausführlichen Darstellungen bei Luce & Suppes, 1965, Colonius, 1984, sowie Falmagne, 1985).

Tatsächlich nahm bereits Fechner (1860, I, S. 104ff.) an, daß sich die Werte bei der Schwellenbestimmung als eine Zufallsvariable auffassen und bestimmen lassen. Diese Idee findet in Thurstones Modell ihre Weiterführung und explizite formale Ausgestaltung: Der durch einen Reiz ausgelöste interne Effekt kann durch eine Zahl beschrieben werden, die jedoch von Darbietung zu Darbietung variieren kann. Die interne Repräsentation eines Reizes x ist also durch Zufallsvariablen U_x, welche die Verteilung der zugehörigen „Diskriminanzprozesse" darstellt, beschreibbar. Diese Zufallsvariablen sollen, so nimmt Thurstone an, Normalverteilungen sein. Die Entscheidungsregel lautet somit: Ein Reiz x wird von einer Vp genau dann als ‚größer' als ein Reiz y beurteilt, wenn der erhaltene Wert der Zufallsvariablen U_x größer als der Wert von U_y. Die Wahrscheinlichkeit $p(x,y)$ für dieses Ereignis ist also

$$p(x,y) = \Pr[U_x \geq U_y] = \Pr[U_x - U_y \geq 0] = \Phi[(\mu_x - \mu_y)/\sigma_{x-y}], \quad (3.4)$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist. Der „modale Diskriminanzprozeß" μ_x bzw. μ_x kann somit als Skalenwert einer internen Skala S betrachtet werden und die Varianz als Maß für die Größe der Unterschiedsschwelle. Die klassische Frage nach der Beziehung zwischen physikalischem Reiz und Empfindung übersetzt sich dann in die Frage nach der Abhängigkeit der entsprechenden Erwartungswerte einer Repräsentation von den physikalischen Reizwerten. Zwischen dem Erwartungswert dieser Zufallsvariablen und ihrer Varianz gibt es jedoch keine zwingende Beziehung. Thurstone wollte, wie später Stevens, sein Modell nicht an die Fechnersche Annahme binden, daß alle Unterschiedsschwellen durch einen konstanten Wert repräsentiert werden. Lediglich in dem als „Case V" bezeichneten Spezialfall seines Modells nimmt er alle Varianzen als gleich an; in diesem Fall vereinfacht sich das Modell zu $p(x,y) = \Phi[S(x) - S(y)]$. Modelle der Art

$$p(x,y) = D[u(x) - u(y)] \tag{3.5}$$

mit beliebiger Verteilungsfunktionen D werden als Fechner-Modelle bczeichnet (Luce & Suppes, 1965, S. 334; Falmagne, 1985. Die Verteilungsfunktion D hat hier die Rolle der psychometrischen Funktion, durch die Differenzen auf der psychophysikalischen Skala in Dominanzwahrscheinlichkeiten übersetzt werden.

Mit dieser Modellgleichung bietet sich eine Möglichkeit, einen probabilistischen Ansatz mit dem algebraischen Konzept der additiv-verbundenen Messung zu verknüpfen, indem man definiert:

$$(x,y) \succcurlyeq (v,w) \Leftrightarrow p(x,y) \geq p(v,w).$$

Hierbei ist vorauszusetzen, daß die Dominanzwahrscheinlichkeiten stets von 0 und 1 verschieden sind. Die Zufallsfluktuationen werden hier gleichsam zum strukturellen Kern der Theorie, indem man als eine Skala gerade das definiert, was den Fehler regularisiert (ein ähnlicher Gedanke findet sich in Levine, 1970). Dies wird besonders deutlich, wenn man beachtet, daß hier zu einer gegebenen Familie von reell-wertigen Zufallsvariablen eine Transformation gesucht wird, die die Zufallsvariablen in solche überführt, die bis auf Verschiebungen um eine additive Konstante identisch verteilt sind. Diese Transformation der physikalischen Reizskala ergibt dann gerade die gesuchte psychophysikalische Skala u, welche dann die Eigenschaft hat, die psychometrischen Funktionen zu parallelisieren. An diesem Beispiel wird bereits die Problematik der Unterscheidung ‚lokal' vs. ‚global' erkennbar.

Läßt sich in einem Fechner-Modell D wiederum als Differenzverteilung zweier unabhängiger und identisch verteilter, stetiger Zufallsvariablen schreiben, so erhält man ein random utility-Modell. Wählt man für D die logistische Funktion, so ist für Paarvergleichsmodelle die Verteilung der Diskriminanzprozesse nicht mehr eindeutig festgelegt (vgl. Luce & Krumhansl, 1989, S. 27). Eine empirische Entscheidung für eine bestimmte Verteilungsfunktion D läßt sich in der Regel nicht treffen, doch läßt sich in Analogie zu bestimmten neurophysiologischen Vorstellungen über die Art der Aggregation unabhängiger und identisch verteilter Signale die Verteilungswahl aus entsprechenden Grenzwertsätzen motivieren (Wandell & Luce, 1978; vgl. Tack, 1983). Weitere Aspekte von Thurstone-Modellen werden in Luce (1977) und Suppes et al. (1989, S. 421 ff.) behandelt.

An dieser Stelle müssen wir auf die schon angedeuteten Probleme der Unterscheidung lokaler und globaler Verfahren zurückkommen. Das oben beschriebene Vorgehen, die Skala über die psychometrische Funktion zu bestimmen, zeigt bereits, wie im probabilistischen Fall die Grenze zwischen ‚lokal‘ und ‚global‘ verschwimmt. Die psychometrische Funktion wird nämlich idealisiert als eine Funktion angesehen, die über dem gesamten betrachteten, d. h. über dem globalen, Reizbereich stets ungleich 0 oder 1 ist; folglich könnte theoretisch die Skalenkonstruktion auf einem fest gewählten Reiz basieren, durch dessen Vergleich mit allen anderen Reizen sich die globale Skala ergäbe. Die psychometrische Funktion gibt gerade an, wie sich das ‚Lokale‘ zu ‚Globalem‘ verbindet. Auf einer so weitgehenden Idealisierung wird man aber die Skalenkonstruktion nicht gründen wollen, und so nimmt man an, daß jeder Reiz der Skala Vergleichsreiz für die Reize seiner Umgebung ist, d. h., daß es eine Familie psychometrischer Funktionen gibt. Nun benötigt man aber eine (meist nicht explizit ausgesprochene) Miniatur-Theorie darüber, wie Schwellenverhalten mit ‚globalen‘ Urteilen zusammenhängt. Diese Miniatur-Theorie ist im Falle der durch Gleichung (5) beschriebenen Fechner-Modelle in der Forderung versteckt, daß auf der gesuchten Skala u alle psychometrischen Funktionen parallel sind, d. h. durch Verschiebung auf der Abszisse auseinander hervorgehen. Der empirische Gehalt dieser postulierten theoretischen Beziehung zwischen einer stückweise aus lokalen Urteilen erstellten Skala und Urteilen über globale Skalenbereiche wird an folgendem Beispiel deutlich: Eine Skala, die nach Fechners Vorschlag durch Kumulieren von Unterschiedsschwellen (über der absoluten Schwelle) erzeugt wird, führt (unter bestimmten Annahmen) zu einer logarithmischen psychophysikalischen Funktion. Die durch Logarithmieren der physikalischen Werte bestimmte psychophysikalische Skala hat die Eigenschaft, daß physikalisch gleichen Reizverhältnissen gleiche ‚Empfindungsdifferenzen‘ entsprechen. Die in der Akustik verwendete Dezibel-Skala (dB) der ‚Lautstärke‘ ist eine solche Skala, da sie als Logarithmus des physikalischen Schalldrucks (bezogen auf den Schwellenwert) definiert ist. Gleiche Abstände auf der Dezibel-Skala sollen also perzeptuell auch gleich erscheinen. Ein 1000 Hz-Sinuston von 50 dB wird aber nicht als halb so laut empfunden wie ein solcher von 100 dB. Die postulierte Beziehung zwischen lokalem und globalem Urteilsverhalten ist hier also nicht erfüllt und ‚lokaler‘ und ‚globaler‘ ‚Empfindungs‘-Begriff fallen auseinander.

Auf eine weitere Schwierigkeit in der Unterscheidung von ‚global‘ und ‚lokal‘ werden wir am Ende des folgenden Abschnitts zu sprechen kommen.

3.5 Globale, probabilistische Modelle

Diese Kategorie soll an zwei Ansätzen von Falmagne veranschaulicht werden, die versuchen, im Rahmen der oben behandelten additiv-verbundenen Strukturen Skalen unter Berücksichtigung von Zufallsfluktuationen zu konstruieren.

Die erste Möglichkeit basiert darauf, ein algebraisches Meßkonzept dadurch zu probabilisieren, daß die zugrunde gelegten Relationen unter Bezugnahme auf Urteilswahrscheinlichkeiten definiert werden. Anders als im Fall der Fechner-Struktur gehen wir dabei nicht von einfachen Dominanzwahrscheinlichkeiten aus, sondern von einer vierstelligen Dominanzrelation. Dabei wird die Wahrscheinlichkeit, die Reizkombination (a,p) als größer zu beurteilen als die Kombination (b,q), mit p[(a,p),(b,q)] bezeichnet. Eine Modellgleichung ergibt sich durch folgende Forderung

$$p[(a,p),(b,q)] \leq 0.5 \Leftrightarrow f(a) + g(p) \leq f(b) + g(q). \tag{3.6}$$

Die Relation $(A \times P, \succcurlyeq_{0.5})$ läßt sich nun definieren durch

$$(a,p) \succcurlyeq_{0.5} (b,q) \Leftrightarrow p[(a,p),(b,q)] > 0.5.$$

Auf diese Weise können alle Axiome der Struktur in Aussagen über Wahrscheinlichkeiten übersetzt werden, etwa die Transitivität der Ordnung in die sog. *schwache stochastische Transitivität*. Diese Übersetzung der probabilistischen in eine algebraische Struktur ist jedoch sehr grob; zumeist nimmt man darüber hinaus an oder wünscht, daß Unterschiede in den Dominanzwahrscheinlichkeiten Differenzen in den Skalenwerten widerspiegeln. Man kann daher obige Modellgleichung verschärfen zu

$$p[(a,p),(b,q)] = F[f(a) + g(p), f(b) + g(q)] \tag{3.7}$$

wobei F streng monoton steigend in der ersten und streng monoton fallend in der zweiten Variablen ist. Ein spezieller Fall hiervon ist wiederum

$$p[(a,p),(b,q)] = D[f(a) + g(p) - f(b) - g(q)] \tag{3.8}$$

mit D streng monoton steigend. Mit dieser Modellgleichung ist für den Fall einer Komponentenstruktur AxP der Anschluß an die Modellgleichung (3.5) der Fechner-Modelle hergestellt.

Für eine detaillierte Behandlung dieser Klasse von Modellen sowie von Bedingungen probabilistischer Konsistenz siehe Luce & Suppes (1965), Falmagne (1979), Roberts (1979), Colonius (1984) sowie Suppes et al. (1989, S.388ff.).

In dem zweiten, von Falmagne (1976) vorgeschlagenen Ansatz zur Behandlung von Zufallsfluktuationen werden statt Urteilen der Art ‚$(a,p) \succcurlyeq (b,q)$', Gleichheitsurteile der Art ‚$(a,p) \sim (b,q)$' betrachtet. Hier wäre etwa an Einstellungsmethoden zu denken, bei welchen die von der Vp manipulierte Variable als Zufallsvariable aufgefaßt wird: Wird bei vorgegebenen a,p und q ein Gleichheitsurteil durch Einstellen eines Wertes b abgegeben, so sei die Verteilung der entsprechenden Zufallsvariablen mit $\mathbf{B}(a,p;q)$ bezeichnet. (Hierbei wird implizit jede Reizkomponente mit ihrem zugeordneten reellen physikalischen Skalenwert identifiziert.) Postuliert man die Gültigkeit einer additiven Kompositionsregel für die Effekte der Variablen a,p und q und nimmt man zudem ein Fehlermodell mit (symmetrischer) Fehlerverteilung $\varepsilon(a,p;q)$ an, so gelangt man zu der Modellgleichung

$$f(\mathbf{B}(a,p;q)) = f(a) + g(p) - g(q) + \varepsilon(a,p;q). \qquad (3.9)$$

Den Anschluß an die Modellgleichung (3.2) für die additiv-verbundene Messung gewinnt man, wenn man bei dem betrachteten Gleichheitsurteil als Wert für b eine Statistik der Zufallsvariablen $\mathbf{B}(a,p;q)$ wählt, die invariant unter monotonen Transformationen ist. Falmagne (1976) benutzte zur Formulierung seines „random conjoint measurement" den Median Md. Da für jede streng monotone Transformation h gilt

$$h(\mathrm{Md}[\mathbf{B}(a,p;q)] = \mathrm{Md}[h(\mathbf{B}(a,p;q))]$$

und da für die als symmetrisch angenommene Fehlerverteilung

$$\mathrm{Md}[\varepsilon(a,p;q)] = 0$$

gilt, erhält man aus (9)

$$\mathrm{Md}[\mathbf{B}(a,p;q)] = f^{-1}[f(a) + g(p) - g(q)]. \qquad (3.10)$$

Mit $b = \mathrm{Md}[\mathbf{B}(a,p;q)]$ ergibt sich dann gerade Gleichung (3.2).

Die Bedingung (3.10) führt zu einer Reihe von qualitativen testbaren Bedingungen, die sich nur auf die experimentell erhobenen Werte $\mathrm{Md}[\mathbf{B}(a,p;q)]$ beziehen. So folgt aus (3.10) etwa die Bedingung

$$\mathrm{Md}[\mathbf{B}(a,p;r)] = \mathrm{Md}[\mathbf{B}(\mathrm{Md}[\mathbf{B}(a,q;r)],p;q)].$$

Diese Bedingung ist das Analogon zur Doppelkürzbarkeitsbedingung, und sie läßt sich durch einen Median-Test statistisch prüfen, wie Falmagne (1976) zeigt. (Gigerenzer und Strube (1983) untersuchten mit diesem Modell noch einmal die Hypothese binauraler Additivität und fanden sie, anders als Levelt et al. (1972), nicht bestätigt.)

Die beidcn Ansätze von Falmagne lassen sich formal wieder in Beziehung setzen. Unter naheliegenden Zusatzannahmen ergibt sich nämlich $(a,p) \succcurlyeq_{0.5} (\mathbf{B}(a,p;q),q)$.

Warum nun, und damit kommen wir noch einmal auf die Problematik ,lokal' vs. ,global' zurück, haben wir diese beiden Zugangsweisen als globale klassifiziert? Offensichtlich wird bei diesen Verfahren die Existenz einer Skala durch strukturelle Bedingungen garantiert, die sich auf den gesamten betrachteten Reizbereich beziehen: nämlich gerade durch die Ordnungsrestriktionen einer additiv-verbundenen Struktur. Besonders deutlich wird dies bei der erstgenannten Zugangsweise: Mit der durch die Bedingung

$$(a,p) \succcurlyeq (b,q) \Leftrightarrow p[(a,x),(b,q)] > 0.5.$$

festgelegen Klassifikation der Dominanzwahrscheinlichkeiten wird die psychometrische Funktion gleichsam wieder deterministisch betrachtet, und die Modellrestriktionen beziehen sich lediglich auf die Ordnung der so klassifizierten Dominanzwahrscheinlichkeiten. Man könnte nun einwenden, daß die erhaltene Skala dennoch auf lokalen Informationen beruhe, da die zugrunde liegenden Urteile Diskriminations- bzw. Dominanzurteile sind und somit in einem Konfusionsbereich liegen müssen. Wir wollen daher zwischen den Konzepten ‚lokal' und ‚Diskriminationsurteile' folgende Unterscheidung treffen: Als lokale Eigenschaften einer Skala seien solche bezeichnet, die sich auf einzelne Elemente der Skala und deren Umgebung im Sinne eines Konfusionsbereiches beziehen. Diskriminationsurteile seien abstrakt durch eine Unterscheidbarkeitsrelation $\succ$ beschrieben; diese kann sich wiederum unmittelbar auf die Elemente der Skala beziehen, d. h. eine zweistellige Relation der Art $a \succ b$ sein, oder sie kann, wie bei additiv-verbundenen Strukturen, höherstufig sein, also $(a,x) \succ (b,y)$. Würde man eine Skala erstellen wollen, deren Elemente Paare der Art (a,x) sind, d. h. ist man nicht an den Elementen der vierstelligen Relation selbst interessiert, so könnte man Diskriminationsurteile der Art $(a,x) \succ (b,y)$ wieder als ‚lokal' bezeichnen. Die üblichen psychophysikalischen Skalen beziehen sich aber, veranlaßt durch die korrespondierenden eindimensionalen physikalischen Größen, gerade auf die einzelnen Reizkomponenten $a,b,...$ der Relation. Folglich ist die probabilistische additiv-verbundene Messung als globales Verfahren anzusehen, das auf komplexeren Diskriminationsurteilen basiert.

Diese Beispiele mögen für eine Veranschaulichung unterschiedlicher Arten der Skalenkonstruktion in der Psychophysik genügen. Weiterführende Darstellungen finden sich in Luce & Galanter (1963a,b), Bock & Jones (1968), Krantz (1972b), Holman & Marley (1974), Luce & Green (1974), Luce (1977) sowie Laming (1986).

3.6 Vergleichende Zusammenfassung

Die zuvor genannten meßtheoretischen Zugangsweisen wollen wir noch einmal kurz unter dem Aspekt der Beziehung von theoretischer Struktur und empirischen Daten betrachten.

Eine Auffassungsweise ist, wesentliche algebraische Bedingungen, die eine Klasse psychophysikalischer Struktur charakterisieren, unmittelbar als empirische Hypothesen aufzufassen, indem man die theoretischen Relationen als empirische ansieht. Durch Beschränkung auf ordinale Daten können bei einer geeignetenden Verteilung der Datenpunkte Zufallsfluktuationen unbemerkt bleiben in dem Sinne, daß keine Inkonsistenzen zwischen Daten und Theorie auftreten. Treten Inkonsistenzen auf, so gibt es zwei prinzipielle Strategien zu ihrer Behandlung: Man kann sie auf Seiten der numerischer Repräsentation durch ein globales Verletzungsmaß wie den oben genannten Stress-Index ‚bewerten'. Oder man sucht Verletzungen auf der Ebene qualitativer (meist universeller) Axiome (oder auch deren Folgerungen) zu beschreiben. Dieses Vorgehen verknüpft sich häufig mit dem Anliegen einer Diagnose systematisch-strukturellen Fehlanpassung zwischen Daten und Theorie. (Die Begrenztheit eines solchen diagnostischen Vorgehens zeigt sich u.a. darin, daß eine bestimmte Fehlanpassung zwischen Daten und Theorie bei äquivalenten Axiomensystemen gleichsam auf verschiedene Axiome projiziert werden kann!) Für die Diagnose und Bewertung der Modellgüte gibt das Auszählen von Axiomverletzungen – ebenso wie der Stress-Index – allenfalls vage Kriterien an die Hand.

Ein anderer Weg, die Beziehung einer meßtheoretischen Struktur zu ihrem empirischen Gegenstandsbereich zu behandeln, liegt darin, bestimmte Arten von strukturellen ‚Abweichungen' als theoretisch interessant zu betrachten und sie bereits innerhalb der deterministischen Theorie bei der Formulierung der Axiome zu berücksichtigen. Beispielsweise kann die Ordnungsrelation der Struktur zu einer Semiordnung abgeschwächt werden, um die Intransitivität von Indifferenzcn als Thcoriebestandteil zu berücksichtigen. Dieses Vorgehen führt zu den oben genannten Schwellenrepräsentationen (vgl. Gleichung (3.3)). Oder die Struktur selbst wird als Struktur mit maximalen Elementen angesehen; bestimmte existentielle Axiome können etwa dadurch verletzt sein, daß maximale Elemente vorliegen, was zu einer entsprechend verfeinerten Axiomatisierung Anlaß geben kann.

Einen Ubergang von der deterministischen zu einer probabilistischen Betrachtungsweise ergibt sich, wenn man die Schwelle mittels der psychometrischen Funktion über ein Wahrscheinlichkeitskriterium definiert. Dies entspricht der allgemeinen Strategie deterministische Relationen unter Bezug auf Wahrscheinlichkeiten zu definieren, die sie etwa dem in den Gleichungen (3.6–3.8) beschriebenen Vorgehen von Falmagne (1979) zugrunde liegt. Die Anbindung an die Daten erfolgt hier über relative Häufigkeiten ordinaler Urteile.

Eine andersartige Stochastisierung der Struktur liegt vor, wenn ihre Elemente selbst, wie bei den random *utility*-Modellen, als Zufallsvariablen aufgefaßt werden. über deren Parameter sich wiederum eine Skala definieren laßt. Die algebraische Theorie wird gleichsam probabilistisch ‚verwackelt', wodurch einige der algebraischen Bedingungen zu statistisch testbaren Hypothesen werden (vgl. Iverson, 1991).

Ein sehr allgemeiner Ansatz, auf theoretischem Wege zu einer Probabilisierung deterministischer Meßstrukturen zu gelangen, wurde von Heyer & Niederée (1992; vgl. Luce & Narens, 1993) entwikelt. Er beruht auf dem Konzept ‚probabilistischer Mischungen' qualitativer deterministischer Meßstrukturen (z.B. additiv-verbundener Strukturen). Dieser Ansatz verallgemeinert in natürlicher Weise das aus der Nutzentheorie bekannte Konzept des binary choice *system* induced *by ranking* (vgl. Fishburn, 1992). Für diesen Fall ist bekannt, daß sich für endliche Trägermengen fester Größe die resultierenden ‚Mischungen' durch ein endliches System linearer Ungleichungen charakterisieren lassen. Heyer und Niederée zeigen, daß ein entsprechendes Resultat auch im allgemeinen Fall gilt, und setzen das genannte Konzept zu einer Verallgemeinerung des random *utility*-Konzeptes in Beziehung.

Mit diesen Vorgehensweisen ist ein Spektrum theoretischer Konzeptualisierungen von Zufallsfluktuationen skizziert, zwischen dessen Extrempunkten sich die verschiedenen Möglichkeiten, den Zufallsfehler bereits bei der Formulierung der theoretischen Struktur zu berücksichtigen, spannen.

Im folgenden Abschnitt werden wir das Problem der Bestimmung eines ‚Fehler'-Konzeptes, das zumeist implizit die bisherigen Abschnitte zur Skalenkonstruktion durchzog, aus allgemeinerer Perspektive behandeln.

4. *Anmerkungen zum Fehlerbegriff*

Im letzten, eher technischen, Abschnitt haben wir idealtypisch zwei Zugangsweisen zur Behandlung des ‚Fehlerproblems' an die Unterscheidung von deterministischen und probabilistischen Modellen geknüpft. Dieser Gegenüberstellung entspricht eine Unterscheidung von ‚Fehlerkonzepten', die sich bezogen auf die Theorie als intern bzw. extern auffassen lassen. Einige Überlegungen grundsätzlicher Art seien nun nachgetragen. Dabei werden wir das Augenmerk auch auf die ideengeschichtliche Entwicklung entsprechender Konzepte richten.

Beginnen wir mit theorieexternen Fehlerkonzepten. Das sich dabei ergebende Problem, den ‚eigentlichen' Meßwert von störenden ‚Nebeneffekten' zu bereinigen, stellte sich nicht erst mit der psychologischen Skalenkonstruktion. Jede Messung, auch die physikalische, ist mit ihm konfrontiert. Was aber als eigentlicher Meßwert und was als Meß- oder Beobachtungsfehler anzusehen ist, kann nur durch den Bezug auf eine Theorie oder eine Vorhersagetechnik festgelegt werden: Theoriekonstruktion und ‚Fehlertheorie' bzw. Techniken der Fehlerbehandlung gehen Hand in Hand.

Bei den theoricexternen Fehlerkonzepten in der Psychophysik wird der Zusammenhang zwischen den betrachteten theoretischen Größen als deterministisch betrachtet, eine Vorgehensweise, die derjenigen der klassischen Physik gleicht. Fehlerkonzepte werden dann eingeführt, um zwischen den beobachteten und theoretischen Größen zu vermitteln (wobei bereits die Einführung des Begriffes ‚theoretisch' gleichsam das notwendige Komplement des Begriffes ‚Fehler' ist).

Es ist bemerkenswert, daß die Bemühungen um ein Verständnis des ‚Beobachtungsfehlers' in der Physik historisch entscheidende Anstöße zur Entwicklung der Psychophysik geliefert haben. Damit lassen sich umgekehrt Teile der Psychophysik vom Standpunkt der Physik aus betrachtet als Bausteine einer physikexteren Fehlertheorie auffassen, ist es doch die menschliche Wahrnehmung, die lange Zeit in der physikalischen Messung das entscheidende letzte Glied des Meßprozesses bildete. Zwar wußte man seit altersher um die Unzuverlässigkeit der Wahrnehmung als eines Erkenntnisinstruments, doch wurde erst im 18.Jahrhundert die Eigengesetzlichkeit dieses Erkenntnisinstruments zum Objekt systematischer wissenschaftlicher Betrachtung. Der Begriff des Beobachtungsfehlers entstand, und mit den Grundlagen seiner formalen Behandlung wurden zugleich die Grundlagen der Wahrscheinlichkeitstheorie gelegt. Um also zuverlässige physikalische Werte zu gewinnen, mußten Einflußfaktoren eliminiert werden, die sich nicht auf physikalische Faktoren zurückführen lassen, sondern sich aus Eigenschaften der menschlichen Wahrnehmung ergeben. Vor allem in der Astronomie spielte die Reliabilität visueller Wahrnehmungsleistungen eine große Rolle: Bei der Helligkeitsmessung von Sternen mußten vorhandene Lichtstärken unterschieden werden und physikalisch gleiche Lichtintensitäten auch als gleich beurteilt werden. So genau sich das Auge, verglichen mit der physikalischen Bestimmung, auch erwies, so zeigten sich doch in systematischer Weise Variabilitäten des Urteils. In seinen Untersuchungen zur Photometrie von 1760 unterschied Lambert drei Fehlerquellen: die unvermeidliche Unbestimmtheit im Urteil des Auges, die Unachtsamkeit des Beobachters und schließlich die Beschaffenheit des Instruments einschließlich anderer, vom Beobachter unabhängiger Umstände. Zudem findet er, daß für die beiden ersten Fehlerarten die Größe des Fehlers mit seiner Auftretenswahrscheinlichkeit zusammenhängt. Dem arithmetischen Mittel der beobachteten Meßwerte kommt, wie Lambert meint, die „größte Wahrscheinlichkeit" zu. Bernoulli, Gauß und Laplace entwickelten dann eine formale Theorie der Beobachtungsfehler, die in der Herleitung der Normalverteilung und der Entwicklung der Methode der kleinsten Quadrate gipfelte.

Astronomische Beobachtungen führten auch zu einer weiteren Einsicht in die Systematik von Beobachtungsfehlern: zur „persönlichen Gleichung". Bei astronomischen Beobachtungen nach einem komplexen Urteilsverfahren, an dem gleichzeitig Auge und Ohr beteiligt waren, hatte 1796 der englische Astronom Maskelyne entdeckt, daß sein Assistent ein bestimmtes Ereignis systematisch eine halbe Sekunde später als er berichtete (woraufhin er ihn entließ). Diese interindividuellen Differenzen machte 1820 der Königsberger Astronom Bessel zum Gegenstand einer eigenständigen Untersuchung, die zu dem Ergebnis führte, daß sich zwischen je zwei Beobachtern eine im Mittel relativ konstante Differenz der Beobachtungszeiten feststellen ließ. Dadurch war es möglich, für jeden Beobachter eine individuelle Fehlergleichung, die „persönliche Gleichung", zu erstellen.

In der Theorie der Beobachtungfehler gelangte man schließlich zu folgender Klassifikation: (i) Fehler des Meßinstrumentes selbst, (ii) systematische interindividuelle Differenzen („persönliche Gleichung"), (iii) systematische Fehler (etwa wenn Beobachtungen nicht unter ‚Normbedingungen' durchgeführt wurden) und (iv) Zufallsfehler, die durch alle unbekannten Faktoren der Variabilität hervorgerufen werden.

Fechner erkannte die Bedeutung der formalen Methoden der Fehlerbehandlung, wie sie im Zusammenhang mit der Physik und Astronomie entwickelt worden waren, für die Psychophysik. Der für ihn so wichtige theoretische Begriff der Schwelle kann nämlich nur dann zu experimentellen Beobachtungen in Beziehung gesetzt werden, wenn man ein Verfahren angibt, aus den über Replikationen variierenden Urteilen die Schwelle als eigentlichen Meßwert zu gewinnen. Zu genau diesem Zwecke schlägt Fechner (1860, Bd. I, S. 104ff.) vor, bei der Methode der richtigen und falschen Fälle, dem Vorläufer des Paarvergleichsverfahrens, die Gaußsche Fehlerverteilung zu verwenden, die damit ihre erste Anwendung in der Psychologie erfährt. Diese Anwendung hatte freilich bei Fechner den Charakter eines reinen Hilfsmittels bei der Schwellenbestimmung und stand noch nicht in einem Zusammenhang mit der Konstruktion einer psychologischen Skala. Wundt (1908, S. 536) kritisierte daher Fechners Übertragung physikalischer Fehlerrnethoden in die Psychologie: „Was für den Physiker ein nicht näher zu untersuchendes störendes Moment ist, das wird also für den Psychologen zur Hauptsache; und was für den Physiker die Hauptsache, der endgültige Mittelwert, das ist für den Psychologen zumeist nur ein nebensächliches Orientierungsmittel. In der psychologischen Verwendung jener physikalischen Methoden der Fehlerelimination von ‚Fehlern' zu reden, ist daher streng genommen selbst ein Fehler: der physikalische Beobachtungsfehler ist vielmehr, psychologisch betrachtet, eine Abweichung von irgend einem mittleren Verhalten, deren individuelle psychologische Bedingungen das eigentliche Objekt der Untersuchung bilden." Für die Konstruktion psychologischer Skalen impliziert dies die Forderung, Variabilitäten bereits bei der Entwicklung von Theorien, aus denen sich die Skala ergibt, und nicht erst beim eigentlichen Akt der Messung im Sinne externer Fehlerkonzepte Rechnung zu tragen. Die Variabilität wird damit selbst zum Gegenstand der Modellbildung. Dieser Schritt wurde formal 1927 durch Thurstone geleistet: Er verlagerte die Normalverteilung gleichsam von außen nach innen. Mit der Annahme von internen Diskriminanzprozessen, die als normalverteilt angesehen werden, postulierte Thurstone einen expliziten psychologischen Mechanismus zur Erklärung von Urteilsvariabilität. Dies führte zum oben bereits behandelten Konzept der psychometrischen Funktion; für die Psychophysik stellt dieses Konzept eine der wichtigsten theoretischen Errungenschaften zur Behandlung probabilistischer Urteilsfluktuationen dar; zudem fand es in der Testtheorie sein Gegenstück in der Item-Response-Funktion.

Anders als in der Physik ist man also in der psychologischen Skalierung nicht vorrangig an den von allen variablen und konstanten Fehlern bereinigen Mittelwerten einer Messung interessiert, sondern gerade an den Gesetzmäßigkeiten, die der Variabilität von Urteilsprozessen zugrunde liegt. Demzufolge kommt hier probabilistischen Vorgehensweisen eine besondere Bedeutung zu; diese Haltung wird nachdrücklich von Falmagne (1976, S. 66) vertreten: „In the behavioral sciences, the only regularities are statistical ones. This erratic nature of behavioral data makes it almost mandatory that the theories be probabilistic."

So berechtigt eine solche Haltung im Hinblick auf bestimmte Fragestellungen der Psychologie grundsätzlich ist und so bedeutend die Rolle möglicher probabilistischer Modelle sein kann, so viel Zurückhaltung ist auf der anderen Seite geboten im Hinblick auf ‚Routine-Probabilisierungen' deterministischer Theorien ohne ausreichenden theoretischen Rückhalt (wenn sich etwa hinter umfasssenden wahrscheinlichkeitstheoretischen Betrachtungen der magere strukturelle Gehalt einer Rating-Skala verbirgt). Eine solche spiegelt sich in besonders extremer Weise in dem gelegentlich geäußerten Gedanken einer allgemeinen (gleichsam theorieinvarianten) formalen Behandlung des Fehlerproblems wider. So wurde in den 70er Jahren vorgeschlagen, die Meßstruktur im Rahmen boolschwertiger oder probabilistischer Modelltheorie zu formulieren, wodurch man jeder Aussage einen Wahrheitsgrad zuordnen kann. Dies führt aber, wie Heyer & Mausfeld (1987) zeigten, zu einer Trivialisierung des Fehlerproblems, da eine Aussage genau dann im boolsch-wertigen oder probabilistischen Sinne gültig ist, wenn sie von jeder Vp deterministisch erfüllt wird.

Oft stellt sich auch die Forderung der Probabilisierung deterministischer Strukturen als ein Reflex auf ein empiristisches Mißverständnis deterministischer Theorien dar. Auch wenn das Folgende nicht als ein grundsätzliches

Plädoyer für ein deterministisches Vorgehen angesehen werden soll, so scheint uns dieses Problem doch eines Exkurses wert.

Einer naiv-empiristischen Lesart zufolge müßten in aller Regel sowohl physikalische Gesetze wie auch die Axiome einer deterministischen Meßstruktur als falsifiziert angesehen werden. Dem liegt ein Mißverständnis zugrunde, welches in meßtheoretischem Zusammenhang seinen Ausdruck bereits in der Verwendung der Bezeichnung ***empirische Struktur*** findet. Die Relationen, die eine qualitative Struktur im Sinne der Meßtheorie charakterisieren, erweisen sich nämlich bei näherer Betrachtung nicht als Beobachtungsprädikate im naiven Sinne, sondern als theoretische (Dispositions)-Prädikate. Suppes et al. (1989, S. 300) stellen unmißverständlich fest: „A relational Statement, such as $a \succcurlyeq b$, is not considered to be the record of a particular record or experiment but is a theoretical assertion inferred from the data." Meßtheoretische Axiome als solche sind also keineswegs direkt empirisch überprüfbar. Gemessen an den entsprechenden empirischen Relationen konstituieren sie theoretische Idealisierungen, in denen sich der jeweils erreichte Stand der theoretischen Perspektive niederschlägt. Auf numerischer Seite findet diese theoretische Idealisierung ihre Entsprechung in der Verwendung der reellen Zahlen, mit denen das Bild einer idealiter beliebig genauen Messungen einhergeht; quantitative Vorhersagen aus der Theorie können also mit beliebiger Exaktheit formuliert werden, so daß von vornherein keine perfekte Übereinstimmung zwischen Theorie und Daten zu erwarten ist. Die Anbindung theoretischer Aussagen an Beobachtungsaussagen kann folglich nur durch theoretische Bindeglieder, Normen oder etablierte Bewertungskriterien erfolgen, die festlegen, wie groß und von welcher Art zulässige Spannungen zwischen theoretischen und empirischen Aussagen sein dürfen. Die Beziehung einer Meßstruktur zu experimentellen Beobachtungen spiegelt also das in der Wissenschaftstheorie umfassend diskutierte Problem von Beobachtung und Theorie wider. Vor diesem Hintergrund erweist sich eine Bewertung der Spannung zwischen den qualitativen Axiomen einer Meßstruktur und den entsprechenden experimentellen Beobachtungen als sehr subtiles Problem (für eine tiefgehende konzeptuelle Analyse sei auf Niederée, 1992a, b, verwiesen).

Deterministischer und probabilistischer Betrachtungsweisen in der Forschung sowie Auffassungen darüber, was man als ein geeignetes Experiment ansieht und was ‚Daten' sind, verknüpfen sich zusammen mit Bewertungskriterien und Bindeglieder, durch die erst eine theoretische und eine Beobachtungsaussage miteinander verbunden werden können, zu einem komplexen Wirkungsgefüge. Die Bindeglieder zwischen dem Theoriekern und seinem Gegenstandsbereich sind selbst in einem weiteren Sinne Bestandteil einer inhaltlichen Theorie und in ihren paradigmatischen Kontext eingebunden. Ein Theoriebereich besteht nämlich nicht nur aus dem idealisierten Theoriekern, etwa im Sinne der Miniatur- Theorie einer Meßstruktur, sondern er beeinhaltet stets auch die Formulierung zahlreicher Anfangs- und Randbedingungen sowie Hilfshypothesen, durch die erst die Theorie mit einem Experiment in Beziehung gesetzt werden kann. Im Prozeß der Entstehung einer Theorie bildet sich mit der (zumeist impliziten) Formulierung solcher Hilfshypothesen eine Ubereinkunft heraus, was als vernünftige Übereinstimmung zwischen Daten und Theorie anzusehen ist und was nicht. Dieser Konsens, der seinen prototypischen Niederschlag in Lehrbuchtabellen der für bestimmte Gesetze ,typischen' Meßwerte findet, muß als wesentlicher Bestandteil des paradigmatischen Kontextes einer Theorie angesehen werden, d. h. „without the tables, the theory would be essentially incomplete“ (Kuhn, 1961, S. 36). Solange es indes keine solchen aus der inhaltlichen Theorie und ihrem paradigmatischen Kontext erwachsenden Kriterien darüber gibt, welche Spannung man zwischen theoretischer Aussage und Beobachtungsaussage zu tolerieren gewillt ist, bleibt eine Entscheidung hierüber willkürlich.

Zweck einer Probabilisierung kann nicht sein, die Theorie gegen Verletzungen zu immunisieren und so ihren empirischen Gehalt zu reduzieren. Mit der Stochastisierung einer psychologischen Theorie muß vielmehr das Ziel verbunden sein, eine inhaltliche Theorie über die internen Mechanismen zu formulieren, welche für die Urteilsfluktuationen verantwortlich sind. Eine Stochastisierung, die nicht an hinreichend viele und hinreichend strenge empirische Restriktionen des betrachteten Gegenstandsbereiches angebunden ist, läuft Gefahr, zum Selbstzweck zu werden und die Anbindung an die substantielle Theoriebildung des jeweiligen Gegenstandsbereiches zu verlieren. Auf die Gefahren probabilistischer Modelle hat insbesondere Krantz (1972c) in seiner Diskussion der meßtheoretischen Perspektive hingewiesen. Ganz anders stellt sich die Behandlung von Beobachtungsfehlern dar, wenn sie nicht lediglich im Kontext einer isolierten Skalenkonstruktion erfolgt, sondern wenn die betrachtete Urteilsaufgabe in eine umfassendere inhaltliche Theorie eingebettet ist. Heyer & Mausfeld (1987) diskutieren unter diesem Aspekt zwei Beispiele aus der Psychophysik.

Damit kommen wir resümierend wieder zu der eingangs gemachten Feststellung zurück, daß Fehlertheorie und substantielle Theorie Hand in Hand gehen müssen. So wichtig es ist, eine Vielfalt formaler Techniken zur ,Fehler'-Behandlung verfügbar zu haben – und hierzu haben die Psychophysik und die Abstrakte Meßtheorie entscheidende Anstösse geliefert –, so unsinnig ist es, in aprioristischer Weise die Entwicklung einer allgemeinen Fehlertheorie für psychologische Urteilsprozesse zu verlangen. Es gibt nicht das Fehlerproblem in der Psychophysik oder Psychologie, sondern jeder Theoriebereich wird, ist er nur reichhaltig genug, eigene Arten der Fehlerbehandlung hervorbringen. Sieht man als Ziel einer naturwissenschaftlich orientierten Psychologie die Entwicklung reichhaltiger substantiellen Theorien in spezifischen psychologischen Phänomenbereichen, so stellt die Behandlung von Beobachtungsfehlern kein eigenständiges Problem mehr dar, wie beispielsweise ein Blick in die verschiedenen Bereichen einer Psychophysik der perzeptuellen Informationsverarbeitung erkennen läßt.

5. *Zur Rolle von Invarianzkonzepten bei der Konstruktion psychologischer Skalen*

Die durch die oben beschriebene Verfahren der Skalenkonstruktion erstellten Skalen zielen ursprünglich auf eine Messung von ‚Empfindungen'. Könnte man diese unabhängig messen, ergäbe sich im Falle eindimensionaler physikalischer Reize die Frage nach der Beziehung einer solchen Empfindungsskala u zu einer vorgegebenen zugehörigen physikalischen Skala f (unter der stillschweigenden Annahme, daß mit jedem Reize genau eine Empfindung assoziiert ist). Doch ist eine Skalenkonstruktion auf psychologischer Seite unter direkter Bezugnahme auf ‚Empfindungen' nicht möglich. Die Konstruktion einer solchen Skala wird sich daher auf gewisse psychologische Relationen beziehen müssen, welche mit dem Urteilsverhalten von Personen in Zusammenhang stehen, und ‚Empfindungen' als theoretische Begriffe (in Anlehnung an alltagssprachliche Vorstellungen) rekonstruieren und zu dem Urteilsverhalten in Beziehung setzen. Wir haben oben mehrere Verfahren, die Skala u zu konstruieren, kennengelernt, die teilweise auf rein psychologischen Relationen beruhen (z. B. Thurstone-Skalen), teilweise auf einem Zusammenspiel psychologischer und physikalischer Relationen (z. B. im Falle des Weberschen Gesetz sowie die Repräsentation einer Grassmann-Struktur). Die Transformation, die den Zusammenhang von physikalischer Skala f und psychophysikalischer Skala u beschreibt, bezeichnet man als psychophysikalische Funktion. (Freilich gibt es nicht zu einer Sinnesmodalität ‚die' psychophysikalische Funktion, da diese von f und u abhängt.) Häufig spiegeln sich in der Form der psychophysikalischen Funktion qualitativ-strukturelle Beziehungen zwischen den zugrunde liegenden psychologischen und physikalischen Relationen wider. In diese Beziehungen kann sich eine neue Einsicht ausdrücken, die über die bei der Skalenkonstruktion bereits vorausgesetzten Prinzipien hinausgeht, wie es in der Regel der Fall ist, wenn die Skala u lediglich auf psychologischen Relationen beruht. Im anderen Fall kann es geschehen, daß sich hier nur Prinzipien widerspiegeln, welche bereits der Skalenkonstruktion bereits zugrunde gelegt wurden (wie z. B. bei der mittels des Weberschen Gesetz gewonnenen Fechnerschen logarithmischen Skala).

Als besonders aufschlußreich bei der Analyse derartiger qualitativ-strukturellen Beziehungen erweisen sich Konzepte der Invarianz und Symmetrie, wclchen auf numerischer Seite die Konzepte Skalenniveau, zulässige *Transforma-*

tion und *meaningfulness* entsprechen. Hierdurch lassen sich mitunter Aufschlüsse über die Form der psychophysikalischen Funktion gewinnen. Der folgende Abschnitt beschäftigt sich mit diesen Konzepten; alle Betrachtungen werden nur für den technisch einfachsten Fall durchgeführt, damit die grundsätzlichen Perspektiven, soweit sie hier von Interesses sind, deutlich werden .werden.

5.1 Skalenniveau, zulässige Transformationen und *meaningfulness*-Konzepte

Während der 30er Jahre fand im Umfeld logisch-empiristischer Wissenschaftstheorie eine lebhafte Diskussion um die Rolle klassifikatorischer, komparativer und quantitativer Begriffe für die empirische Theoriebildung statt. Dabei war man besonders daran interessiert, die mit den Begriffen der Messung und der quantitativen Aussage zusammenhängenden konzeptuellen Fragen zu klären. Hierzu gehörten Fragen nach dem empirischen Gehalt quantitativer Aussagen und die Suche nach Kriterien, wann eine quantitative Aussage als sinnvoll oder als sinnlos anzusehen ist. Cohen & Nagel (1934) unterschieden drei Typen numerischer Skalen: Skalen, deren Zahlen nur zur Identifikation empirischer Objekte dienen, Skalen, die eine empirische Ordnung widerspiegeln, und schließlich Skalen, die eine quantitative empirische Beziehung widerspiegeln. Diese Klassifikation wurde durch Carnap, Hempel und vor allem durch den Mathematiker G.H. Birkhoff präzisiert. Diese Überlegungen wurden innerhalb der Psychologie von S.S. Stevens aufgegriffen und als Theorie der Skalenniveaus bekannt gemacht. Seitdem wurden und werden derartige Fragen vor allem in der Psychologie diskutiert, wo sie als besonders bedrängend empfunden werden.

Die mit der Stevensschen Klassifikation verbundenen Probleme seien an einem einfachen Fall verdeutlicht. Dazu betrachten wir statt quantitativer Aussagen numerische Relationen (Invarianzuntersuchungen von quantitativen Aussagen werden zumeist auf solche der in ihnen enthaltenen numerischen Relationen zurückgeführt, da der Begriff ‚quantitative Aussage' sehr viel schwerer präzise formal zu klären ist). Sei mit $\mathbf{A}$ eine Menge von ‚empirischen Objekten' bezeichnet und mit N eine Menge von ‚numerischen Objekten' (etwa $N \subset Re$). R sei eine n-stellige Relation auf N. Den Objekten in $\mathbf{A}$ sollen nun Zahlen aus N zugeordnet werden, welche die Objekte in bestimmten vorgegebenen Aspekten charakterisieren sollen. Eine Abbildung, die dies leistet, heißt eine Skala. Die Menge aller Funktionen, die dies in ‚analoger Weise' leisten, heißt Skalenfamilie (bzgl. $\mathbf{A}$ und N) und sei mit F bezeichnet. Nun können wir Kriterien für die Invarianz von Relationen formulieren:

(M1) Sei R eine n-stellige Relation auf N und sei T eine Menge von Funktionen t: $\mathbf{N}\rightarrow\mathbf{N}$. Dann heißt R *T-invariant* genau dann, wenn für alle $t \in T$ und alle $x_1,...,x_n \in \mathbf{N}$ gilt:

$$(x_1,...,x_n) \in R \Leftrightarrow (t(x_1),...,t(x_n)) \in R.$$

(M2) R heißt *F-invariant* genau dann, wenn für alle $f,g \in F$ und alle $a_1,...,a_n$ **A** gilt:

$$(f(a_1),...,f(a_n)) \in R \Leftrightarrow (g(a_1),...,g(a_n)) \in R$$

Das meaningfulness-Kriterium (M2), von Adams, Fagot & Robinson (1965, S. 105f.) als „reference invariance" bezeichnet, ist zunächst intuitiv verständlicher als (M1), da es gerade die Forderung beinhaltet, daß eine numerische Relation R invariant unter allen ‚Repräsentationen' der empirischen Objekte sein soll, die hinsichtlich der geforderten Charakterisierung ‚gleichermaßen' geeignet sind. Das Kriterium (M1) führt dagegen unmittelbar zu der Frage, wie sich eine geeignete Menge T charakterisieren läßt. Eine einzelne Abbildung t mit der Eigenschaft $t \circ f \in F$ für alle f (wobei mit o die Komposition von Abbildungen bezeichnet wird), heißt *zulässige Transformation* (bzgl. F). Wie läßt sich nun die Menge der zulässigen Transformationen bestimmen? Diese Frage ist innerhalb des Stevensschen Ansatzes nicht in befriedigender Weise zu klären. Wie die beiden genannten Kriterien zusammenhängen, wird in dem Fall deutlich, daß für ein beliebiges $f_0 \in F$ gilt: $F = \{t \circ f_0 | t \in T\}$, d. h. mittels T läßt sich aus jeder beliebigen Skala ganz F erzeugen. Die T-Invarianz erweist sich aus dieser Perspektive als verkappte F-Invarianz; diese ist der grundlegendere Begriff, auch wenn in der Literatur häufig mit der Betonung *Invarianz unter den zulässigen Transformationen* der gegenteilige Eindruck erweckt wird. Erst im Repräsentationsansatz der Messung werden die hier angesprochenen Probleme in befriedigender Weise gelöst. Die Skalenfamilie F ist dann gerade die Menge der strukturerhaltenden Abbildungen (Homomorphismen bzw. Isomorphismen) zwischen qualitativer und numerischer Struktur, wodurch sich die obige Beschreibung, daß die Abbildungen in F das Geforderte ‚in analoger Weise' leisten, präzise fassen läßt. Das Skalenniveau ergibt sich dabei (bis auf Konjugation von Skalenfamilien, die dem Übergang von einer numerischen Repräsentationsstruktur zu einer anderen entsprechen; vgl. Krantz et al., 1971, S. 99ff., sowie Falmagne & Narens, 1983) aus bestimmten Annahmen über die betrachtete qualitative Struktur. (Hierbei wird auch deutlich, daß die beiden obigen Kriterien nicht notwendig äquivalent sein müssen, da es Situationen geben kann, in denen zwei Homomorphismen f und g durch ‚zulässige Transformationen' nicht ineinander überführt werden können.) Ausführliche Darstellungen finden sich in Adams, Fagot & Robinson (1965), Pfanzagl (1971), sowie Luce et al. (1990, Kap.22).

Während damit die Bestimmung der Menge der zulässigen Transformationen, d. h. des Skalenniveaus, für die Repräsentation einer qualitativen Struktur geklärt ist, stellt sich darüber hinaus die Frage nach einer qualitativen Charakterisierung des Sachverhalts, daß eine Struktur mit einem bestimmten Skalenniveau repräsentiert werden kann. Eine wesentliche Rolle in diesem Zusammenhang spielen Symmetriekonzepte, wie das des Automorphismus einer Struktur (als eines strukturinternen Isomorphismus). Ein Automorphismus α einer Struktur $\boldsymbol{A} = \langle A, S_1, \ldots, S_n \rangle$ ist eine bijektive Abbildung von A auf sich selbst mit der Eigenschaft, daß für alle $a_1, \ldots, a_{k(j)} \in A$ und für alle $j = 1, \ldots, n$ gilt:

$$(a_1, \ldots, a_{k(j)}) \in S_j \Leftrightarrow (\alpha(a_1), \ldots, \alpha(a_{k(j)})) \in S_j, \tag{5.1}$$

wobei $k(j)$ die Stelligkeit der Relation S_j bezeichne.

Betrachten wir nun das folgende Beipiel: Sei $\mathbf{A}$ eine extensive Struktur, d. h. $\boldsymbol{A} = (A, \succcurlyeq, \bigcirc)$, $\Re = (Re+, \geq, +)$ und f ein Isomorphismus zwischen beiden Strukturen. Dann läßt sich zeigen, daß auch $r \cdot f$ $(r > 0)$ ein Isomorphismus ist und daß je zwei Isomorphismen in dieser Weise zusammenhängen. Wir betrachten nun die Abbildung $\alpha_r = f^{-1}rf$, d. h. $\alpha_r(a) = f^{-1}(rf(x))$; diese überführt $\boldsymbol{A}$ in $\mathbf{A}$ und ist ein Autornorphismus von A, dem auf numerischer Seite der durch $t(x)=rx$ definierte Automorphismus von $\Re$ entspricht. Umgekehrt gilt: Ist α ein Automorphismus von $\mathbf{A}$, dann ist die Abbildung $f \circ \alpha$ ein Isomorphismus von $\boldsymbol{A}$ nach $\Re$. Derartige Betrachtungen führen, wie erstmals Narens (1981) gezeigt hat, zu einem tieferen qualitativen Verständnis der Skalenniveaus: Jedes Skalenniveau auf numerischer Seite spiegelt interne Symmetrien der qualitativen Struktur wider. Von besonderer Bedeutung in diesem Zusammenhang ist die Narenssche Klassifikation der Automorphismen einer qualitativen Struktur nach ihrer Reichhaltigkeit und Redundanz. (Die Reichhaltigkeit der Automorphismengruppe wird durch Narens' Konzept der „M-homogeneity" beschrieben, die angibt, ob es einen Automorphismus gibt, der zwei vorgegebene M-Tupel geordneter Elemente ineinander überführt; die Redundanz wird durch die „N-uniqueness“ erfaßt, die angibt, ob zwei Automorphismen, die an N Punkten übereinstimmen, notwendigerweise identisch sind.) Aufbauend auf dieser Klassifikation weisen Narens und Alpern in mathematisch tiefliegenden Resultaten insbesondere nach, daß (für geordnete Strukturen, die ein Kontinuum bilden) die üblichen Skalentypen fast vollständig die möglichen Symmetrien beschreiben, obgleich auf numerischer Seite a priori beliebige Arten zulässiger Transformationen möglich sein könnten (siehe hierzu Luce et al., 1990, Kap. 20).

In entsprechender Weise lassen sich numerischen Invarianzkonzepten solche qualitativer Art gegenüberstellen. Diese beruht auf dem folgenden Begriff der Struktur-Invarianz.

(M3) Sei A eine Struktur mit Trägermenge A und S eine Relation auf **A.** S heißt *strukturinvariant* (genauer: A-invariant) genau dann, wenn für alle Automorphismen a der Struktur und alle $a_1,\ldots,a_n \in A$ gilt:

$$((a_1),\ldots,(a_n)) \in S \Leftrightarrow (\alpha(a_1),\ldots,\alpha(a_n)) \in S.$$

Diese Bedingung besagt gerade, daß S invariant ist unter den Automorphismen der Struktur. Der Zusammenhang zu den zuvor formulierten Invarianz-Konzepten (M1) und (M2) soll wiederum an einem Beispiel verdeutlicht werden. Sei A eine qualitative Struktur, $\mathfrak{R}$ eine isomorphe numerische Struktur und f ein Isomorphismus zwischen beiden. Betrachten wir nun eine ‚neue' qualitative Relation S auf A und die zugeordnete numerische Relation R = f(S). Dann ist S genau dann strukturinvariant, wenn R F-invariant ist, wobei F die Menge aller Isomorphismen zwischen den beiden Strukturen ist. (Und dies ist wiederum genau dann der Fall, wenn R T-invariant ist, wobei T die Menge aller Automorphismen von 32 ist.) Für die allgemeinen Beziehungen der aufgeführten Invarianzkriterien zueinander siehe Luce et al. (1990, Kap. 22) sowie Narens (1985, S. 156ff.).

Betrachten wir als Beispiel die im Abschnitt 3.3 im Zusammenhang mit dem Weberschen Gesetz eingeführte qualitative Semiordnung $\succ_m$. Sei, wie oben, A = (A, $\succcurlyeq$, **O**) ein physikalisches Kontinuum und sei $\mathfrak{R}$ = (Re', $\geq$, +). Unter Bezugnahme auf die in Gleichung **(3)** dargestellte Beziehung

$$a \succ_m b \Leftrightarrow f(a) > f(b) + \Delta_f(b)$$

läßt sich die zugeordnete numerische Relation R schreiben als

$$R = \{(x,y) \mid x > y + \Delta(y)\},$$

(wobei $\Delta(f(b)) = \Delta_f(b)$).

f ist in diesem Fall eine Verhältnisskala, R somit F-invariant, wenn es invariant unter linearen Abbildungen ist. Dies wiederum ist genau dann der Fall, wenn $\Delta(y) = k \cdot y$, d. h., wenn das Webersche Gesetz gilt. Dies nun ist wiederum äquivalent dazu, daß $\succ_m$ strukturinvariant ist. Mit anderen Worten: Die Gültigkeit des Weberschen Gesetzes ist dazu äquivalent, daß die Semiordnung $\succ$ bezüglich der zugrunde liegenden extensiven physikalischen Struktur invariant ist.

Die Grenzen dieser Betrachtungsweise werden deutlich, wenn man Varianten des Weberschen Gesetzes betrachtet, welche der Existenz einer absoluten Schwelle Rechnung tragen: Die Einführung einer absoluten Schwelle ist nämlich mit der Annahme der Strukturinvarianz unverträglich und macht eine geeignete Verfeinerung dieses und verwandter Konzepte erforderlich.

5.2 Die ‚möglichen Formen' psychophysikalischer Funktionen und Gesetze

In einer klassischen Arbeit beschäftigte sich Luce (1959) mit Bedingungen, unter denen sich die Form der psychophysikalischen Funktion bestimmen läßt. Restriktionen zur Bestimmung ihrer Form gewinnt Luce aus Voraussetzungen eines Skalenniveaus der mit der physikalischen Skala f und der psychophysikalischen Skala u verbundenen Skalenfamilien und aus einer metatheoretischen Forderung, die er als „principle of theory construction" bezeichnet (Luce, 1959, S. 85). Jede psychophysikalische Funktion muß danach die Eigenschaft haben, daß eine zulässige Transformation der unabhängigen Variablen (hier also: der physikalischen) zu einer zulässigen Transformation der abhängigen Variablen (hier also: der psychophysikalischen) führt.

Wird die psychophysikalische Funktion (bzgl. f und u) mit Φ bezeichnet, d. h. besteht für alle Reize a die Beziehung $u(a) = \Phi(f(a))$, so bedeutet dies gerade, daß zu jeder zulässigen Transformation t der physikalischen Skala eine zulässige Transformation t' der psychologischen Skala u existiert, so daß gilt:

(M4) $\Phi(t(f(a))) = t'(u(a)).$

Setzt man beispielsweise voraus, daß die physikalische Skala (genauer Skalenfamilie) ebenso wie die psychophysikalische eine Verhältnisskala ist, so ergibt sich aus (M4) eine Funktionalgleichung, aus welcher (unter bestimmten Stetigkeitsannahmen) folgt, daß Φ eine Potenzfunktion sein muß. Setzt man statt dessen voraus, daß u eine Intervallskala ist, so ist auch eine logarithmische Form von Φ möglich. In diesem Sinne behandelt Luce eine Reihe von Kombinationen verschiedener Skalenniveaus. Dies setzt selbstverständlich eine angemessene Festlegung insbesondere des psychophysikalischen Skalenniveaus voraus.

Die Bedingung (M4) läßt sich zu den oben eingeführten qualitativen Invarianzkonzepten in folgender Weise in Beziehung setzen: Sei **A** die Menge der physikalischen Reize, und nehmen wir an, daß die betrachtete physikalische Skalenfamilie gerade die Menge aller Isomorphismen zwischen einer qualitativen physikalischen Struktur $\mathcal{A}$ mit Trägermenge **A** und einer numerischen Struktur $\Re_1$ ist und die psychophysikalische Skalenfamilie die Menge der Isomorphismen zwischen einer psychophysikalischen Struktur B mit Trägermenge A und einer numerischen Struktur $\Re_2$. (M4) erweist sich dann gerade als äquivalent zu der Annahme, daß die Relationen der psychophysikalischen Struktur A-invariant sind.

Vergleichbare und über die Luceschen hinausgehenden Resultate erzielten Roberts & Rosenbaum (1986), die sich jedoch dem Problem aus einer anderen Perspektive nähern als Luce.

Die psychophysikalische Funktion ist nur ein spezielles Beispiel einer Beziehung zwischen Skalen in der Psychophysik. So wurden von Falmagne & Narens (1983) psychophysikalische Gesetze untersucht, die sich auf zwei numerische Input-Codes und einen numerischen Output-Code beziehen. Für die Beziehungen zwischen den Input- und Output-Codes formulieren sie verschiedene Invarianzprinzipien und untersuchen deren logische Beziehung sowie die Einschränkungen, die sich (unter gewissen strukturellen Zusatzannahmen) aus diesen Invarianzprinzipien für den Zusammenhang von Input- und Output-Codes ergeben.

Invarianz-Argumente dieser Art können also dazu beitragen, von vornherein die Klasse der mit ihnen verträglichen numerischen Gesetze deutlich einzugrenzen. Ein Beispiel hierzu aus der Psychoakustik findet sich in Falmagne & Narens (1983, S. 314ff.; s. a. Falmagne, 1985, S. 339ff.). Ein bekannter Fall, durch derartige Betrachtungen unzulässige Formen von Gesetzen auszuschließen, ist die Dimensionsanalyse der Physik, zu der die genannten Invarianzbetrachtungen eine Verwandtschaft aufweisen. Die Dimensionsinvarianz als Forderung auf numerischer Seite ist äquivalent mit der Struktur-Invarianz auf Seiten der qualitativen Struktur physikalischer Größen, wie Luce (1978; siehe auch Ramsey, 1976) gezeigt hat.

5.3 Wie sinnvoll sind *meaningfulness*-Betrachtungen?

Die genannten Invarianz-Kriterien basieren auf metatheoretischen Intuitionen, denen in der Physik als Dimensionsanalyse und Symmetriebetrachtungen und in der Mathematik als ein Ordnungsschema auf der Grundlage der Invarianz geometrischer Konzepte unter bestimmten Transformationsgruppen (Kleinsches Programm) eine große Bedeutung zukommt. In der Physik erfüllen fast alle bekannten Gesetze diese Forderungen, und die auf Invarianzkonzepten beruhenden Techniken haben sich als fruchtbar sowohl zum Auffinden möglicher wie auch zur Elimination unbrauchbarer Gesetzesformen erwiesen. Im Kontext der Psychophysik bringen die obigen Invarianzforderungen zudem zum Ausdruck, daß ein System gesetzmäßiger Beziehungen zwischen physikalischen und psychologischen Variablen gewisse Kohärenzeigenschaften und Symmetrien haben soll.

Betrachten wir die Tragweite solcher Invarianzforderungen am Beispiel des Weberschen Gesetzes: Invarianzbetrachtungen der genannten Art führen, wie wir oben gesehen haben, zu einer Auszeichnung des Weberschen Gesetzes unter den möglichen Gesetzen, die das Verhalten der Unterschiedsschwelle in Abhängigkeit vom physikalischen Reiz beschreiben könnten; diese wollen wir kurz als Sensitivitätsgesetze bezeichnen. Dies kann jedoch nicht heißen, daß das Webersche Gesetz das einzig mögliche Sensitivitätsgesetz ist, wie es eine aprioristische Lesart von Invarianzbetrachtungen impliziert, die in prototypischer Weise der Begriff *(non-)meaningfulness* und das Lucesche „principle of theory construction" zum Ausdruck bringen. In der Tat sind nicht-lineare Sensitivitätsgesetze in der Psychophysik häufig zu finden. In entsprechender Weise kritisierte Rozeboom (1962) die Anwendung des Luceschen Prinzips bereits für die Physik unter Verweis auf das exponentielle Gesetz des radioaktiven Zerfalls. Dies veranlaßte Luce (1962), seine Position entsprechend zu modifizieren.

Sensitivitätsgesetze, wie etwa „Δs ist proportional zu s^2", die nicht-invariante Relationen $\succ_m$ beschreiben, sind nicht in einem absoluten Sinne „non-meaningful". Anders als im Falle der invarianten Gesetze erfordert die präzise numerische Spezifikation eines eine solche Relation beschreibenden Gesetzes, d. h. der obigen Relation R, den Bezug auf eine konkrete Skala bzw. Einheit. Unter Zugrundelegung einer bestimmten extensiven Skala f ergibt sich für das soeben genannte Gesetz eine Spezifikation der Art $\Delta_f(a) = c_f f(a)^2$. Die Nicht-Invarianz drückt sich nun darin aus, daß der Wechsel zu einer anderen Skala mit einer Anderung der Konstanten c_f einhergeht. Derartige Skalenabhängigkeiten numerischer Spezifikationen von Gesetzen werden durch Pfanzagls (1971, S. 50) Konzept der *meaningful parametrization* erfaßt.

(M5) Eine Familie von Relationen $(R_f)_{f\in F}$ heißt *meaningfully parametrized* genau dann, wenn für alle $f,g \in F$ und alle $a_1,\dots,a_n \in A$ gilt:

$$(f(a_1),\dots,f(a_n)) \in R_f \Leftrightarrow (g(a_1),\dots,g(a_n)) \in R_g.$$

(Siehe hierzu auch Luce et al., 1990, S.278). Im obigen Beispiel ist $R_f = \{(x,y) \mid x = y + c_f y^2\}$. Die Möglichkeit, Skalenabhängigkeiten in der Art von (M5) zuzulassen, von der auch die Physik Gebrauch macht, zeigt die Grenzen eines aprioristischen *meaningfulness*-Verständnisses.

Dies bedeutet jedoch nicht, daß den genannten Invarianzkonzepte und verwandten Intuitionen in den Wissenschaften nicht dennoch eine erhebliche Rolle zukommt, wie es bereits das Beispiel der Physik lehrt (vgl. v. Fraasen, 1989). Sie bilden ein wesentliches Element theoretischen Denkes jenseits eines reinen *curve fittings*, in dem sich die Haltung spiegelt, ,intrinsisch Gesetzhaftes' von ,kontingenten empirischen Generalisierungen' scheiden zu wollen. Diese Haltung ist oftmals mit Auffassungen einer ,Theorie-Asthetik' verbunden, wie sie seit Pythagoras die Naturwissenschaften insbesondere in der Form von Symmetriebetrachtungen prägt (vgl. Guthrie, 1962, S. 236ff.; v. Fritz, 1981, S. 47ff.).

Invarianzkonzepte weisen einen engen Zusammenhang mit Definierbarkeitskonzepten auf. So sind insbesondere alle in einer Struktur A (etwa im Sinne Prädikatenlogik erster Stufe) definierbaren Relationen A-invariant. (Die Umkehrung gilt jedoch nur in einem eingeschränkten Sinne.) Läßt sich also eine neue Relation unter Verwendung der Grundrelationen einer Struktur definieren, so ist sie strukturinvariant. Dies eröffnet eine interessante neue Perspektive für *meaningfulness-Betrachtungen:* Man geht von einer vorgegebenen Struktur aus und betrachtet deren Grundrelationen als *meaningful*; sodann definiert man relativ zu dieser gegebenen Struktur solche Relationen als *meaningful*, die sich (in einem zu präzisierenden Sinne) aus den Grundrelationen definieren lassen. Diese Art des Zugangs wurde von Narens (1988) vorgeschlagen. Auch hier ist wieder vor einem aprioristischen Mißverständnis zu warnen, da es natürlich keine absoluten Kriterien dafür geben kann, welche Struktur als ‚Fundament' derartigen Betrachtungen zugrunde zu legen ist. Narens & Mausfeld (1992) ziehen Definierbarkeitskonzepte heran, um im Rahmen eines erweiterten Invarianz-Klassifikationskriteriums für die Psychophysik Kriterien dafür anzugeben, wann zwei physikalische Situationen als äquivalent angesehen werden sollen. Hat man ein solches Kriterium gewonnen, so trägt das folgende psychophysikalische Äquivalenzprinzip der asymmetrischen Beziehung zwischen physikalischen und psychologischen Relationen Rechnung: Eine notwendige Voraussetzung dafür, daß eine psychophysikalische Aussage einen psychologischen Gehalt hat, ist, daß ihr Wahrheitswert unabhängig von der Wahl äquivalenter physikalischer Systeme ist. Dem von Narens & Mausfeld (1992) entwickelten Kriterium zufolge sind zwei physikalischer Strukturen mit gleicher Trägermenge genau dann äquivalent, wenn sie isomorph sind und sich jede Relation der einen Struktur durch eine Relation der anderen Struktur (in einer hinreichend reichhaltigen logischen Sprache) definieren läßt. Im Falle zweier stetiger extensiver Strukturen mit Repräsentationen f_1 bzw. f_2 ist dies äquivalent zu der Bedingung, daß es ein $r \in \mathrm{Re}^+$ gibt, so daß für die beiden Repräsentationen gilt:

$$f_1^{-1}[f_1(x) + f_1(y)] = f_2^{-1}[f_2(x)^r + g_2(y)^r]^{1/r}.$$

Ersetzt man also in einer extensiven Struktur mit Operation $\bigcirc_1$ und Repräsentation f_1 die extensive Operation durch eine Operation $\bigcirc_2$, für die gilt:

$$x \bigcirc_2 y = f_1^{-1}[f_1(x)^r + f_1(y)^r]^{1/r},$$

so erhält man eine physikalisch äquivalente Struktur. Man kann leicht zeigen, daß physikalisch äquivalente Strukturen dieselbe Automorphismengruppe haben (da eine Transformation der Art $f \rightarrow \alpha f$ die beiden obigen Ausdrücke nicht ändert).

Gegenüber der klassischen *meaningfulness*-Konzeption, wie sie insbesondere durch die obigen Bedingungen (M1)-(M3) beschreiben wurde, bezieht das Äquivalenzprinzip nicht nur die numerische Repräsentation, sondern die qualitativen physikalischen Operationen selbst mit in die Invarianzforderung ein und verschärft damit die Bedingungen, unter denen quantitative Aussagen und numerische Relationen in einer psychophysikalischen Struktur als psychologisch relevant anzusehen sind.

Im Lichte des Aquivalenzprinzips erweisen sich eine Reihe quantitativer Aussagen der Psychophysik als nicht psychologisch relevant: Wie leicht erkennbar ist, ergibt sich beispielsweise aus dem Äquivalenzprinzip, daß die durch Größenschätzung erhaltenen Exponenten nicht die untersuchte Sinnesmodalität charakterisieren, sondern ebenfalls von der verwendeten physikalischen Operation abhängen (im Beispiel der Lautstärke verdoppelt sich der Exponent, wenn man den physikalischen Reiz in Einheiten des Schalldrucks statt der Schallenergie mißt). Sinnesmodalitäten, deren Exponent größer als 1 ist als „expanders“ und solche, deren Exponent kleiner als 1 ist als „compressors“ zu klassifizieren, wie Stevens (e. g. 1974, S. 370) dies unternimmt, ist dem Äquivalenzprinzip zufolge nicht sinnvoll, da einer solchen Klassifikation kein empirisch-psychologischer Gehalt zukommt.

Die obigen Ausführungen legen nahe, Invarianz-Forderungen in Psychophysik und Psychologie nicht als strikte aprioristische ‚Sinnkriterien‘ zu betrachten, sondern als ein heuristisches Forschungsmittel. Beide Arten von *meaningfulness*-Betrachtungen, die auf Symmetrieintuitionen basierenden Invarianzbetrachtungen und die durch eine logisch-reduktionistische Perspektive motivierten Definierbarkeitsbetrachtungen sind aprioristische Metaprinzipien, durch die Willkürliches bei der Theoriebildung ausgeschlossen werden soll und die angeben, welche Konzepte im weitesten Sinne in der theoretischen Sprache formulierbar bzw. ausdrückbar sind. Der Wert eines *meaningfulness*-Kriteriums hängt natürlich in entscheidender Weise von der gewählten theoretischen Sprache bzw. von der Ausgangsstruktur ab (für deren Wahl sich andererseits kaum aprioristische Kriterien finden lassen). So können derartige Betrachtungen nur eine Art Faustregel sein, denn zu diesen Prinzipien in Widerspruch stehende Fälle können, wie Rozebooms (1962) Beispiel des exponentiellen Zerfallgesetzes zeigt, stets durch Zusatzannahmen eingeführt werden; doch weisen *meaningfulness*-Betrachtungen darauf hin, daß solche Zusatzannahmen explizit gemacht werden müssen, und erhöhen dadurch gleichsam die Hürde für Willkürelemente.

Daß eine zu strikte Handhabung von meaningfulness-Forderungen einer angemessenen Theorienentwicklung gar entgegenstehen kann, läßt sich an einem Beispiel aus der Farbwahrnehmung verdeutlichen: Die empirisch gefundene Nicht-Linearität des sog. Blau-Gelb-Kanals der Opponenten Theorie der Farbcodierung bedeutet, daß die Einführung sog. Urfarben in die Grassmann-Struktur *non-meaningful* ist, da sie deren Automorphismen zerstört. Diese *non-meaningfulness* auf der Ebene der Grassmann-Struktur bedeutet aber einen echten Erkenntnisgewinn.

Dieses Beispiel macht deutlich, daß unter dem Aspekt von Forschungsheuristiken meaningfulness-Betrachtungen notwendigerweise in einem gewissen Sinne konservativ sind. Jeder echte Zuwachs einer Theorie an empirischem Gehalt kann dazu führen, daß die neue Theorie weniger Automorphismen aufweist als die alte (würde man wie in einem Prosastück von J. L. Borges eine Theorie zulassen, die eine 1-1-Darstellung der Welt wäre, so hätte diese bis auf die Identität keine Automorphismen mehr). Zwischen einem Erkenntnisfortschritt und meaningfulness-Betrachtungen besteht also eine gewisse Spannung, und nur bezogen auf die jeweilige Stufe der Theoriebildung können diese ein Instrument zur Klärung methodologischer Probleme sein. In jedem Fall ist die Bezeichnung *meaningfulness*, die sich historisch aus den semantischen Sinnkriterien des logischen Empirismus herleitet, eine sehr unglückliche Wortwahl (bestenfalls könnte man von einer konditionalen *meaningfulness* sprechen); das Konzept hat mit Attributen wie ‚sinnvoll' und ‚bedeutsam' wenig gemein, kann ihnen in einem Prozeß der Theorieentwicklung gar entgegenstehen.

Ihrem Ursprunge nach spiegeln sich in Invarianzkriterien physikalische Sichtweisen wider. Somit stellt sich die Frage nach der Angemessenheit der Ubertragung einer solchen Perspektive in die Psychophysik. Diese Frage muß als offen angesehen werden, steht doch, so scheint es, die Psychophysik in Aspekten der Theoriebildung der Biologie näher als der Physik. Aus neurophysiologischer Perspektive ist eine strenge und globale Erfülltheit von Invarianzkriterien nicht zu erwarten, was indes approximative Gültigkeit nicht ausschließt. Nähert man sich der Psychophysik aus funktionalistischer Sicht (computational approach), so können geeignet verfeinerte und inhaltlich eingebundene Invarianzkonzepten wiederum von entscheidender Bedeutung sein. Hierauf verweisen schon, wenn auch in subtiler Weise, die Intuitionen, welche an den Begriff der Invarianten gebundenen sind (vgl. Mausfeld, in diesem Band, Kap. 4), die wiederum mit den Konstanzleistungen des Wahrnehmungssystems zusammenhängen.

Literatur

Adams, E. W., Fagot, R. F. & Robinson, R. **(1965).** A theory of appropriate statistics. Psychometrika, 30, **99–127.**

Anderson, N. H. **(1982).** Methods of information integration theory. New York: Academic Press.

Arbuckle, J. & Larimer, J. **(1976).** The number of two-way tables satisfying certain additivity axioms. Journal of Mathematical *Psychology*, 13, **89–1**00.

Baird, J. C. & Noma, E. **(1978).** Fundamentals of *scaling and psychophysics*. New York: Wiley.

Bock, R. D. & Jones, L. V. (1968). *The measurement and prediction of judgement and choice.* San Francisco: Holden-Day.

Böhme, G., van den Daele, W. & Krohn, W. (1974). Die Finalisierung der Wissenschaft. In W. Dietrich (Hrsg.) *Theorien der Wissenschaftsgeschichte* (S.276–311). Frankfurt: Suhrkamp.

Bourbaki, N. (1966). *General Topology. Vol. II.* Paris: Hermann.

Brentano, F. (197411924). *Psychologie vom empirischen Standpunkt.* Bd. I. Leipzig.

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science, 3,* 186–190.

Cohen, M. R. & Nagel, E. (1934). *An introduction to logic and scientific method.* London: Routledge & Kegan Paul.

Colonius, H. (1984). *Stochastische Theorien individuellen Wahlverhaltens.* Berlin: Springer.

Debreu, G. (1959). Topological rnethods in cardinal utility theory. In K.J. Arrow, S. Karlin & P. Suppes (Eds.), *Mathematical models in the social sciences.* (S.16–26). Stanford: Stanford University Press.

Drösler, J. (1989). Quantitative psychology. Göttingen: Hogrefe.

Falmagne, J.-C. (1976). Random conjoint measurement and loudness summation. *Psychological Review, 83,* 65–79.

Falmagne, J.-C. (1979). On a class of probabilistic conjoint measurement models: Some diagnostic properties. *Journal of Mathematical Psychology, 19,* 73–88.

Falmagne, J.-C. (1985). *Elements of psychophysical theory.* Oxford: Clarendon.

Falmagne, J.-C. (1986). Psychophysical measurement and theory. In K.R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance. Vol. I. Sensory processes and perception.* New York: Wiley.

Falmagne, J.-C. & Narens, L. (1983). Scales and meaningfulness of quantitative laws. *Synthese, 55,* 287–325.

Fechner, G. Th. (1860). *Elemente der Psychophysik. Bd. I. & II.* Leipzig: Breitkopf & Härtel.

Fechner, G. Th. (1877). *In Sachen der Psychophysik.* Leipzig.

Fechner, G. Th. (1888). Uber die psychischen Maßprinzipien und das Webersche Gesetz. *Philosophische Studien,* 4, 161–230.

Fishburn, P. C. (1992). Induced binary probabilities and the linear ordering polytope: A status report. *Mathematical Social Sciences,* **23,** 67–80.

v. Fritz, K. (1971). *Grundprobleme der Geschichte der antiken Wissenschaft.* Berlin: de Gruyter.

Gigerenzer, G. & Strube, G. (1983). Are there limits to binaural additivity of loudness? *Journal of Experimental Psychology: Human Perception and Performance,* **9,** 126–136.

Gould, S.J. (1981). *The mismeasure of man.* New York: Norton.

Grassmann, H. (1853). Zur Theorie der Farbmischung. *Poggendorffs Annalen der Physik und Chemie, 89,* 69–84.

Guthrie, W.K.C. (1962). *A history of greek philosophy. The earlier Presocratics and the Pythagoreans.* Cambridge: Cambridge University Press.

Helmholtz, H.v. (1887). Zählen und Messen erkenntnistheoretisch betrachtet. In *Philosophische Aufsätze Eduard Zeller gewidmet.* Leipzig: Fues.

Heyer, D. & Mausfeld, R. (1987). On errors, probabilistic measurement and boolean valued logic. *Methodika, 1,* 113–138.

Heyer, D. & Niederée, R. (1992). Generalizing the concept of binary choice systems induced by rankings: One way of probabilizing deterministic measurement structures. *Mathematical Social Sciences,* 23, 31-44.

Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Mass. *Berichte und Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipig, Mathematisch-Physische Classe, 53,* 1–46.

Holman, E.W. & Marley, A.A.J. (1974). Stimulus and response measurement. In E.C. Carterette & M.P. Friedman (Eds.), *Handbook ofperception. Vol. II.* (S.173–213). New York: Academic Press.

Iverson, J. (1991). Probabilistic measurement theory. In J.-P. Doignon & J.-C. Falmagne (Eds.), *Mathematical Psychology. Current developments* (S.135–155). Berlin: Springer.

Kuhn, T.S. (1961). The function of measurement in modern science. In H. Wolf (Ed.), *Quantification. A history of the meaning of measurement in the natural and social sciences.* (S.31–63). Indianapolis: Bobbs-Merrill.

Krantz, D.H. (1972a). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology, 9,* 168–199.

Krantz, D.H. (1972b). Visual scaling. In L.M. Hurvich & D. Jameson (Eds.), *Handbook of sensory physiology, VII/4, Visual psychophysics.* (S.660-689). Berlin: Springer.

Krantz, D.H. (1972c). Measurement structures and psychological laws. *Science, 175,* 1427–1435.

Krantz, D.H., Luce, R.D., Suppes, P. & Tversky, A. (1971). *Foundations of measurement. Vol. I. Additive and polynomial representations.* New York: Academic Press.

von Kries, J. (1882). Über die Messung intensiver Größen und über das sogenannte psychophysische Gesetz. *Vierteljahresschrift für wissenschaftliche Philosophie,* 6, 257–294.

Krull, W. (1960). Über die Endomorphismen von total geordneten Archimedischen Abelschen Halbgruppen. *Mathematische Zeitschrift,* 74, 81–90.

Laming, D.R.J. (1986). *Sensory analysis.* New York: Academic Press.

Levelt, W.J.M., Riemersma, J.B. & Bunt, A.A. (1972). Binaural additivity in loudness. *British Journal of Mathematical and Statistical Psychology,* 21, 51–68.

Levine, M.V. (1970). Transformations that render curves parallel. *Journal of Mathematical Psychology, 7,* 410-443.

Luce, R.D. (1959). On the possible psychophysical laws. *Psychological Review,* 66, 81–95.

Luce, R.D. (1962). Comments on Rozeboom's criticisms of ,On the possible psychophysical laws'. *Psychological Review,* 69, 548–551.

Luce, R. D. (1977). Thurstone's discriminal processes fifty years later. *Psychometrika*, 42, 461–489.

Luce, R. D. (1978). Dimensionally invariant numerical laws correspond to meaningful qualitative relations. *Philosophy of Science, 45,* 1–16.

Luce, R.D. (1990). „On the possible psychophysical laws" revisited: Remarks on cross-modal matching. *Journal of Mathematical Psychology, 97,* 66–77.

Luce, R. D. & Galanter, E. (1963a). Discrimination. In R. D. Luce. R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology, Vol. I* (s. 191–243). New York: Wiley.

Luce, R. D. & Galanter, E. (1963b). Psychophysical scaling. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology, Vol. I* (S. 245–307). New York: Wiley.

Luce, R.D. & Green, D.M. (1974). Detection, discrimination, and recognition. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception. Psychophysical judgment and measurement. Vol. II.* (S. 300–342). New York: Academic Press.

Luce, R. D., Krantz, D. H., Suppes, P. & Tversky, A. (1990). *Foundations of measurement, Vol. III. Representation, axiomatization, and invariance.* New York: Academic Press.

Luce, R. D. & Krumhansl, C. L. (1988). Measurement, scaling, and psychophysics. In R. C. Atkinson, R.J. Herrnstein, G. Lindzey & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology. Vol. I. Perception and motivation.* (S. 3–73). New York: Wiley.

Luce, R.D. & Narens, L. (1993). Fifteen problems concerning the representational theory of measurement. In P. Humphreys (Ed.), *Essays in honor of Patrick Suppes.* Amsterdam: Kluwer.

Luce, R. D. & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology. Vol. III.* (S.249–410). New York: Wiley.

Martin, G. (1956). *Klassische Ontologie der Zahl.* Köln: Kölner Universitätsverlag.

McClelland, G. (1977). A note on Arbuckle and Larimer ,The number of two-way tables satisfying certain additivity axioms'. *Journal of Mathematical Psychology,* 15, 292–295.

McClelland, G.H. & Coombs, C. H. (1975). Ordmet: A general algorithm for constructing all numerical solutions to ordered metric data. *Psychometrika, 50,* 269–290.

Merkel, J. (1888). Die Abhängigkeit zwischen Reiz und Empfindung. *Philosophische Studien,* 4, 541–594.

Mittelstraß, J. & Schroeder-Heister, P. (1986). Zeichen, Kalkül, Wahrscheinlichkeit. Elemente einer Mathesis universalis bei Leibniz. In H. Stachowiak (Hrsg.), *Handbuch pragmatischen Denkens I.* Hamburg: Meiner.

Narens, L. (1981). On the scales of measurement. *Journal of Mathematical Psychology,* 24, 249–275.

Narens, L. (1985). *Abstract measurement theory.* Cambridge: MIT Press.

Narens, L. (1988). Meaningfulness and the Erlanger program of Felix Klein. *Mathématiques Informatique et Sciences Humaines, 101*, 61–72.

Narens, L. (1993). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology* (to appear).

Narens, L. & Luce, R. D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin,* 99, 166–180.

Narens, L. & Mausfeld, R. (1992). On the relation of the psychological and the physical in psychophysics. *Psychological Review*, 99, 467–479.

von Neumann, J. & Morgenstern, O. (1947). *Theory of games and economic behavior*. 2nd Ed. Princeton: Princeton University Press.

Niederbe, R. (1987). On the reference to real numbers in fundamental measurement: A model-theoretic approach. In E. Roskam & R. Suck (Eds.) *Progress in mathematical psychology* (S. 3–24). Amsterdam: North-Holland.

Niederbe, R. (1992a). *Maß und Zahl. Logisch-rnodelltheoretische Untersuchungen zur Theorie fundamentaler Messung.* Frankfurt: Lang.

Niederbe, R. (1992b). What do numbers measure? A new approach to fundamental measurement. *Mathernatical Social Sciences, 24,* 237–276.

Pfanzagl, J. (1971). *Theory of measurement.* 2nd ed. Wien: Physica.

Plateau, M.J. (1872). Sur la Measure des Sensations physiques, et sur la Loi qui Lie l'Intensité des ces Sensations à l'Intensité de la Cause Excitante. *Bulletin de l'Academie Royale des Sciences, des Lettres et des Beaux Arts de Belgique,* 33, 376–388.

Ramsey, F.P. (1931). *The foundations of mathematics and other logical essays.* New York: Harcourt Brace.

Ramsey, J.O. (1976). Algebraic representations in physical and behavioral sciences. *Synthese,* 32, 419–453.

Restle, F. & Greeno, J. G. (1970). *Introduction to mathematical psychology*. Reading, Mass.: Addison-Wesley.

Roberts, F. S. (1979). *Measurement theory with applications to decision making, utility, and the social sciences.* Reading, Mass.: Addison-Wesley.

Roberts, F. S. & Rosenbaum, Z. (1986). Scale type, meaningfulness, and the possible psychophysical laws. *Mathematical Social Sciences, 12*, 77–95.

Robertson, T., Wright, F.T. & Dykstra, R. L. (1988). *Order restrictions in statistical inference.* Chichester: Wiley.

Roskam, E. (1983). Allgemeine Datentheorie. In H. Feger & J. Bredenkamp (Hrsg.), *Enzyklopädie der Psychologie. Messen und Testen.* (S. 1–135). Göttingcn: Hogrefe.

Rozeboom, W. W. (1962). The untenability of Luce's principle. *Psychological Review,* 69, 542–547.

Schneider, B., Parker, S. & Stein, D. (1974). The measurement of loudness using direct comparisons of sensory intervalls. *Journal of Mathematical Psychology, 11,* 259–273.

Stevens, S.S. (1957). On the psychophysical law. *Psychological Review,* 64, 153–181.

Stevens, S.S. (1974). Perceptual magnitude and its measurement. In E.C. Carterette & M.P. Friedman (Eds.), *Handbook of perception. Vol. II. Psychophysical judgment and measurement.* (S. 361–389). New York: Academic Press.

Suppes, P. (1962). Models of data. In E, Nagel, P. Suppes & A. Tarski (Eds.), *Logic, methodology and philosophy of science* (S. 252–261). Stanford: Stanford University Press.

Suppes, P., Krantz, D.H., Luce. R.D. & Tversky, A. (1989). *Foundations of measurement. Vol. II. Geometrical, threshold, and probabilistic representations.* New York: Academic Press.

Tack, W.H. (1983). Psychophysische Methoden. In H. Feger & J. Bredenkamp (Hrsg.), *Enzyklopädie der Psychologie. Messen und Testen.* (S. 346–426). Göttingen: Hogrefe.

Wandell, B. & Luce, R.D. (1978). Pooling peripheral information: Averages versus extrem values. *Journal of Mathematical Psychology, 17,* 220–235.

Wundt, W. (1908). *Grundzüge der physiologischen Psychologie. Bd. I.* (6. Aufl.). Leipzig: Engelmann.