Zuerst erschienen in: Probabilistische Modelle in der Sozialpsychologie / hrsg. von W. F. Kempf. Bern: Huber, 1974, pp. 13-55

ERSTES KAPITEL

Dynamische Modelle zur Messung sozialer Verhaltensdispositionen*

WILHELM F. KEMPF

1.1 Die Formalisierung psychologischer Konzepte und ihr Einfluss auf die psychologische Theorienbildung

Die Geschichte der Anwendung mathematischer Modelle in der Psychologie ist älter als zumeist vermutet wird. Bereits im Jahre 1837 hat *Herbart* eine mathematische Formulierung psychologischer Theorien gefordert und den Versuch unternommen, die Gesetze der *Newton*schen Mechanik auf die Psychologie zu übertragen. Obwohl dieser Versuch von vorneherein zum Scheitern verurteilt war, verdient er doch festgehalten zu werden, denn er ist typisch für die anfängliche Orientierung der psychologischen Methodologie an den Idealen der klassischen Physik.

Nur wenige Jahre später gelang es *Gustav Theodor Fechner*, mit dem *Weber*schen Gesetz zum ersten Mal eine mathematische Beziehung zwischen psychischen Phänomenen zu formulieren, die auch im Angesicht empirischer Daten lange Zeit bestehen bleiben konnte. Der wesentlichste Fortschritt, den das *Weber*sche Gesetz mit sich brachte, liegt darin, dass es auf experimentalpsychologischen Beobachtungen beruhte, während *Herbart* die Möglichkeit einer experimentellen Psychologie noch negiert hatte. Mit der Entwicklung der experimentellen Psychologie und unter dem raschen Aufschwung, den die mathematische Statistik in den ersten Jahrzehnten des 20. Jahrhunderts genommen hat, wurde die Anwendung mathematischer Modelle in der Psychologie bald zu einer Selbstverständlichkeit und erreichte mit der klassischen Testtheorie und der Faktorenanalyse ihre ersten Höhepunkte.

* Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft im Rahmen des Sonderforschungsbereichs 22 „Sozialisation und Kommunikation" am Sozialwissenschaftlichen Forschungszentrum der Universität Erlangen-Nürnberg und vom Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel unterstützt.

Diese erste Phase der Entwicklung war dadurch gekennzeichnet, dass die jeweiligen Modellannahmen in erster Linie aus Gründen der mathematischen Einfachheit gewählt wurden und sich nur sekundär an den Erfordernissen des psychologischen Forschungsgegenstandes orientierten. Zugleich waren diese Modelle auf eine universelle Anwendbarkeit hin ausgelegt und wurden infolgedessen häufig als blosse Methoden missverstanden. So hat z. B. erst KALVERAM (1970a) eine klare Trennung zwischen Modell und Methoden der Faktorenanalyse vorgenommen. Dennoch kann die Bedeutung dieser ersten Phase der Anwendung mathematischer Modelle in der Psychologie nicht hoch genug eingeschätzt werden, denn erst durch sie konnte sich die Psychologie endgültig als experimentell-empirische Wissenschaft etablieren.

Die zweite Phase der Entwicklung setzte um das Ende des zweiten Weltkrieges ein und entstand zunächst aus der Kritik an der klassischen Testtheorie. Sie ist mit Namen wie *Guttman, Lazarsfeld* und *Rasch* verknüpft. Seit dem Erscheinen von *Fischers* Psychologischer Testtheorie kommt ihr auch in der deutschsprachigen Psychologie immer mehr Bedeutung zu. Ihr wesentlichstes Merkmal besteht in der zunehmenden Betonung erkenntniskritischer und wissenschaftstheoretischer Überlegungen. Dies hängt auch mit dem hohen Entwicklungsstand zusammen, welchen die allgemeine Messtheorie inzwischen erreicht hat (vgl. SUPPES & ZINNES 1963; PFANZAGL 1968).

Von dem unbezweifelbaren Erfolg der psychologischen Testtheorie abgesehen, blieb die mathematische Psychologie jedoch bis vor wenigen Jahren die Domäne einiger weniger Spezialisten. Obwohl es heute fast keine psychologische Untersuchung mehr gibt, welche ohne die Anwendung mathematischer Modelle auskommt, beschränkt sich die Mehrheit der Untersuchungen immer noch auf sehr einfache Modelle, deren modellhafter Charakter zudem oft übersehen wird. Die mathematischen Modellannahmen, welche den verbreiteten statistischen Methoden zugrunde liegen, werden in der Psychologie nur allzu selten reflektiert und in Relation zu den inhaltlichen Theorien gesetzt, welche mit Hilfe dieser Methoden überprüft werden. So findet z. B. der Gedanke, dass die Signifikanz der Wechselwirkungen in einer mehrfachen Varianzanalyse als Falsifikation der Annahme der Additivität der Treatmenteffekte zu gelten hat, in der psychologischen Literatur nur relativ selten eine psychologische Interpretation. Wahrscheinlich ist dies darauf zurückzuführen, dass die Linearitätsannahme auch in der Varianzanalyse vornehmlich aus Gründen der mathematischen Einfachheit getroffen wird, während psychologische Theorien meist verbal formuliert und nur selten so präzise sind, um z. B. die Additivität von Treatmenteffekten begründen zu können. Der empirische Gehalt psychologischer Theorien geht oft über ein „Mehr oder Weniger" nicht hinaus.

Die daraus resultierende Diskrepanz zwischen den Theorien und Methoden der Psychologie zeigt sich z.B. an LAUTMANNS (1971) Analyse des Begriffs der sozialen Norm. Ungeachtet der Tatsache, dass sowohl die soziologische als auch die sozialpsychologische Forschung soziale Normen fast ausschliesslich durch Gruppenmittelwerte operationalisieren, unterscheidet Lautmann drei verschiedene Normenbegriffe:

– einen soziologischen Normenbegriff: die Norm als Richtschnur,
– einen sozialpsychologischen Normenbegriff: die Norm als Bezugsrahmen,
– einen deskriptiven Normenbegriff: die Norm als Durchschnitt.

Im vierten Kapitel dieses Buches werde ich ein probabilistisches Modell zur Messung sozialer Normen vorstellen, welches den Versuch unternimmt, eine eindeutige Beziehung zwischen Methoden zur Messung sozialer Normen und einem sozialpsychologischen Normenbegriff herzustellen. Wie sich zeigen wird, führt dieser Versuch auch zu einer Abklärung der bislang theoretisch wie empirisch weitgehend unbestimmten Relation zwischen den Konzepten „Norm" und „Konformität" (vgl. BRANDT & KÖHLER 1971). In ähnlicher Weise hat auch schon die Anwendung der Graphentheorie auf *Festinger*s „Theory of Cognitive Dissonance" und *Heider*s „Theory of Cognitive Organization" zu einer Präzisierung sozialpsychologischer Konzepte beigetragen (vgl. BOUDON 1973).

Damit sind wir bei der Frage angelangt, was die Formalisierung psychologischer Konzepte für die psychologische Theorienbildung zu leisten vermag und welche Anforderungen an die mathematischen Modelle der Psychologie gestellt werden müssen. Dabei erweist es sich als zweckmässig, zwischen theoretischen und empirischen Modellen zu unterscheiden, wobei letztere noch in explikative und prognostische Modelle unterteilt werden können. Deskriptive Modelle als eine eigene Modellklasse herauszustellen erscheint aus zweierlei Gründen nicht erforderlich. Erstens werden deskriptive Funktionen auch von den oben genannten Modellen ausgeübt und zweitens ist es fraglich, ob es rein deskriptive Modelle in der Psychologie überhaupt gibt. Die Messmodelle jedenfalls, die häufig als Beispiel deskriptiver Modelle angeführt werden, sind im allgemeinen den explikativen Modellen zuzuordnen. Wie ich an anderer Stelle (KEMPF 1974) gezeigt habe, transzendiert (fast) jede Messung die Erfahrung und kann daher nicht als blosse Deskription angesehen werden.

Ein Beispiel für ein theoretisches Modell ist die auf *von Neumann* und *Morgenstern* zurückgehende Spieltheorie, welche weder die Erklärung noch die Vorhersage des Spielverhaltens anstrebt, sondern vielmehr die Eigenschaften von Spielen untersucht, die durch bestimmte Axiome definiert sind. Die Bedeutung der Spieltheorie für die Sozialpsychologie besteht vor allem darin, dass sie eine klare Definition und Analyse von Konfliktsituationen erlaubt und dadurch das Verhalten in solchen Konfliktsituationen einer psychologischen Analyse erst zugänglich macht[1]. In ähnlicher Funktion wird z. B. auch die Informationstheorie in der Psychologie angewendet.

Prognostische Modelle verfolgen das Ziel, Verhalten vorauszusagen. Sie werden meist dann angewendet, wenn bestimmte Entscheidungen, wie z. B. die Wahl von Behandlungsmethoden in der Klinischen Psychologie, zu treffen sind, aber die theoretischen Erklärungsmöglichkeiten des entscheidungsrelevanten Verhaltens für eine Prognose nicht ausreichen. Als Beispiele können die multiple Regression und die mit ihr verwandte multivariate Analyse (COLEMAN 1964) genannt werden. Auch das Validitätskonzept der klassischen Testtheorie ist in diesem Sinne zu verstehen. Validität bedeutet nicht Erklärbarkeit, sondern Vorhersagbarkeit mithilfe linearer Regressionsmethoden. Wie entscheidungstheoretische Untersuchungen von CRONBACH & GLESER (1965) gezeigt haben, können prognostische Modelle allerdings nicht losgelöst von dem jeweiligen Entscheidungszusammenhang gesehen werden. Zumindest das klassische Validitätskonzept erweist sich dadurch als revisionsbedürftig, und es scheint notwendig zu sein, die Funktion prognostischer Modelle in der Psychologie neu zu überdenken.

Verhaltensprognosen können auch aus explikativen Modellen abgeleitet werden. Im Gegensatz zu den prognostischen Modellen, deren Zweckbestimmung auf die Vorhersage von Verhalten beschränkt ist, erheben explikative Modelle jedoch den Anspruch, Verhalten erklären können. Explikative Modelle sind formalisierte Theorien und müssen daher viel strengeren Kriterien genügen als die prognostischen Modelle. Die wichtigsten Aufgaben, welche sie innerhalb des psychologischen Forschungsprozesses zu erfüllen haben, sind:

- die Präzisierung psychologischer Konzepte,
- die Herstellung einer eindeutigen Zuordnung zwischen inhaltlich-psychologischen Theorien und den Methoden ihrer Überprüfung.

Im Idealfall wird es dabei möglich sein, bestimmten psychologischen Konzepten in eindeutiger Weise bestimmte Klassen von mathematischen Modellen zuzuordnen, über deren Vergleich dann Aufschlüsse darüber

[1] Anmerkung des Herausgebers: Vgl. auch das 6. Kapitel dieses Buches.

gewonnen werden können, in welcher Relation die zugrundeliegenden psychologischen Konzepte zueinander stehen. Zu so starken psychologischen Aussagen kann die mathematische Psychologie allerdings nur dann führen, wenn ihre Modelle eine Reihe von Voraussetzungen erfüllen, die über blosse mathematische Widerspruchsfreiheit weit hinausgehen.

Eine der wichtigsten Forderungen betrifft die eindeutige Interpretierbarkeit der Modellstruktur und der Modellparameter. Als formalisierte Theorie kann ein Modell nur dann gelten, wenn es keinerlei Annahmen enthält, welche nicht auch ein inhaltlich-psychologisches Äquivalent besitzen. Nur dann ist es möglich, eine eindeutige Zuordnung zwischen Theorie und Modell zu treffen und aus der Überprüfung des Modells Rückschlüsse über die Geltung der Theorie zu ziehen. Dennoch wird diese Forderung oft nur unzureichend erfüllt, z.B. wenn man in ein mathematisches Modell zusätzliche Modellparameter einführt, welche als einzige Aufgabe die Funktion erfüllen, die Anpassung des Modells an die empirischen Daten zu verbessern. Andere Modelle sind so komplex, dass die Interpretation der Ergebnisse ihrer Anwendung die Leistungsfähigkeit des Psychologen übersteigt. So scheitert z.B. eine methodisch befriedigende Interpretation faktorenanalytischer Ergebnisse meist daran, dass sie alle zulässigen Skalentransformationen, d.h. alle Rotationen der Ladungsmatrix mit einschliessen müsste (vgl. KEMPF 1972b).

Die zweite Forderung, welche hier erhoben werden soll, betrifft die mathematische Handhabbarkeit des Modells. Darunter ist dreierlei zu verstehen:

- die einwandfreie Bestimmbarkeit der Modellparameter,
- die einwandfreie Vergleichbarkeit der Modellparameter (sofern die zu formalisierende Theorie Aussagen über Relationen zwischen Modellparametern trifft),
- die einwandfreie Prüfbarkeit der Modellstruktur.

Auch diese Kriterien können wieder anhand der Faktorenanalyse erläutert werden. Wie KALVERAM (1970a) zeigen konnte, erlaubt das allgemeine Faktorenmodell mit $p \leqslant n$ gemeinsamen und q spezifischen Faktoren die praktische Berechnung einer Faktorenanalyse nur unter der (empirisch nicht überprüfbaren) Voraussetzung, dass die spezifischen Faktoren untereinander und mit den gemeinsamen Faktoren unkorreliert sind, wobei auch dann nicht einmal die Auffindbarkeit einer vorgegebenen Faktorenstruktur gesichert ist, da die Kommunalitäten durch die Kovarianzmatrix nicht eindeutig bestimmt sind. Dennoch führt die Berechnung einer Faktorenanalyse stets zu irgendeinem Ergebnis, weshalb die Faktorenanalyse von vielen Autoren als tautologisches Modell angesehen wird.

„Man kann in der Tat zahlreiche Beispiele nicht falsifizierbarer Modelle angeben, zum Beispiel die Multifaktorenanalyse in der Psychologie. Was auch immer die verfügbaren Daten sein mögen, eine multifaktorielle Analyse ergibt immer eine Lösung, die man entweder verwirft oder akzeptiert aufgrund von Kriterien, die nichts mit der Kongruenz mit der Beobachtung zu tun haben, denn diese ist ja per definitionem immer vorhanden" (BOUDON 1973, S. 50).

Wie sich zeigen lässt, beruht diese Ansicht auf einer Verwechslung zwischen faktorenanalytischen Modellen und faktorenanalytischen Methoden. So ist z. B. die oft kritisierte Stichprobenabhängigkeit faktorenanalytischer Ergebnisse (FISCHER 1968) nicht im Modell begründet, sondern ergibt sich erst dann, wenn die Modellvoraussetzungen nicht erfüllt sind. Zu den Modellvoraussetzungen sind dabei auch jene Annahmen zu rechnen, durch welche das Kommunalitätenproblem überbrückt wird (vgl. FISCHER & ROPPERT 1965; KALVERAM 1970b; KEMPF 1972b). Die Stichprobenunabhängigkeit faktorenanalytischer Ergebnisse kann somit als Prüfkriterium für die Geltung der Modellannahmen herangezogen werden, wobei allerdings zu berücksichtigen ist, dass das Kommunalitätenproblem im allgemeinen Faktorenmodell immer noch ungelöst ist, während eingeschränkte Faktorenmodelle, welche das Problem überwinden, von äusserst strengen Voraussetzungen ausgehen. In der Praxis ist kaum damit zu rechnen, dass sie einem Falsifikationsversuch standhalten. Diese eingeschränkten Faktorenmodelle, die ich an anderer Stelle (KEMPF 1972b) einer ausführlichen Diskussion unterzogen habe, erweisen sich daher als weitgehend „unrealistisch", womit wir einen Terminus aufgreifen, der auf FISCHER (1968) zurückgeht.

Als „unrealistisch" wird ein Modell bezeichnet, das von nahezu beliebigen Daten widerlegt werden kann. „Realistisch" ist ein Modell dann, wenn sich Datenbereiche finden lassen, in denen es nicht verworfen werden muss. Der Fortschritt, den die Formalisierung psychologischer Konzepte für die psychologische Theorienbildung bringen kann, wird davon abhängen, ob es gelingt, realistische Modelle zu konstruieren. Das Dilemma, in dem sich die (mathematische) Psychologie befindet, besteht darin, Modelle zu finden, welche überprüfbar sind, aber zugleich auch eine reelle Chance haben, Falsifikationsversuchen standzuhalten. Dies hat dazu geführt, dass probabilistische Modelle in der Psychologie gegenüber deterministischen Modellen eindeutig bevorzugt werden, was aber nicht den Blick davon ablenken sollte, dass es in der Psychologie auch deterministische Theorien gibt, die sich – zumindest in sehr engen Datenbereichen – bewähren konnten wie z. B. SCANDURAS (1973) „Theory of Structural Learning". Insgesamt ist in der Psychologie derzeit ein Trend zu eher restriktiven Modellen zu verzeichnen, wozu nicht nur deterministische Ansätze zu zählen sind, sondern auch neuere Entwicklungen in der probabilistischen Testtheorie (vgl. FISCHER 1968), die sich inzwischen zu einer allgemeinen probabilistischen Theorie experimentalpsychologischer Beobachtungen weiterentwickelt hat (vgl. MICKO 1970; SCHEIBLECHNER 1971a; FISCHER 1972; KEMPF 1972a). Im Gegensatz zu den oben erwähnten eingeschränkten Faktorenmodellen ergibt sich die Restriktivität dieser Modelle aber nicht nur aus mathematischen Notwendigkeiten innerhalb eines willkürlich gewählten Modellansatzes, sondern vor allem auch aus psychologischen und wissenschaftstheoretischen Überlegungen. Je besser aber eine restriktive Modellannahme begründet werden kann, desto geringer ist die Gefahr, dass sie sich als unrealistisch erweist: desto grösser ist aber auch der Erkenntnisgewinn, der aus einer eventuellen Falsifikation des Modells folgt. In Modellen, die auf willkürlichen Modellannahmen beruhen, ist die empirische Prüfung der Modellstruktur ein notwendiges Übel, das über die Anwendbarkeit des Modells entscheidet. In Modellen dagegen, deren Modellannahmen psychologische Konzepte repräsentieren, wird die Prüfung der Modellstruktur zu einem zentralen Instrument psychologischer Theorienprüfung.

1.2 Differentialpsychologische Konzepte in der Sozialpsychologie, diskutiert am Beispiel der „Aggressivität"

Die Annahme individueller „Verhaltensdispositionen", wie sie der differentiellen Psychologie und der Persönlichkeitspsychologie zugrunde liegt, ist einer der ältesten und am meisten verbreiteten Versuche zur Erklärung menschlichen Verhaltens. In der Sozialpsychologie finden sich solche Erklärungsansätze unter anderem in Konzepten wie „Einstellung" oder „Aggressivität" und in dem (letztlich gescheiterten) Versuch einer persönlichkeitspsychologischen Lösung des Führerschaftsproblems.

Trotz ihrer Gebräuchlichkeit sind die psychologischen Dispositionskonzepte jedoch immer noch umstritten, und manche Kritiker (vgl. ROHRACHER 1963; KRISTOF 1968) sehen in ihnen eine „Auferstehung der Vermögenspsychologie"; TRAXEL (1964) spricht von „begriffsrealistischen Fiktionen" und HOLZKAMP (1964, 1965) weist auf die Gefahr einer „unzulässigen Realitätsverdoppelung" hin, welche nur dann vermieden werden könne, wenn man die psychologischen Eigenschaftsbegriffe konsequent als theoretische Konstruktionen auffasse, die weder etwas unmittelbar Vorfindbares bezeichnen, noch als Hinweis auf eine verborgene

Realität verstanden werden dürfen. Letztlich weisen alle diese Kritiken in dieselbe Richtung: Die Willkürlichkeit der Annahme von Verhaltensdispositionen kann durch blosse „operationale Definitionen", wie sie in der differentiellen Psychologie gebräuchlich sind, nicht behoben werden. Wie JÜTTEMANN (1972) zu Recht bemerkt hat, bewirkt aber auch die Rede von „theoretischen Konstrukten" bloss eine Verschleierung der Tatsache, dass die Eigenschaftsbegriffe in der psychologischen Wissenschaftssprache nur unzureichend bestimmt sind. Sie kann daher nichts zur Klärung der Frage beitragen, unter welchen Voraussetzungen die Annahme von Verhaltensdispositionen wissenschaftstheoretisch gerechtfertigt werden kann.

Andere Autoren wie z.B. MERZ (1960) versuchen ohne den Begriff der „theoretischen Konstrukte" auszukommen und vertreten ausdrücklich die Auffassung, die Rede von der „Aggressivität" z.B. bedeute, dass man Personen nach dem Grad ordnen kann, in welchem ihnen diese Eigenschaft zukommt. KEMPF (1974) hat diesen Ansatz weiter verfolgt und gefordert, dass Verhaltensdispositionen konsequent als metrische Begriffe zu betrachten seien, deren Einführung das Vorliegen einer Skala bereits voraussetze. Die Annahme von Verhaltensdispositionen könne nur dann gerechtfertigt werden, wenn diese auch messbar sind. Dabei muss aber vorausgesetzt werden, dass Messung nicht bloss als „die Kunst, Phänomenen Zahlen zuzuordnen" verstanden wird, sondern – im Sinne der allgemeinen Messtheorie – als eine Abbildung von Messobjekten und zwischen ihnen empirisch feststellbaren Relationen auf (reelle) Zahlen und Relationen zwischen diesen. Als Mindestvoraussetzung für die Annahme von Verhaltensdispositionen ergibt sich dann die Auffindung einer empirischen Ordnungsrelation. Im Falle der „Aggressivität" z.B. wäre diese durch Einführung eines zweistelligen Prädikators „aggressiver als" zu definieren, etwa indem wir vereinbaren, dass von zwei Personen die eine genau dann als aggressiver gelten soll, wenn sie stärker zu aggressiven Handlungen tendiert als die andere. Was mit einer solchen Definition gemeint ist, wird später noch zu präzisieren sein. In jedem Falle muss dabei aber auch die empirisch wohlbegründete Auffassung mit berücksichtigt werden, dass die Situationen, in welchen aggressives Verhalten stattfinden kann, oft unterschiedlichen Aggressionsanreiz aufweisen. Ein Vergleich von Personen bezüglich ihrer Aggressivität ist daher nur dann möglich, wenn er von den situativen Faktoren nicht verzerrt wird. Entsprechend der methodologischen Tradition der Psychologie kann dies erreicht werden, wenn der Vergleich zweier Personen jeweils auf eine bestimmte Situation (S) bezogen und derart der Prädikator „aggressiver als" nur relativ zu dieser Situation eingeführt wird. Durch eine solche bedingte Definition allein wird der Einfluss situativer Faktoren jedoch noch nicht ausgeschaltet, sondern vielmehr für jede Situation eine spezifische Aggressivitätsform postuliert. Die Ausschaltung situativer Faktoren gelingt erst dann, wenn man zusätzlich voraussetzt, dass die Aggressivität innerhalb eines definierten Universums von Situationen (Σ) invariant bleibt, so dass die Aussagen, welche über die Relation zweier Personen getroffen werden, nicht davon abhängen, welche Situation S aus Σ ausgewählt und dem Vergleich der Personen zugrunde gelegt wird. Gibt es eine Situation aus Σ, in der von zwei Personen P_1 und P_2 die eine aggressiver ist als die andere, so muss diese Beziehung auch für alle anderen Situationen aus Σ gelten:

(1.2.1) $(\exists S \epsilon \Sigma\colon P_1 \text{ aggressiver als } P_2) \rightarrow$
$\rightarrow (\forall S \epsilon \Sigma\colon P_1 \text{ aggressiver als } P_2)$.

Mit der in (1.2.1.) formulierten Annahme der „Σ-Invarianz" der Aggressivität kommen zugleich auch neue Ordnungsgesichtspunkte ins Spiel: Aussagen über die Aggressivität von Personen bedeuten nun nicht mehr bloss eine tautologische Umformulierung einer Beziehung, welche in einer spezifischen Situation beobachtet wurde, sondern sie haben einen höheren Allgemeinheitsgrad und können auf andere Situationen generalisiert werden, wobei das Universum Σ die Klasse der Situationen festlegt, auf welche generalisiert werden darf. Für sich allein genommen ist die Annahme der Σ-Invarianz allerdings unzulässig, da sie keinerlei falsifizierbare Konsequenzen aufweist. Denn eine Person kann sich nicht im selben Zeitpunkt in einer Situation S befinden und zugleich in einer anderen Situation S'. Die Annahme der Σ-Invarianz machte es somit erforderlich, zugleich auch eine relative, zeitliche Invarianz der Aggressivität zu postulieren, wonach die Relationen, in welchen je zwei Personen P_1 und P_2 bezüglich ihrer Aggressivität zueinander stehen, innerhalb definierte Zeitabschnitte τ_1 und τ_2 erhalten bleiben, d.h. für beliebige Zeitpunkte t_1 und t_2 aus diesen Zeitabschnitten gelten:

(1.2.2) $(\exists t_1 \epsilon \tau_1, t_2 \epsilon \tau_2, S \epsilon \Sigma\colon P_1 \text{ aggressiver als } P_2) \rightarrow$
$\rightarrow (\forall t_1 \epsilon \tau_1, t_2 \epsilon \tau_2, S \epsilon \Sigma\colon P_1 \text{ aggressiver als } P_2)$.

Neben die Generalisierbarkeit über Situationen aus Σ tritt dann auch noch die Generalisierbarkeit über Zeitpunkte innerhalb definierter Zeitabschnitte. Eine Voraussetzung, welche unserem psychologischen Vorverständnis über „Aggressivität" durchaus entspricht (vgl. z.B. SELG 1968). Wir fassen also zusammen, dass eine methodisch befriedigende

Einführung von Verhaltensdispositionen wie „Aggressivität" nur dann möglich ist, wenn diese als relativ überdauernde Eigenschaften von Personen aufgefasst werden, welche nicht nur in einzelnen Situationen, sondern in genau angebbaren Klassen von Situationen sich manifestieren. Erst wenn diese Voraussetzungen erfüllt sind, können Aussagen über Verhaltensdispositionen generalisiert werden und erst dann kommt ihnen ein psychologischer Erklärungswert zu[2]. Dafür muss aber erst noch geklärt werden, was unter der Aussage zu verstehen sein soll, dass eine Person P_1 (im Zeitpunkt t_1) in einer gegebenen Situation S stärker zu aggressiven Handlungen tendiert als eine andere Person P_2 (im Zeitpunkt t_2). Soweit mir bekannt ist, wurden diesbezüglich in der Psychologie überhaupt nur zwei Möglichkeiten ernsthaft diskutiert. Die erste besteht darin, den Prädikator „aggressiver als" durch eine feste Beziehung zwischen den manifesten Verhaltensweisen zu definieren, welche die miteinander zu vergleichenden Personen in der Situation S zeigen. KEMPF (1974) hat diese Möglichkeit einer eingehenden Analyse unterzogen, aus der hervorgeht, dass in diesem Falle die Annahmen der Σ-Invarianz und der relativen zeitlichen Invarianz der Aggressivität schon vorausgesetzt werden müssen, damit überhaupt eine Rangordnung festgelegt werden kann. Denn geht man davon aus, dass von zwei Personen in der Situation S jene als „aggressiver" gelten soll, welche faktisch aggressiv handelt, während sich die andere nicht-aggressiv verhält

$$(1.2.3)\qquad \left\{\begin{array}{l}P_1 \text{ handelt in S aggressiv}\\ \text{und } P_2 \text{ handelt in S}\\ \text{nicht-aggressiv}\end{array}\right\} \Rightarrow \left\{\begin{array}{l}P_1 \text{ ist in S aggressiver}\\ \text{als } P_2\end{array}\right\}$$

so liefert eine einzelne Situation S allenfalls eine binäre Klasseneinteilung in „aggressive" Personen, welche die Situation mit einer Aggression beantworten, und „nicht-aggressive" Personen, deren Antwort auf die Situation keine Aggression ist. Gemeinsam mit den in (1.2.2) zusammengefassten Voraussetzungen der Σ-Invarianz und der relativen, zeitlichen Invarianz der Aggressivität dagegen, definiert die Beziehung (1.2.3) eine GUTTMAN-Skala (1950), durch welche die Personen P_v bezüglich ihrer Aggressivität ξ_v in eine eindeutige Rangreihe gebracht werden und zu-

[2] Auf den ersten Blick müssen die getroffenen Voraussetzungen als zu streng erscheinen, da Aussagen über die Aggressivität einer Person z. B. offensichtlich auch dann generalisiert werden können, wenn die Aggressivität einem gesetzmässigen Wandel unterliegt. Auch in diesem Fall können wir aber von einer invarianten „Ausgangsaggressivität" sprechen, welche von einem gesetzmässigen Veränderungsprozess überlagert wird.

gleich die Situationen S_i aus Σ bezüglich ihres „negativen Aggressionsanreizes" σ_i, d.h. bezüglich der „Aggressionshemmung", welche von der Situation ausgeht:

$$(1.2.4) \quad P_v \text{ handelt in } S_i \begin{cases} \text{aggressiv} & \text{wenn } \xi_v \geqslant \sigma_i \\ \text{nicht-aggressiv} & \text{wenn } \xi_v < \sigma_i \end{cases}$$

Je grösser die von einer Situation ausgehende „Aggressionshemmung" sein darf, ohne bei einer Person eine aggressive Reaktion zu verhindern, desto stärker tendiert die Person zu aggressiven Handlungen. Je grösser die Aggressivität einer Person sein muss, um die aggressive Beantwortung einer Situation zu bedingen, desto stärker ist die „Aggressionshemmung", welche von der Situation ausgeht. Je zwei Personen P_1 und P_2 mit $\xi_1 \geqslant \sigma_i > \xi_2$ werden durch die Situation S_i eindeutig in eine Rangreihe gebracht. Jede andere Situation S_j mit $\xi_1 \geqslant \sigma_j > \xi_2$ führt zu demselben Ergebnis. Werden zwei Personen P_1 und P_2 innerhalb der Zeitabschnitte τ_1 und τ_2 in derselben Situationen $S_1, S_2, ..., S_i, ..., S_k$ aus Σ beobachtet, so gilt jene Person als aggressiver, die häufiger aggressiv reagiert. Zeigen zwei Personen gleich viele aggressive Reaktionen, so wird keine Aussage darüber getroffen, welche der beiden Personen aggressiver ist.

Auf den ersten Blick scheint diese Beziehung mit einer herkömmlichen operationalen Definition der Aggressivität übereinzustimmen. Tatsächlich ist dies jedoch nicht der Fall. Denn operationale Definitionen legen den Prädikator „aggressiver als" durch einen bestimmten Test, d.h. durch eine bestimmte Auswahl von Situationen fest. In Guttman's Quasiskala dagegen sind die Aussagen, welche über die Relation zweier Personen getroffen werden, davon essentiell unabhängig, welche Situationen aus Σ dem Vergleich zugrunde gelegt wurden. Alle nur denkbaren Teilmengen von Situationen aus Σ führen zu demselben Ergebnis. Allenfalls kann die Genauigkeit der Aussagen variieren, z.B. wenn ein „Test" keine Situation S_i mit $\xi_1 \geqslant \sigma_i > \xi_2$ enthält und daher keine Entscheidung darüber erlaubt, welche der beiden Personen P_1 und P_2 aggressiver ist. Es gibt aber keine Teilmenge von Situationen, welche zu einer Umkehr von Rangpositionen führen könnte. Ist die Rangordnung der Situation bezüglich ihres „negativen Aggressionsanreizes" bereits bekannt, so erlaubt die Guttman-Skala sogar den Vergleich von Personen, welche in verschiedenen Situationen aus Σ beobachtet wurden. Denn aufgrund der Beziehungen (1.2.2–1.2.4) wird durch die aggressive Beantwortung einer bestimmten Situation immer auch die aggressive Beantwortung aller jener Situationen bedingt, von welchem eine weniger starke „Aggressionshemmung" ausgeht.

Damit werden aber zugleich auch ausserordentlich strenge Anforde-

rungen an die empirischen Daten gestellt, und die Guttman-Skala gilt daher allgemein als ein unrealistisches Modell. Meines Erachtens trifft diese Bewertung weniger deshalb zu, weil die Guttman-Skala von ihrer Struktur her deterministisch ist, als vor allem aufgrund der Tatsache, dass sie deterministische Voraussetzungen an Teilbereiche der Psychologie heranträgt, die weit davon entfernt sind, irgendwelche Vorhersagen von manifestem Verhalten auch nur annähernd „mit Sicherheit" treffen zu können. Ob es z. B. in der psychologischen Aggressionsforschung je gelingen wird, deterministische Aussagen über die Bedingungen aggressiven Verhaltens zu treffen, braucht hier nicht zum Gegenstand von Spekulationen gemacht zu werden. Es steht jedenfalls fest, dass solche deterministischen Aussagen einen ausserordentlich hohen Erkenntnisstand voraussetzen würden, an dem gemessen das Konzept der „Aggressivität" als vergleichsweise grob und aussageschwach erachtet werden muss.
Dies gilt insbesondere dann, wenn der Aggressivität bloss eine Guttman-Skala zugrunde gelegt werden kann, die

- nur eine Rangordnung der Personen festlegt und
- deren Aussagen nur innerhalb sehr enger und spezifischer Klassen von Situationen generalisiert werden können.

Es ist daher naheliegend, die bisherige Vereinbarung des Prädikators „aggressiver als" durch eine probabilistische Definition zu ersetzen und zu vereinbaren, dass eine Person P_1 (im Zeitpunkt t_1) in einer gegebenen Situation S genau dann stärker zu aggressiven Handlungen tendiert als eine andere Person P_2 (im Zeitpunkt t_2), wenn sie die Situation mit grösserer Wahrscheinlichkeit aggressiv beantwortet:

$$(1.2.5) \qquad \left\{\begin{array}{l}P_1 \text{ handelt in S mit grösserer} \\ \text{Wahrscheinlichkeit aggressiv} \\ \text{als } P_2\end{array}\right\} \Leftrightarrow \left\{\begin{array}{l}P_1 \text{ ist in S aggressiver} \\ \text{als } P_2\end{array}\right\}$$

Aus dieser Definition, die in ähnlicher Weise z. B. von KAUFMANN (1965) vertreten wird, folgt dann für jede beliebige Person P_v und jede Situation S_i aus Σ eine streng monoton wachsende Beziehung zwischen der Aggressivität der Person und der Wahrscheinlichkeit p_{vi}, mit welcher sie die Situation aggressiv beantwortet. Die Voraussetzungen der Σ-Invarianz und der relativen zeitlichen Invarianz der Aggressivität nehmen die Form einer probabilistischen Gesetzmässigkeit an:

$$(1.2.6) \qquad (\exists t_1 \epsilon \tau_1, t_2 \epsilon \tau_2, S_i \epsilon \Sigma: p_{1i} > p_{2i}) \rightarrow$$
$$\rightarrow (\forall t_1 \epsilon \tau_1, t_2 \epsilon \tau_2, S_i \epsilon \Sigma: p_{1i} > p_{2i}).$$

Stochastische Gesetzesaussagen können jedoch im strengen Sinne weder verifiziert noch falsifiziert, sondern nur aufgrund eines methodologischen Beschlusses entschieden werden. Da aus probabilistischen Modellen immer nur probabilistische Aussagen deduzierbar sind, kann die statistische Falsifikation eines stochastischen Modells nicht mit Sicherheit, sondern nur unter Absehung von einer gewissen – allerdings genau angebbaren – Fehlerwahrscheinlichkeit erfolgen. Die Vorgangsweise ist bekannt: zunächst wird aus dem zu falsifizierenden Modell (das unter anderem die sogenannte Null-Hypothese enthält) die Verteilung einer Prüfgrösse deduziert. Sodann wird die Abweichung des empirisch realisierten Wertes der Prüfgrösse von dem aufgrund des Modells erwarteten „idealen" Wert festgestellt. Schliesslich wird das Modell verworfen, wenn die Wahrscheinlichkeit einer so starken oder stärkeren Abweichung bei Geltung des Modells kleiner ist als eine vereinbarte Zahl.

Unter der Definition (1.2.5) kann auch die Relation zweier Personen bezüglich ihrer Aggressivität nur mithilfe statistischer Verfahren geschätzt werden, wofür eine Stichprobe von Beobachtungen vorliegen muss. Die ursprünglich aufgrund anderer Überlegungen eingeführten Annahmen der Σ-Invarianz und der relativen zeitlichen Invarianz der Aggressivität erweisen sich daher als notwendige Voraussetzungen dafür, dass die Rangordnung der Personen bezüglich ihrer Aggressivität überhaupt geschätzt werden kann. Um eine leistungsfähigere Skala der „Aggressivität" zu erhalten, müssen über (1.2.5) und (1.2.6) hinaus auch noch mehrstellige Relationen zwischen den Personen eingeführt werden, z.B. indem man bestimmte Annahmen über die Form des funktionalen Zusammenhanges zwischen der Aggressivität und den Wahrscheinlichkeiten p_{vi} trifft.

Wie man heute zeigen kann, müssen diese Annahmen allerdings bestimmten wissenschaftstheoretischen und mathematisch-statistischen Kriterien genügen und können daher nicht willkürlich getroffen werden. So ist es aufgrund des Postulates (1.2.6) theoretisch belanglos, auf welche Situationen aus Σ der Vergleich zweier Personen gestützt wird, und es muss daher gefordert werden, dass auch die Verteilung der Schätzfunktionen, aus denen die Relation zweier Personen bestimmt wird, von der Stichprobe der Situationen unabhängig ist. Denn ist diese Voraussetzung nicht erfüllt, so stellt die Stichprobenauswahl eine „störende Bedingung" im Sinne von HOLZKAMP (1968) dar, und es ist möglich, jede beliebige Aussage über die Relation zweier Personen gegen die empirischen Befunde zu exhaurieren. Wie der dänische Statistiker GEORG RASCH (1965) gezeigt hat, führt bereits diese Minimalforderung, die er spezifische Objektivität[3] nennt, notwendigerweise zu einer sehr speziel-

[3] Rasch hat dem Prinzip der spezifischen Objektivität ursprünglich eine noch viel

len Klasse von probabilistischen Messmodellen, d.h. zu ganz bestimmten Voraussetzungen über die Form des Zusammenhangs zwischen der Aggressivität und den Wahrscheinlichkeiten p_{vi}. Ist aber die spezifische Objektivität nicht erfüllt, so ist nicht geklärt, welche Beziehung zwischen einer gegebenen Aussage über die Relation zweier Personen und gegebenen Beobachtungsdaten besteht. Die Holtzkamp'sche Auffassung, wonach Verhaltensdispositionen blosse gedankliche Konstruktionen ohne jeglichen Realitätsbezug sind, ist dann durchaus gerechtfertigt. Es muss daher gefordert werden, dass die spezifische Objektivität nicht nur als Voraussetzung dafür anzusehen ist, dass die jeweiligen Aussagen über Relationen zwischen Personen als wissenschaftlich gelten können, sondern auch dafür, dass Verhaltensdispositionen überhaupt Gegenstand wissenschaftlicher Aussagen sein können. Begriffe wie „Aggressivität" sind in der psychologischen Wissenschaftssprache nur dann zulässig, wenn ihre Messung dem Prinzip der spezifischen Objektivität genügt.

1.3 Abriss der Theorie der spezifisch objektiven Messung nach Rasch

Die bisherigen Überlegungen haben uns dazu geführt, der Einführung des Begriffs der „Aggressivität" eine streng monoton wachsende Beziehung zwischen der Wahrscheinlichkeit aggressiven Verhaltens und der Aggressivität der Personen zugrunde zu legen. Dabei haben wir uns stets vor Augen gehalten, dass analoge Überlegungen auch für andere Verhaltensdispositionen wie z.B. die Intelligenz oder die Einstellungen von Personen getroffen werden können. Im folgenden werden wir das Beispiel der „Aggressivität" verlassen und uns einer Terminologie bedienen, wie sie in der psychologischen Testtheorie üblich ist. Anstelle von „Situationen" sprechen wir nun von „Items", anstelle von „aggressiven" und „nicht-aggressiven" Reaktionen von „positiver" und „negativer" Itembeantwortung. Den Begriff der „Aggressivität" ersetzen wir durch den allgemeineren Begriff einer „latenten Eigenschaft" (ξ), welche durch ein potentiell unendliches Universum von Items ($i = 1,2, \ldots$) und streng monoton wachsenden Funktionen $f_i(\xi)$ definiert ist, so dass $p(+ \mid v, i) = f_i(\xi_v)$ und $p(- \mid v, i) = 1 - f_i(\xi_v)$. Schreiben wir $a_{vi} = 1$ im Falle einer positiven Itembeantwortung und $a_{vi} = 0$ im Falle einer negativen Itembeantwortung, so können diese Beziehungen in der Form

grundsätzlichere, inhaltliche Interpretation gegeben. Die hier vertretene Definition geht auf KEMPF (1974) zurück und stimmt mit der statistischen Argumentation bei RASCH (1965) überein.

$$(1.3.1) \quad p\{a_{vi}\} = f_i(\xi_v)^{a_{vi}} \cdot (1 - f_i(\xi_v))^{1-a_{vi}}$$

zusammengefasst werden. Die Funktionen $f_i(\xi)$ werden in der psychologischen Testtheorie als „Itemcharakteristiken" bezeichnet. Über ihre Form wollen wir vorläufig noch keinerlei Annahmen treffen, sondern lediglich voraussetzen, dass die Items lokal stochastisch unabhängig voneinander sind, d.h. dass die Wahrscheinlichkeit einer positiven Beantwortung des Items i durch eine Person v in keiner Weise davon abhängt, welche anderen Items die Person ausserdem noch positiv beantwortet und welche nicht. Formal gesehen bedeutet diese Annahme (die in der gesamten probabilistischen Testtheorie getroffen wird), dass die in einer Stichprobe von Personen beobachtbaren Iteminterkorrelationen ausschliesslich durch die Verknüpfung der Items mit der latenten Dimension ξ zustandekommen und daher für konstantes ξ verschwinden. Psychologisch gesehen bedeutet die lokale stochastische Unabhängigkeit, dass die Beantwortung eines Items in keiner Weise durch die Reaktionen des Probanden auf andere Items beeinflusst werden soll. Übungs-, Ermüdungs- oder Positionseffekte scheinen dem Prinzip der lokalen stochastischen Unabhängigkeit daher zunächst zu widersprechen. Setzt man aber voraus, dass es sich dabei um generelle Effekte handelt, die bei allen Vpn gleich stark ausgeprägt sind, so verschwindet dieser Widerspruch und kann durch die Annahme ersetzt werden, dass die Form der Itemcharakteristiken nicht nur von den relevanten Eigenschaften der Items per se, sondern auch von deren Positionen innerhalb des Tests bestimmt wird (KEMPF 1970). Trotzdem bewirkt die Annahme der lokalen stochastischen Unabhängigkeit eine nicht unerhebliche Einschränkung der Theorienbildung, und es gibt in der Psychologie eine ganze Reihe von theoretischen Ansätzen wie z. B. die Konzepte der „Katharsis" und des „erfolgskontingenten Lernens", denen die Annahme der lokalen stochastischen Unabhängigkeit widerspricht. In den beiden letzten Abschnitten dieses Kapitels werde ich zwei probabilistische Messmodelle vorstellen, welche ohne die Annahme der lokalen stochastischen Unabhängigkeit auskommen und somit auch die Formalisierung dynamischer Theorien erlauben.

Werden über die Itemcharakteristiken keine Annahmen getroffen, welche über die Monotonieannahme hinausgehen, so kann ein Vergleich von Personen bezüglich der latenten Dimension ξ nur so vorgenommen werden, dass man die Personen aufgrund der Häufigkeit in eine Rangreihe bringt, mit welcher sie die Items eines gegebenen Tests positiv beantworten. Sei k die Anzahl der im Test enthaltenen Items, so definieren wir

$$(1.3.2) \qquad a_{vo} = \sum_{i=1}^{k} a_{vi}$$

und erhalten für den Erwartungswert des Testscores a_{vo} die Beziehung

$$(1.3.3) \qquad E(a_{vo}) = \sum_{i=1}^{k} E(a_{vi}) = \sum_{i=1}^{k} f_i(\xi_v),$$

welche ebenfalls eine streng monoton wachsende Funktion der latenten Dimension ξ ist, da voraussetzungsgemäss jede einzelne der Funktionen $f_i(\xi)$ streng monoton wächst. Wir sehen also, dass die im vorangegangenen Abschnitt eingeführten Voraussetzungen für die Annahme von Verhaltensdispositionen in jedem Falle einen erwartungstreuen Vergleich von Personen erlauben. Dennoch kann ein Vergleich von Personen aufgrund der Anzahl der positiv beantworteten Items in einem Test mitunter nicht sehr brauchbar sein, z.B. dann, wenn der Testscore a_{vo} nur einen Bruchteil der Information enthält, welche in den Antworten a_{vi} der Personen enthalten ist. Sollen Personen aufgrund ihres Summenscores a_{vo} miteinander verglichen werden, so wird man daher fordern, dass dieser Score die gesamte Testinformation bezüglich des unbekannten Personenparameters ξ_v enthalten soll, d.h. dass er eine erschöpfende Statistik für den Personenparameter ist.

Mathematisch formuliert bedeutet diese Forderung, dass der Antwortvektor $(a_{vi}) = (a_{v1}, \ldots, a_{vk})$ einer Person keinerlei Information über den Personenparameter ξ_v enthalten darf, welche durch die Statistik a_{vo} nicht voll ausgeschöpft wird, so dass die bedingte Wahrscheinlichkeit

$$(1.3.4) \qquad p\{(a_{vi}) \mid a_{vo}\} = \frac{p\{(a_{vi})\}}{p\{a_{vo}\}}$$

nicht mehr von ξ_v abhängig ist. BIRNBAUM (1968) hat nachgewiesen, dass die Beziehung (1.3.4) bei lokaler stochastischer Unabhängigkeit der Items mit Notwendigkeit zu RASCHS (1960) speziellem logistischen Testmodell führt, dessen Itemcharakteristiken (1.3.5) alle Funktionen des gleichen Typs sind und sich nur durch einen „Item-Leichtigkeitsparameter" ϵ_i voneinander unterscheiden

$$(1.3.5) \qquad f_i(\xi_v) = \frac{\xi_v \epsilon_i}{1 + \xi_v \epsilon_i} \qquad (i = 1, 2, \ldots).$$

Wir fassen also zusammen, dass bei lokaler stochastischer Unabhängigkeit der Items ein Vergleich von Personen aufgrund der Anzahl der positiv beantworteten Items nur dann ohne Informationsverlust vorgenommen werden kann, wenn das Modell (1.3.5) erfüllt ist. Ist das Modell nicht erfüllt, so erlaubt diese Vorgangsweise immer noch eine erwartungstreue Schätzung der Rangpositionen der Probanden auf der latenten Dimension ξ.

Im folgenden werden wir zeigen, wie das Prinzip der erschöpfenden Statistiken dazu ausgenutzt werden kann, spezifisch objektive Vergleiche von Personen bzw. von Items vorzunehmen. Dabei gehen wir davon aus, dass n Personen mit den Personenparametern $\xi_1, \ldots, \xi_n$ k Items mit den Item-Leichtigkeitsparametern $\epsilon_1, \ldots, \epsilon_k$ unabhängig voneinander beantwortet haben, so dass die Gesamtheit der Antworten a_{vi} in Form einer n-zeiligen und k-spaltigen Antwortmatrix $((a_{vi}))$ dargestellt werden kann, deren Likelihood aufgrund der Beziehungen (1.3.1) und (1.3.5) die Form

$$(1.3.6) \qquad p\{((a_{vi}))\} = \prod_{v=1}^{n} \prod_{i=1}^{k} p\{a_{vi}\} = \frac{\prod_{v=1}^{n} \prod_{i=1}^{k} (\xi_v \epsilon_i)^{a_{vi}}}{\prod_{v=1}^{n} \prod_{i=1}^{k} (1 + \xi_v \epsilon_i)}$$

hat. Wir stellen fest, dass der Parameter ξ_v im Zähler dieses Ausdrucks an k Stellen, jedesmal mit dem Exponenten a_{vi} vorkommt, insgesamt also mit dem Exponenten a_{vo}. Der Parameter ϵ_i kommt an n Stellen vor, ebenfalls mit dem Exponenten a_{vi} und daher insgesamt mit den Exponenten $a_{oi} = \sum_{v=1}^{n} a_{vi}$. Wir können (1.3.6) daher auch in der einfachen Form

$$(1.3.7) \qquad p\{((a_{vi}))\} = \frac{\prod_{v=1}^{n} \xi_v^{a_{vo}} \cdot \prod_{i=1}^{k} \epsilon_i^{a_{oi}}}{\prod_{v=1}^{n} \prod_{i=1}^{k} (1 + \xi_v \epsilon_i)}$$

anschreiben, woraus wir unmittelbar ersehen, dass die Likelihood der Datenmatrix $((a_{vi}))$ nur von den Randvektoren $(a_{vo}) = (a_{1o}, \ldots, a_{no})$ und $(a_{oi}) = (a_{o1}, \ldots, a_{ok})$ abhängt. Alle möglichen Datenmatrizen mit übereinstimmenden Randvektoren (a_{vo}) und (a_{oi}) sind daher gleich wahrscheinlich, so dass

$$(1.3.8) \qquad p\left\{(a_{vo}), (a_{oi})\right\} = \begin{bmatrix} (a_{vo}) \\ \\ (a_{oi}) \end{bmatrix} \cdot p\left\{((a_{vi}))\right\},$$

worin

$$(1.3.9) \qquad \begin{bmatrix} (a_{vo}) \\ \\ (a_{oi}) \end{bmatrix} = \text{ die Anzahl der möglichen Datenmatrizen bei gegebenen Randvektoren } (a_{vo}) \text{ und } (a_{oi}).$$

Die bedingte Wahrscheinlichkeit einer bestimmten Antwortmatrix $((a_{vi}))$ bei gegebenen Randvektoren (a_{vo}) und (a_{oi}) hängt somit von den Modellparametern nicht mehr ab. Informationen darüber, welche Person welches Item positiv beantwortet hat, bringen keine neuen Aufschlüsse über die Modellparameter und sind daher für die Parameterschätzung irrelevant. Die Randvektoren (a_{vo}) und (a_{oi}) sind erschöpfende Statistiken für die Parametervektoren (ξ_v) und (ϵ_i). Wird nur einer der Randvektoren konstant gehalten, so erhalten wir bedingte Likelihoodfunktionen der Datenmatrix $((a_{vi}))$, welche nur noch von jenen Modellparametern abhängig sind, deren erschöpfende Statistiken nicht festgehalten werden. Wie RASCH (1966c) zeigt, gilt

$$(1.3.10) \qquad p\left\{((a_{vi})) \mid (a_{vo})\right\} = \frac{\prod_{i=1}^{k} \epsilon_i^{a_{oi}}}{\prod_{v=1}^{n} \gamma_{a_{vo}}(\epsilon_i)}$$

worin $\gamma_{a_{vo}}(\epsilon_i)$ die symmetrische Grundfunktion a_{vo}-ter Ordnung der Parameter $\epsilon_1, \ldots, \epsilon_k$ bezeichnet:

$$(1.3.11)\quad \begin{aligned} \gamma_0(\epsilon_i) &= 1\\ \gamma_1(\epsilon_i) &= \epsilon_1 + \epsilon_2 + \ldots + \epsilon_k\\ \gamma_2(\epsilon_i) &= \epsilon_1\,\epsilon_2 + \epsilon_1\,\epsilon_3 + \ldots + \epsilon_{k-1}\,\epsilon_k\\ &\ldots\ldots\ldots\ldots\ldots\\ \gamma_k(\epsilon_i) &= \epsilon_1\,\epsilon_2 \ldots \epsilon_k, \end{aligned}$$

und in analoger Weise

$$(1.3.12)\quad p\left\{((a_{vi})) \mid (a_{oi})\right\} = \frac{\prod_{v=1}^{n} \xi_v^{a_{vo}}}{\prod_{i=1}^{k} \gamma_{a_{oi}}(\xi_v)},$$

worin $\gamma_{a_{oi}}(\xi_v)$ die symmetrische Grundfunktion a_{oi}-ter Ordnung der Parameter $\xi_1, \ldots, \xi_n$ bezeichnet. Die Separierbarkeit der Modellparameter – das formale Äquivalent der spezifischen Objektivität – ist damit nachgewiesen. Aufgrund der Beziehung (1.3.12) kann der Vergleich der Personen bezüglich der latenten Eigenschaft ξ auf eine bedingte Wahrscheinlichkeitsverteilung der Beobachtungsdaten $((a_{vi}))$ gestützt werden, welche nur noch von den zu vergleichenden Parametern abhängig ist.

Zur Schätzung der Modellparameter hat RASCH (1966b) vorgeschlagen, die Methode der Maximum-Likelihood auf die bedingten Likelihoodfunktionen (1.3.10) bzw. (1.3.12) anzuwenden, und ANDERSEN (1973a) hat diese Vorgangsweise durch eine umfangreiche mathematische Theorie der „bedingten Inferenz" gerechtfertigt. Die daraus resultierenden Schätzfunktionen für die Parameter des Modells (1.3.5) sind Spezialfälle der im zweiten Kapitel dieses Buches diskutierten Algorithmen für das mehrkategorielle Testmodell von RASCH (1961)

$$(1.3.13)\quad p\{h \mid v, i\} = \frac{\xi_{vh}\,\epsilon_{ih}}{1 + \sum_{q=1}^{m-1} \xi_{vq}\,\epsilon_{iq}}$$

worin $p\{h \mid v, i\}$ die Wahrscheinlichkeit beschreibt, mit welcher die Vp v das Item i in der h-ten Antwortkategorie von m möglichen Anwortkategorien beantwortet.

Computerprogramme für die numerische Lösung der Schätzgleichungen finden sich bei FISCHER & ALLERUP (1968) sowie FISCHER & SCHEIBLECHNER (1970) für das zweikategorielle Testmodell (1.3.5) und bei

SCHEIBLECHNER (1971b) sowie ANDERSEN (1972) für das mehrkategorielle Testmodell (1.3.13).

Um die Bedeutung der bedingten („conditional") Maximum-Likelihood-Methode (CML-Methode) für die psychologische Statistik aufzuzeigen, greifen FISCHER (1971) und ANDERSEN (1973b) auf die von NEYMAN & SCOTT (1948) eingeführte Unterscheidung zwischen strukturellen und inzidentellen Modellparametern zurück. Sie beruht im wesentlichen auf der Relation zwischen der Anzahl der Parameter, welche in einem probabilistischen Modell enthalten sind, und der Anzahl der Beobachtungen, welche von diesen Parametern abhängen.

- Strukturparameter sind von endlicher Anzahl, und es können zu jedem einzelnen von ihnen potentiell unendlich viele Beobachtungen erhoben werden, welche von diesem abhängig sind.
- Inzidentelle Parameter dagegen gibt es potentiell unendlich viele, wobei zu jedem einzelnen von ihnen immer nur eine endliche Anzahl von Beobachtungen gegeben ist.

In der psychologischen Testtheorie z. B. sind die Itemparameter als Strukturparameter aufzufassen. Jeder Test enthält immer nur eine begrenzte Anzahl von Items, und durch Vergrösserung der Personenstichprobe ist es möglich, beliebig viele Beobachtungen zu erhalten, welche von den Itemparametern abhängen. Jede Vergrösserung der Personenstichprobe jedoch bewirkt die Einführung eines weiteren Personenparameters, wobei dieselbe Person stets nur eine relativ kleine Anzahl von Items bearbeitet. Die Personenparameter sind daher als inzidentelle Parameter aufzufassen. FISCHER (1971) hat darauf hingewiesen, dass auch die Umkehrung denkbar ist, dass individuelle Parameter die Eigenschaft von Strukturparametern annehmen, z. B. wenn man in der Psychophysik von der Annahme ausgeht, dass man an einer einzigen Versuchsperson beliebig viele Schwellenmessungen vornehmen kann. Im vierten Kapitel unseres Buches werden wir ein Modell zur Messung sozialer Normen diskutieren, das die Eigenschaft hat, dass inhaltlich völlig gleich definierte Parameter – nämlich die Normierungsleistungen von Teilgruppen einer Gesellschaft relativ zu den Normen des übergeordneten sozialen Systems – einmal als Strukturparameter und das andere Mal als inzidentelle Parameter aufzufassen sind, je nachdem ob die Normen grosser Gruppen oder ob Kleingruppennormen untersucht werden.

Zusammenfassend kann gesagt werden, dass fast jede psychologische Theorie inzidentelle Parameter enthält, zumindest dann, wenn sie individuelle Unterschiede zwischen den Personen zulässt. Wie aus den Arbeiten von NEYMAN & SCOTT (1948) und ANDERSEN (1973a, b) hervorgeht, ergeben sich daraus grundsätzliche statistische Schätzprobleme. Denn die herkömmlichen Parameterschätzmethoden versagen, wenn die Anzahl der zu schätzenden Parameter nicht gegen einen festen Wert konvergiert, während die Anzahl der Beobachtungen gegen unendlich strebt. In einem solchen Fall liefert die Maximum-Likelihood-Methode nicht einmal konsistente Schätzfunktionen. Wendet man die Methode der Maximum-Likelihood dagegen auf bedingte Likelihoodfunktionen an, in denen (minimal) erschöpfende Statistiken für die inzidentellen Parameter als bekannt vorausgesetzt werden, so können die Strukturparameter geschätzt werden, ohne dass zugleich auch eine Schätzung der inzidentellen Parameter erforderlich ist. Wie wir am Beispiel des Modells (1.3.5) bereits gesehen haben, sind diese bedingten Likelihoodfunktionen von den inzidentellen Parametern unabhängig, und nach ANDERSEN (1970) hat die CML-Methode unter sehr allgemeinen Voraussetzungen dieselben Eigenschaften wie die direkte Maximum-Likelihood-Methode bei Fehlen inzidenteller Parameter: sie liefert konsistente Schätzfunktionen, die asymptotisch normalverteilt und erwartungstreu sind und deren Fehlervarianz bekannt ist. Auf der Grundlage der CML-Schätzungen der Strukturparameter kann die empirische Prüfung

- der Modellstruktur und
- der psychologischen Hypothesen über Relationen zwischen den Strukturparametern des Modells

in Form von bedingten Likelihood-Quotienten-Tests (ANDERSEN 1971, 1973c) ausgeführt werden[4].

ANDERSEN (1973b) hat untersucht, welchen formalen Bedingungen ein Testmodell genügen muss, damit die Anwendbarkeit der CML-Methode gewährleistet ist, und gezeigt, dass die Testmodelle (1.3.5) und 1.3.13) bei lokaler stochastischer Unabhängigkeit der Items die einzig möglichen Modelle sind, welche CML-Parameterschätzungen erlauben. Ist die lokale stochastische Unabhängigkeit der Items gewährleistet, so ist das Prinzip der CML-Parameterschätzung notwendig und hinreichend für die Testmodelle von Rasch.

1.4 Ein dynamisches Messmodell mit separierbaren Parametern

Eine wesentliche Verallgemeinerung der probabilistischen Testtheorie besteht im Verzicht auf die Annahme der lokalen stochastischen Unabhängigkeit der Items. Für die Messung der Aggressivität einer Person z. B.

[4] Anmerkung des Herausgebers: Vgl. das 2. Kapitel dieses Buches.

ist diese Erweiterung des Modellansatzes deshalb wesentlich, weil in der Katharsis-Hypothese behauptet wird, dass der Vollzug eines Aggressionsaktes die Wahrscheinlichkeit weiterer Aggressionen vermindere. In neueren Formulierungen der Katharsis-Hypothese (vgl. FESHBACH 1956; BRAMEL et al. 1968) wird dabei vorausgesetzt, dass die Person zuvor „zu aggressivem Verhalten provoziert worden ist; d.h. das Verhalten eines anderen habe Aggressionstendenzen in ihr geweckt, die primär gegen den Urheber der Provokation gerichtet sind" (DANN 1971, S. 60). Gehen wir von diesen neueren Formulierungen der Katharsis-Hypothese aus, so ergibt sich die Feststellung, dass unsere bisherigen Überlegungen zur Einführung des Begriffs der „Aggressivität" nur auf nichtprovozierte Aggressionstendenzen von Personen anwendbar sind. Soll provozierte „Aggressivität" gemessen werden, so steht das Konzept der Katharsis im Widerspruch zur Annahme der lokalen stochastischen Unabhängigkeit der Situationen und damit auch im Widerspruch zum Postulat der relativen zeitlichen Invarianz der durch (1.2.5) definierten Relation „aggressiver als". Wie sich zeigt, steht das Konzept der Katharsis aber nicht im Widerspruch zur Annahme der relativen zeitlichen Invarianz der Aggressivität, wenn man die Definition (1.2.5) in geeigneter Weise abschwächt, indem man für den Vergleich zweier Personen aufgrund der i-ten Situation nach der Provokation voraussetzt, dass sich die beiden Personen in den vorangegangenen Situationen gleich bzw. in einem wohldefinierten Sinne äquivalent verhalten haben.

Formal kommt die dynamische Erweiterung der probabilistischen Testtheorie darin zum Ausdruck, dass die Wahrscheinlichkeit eines bestimmten Antwortvektors (a_{vi}) nun nicht mehr in der Form

$$(1.4.1) \qquad p\{(a_{vi})\} = \prod_{i=1}^{k} p\{a_{vi}\}$$

dargestellt werden kann, sondern die Gestalt

$$(1.4.2) \qquad p\{(a_{vi})\} = \prod_{i=1}^{k} p\{a_{vi} \mid s_{vi}\}$$

hat, worin s_{vi} den partiellen Antwortvektor $(a_{v1}, \ldots, a_{vi-1})$ bezeichnet. An die Stelle der Itemcharakteristiken $f_i(\xi)$ treten daher nun bedingte Itemcharakteristiken

$$(1.4.3) \qquad f_{i \cdot s_{vi}}(\xi) = p\{a_{vi} = 1 \mid (a_{v1}, \ldots, a_{vi-1}) = s_{vi}\},$$

wobei wir im Einklang mit der oben angegebenen Abschwächung der Definition (1.2.5) voraussetzen wollen, dass jede einzelne der Funktionen $f_{i.s_{vi}}(\xi)$ mit der latenten Dimension ξ streng monoton wachsen soll. Sind die bedingten Itemcharakteristiken $f_{i.s_{vi}}(\xi)$ für festes i alle gleich, so ergibt sich die lokale stochastische Unabhängigkeit der Items als Spezialfall des allgemeineren Modellansatzes (1.4.2–1.4.3).

Als speziellen Modellansatz treffen wir die Annahme, dass die bedingten Itemcharakteristiken $f_{i.s_{vi}}(\xi)$ von der Beantwortung der vorangegangenen Items ausschliesslich über die Anzahl

$$(1.4.4) \qquad r_{vi} = \begin{cases} 0 & \text{für } i = 1 \\ \sum_{j=1}^{i-1} a_{vj} & \text{für } i = 2, 3, \ldots, k \end{cases}$$

der positiv beantworteten, vorangegangenen Items abhängt, so dass

$$(1.4.5) \qquad f_{i \cdot s_{vi}}(\xi) = f_{i \cdot r_{vi}}(\xi)$$

für alle partiellen Antwortvektoren s_{vi} mit gleichem Summenscore r_{vi} erfüllt ist. Unter den Modellvoraussetzungen (1.4.4–1.4.5) sind somit alle möglichen partiellen Antwortvektoren mit gleichem Summenscore r_{vi} bezüglich ihres Einflusses auf die Wahrscheinlichkeit einer positiven Beantwortung des i-ten Items als äquivalent zu betrachten.

Für die bedingten Itemcharakteristiken $f_{i.r_{vi}}(\xi)$ schliesslich postulieren wir die Modellstruktur

$$(1.4.6) \qquad f_{i \cdot r_{vi}}(\xi) = \frac{\xi_v + \psi_{r_{vi}}}{\xi_v + \sigma_i}$$

worin $\psi_{r_{vi}} < \sigma_i$ vorausgesetzt ist und worin σ_i die Schwierigkeit des i-ten Items repräsentiert: je grösser σ_i ist, desto geringer ist die Wahrscheinlichkeit einer positiven Beantwortung des i-ten Items (vgl. Abbildung 1.4.1).

Der Einfluss, welchen die positive Beantwortung von r_{vi} vorangegangenen Items auf die Lösungswahrscheinlichkeit des i-ten Items ausübt, kommt in den Parametern $\psi_{r_{vi}}$ zum Ausdruck, die wir daher im folgenden als Transfer-Parameter bezeichnen wollen. Je grösser der numerische Wert des Transfer-Parameters $\psi_{r_{vi}}$ ist, desto grösser ist die (bedingte) Lösungswahrscheinlichkeit des i-ten Items (vgl. Abb. 1.4.2).

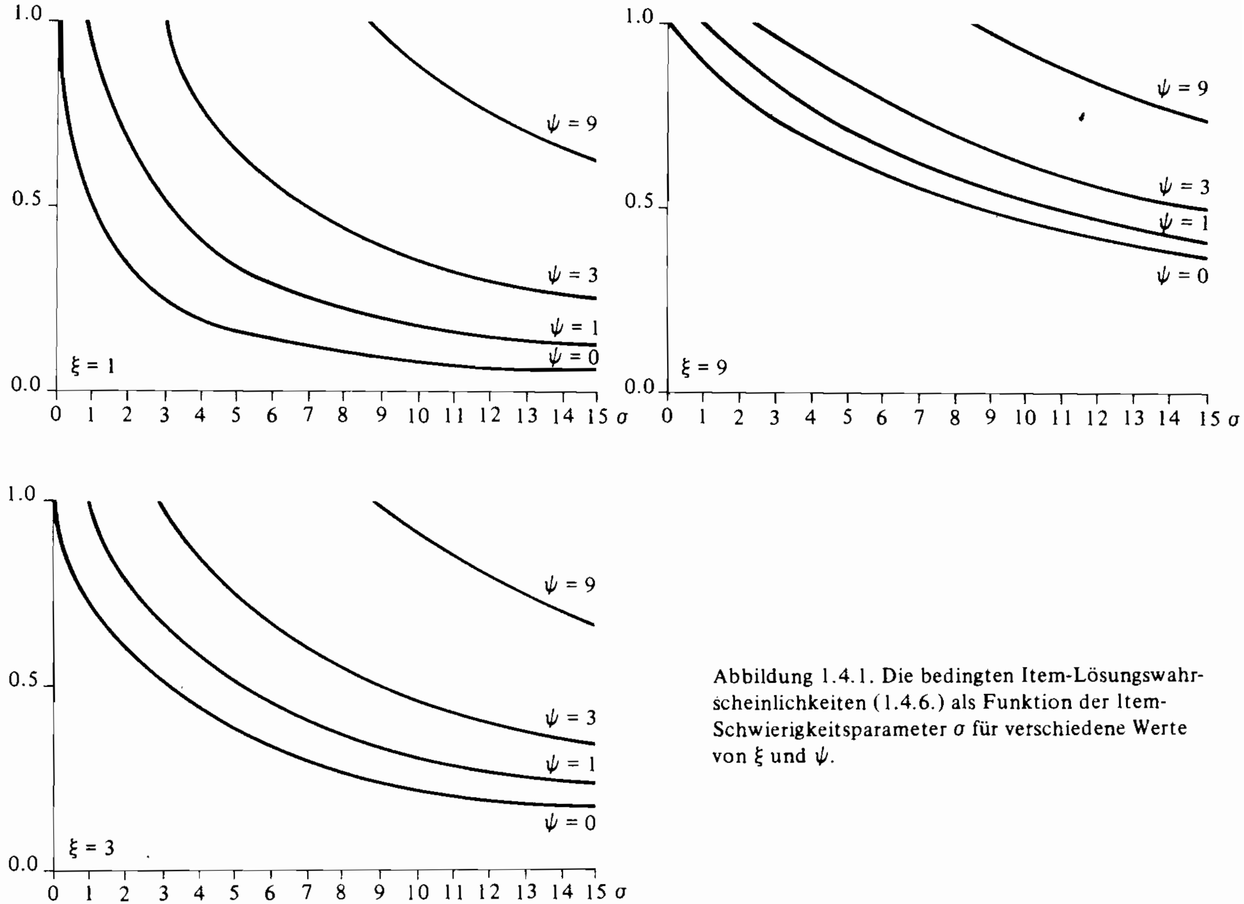
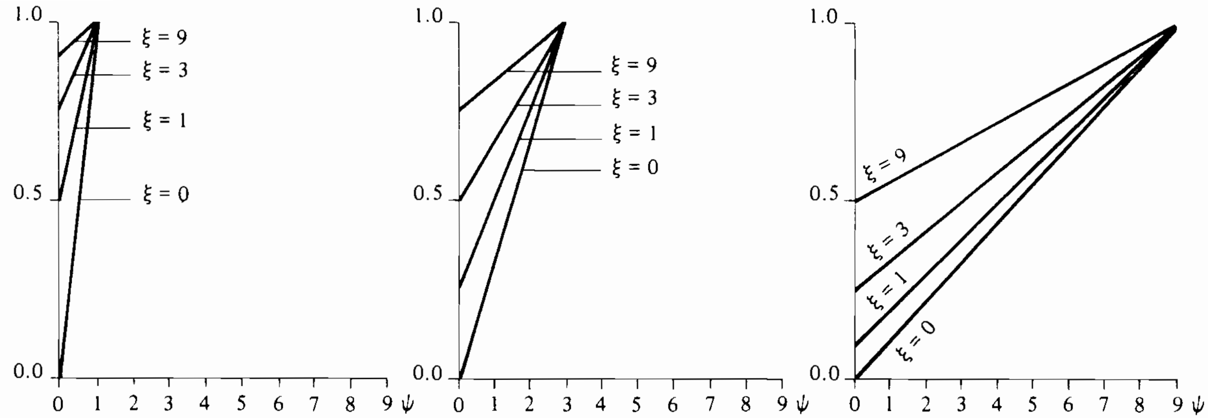


Abbildung 1.4.1. Die bedingten Item-Lösungswahrscheinlichkeiten (1.4.6.) als Funktion der Item-Schwierigkeitsparameter σ für verschiedene Werte von ξ und ψ.



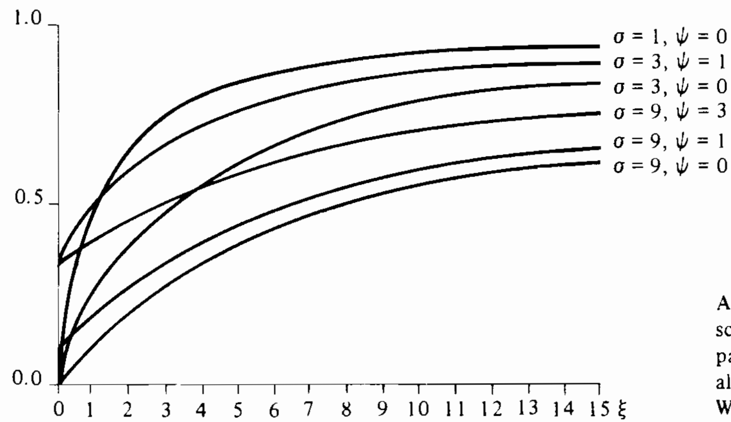


Abbildung 1.4.2. Die bedingten Item-Lösungswahrscheinlichkeiten (1.4.6.) als Funktion der Transferparameter für verschiedene Werte von ξ und σ, sowie als Funktion der latenten Dimension ξ für verschiedene Werte von σ und ψ.

- Steht der numerische Wert der Transfer-Parameter $\psi_{r_{vi}}$ in einem monoton wachsenden Zusammenhang mit r_{vi}, so kann der Transfer daher als „Lerngewinn" interpretiert werden.
- Ist die Abhängigkeit der Transfer-Parameter $\psi_{r_{vi}}$ von r_{vi} dagegen monoton fallend, so ist der Transfer als „Reaktionshemmung" zu interpretieren. (In diesem Sinne kann z.B. die Katharsis als eine Reaktionshemmung für Aggressionen verstanden werden.)
- Ist der Zusammenhang zwischen $\psi_{r_{vi}}$ und r_{vi} nicht monoton, so sprechen wir von einer „Fluktuation", welche durch gleichzeitig stattfindende Lern- und Hemmungsprozesse erklärt werden kann, die mit unterschiedlicher Beschleunigung ablaufen.

Durch die Modellgleichung (1.4.6) sind die Parameter des Modells jedoch noch nicht vollständig bestimmt, d.h. die Geltung von (1.4.6) und die Bedeutung der Modellparameter bleiben auch dann erhalten, wenn wir die Modellparameter mit einer positiven Konstanten multiplizieren und zu den Personenparametern ξ_v eine beliebige andere Konstante addieren, welche wir zugleich von den Transfer-Parametern $\psi_{r_{vi}}$ und den Itemparametern σ_i subtrahieren. Die Parameter des dynamischen Testmodells (1.4.6) werden somit auf Intervallskalen gemessen und es ist erforderlich, eine Skalennormierung vorzunehmen. O.B.d.A. können wir daher voraussetzen, dass

$$(1.4.7) \qquad \mathrm{MIN}\,(\psi_{r_{vi}}) = 0 \quad \text{für } r_{vi} = 0, \ldots, k-1$$

und

$$(1.4.8) \qquad \prod_{i=1}^{k} \sigma_i = 1,$$

womit wir Parameterisierungsbedingungen einführen, welche denen entsprechen, die im speziellen logistischen Testmodell (1.3.5) von RASCH (1960) üblich sind, das sich für $\psi_0 = \psi_1 = \ldots = \psi_{k-1} = 0$ und mit der Definition $\epsilon_i = \sigma_i^{-1}$ als Spezialfall des Modells (1.4.6) ergibt. Wir zeigen, dass das dynamische Testmodell (1.4.6) im wesentlichen dieselben Eigenschaften aufweist wie die Testmodelle von RASCH (1960, 1961): Auch im Modell (1.4.6)

- ist die Anzahl der gelösten Aufgaben a_{v0} eine erschöpfende Statistik für den Parameter der Person v,
- können Vergleiche von Personen (oder Items) in spezifisch objektiver Weise ausgeführt werden und
- existieren CML-Schätzfunktionen für die Strukturparameter.

Erschöpfende Statistiken: Um nachzuweisen, dass die Anzahl der gelösten Aufgaben a_{vo} im Modell (1.4.6) eine erschöpfende Statistik für den Personenparameter ξ_v ist, müssen wir zeigen, dass die bedingte Wahrscheinlichkeit (1.3.4) des Antwortvektors (a_{vi}) der Person bei gegebenem Summenscore a_{vo} nicht mehr von dem Parameter ξ_v abhängt. Zunächst schreiben wir (1.4.6) in gleichwertiger Form

$$(1.4.9) \qquad p\left\{a_{vi} \mid r_{vi}\right\} = \frac{(\xi_v + \psi_{r_{vi}})^{a_{vi}} (\sigma_i - \psi_{r_{vi}})^{1 - a_{vi}}}{\xi_v + \sigma_i}$$

an und erhalten durch Einsetzen in (1.4.2)

$$(1.4.10) \qquad p\left\{(a_{vi})\right\} = \frac{\prod_{i=1}^{k} (\xi_v + \psi_{r_{vi}})^{a_{vi}} (\sigma_i - \psi_{r_{vi}})^{1 - a_{vi}}}{\prod_{i=1}^{k} (\xi_v + \sigma_i)}$$

$$= \frac{\prod_{r=0}^{a_{vo}-1} (\xi_v + \psi_r) \prod_{i=1}^{k} (\sigma_i - \psi_{r_{vi}})^{1 - a_{vi}}}{\prod_{i=1}^{k} (\xi_v + \sigma_i)}$$

Die Wahrscheinlichkeit eines bestimmten Summenscores a_{vo} folgt dann aus (1.4.10) durch Aufsummieren der Wahrscheinlichkeiten $p\left\{(a^*_{vi})\right\}$ über alle möglichen Antwortvektoren (a^*_{vi}), welche denselben Summenscore $\sum_{i=1}^{k} a^*_{vi} = a_{vo}$ haben:

$$(1.4.11) \qquad p\left\{a_{vo}\right\} = \sum_{(a^*_{vi}) \mid a_{vo}} p\left\{(a^*_{vi})\right\}$$

$$= \frac{\prod_{r=0}^{a_{vo}-1} (\xi_v + \psi_r) \sum_{(a^*_{vi}) \mid a_{vo}} \prod_{i=1}^{k} (\sigma_i - \psi_{r^*_{vi}})^{1 - a^*_{vi}}}{\prod_{i=1}^{k} (\xi_v + \sigma_i)}$$

worin $r^*_{vi} = 0$ für $i = 1$ und $r^*_{vi} = \sum_{j=1}^{i-1} a^*_{vi}$ für $i = 2, 3, \ldots, k$.

Für die bedingte Wahrscheinlichkeit $p\left\{(a_{vi}) \mid a_{vo}\right\}$, schliesslich ergibt sich aus (1.4.10) und (1.4.11) der Ausdruck

$$(1.4.12) \qquad p\left\{(a_{vi}) \mid a_{vo}\right\} = \frac{p\left\{(a_{vi})\right\}}{p\left\{a_{vo}\right\}}$$

$$= \frac{\prod_{i=1}^{k} (\sigma_i - \psi_{r_{vi}})^{1-a_{vi}}}{\sum_{(a^*_{vi}) \mid a_{vo}} \prod_{i=1}^{k} (\sigma_i - \psi_{r^*_{vi}})^{1-a^*_{vi}}}$$

der den Personenparameter ξ_v nicht mehr enthält, womit der gewünschte Nachweis erbracht ist.

Spezifische Objektivität: Im folgenden verallgemeinern wir auf n Probanden, welche die k Items unabhängig voneinander bearbeiten, und erhalten aus (1.4.10)

$$(1.4.13) \qquad p\left\{((a_{vi}))\right\} = \prod_{v=1}^{n} p\left\{(a_{vi})\right\}$$

$$= \frac{\left\{\prod_{v=1}^{n} \prod_{r=0}^{a_{vo}-1} (\xi_v + \psi_r)\right\} \cdot \left\{\prod_{i=1}^{k} \prod_{r=0}^{i-1} (\sigma_i - \psi_r)^{n_{ri}}\right\}}{\prod_{v=1}^{n} \prod_{i=1}^{k} (\xi_v + \sigma_i)}$$

worin n_{ri} die Anzahl der Personen mit $r_{vi} = r$ und $a_{vi} = 0$ bezeichnet. Wie wir feststellen, hängt also die Likelihood (1.4.13) der Datenmatrix $((a_{vi}))$ nur von den Häufigkeiten a_{vo} $(v = 1, n)$ und n_{ri} $(i = 1, k;\ r = 0, i-1)$ ab, und alle möglichen Datenmatrizen mit übereinstimmenden Personenrandvektoren (a_{vo}) und Itemrandmatrizen $((n_{ri}))$ sind daher gleich wahrscheinlich, so dass

$$(1.4.14) \qquad p\left\{(a_{vo}), ((n_{ri}))\right\} = \begin{bmatrix} (a_{vo}) \\ ((n_{ri})) \end{bmatrix} \cdot p\left\{((a_{vi}))\right\},$$

worin

$$(1.4.15) \qquad \begin{bmatrix} (a_{vo}) \\ ((n_{ri})) \end{bmatrix} =$$ die Anzahl der möglichen Datenmatrizen bei gegebenem Personenrandvektor (a_{vo}) und gegebener Itemmatrix $((n_{ri}))$.

Die bedingte Wahrscheinlichkeit einer bestimmten Antwortmatrix bei gegebenem (a_{vo}) und $((n_{ri}))$

$$(1.4.16) \qquad p\left\{((a_{vi})) \mid (a_{vo}), ((n_{ri}))\right\} = \frac{p\left\{((a_{vi}))\right\}}{p\left\{(a_{vo}), ((n_{ri}))\right\}}$$

$$= \frac{1}{\begin{bmatrix} (a_{vo}) \\ ((n_{ri})) \end{bmatrix}}$$

ist daher von den Modellparametern unabhängig. Informationen darüber, welche Person welche Items positiv beantwortet hat, bringen keine neuen Aufschlüsse über die Modellparameter und sind daher für die Parameterschätzung irrelevant. Der Personenrandvektor (a_{vo}) und die Itemrandmatrix $((n_{ri}))$ schöpfen die Information über die Modellparameter voll aus.

Als nächstes weisen wir die Separierbarkeit der Modellparameter im dynamischen Testmodell (1.4.6) nach und zeigen, dass das Modell spezifisch objektive Vergleiche von Personen erlaubt. Zu diesem Zwecke leiten wir zunächst einen Ausdruck für die Wahrscheinlichkeit der Itemrandmatrix $((n_{ri}))$ her und erhalten

$$(1.4.17) \qquad p\left\{((n_{ri}))\right\} = \sum_{((a^*_{vi}))\,|\,((n_{ri}))} p\left\{((a^*_{vi}))\right\}$$

$$= \frac{\prod_{i=1}^{k} \prod_{r=0}^{i-1} (\sigma_i - \psi_r)^{n_{ri}} \sum_{((a^*_{vi}))\,|\,((n_{ri}))} \prod_{v=1}^{n} \prod_{r=0}^{a^*_{vo}-1} (\xi_v + \psi_r)}{\prod_{v=1}^{n} \prod_{i=1}^{k} (\xi_v + \sigma_i)}$$

worin $\sum_{((a^*_{vi}))\,|\,((n_{ri}))}$ eine Summenbildung über alle möglichen Antwortmatrizen

$((a^*_{vi}))$ bei gegebener Itemrandmatrix $((n_{ri}))$ bezeichnet, und worin $a^*_{vo} = \sum_{i=1}^{k} a^*_{vi}$. Für die bedingte Wahrscheinlichkeit der Antwortmatrix $((a_{vi}))$ bei gegebener Itemrandmatrix $((n_{ri}))$ folgt dann aus (1.4.13) und (1.4.14) die Gleichung

$$(1.4.18) \qquad p\left\{((a_{vi}))\,|\,((n_{ri}))\right\} = \frac{p\left\{((a_{vi}))\right\}}{p\left\{((n_{ri}))\right\}}$$

$$= \frac{\prod_{v=1}^{n} \prod_{r=0}^{a_{vo}-1} (\xi_v + \psi_r)}{\sum_{((a^*_{vi}))\,|\,((n_{ri}))} \prod_{v=1}^{n} \prod_{r=0}^{a^*_{vo}-1} (\xi_v + \psi_r)},$$

in der die Itemparameter nicht mehr vorkommen. Auf der Grundlage der Beziehung (1.4.18) ist es somit möglich, spezifisch objektive Vergleiche von Personen vorzunehmen. Die Personenparameter ξ_v dürfen daher als latente Eigenschaften der Personen interpretiert werden.

CML-Schätzfunktionen: Abschliessend leiten wir CML-Schätzfunktionen für die strukturellen Item- und Transferparameter her, indem wir (1.4.12) in

$$(1.4.19)\qquad p\left\{((a_{vi}))\mid(a_{vo})\right\} = \prod_{v=1}^{n} p\left\{(a_{vi})\mid a_{vo}\right\}$$

einsetzen und die logarithmischen Ableitungen der bedingten Likelihood

$$(1.4.20)\qquad p\left\{((a_{vi}))\mid(a_{vo})\right\} = \frac{\prod_{i=1}^{k}\prod_{r=0}^{i-1}(\sigma_i-\psi_r)^{n_{ri}}}{\prod_{v=1}^{n}\sum_{(a^*_{vi})\mid a_{vo}}\prod_{i=1}^{k}(\sigma_i-\psi_{r^*_{vi}})^{1-a^*_{vi}}}$$

Null setzen. Die daraus folgenden Bestimmungsgleichungen lauten für die Itemparameter $\sigma_\alpha(\alpha = 1, \ldots, k)$

$$(1.4.21)\qquad \sum_{r=0}^{\alpha-1}\frac{n_{r\alpha}}{\sigma_\alpha-\psi_r} = \sum_{a_{vo}=0}^{k}\frac{n(a_{vo})\cdot\sum_{(a^*_{vi})\mid a_{vo}}(1-a^*_{v\alpha})\prod_{\substack{i=1\\ i\neq\alpha}}^{k}(\sigma_i-\psi_{r^*_{vi}})^{1-a^*_{vi}}}{\sum_{(a^*_{vi})\mid a_{vo}}\prod_{i=1}^{k}(\sigma_i-\psi_{r^*_{vi}})^{1-a^*_{vi}}}$$

und für die Transferparameter $\psi_\beta\,(\beta = 0, \ldots, k-1)$

$$(1.4.22)\qquad \sum_{i=1}^{k}\frac{n_{\beta i}}{\sigma_i-\psi_\beta} =$$

$$= \sum_{a_{vo}=0}^{k}\frac{n(a_{vo})\cdot\sum_{(a^*_{vi})\mid a_{vo}}\sum_{\substack{i=1\\ r^*_{vi}=\beta}}^{k}(1-a^*_{vi})\prod_{\substack{j=1\\ j\neq i}}^{k}(\sigma_j-\psi_{r^*_{vj}})^{1-a^*_{vj}}}{\sum_{(a^*_{vi})\mid a_{vo}}\prod_{i=1}^{k}(\sigma_i-\psi_{r^*_{vi}})^{1-a^*_{vi}}}$$

$n(a_{vo})$ bezeichnet die Anzahl der Personen, deren Antwortvektoren denselben Summenscore a_{vo} aufweisen.

Rechentechnische Fragen der numerischen Lösung der Schätzgleichungen, insbesondere der rekursiven Berechnung der Funktion

$$(1.4.23)\quad G(a_{vo}) = \sum_{(a^*_{vi})\,|\,a_{vo}} \prod_{i=1}^{k} (\sigma_i - \psi_{r^*_{vi}})^{1 - a^*_{vi}}$$

und ihrer partiellen Ableitungen nach den Itemparametern σ_i und den Transferparametern ψ_r, behandeln KEMPF & HAMPAPA (1974). Ein Computerprogramm zur iterativen Berechnung der Schätzfunktionen findet sich bei KEMPF & MACH (1974).

1.5 Die Erklärbarkeit dynamischer Prozesse durch die Veränderung von Verhaltensdispositionen, diskutiert am Beispiel der „Katharsis-Hypothese"

Versucht man die Ergebnisse des letzten Abschnittes zusammenzufassen, so fällt auf, dass das dynamische Messmodell (1.4.6) zwar im wesentlichen dieselben wünschenswerten Eigenschaften aufweist wie die Testmodelle von Rasch, aber doch unter gewissen Einschränkungen.

– Anders als in den Testmodellen von Rasch, deren Parameter beliebige, positive, reelle Zahlenwerte annehmen können, bestehen im dynamischen Modell gewisse Restriktionen bezüglich der Modellparameter untereinander, d.h. die Item- und Transfer-Parameter müssen der Bedingung

$$(1.5.1)\qquad \psi_{r_{vi}} < \sigma_i \qquad \text{für } r_{vi} = 0, 1, \ldots, i-1$$

genügen und

– anders als in den Testmodellen von Rasch und deren mehrfaktoriellen Verallgemeinerungen (MICKO 1970, SCHEIBLECHNER 1971a, KEMPF 1972a), in denen jede Klasse von Parametern unabhängig von allen übrigen Parameterklassen geschätzt werden kann, ist dies im dynamischen Modell nicht möglich. Die Item- und Personenparameter können zwar unabhängig voneinander bestimmt werden, nicht jedoch unabhängig von den Transfer-Parametern.

Aus psychologischer Sicht können solche Abhängigkeiten der Modellparameter bzw. ihrer Schätzfunktionen nur dann gerechtfertigt werden, wenn sie eine eindeutige psychologische Interpretation haben und wenn gezeigt werden kann, dass sie dem Grundprinzip der spezifischen Objektivität nicht widersprechen.

Zu diesem Zweck definieren wir

$$(1.5.2) \qquad \xi_{vr_{vi}} = \xi_v + \psi_{r_{vi}}$$

als den Ausprägungsgrad der latenten Dimension ξ bei der Person v nach r_{vi} „positiven" Reaktionen und

$$(1.5.3) \qquad \sigma_{ir_{vi}} = \sigma_i - \psi_{r_{vi}}$$

als die Schwierigkeit der Aufgabe i nach r_{vi} „positiven"Reaktionen und schreiben das Modell (1.4.6) in gleichwertiger Form

$$(1.5.4) \qquad f_{i \cdot r_{vi}}(\xi) = \frac{\xi_{vr_{vi}}}{\xi_{vr_{vi}} + \sigma_{ir_{vi}}}$$

an. Aufgrund dieser Beziehungen ist es dann möglich, die dynamische Transfer-Komponente des Modells (1.4.6) als Veränderung der latenten Dimension ξ bei gleichzeitiger, gegenläufiger Veränderung der Item-Schwierigkeit zu identifizieren. Der oben festgestellte Tatbestand, wonach weder die Personen- noch die Itemparameter unabhängig von den Transferparametern bestimmt werden können, steht somit nicht im Widerspruch zum Prinzip der spezifischen Objektivität. Denn wenn sich eine Variable während ihrer Messung ständig verändert, so kann auch nicht gefordert werden, dass ihre Messung von der gleichzeitigen Messung ihrer Veränderung unabhängig sein soll.

Zugleich folgt aus der Definition (1.5.3) auch eine eindeutige Interpretation der Beziehung (1.5.1). Sie lautet, dass der maximal-mögliche, positive Transfer der vorangegangenen Itemlösungen auf die Lösungswahrscheinlichkeit des i-ten Items die „Ausgangsschwierigkeit" σ_i des Items nicht erreichen darf, da sonst die Beziehung

$$(1.5.5) \qquad f_{i \cdot r_{vi}}(\xi) > 0$$

nicht mehr erfüllt ist und das Modell (1.5.4) entweder deterministischen Charakter annimmt (falls $\psi_{r_{vi}} = \sigma_i$) oder aber den Axiomen der Wahrscheinlichkeitsrechnung nicht mehr entspricht (falls $\psi_{r_{vi}} > \sigma_i$). Psychologisch gesehen bedeutet die Restriktion (1.5.1) daher, dass eine probabilistische Analyse der Lösungen des i-ten Items nur dann möglich ist, wenn das Item in jedem Falle noch eine gewisse „Restschwierigkeit" $\sigma_{ir_{vi}}$ aufweist, ungeachtet dessen, wieviele andere Items zuvor positiv beantwortet wurden. Interessant ist an dieser Beziehung vor allem, dass sie die Möglichkeit des Überganges von probabilistischen zu deterministischen Aussagen aufweist.

Indem das dynamische Testmodell (1.4.6) den Transfer von früheren auf spätere Item-Lösungen als gleichzeitige Veränderung der latenten Eigenschaft der Personen und der Itemschwierigkeiten beschreibt, setzt seine Anwendung die Definition eines psychologischen Anfangszustandes des dynamischen Prozesses voraus, wie sie z.B. in den oben erwähnten neueren Formulierungen der Katharsis-Hypothese gegeben ist, in denen man davon ausgeht, dass der Vollzug einer Aggressionshandlung nur dann zu einer Verringerung der Wahrscheinlichkeit weiterer Aggressionen führt, wenn zuvor eine Provokation stattgefunden hat. Wird auf die Definition eines psychologischen Anfangszustandes des dynamischen Prozesses verzichtet, wie dies in der „originalen" Katharsis-Hypothese der Fall ist (vgl. DANN 1971), so ist die Anwendbarkeit des Modells (1.4.6) nur dann gegeben, wenn entweder

- die gesamte „Vorgeschichte" der Probanden bekannt ist, d.h. wenn für jeden Probanden bekannt ist, wie oft er vor Beginn der systematischen Verhaltensbeobachtungen schon aggressiv gehandelt hat, oder
- wenn vorausgesetzt wird, dass der Transfer-Effekt ψ_r für wachsendes r gegen einen festen Wert konvergiert und dass jeder einzelne Proband bereits vor Beginn der systematischen Verhaltensbeobachtung häufig genug aggressiv gehandelt hat, so dass der Transfer von den präexperimentellen Verhaltensweisen auf die im Experiment beobachteten, diesen festen Wert bereits sehr gut approximiert.

Während die erste dieser Möglichkeiten in der Praxis der psychologischen Forschung nicht realisierbar ist, führt die zweite Möglichkeit dazu, dass während der systematischen Verhaltensbeobachtung kein Transfer mehr nachweisbar ist, da die ψ_r ihren asymptotischen Endwert voraussetzungsgemäss bereits (näherungsweise) erreicht haben. Das dynamische Testmodell (1.4.6) ist zur Untersuchung der „originalen" Katharsis-Hypothese somit grundsätzlich nicht geeignet. Wollte man die „originale" Katharsis-Hypothese untersuchen, so müsste man vielmehr ein Modell konstruieren, in welchem sich der Transfer von früheren auf spätere Handlungen einer Person ausschliesslich auf die Ausprägung der latenten Dimension auswirkt, d.h. in dem die Katharsis durch eine Veränderung der Aggressivität der Person allein erklärt werden kann, so dass die Aggressivität einer Person als Funktion ihrer früheren Aggressionshandlungen darstellbar ist. Wie sich zeigt, steht ein solcher Modellansatz jedoch in grundsätzlichem Widerspruch zu den Prinzipien der spezifischen Objektivität und der CML-Methode. Um dies nachzuweisen, greifen wir auf die oben erwähnten Theoreme von RASCH (1965) bzw. ANDERSEN (1973b) zurück, wonach die Forderung nach spezifischer Objektivität und die

Methode der CML-Parameterschätzung im Spezialfall der lokalen stochastischen Unabhängigkeit der Items notwendigerweise zu Rasch's speziellem logistischen Testmodell (1.3.5) führen. Jedes dynamische Testmodell, das vergleichbare Eigenschaften aufweisen soll, muss daher in irgendeiner Form als Verallgemeinerung des Modellansatzes von Rasch darstellbar sein. Soll der Transfer von früheren auf spätere Item-Lösungen darüber hinaus auch noch als alleinige Veränderung der latenten Eigenschaft der Personen beschreibbar sein, so ergibt sich daher notwendigerweise der Modellansatz

$$(1.5.6) \qquad f_{i \cdot s_{vi}}(\xi) = \frac{\xi_{vi}\epsilon_i}{1 + \xi_{vi}\epsilon_i}$$

worin ξ_{vi} die Position beschreibt, welche die Person v im Zeitpunkt der Beantwortung des Items i auf der latenten Dimension ξ einnimmt. Das Modell (1.5.6) erlaubt CML-Parameterschätzungen jedoch nur dann, wenn ξ_{vi} von der Beantwortung der vorangegangenen Items unabhängig ist, womit wir zum Spezialfall der lokalen stochastischen Unabhängigkeit der Items zurückkehren.

Zum Beweis genügt ein einfaches Gegenbeispiel, etwa wenn wir von nur zwei Items i und j ausgehen. Soll dann die Anzahl der positiv beantworteten Items eine erschöpfende Statistik für die Personenparameter sein, so darf aufgrund von (1.3.4) die bedingte Wahrscheinlichkeit

$$(1.5.7) \qquad p\left\{(a_{vi}, a_{vj}) \mid a_{vi} + a_{vj}\right\} = \frac{p\left\{(a_{vi}, a_{vj})\right\}}{p\left\{a_{vi} + a_{vj}\right\}}$$

nicht mehr von den Personenparametern abhängen. In den einfachen Fällen $a_{vi} + a_{vj} = 0$ und $a_{vi} + a_{vj} = 2$ ist diese Forderung trivialerweise erfüllt. Für $a_{vi} + a_{vj} = 1$ führt sie dazu, dass

$$(1.5.8) \qquad \frac{p\left\{a_{vi}=1\right\} \cdot p\left\{a_{vj}=0 \mid a_{vi}=1\right\}}{p\left\{a_{vi}=1\right\} \cdot p\left\{a_{vj}=0 \mid a_{vi}=1\right\} + p\left\{a_{vi}=0\right\} \cdot p\left\{a_{vj}=1 \mid a_{vi}=0\right\}}$$

$$= \frac{1}{1 + \dfrac{p\left\{a_{vi}=0\right\} \cdot p\left\{a_{vj}=1 \mid a_{vi}=0\right\}}{p\left\{a_{vi}=1\right\} \cdot p\left\{a_{vj}=0 \mid a_{vi}=1\right\}}} = \text{const.}$$

nicht mehr von den Personenparametern abhängen darf (vgl. FISCHER 1968). Einsetzen von $\xi_{vj} = \xi_{vj \cdot 0}$, falls die Person die Aufgabe i negativ beantwortet hat und $\xi_{vj} = \xi_{vj \cdot 1}$ im Falle einer positiven Beantwortung in (1.5.6) und Einsetzen von (1.5.6) in (1.5.8) ergibt dann nach einfacher algebraischer Umformung die Beziehung

$$(1.5.9) \qquad \frac{\epsilon_j}{\epsilon_i} = c.\frac{\xi_{vi}(1 + \xi_{vj \cdot 0}\epsilon_j)}{\xi_{vj \cdot 0}(1 + \xi_{vj \cdot 1}\epsilon_j)},$$

worin c eine von den Personenparametern unabhängige Konstante ist. Die Beziehung (1.5.9) kann aber nur dann erfüllt sein, wenn entweder $\xi_{vj \cdot 0} = \xi_{vj \cdot 1} = \xi_{vj}$ und ξ_{vi}/ξ_{vj} = const., woraus die lokale stochastische Unabhängigkeit der Items folgt, oder wenn die Itemparameter von den Personenparametern abhängig sind, was der Voraussetzung widerspricht, dass der Transfer die Personenparameter verändern soll, während die Itemparameter konstant bleiben.

Zusammenfassend lässt sich daher sagen, dass die „originale" Katharsis-Hypothese dem Konzept der „Aggressivität" grundsätzlich widerspricht. Andererseits führt die „originale" Katharsis-Hypothese jedoch notwendigerweise zur Annahme interindividueller Unterschiede (vgl. HILKE et al. 1973) und ist somit einer methodisch befriedigenden Überprüfung nicht zugänglich.

Dass es dennoch zulässige Formulierungen der Katharsis-Hypothese geben kann, in welchen die Katharsis mit einer Verringerung der Aggressivität gleichzusetzen ist, zeigen die folgenden Überlegungen, in denen wir davon ausgehen, dass die Unterdrückung provozierter Aggressionstendenzen zu einem „Aggressionsstau" führt, welcher erst dann ausgeglichen wird, wenn die Provokation aggressiv beantwortet würde. Unter einem „Aggressionsstau" ist dabei zu verstehen, dass die Aggressivität einer Person so lange zunimmt, bis die Provokation aggressiv beantwortet wurde. „Katharsis" bedeutet in diesem Zusammenhang den „Ausgleich" des Aggressionsstaus, d.h. das Zurückfallen der Aggressivität auf ihren Ausgangswert.

Versucht man diese Präzisierung der Katharsis-Hypothese zu formalisieren, so kann jeder Provokation i und jeder Person v eine Sequenz von Reaktionen zugeordnet werden, welche so lange andauert, bis die Provokation aggressiv beantwortet wurde. Für die Wahrscheinlichkeit, mit welcher die Reaktionssequenz im j-ten Zeitpunkt nach der Provokation noch nicht abgeschlossen ist, verwenden wir die Abkürzung

$$(1.5.10) \qquad q_{vij} = p\,\{z_{vi} > j\} = \frac{\lambda_{vij}}{1 + \lambda_{vij}} \quad \text{für } j = 1, 2, \ldots$$

worin z_{vi} jenen Zeitpunkt nach der i-ten Provokation bezeichnet, in welchem die Person v die Provokation aggressiv beantwortet und worin λ_{vij} durch

$$(1.5.11) \quad \lambda_{vij} = \frac{q_{vij}}{1 - q_{vij}} \quad \text{für } j = 1, 2, \ldots; \quad q_{vij} < 1$$

definiert ist. Mit der Definition

$$(1.5.12) \quad \omega_{vij} = \frac{\lambda_{vij+1}}{\lambda_{vij}} \quad \text{für } j = 1, 2, \ldots$$

folgt dann für beliebige aufeinanderfolgende Zeitpunkte j und j+1 die Beziehung

$$(1.5.13)^5 \quad p\{z_{vi} > j+1 \mid z_{vi} > j\} = \frac{q_{vij+1}}{q_{vij}} = \frac{\dfrac{\omega_{vij}\lambda_{vij}}{1 + \omega_{vij}\lambda_{vij}}}{q_{vij}} = \frac{\dfrac{\omega_{vij} q_{vij}}{1 - q_{vij}}}{q_{vij}\left(1 + \dfrac{\omega_{vij} q_{vij}}{1 - q_{vij}}\right)} = \frac{\omega_{vij}}{1 + (\omega_{vij} - 1) \cdot q_{vij}},$$

die eine streng monoton wachsende Funktion des Parameters ω_{vij} ist (vgl. Abbildung 1.5.1).

[5] Man beachte die formale Analogie zum Beta-Modell von LUCE (1959).

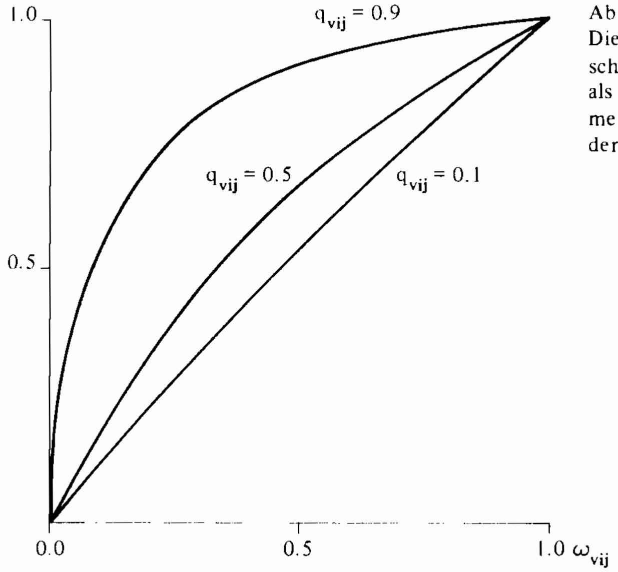


Abbildung 1.5.1
Die bedingte Wahrscheinlichkeit (1.5.13) als Funktion des Parameters ω_{vij} für verschiedene Werte von q_{vij}.

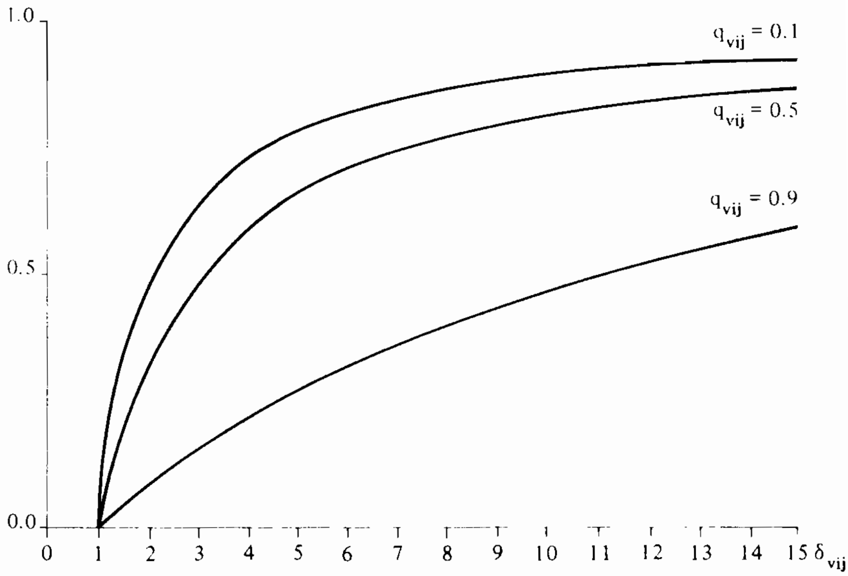


Abbildung 1.5.2. Die bedingte Wahrscheinlichkeit (1.5.14) als Funktion des Parameters $\delta_{vij} = \omega_{vij}^{-1}$ für verschiedene Werte von q_{vij}.

Die bedingte Wahrscheinlichkeit

$$(1.5.14) \qquad p\{z_{vi} = j+1 \mid z_{vi} > j\} = 1 - p\{z_{vi} > j+1 \mid z_{vi} > j\},$$

mit welcher die Person v im Zeitpunkt j + 1 aggressiv reagiert, nachdem sie ihre Aggressivität zuvor unterdrückt hatte, steht daher in einer streng monoton fallenden Abhängigkeit von ω_{vij} bzw. in einer streng monoton wachsenden Abhängigkeit von $\delta_{vij} = \omega_{vij}^{-1}$ (vgl. Abbildung 1.5.2). Der Parameter δ_{vij} soll daher im folgenden als der „Aggressivitätszuwachs" bezeichnet werden, welcher sich für die Person v aus der Unterdrückung ihrer Aggressivität im j-ten Zeitpunkt nach der i-ten Provokation ergibt.

Um zu zeigen, dass diese Interpretation gerechtfertigt ist, müssen wir noch nachweisen, dass δ_{vij} tatsächlich als Veränderung der Aggressivität der Person dargestellt werden kann. Zu diesem Zweck greifen wir auf die oben getroffene Annahme zurück, dass die aggressive Beantwortung einer Provokation den „Aggressionsstau" wieder ausgleichen soll, so dass die Aggressivität der Person nach der aggressiven Beantwortung der Provokation wieder ihren Ausgangswert annimmt und daher abgeschlossene Reaktionssequenzen einer Person als unabhängig voneinander betrachtet werden können. Aufgrund der Überlegungen, welche wir in den Abschnitten 1.2 und 1.3 angestellt haben, kann unsere Rede vom „Ausgangswert" der Aggressivität daher nur dann gerechtfertigt werden, wenn die Wahrscheinlichkeit, mit welcher die Person eine Provokation sofort (d.h. im Zeitpunkt j = 1) aggressiv beantwortet, in der Form

$$(1.5.15) \qquad p\{+ \mid v, i\} = \frac{\xi_v \epsilon_i}{1 + \xi_v \epsilon_i} \quad \text{und} \quad \begin{array}{l} \text{für } i = 1, \dots, k \\ \text{für } v = 1, \dots, n \end{array}$$

dargestellt werden kann, worin ξ_v die „Ausgangsaggressivität" der Person und ϵ_i den „Aggressionsanreiz" der Situation bezeichnet, in welcher sich die Person unmittelbar nach der Provokation befindet. ϵ_i wird im allgemeinen auch von der Art und Stärke der Provokation abhängen.

Einsetzen von (1.5.15) in (1.5.10–1.5.13) ergibt dann die Beziehungen

$$(1.5.16) \qquad q_{vij} = \begin{cases} 1 - p\{+ \mid v, i\} = \dfrac{\sigma_i}{\xi_v + \sigma_i} & \text{falls } j = 1 \\[2ex] (1 - p\{+ \mid v, i\}) \displaystyle\prod_{t=1}^{j-1} \frac{\omega_{vit}}{1 + (\omega_{vit} - 1) \cdot q_{vit}} & \text{falls } j > 1 \end{cases}$$

worin $\sigma_i = \epsilon_i^{\quad 1}$

und

$$(1.5.17) \qquad q_{vij} = \frac{(\xi_v \cdot \prod_{t<j} \delta_{vit})^{1} \cdot \sigma_i}{1 + (\xi_v \cdot \prod_{t<j} \delta_{vit})^{-1} \cdot \sigma_i} \qquad \text{für } j = 1, 2, \ldots$$

aufgrund derer δ_{vit} tatsächlich als Veränderung der Aggressivität der Person darstellbar ist. Dabei ist allerdings zu beachten, dass nur noch die „Ausgangsaggressivität" der Definition (1.2.5) genügt, während der Prädikator „aggressiver als" relativ zum j-ten Zeitpunkt nach einer Provokation durch

(1.5.18) {P_1 ist relativ zum j-ten Zeitpunkt nach der i-ten Provokation aggressiver als P_2} ⇔ {P_1 hat die i-te Provokation im j-ten Zeitpunkt danach mit grösserer Wahrscheinlichkeit bereits aggressiv beantwortet als P_2}

definiert ist. Damit unterscheidet sich die hier verwendete Definition des Konzepts der „Aggressivität" auch in grundsätzlicher Weise von der unter (1.4.3) formulierten Abschwächung der Definition (1.2.5). Die bedingte Wahrscheinlichkeit, mit welcher eine Person im j-ten Zeitpunkt nach einer Provokation aggressiv handelt, nachdem sie ihre Aggressivität zuvor unterdrückt hatte, ist im Modell (1.5.17) nicht nur eine streng monoton wachsende Funktion der „Ausgangsaggressivität" der Person, sondern auch eine streng monoton wachsende Funktion der seit der Provokation aufgestauten, aggressiven Energie $\prod_{t<j} \delta_{vit}$. Wir sehen also, dass die Frage nach einer adäquaten Definition des Begriffs der „Aggressivität" in wesentlicher Weise davon abhängig ist, welchen Gesetzmässigkeiten die Katharsis folgt.

Damit eine methodisch befriedigende Schätzung der Modellparameter in (1.5.17) möglich wird, müssen allerdings noch gewisse Restriktionen eingeführt werden, d.h. es muss vorausgesetzt werden, dass der „Aggressivitätszuwachs" δ_{vij} in zwei multiplikativ verknüpfte Faktoren aufgespalten werden kann, von denen einer (α_{vj}) von i unabhängig ist, während der andere (β_{ij}) nicht von v abhängt.

$$(1.5.18) \qquad \delta_{vij} = \alpha_{vj} \cdot \beta_{ij} \qquad \text{für } \begin{array}{l} i = 1, \ldots, k \\ v = 1, \ldots, n \\ j = 1, 2, \ldots \end{array} \quad .$$

Mit den Definitionen

$$(1.5.19)\qquad \theta_{vj} = (\xi_v \cdot \prod_{t<j} \alpha_{vj})^{-1}$$

und

$$(1.5.20)\qquad \eta_{ij} = \sigma_i \cdot \prod_{t<j} \beta_{ij}$$

folgt dann aus (1.5.17) die Beziehung

$$(1.5.21)\qquad q_{vij} = \frac{\theta_{vj}\eta_{ij}}{1+\theta_{vj}\eta_{ij}} \qquad \text{für } \begin{array}{l} i = 1, \dots, k \\ v = 1, \dots, n \\ j = 1, 2, \dots \end{array}$$

und unter Verwendung der Notation

$$(1.5.22)\qquad x_{vi}^{(j)} = \begin{cases} 1 \text{ wenn } z_{vi} > j \\ 0 \text{ wenn } z_{vi} \leqslant j \end{cases}$$

können die Reaktionssequenzen von n Probanden auf k Provokationen durch eine Sequenz von (n, k)-Matrizen

$$((x_{vi}^{(1)})), ((x_{vi}^{(2)})), \dots, ((x_{vi}^{(j)})), \dots$$

beschrieben werden, wobei alle Beobachtungen $x_{vi}^{(j)}$ innerhalb jeder Datenmatrix $((x_{vi}^{(j)}))$ voneinander lokal unabhängig sind, während jede Zeile $(x_{vi}^{(j)})$ einer gegebenen Datenmatrix von den entsprechenden Zeilen der übrigen Datenmatrizen lokal abhängig ist. Das Modell (1.5.17–1.5.18) erlaubt es somit, die Datenmatrizen $((x_{vi}^{(j)}))$ getrennt voneinander mithilfe bedingter Inferenzmethoden zu analysieren und anschliessend die Parameterschätzungen $\hat{\theta}_{vj}$ und $\hat{\eta}_{ij}$ nach der Methode der kleinsten Quadrate in ihre einzelnen Komponenten zu zerlegen. Denn aufgrund der Beziehung (1.5.21) kann jede einzelne der Datenmatrizen $((x_{vi}^{(j)}))$ durch ein Raschmodell der Form (1.3.5) beschrieben werden.

Literatur

ANDERSEN, E. B.: Asymptotic properties of conditional maximum likelihood estimators. Journ. Royal. Stat. Soc., B *32*, 283, 1970.

The asymptotic distribution of conditional likelihood ratio tests. Journ. Amer. Stat. Ass. *66*, 630, 1971.

A computer program for solving a set of conditional maximum likelihood equations arising in the Rasch model for questionnaires. Research Memorandum, Educational Testing Service, Princeton 1972.

Conditional inference and models for measuring. Mentalhygiejnisk Forlag, Copenhagen 1973a.

Conditional inference for multiple-choice questionnaires. British Journ. Math. Stat. Psychol. *26*, 31, 1973b.

A goodness of fit test for the Rasch Model. Psychometrika *38*, 123, 1973c.

BIRNBAUM, A.: Some latent trait models and their use in inferring an examinee's ability. In: LORD, F. M., NOVICK, M. R.: Statistical theories of mental test scores. Addison-Wesley, Reading, Mass. 1968.

BOUDON, R.: Mathematische Modelle und Methoden. Ullstein, Frankfurt 1973.

BRAMEL, D., TAUB, B., BLUM, B.: On observer's reaction to the suffering of his enemy. Journal of Personality and Social Psychology *8*, 384, 1968.

BRANDT, U., KÖHLER, B.: Norm und Konformität. In: Handbuch der Psychologie, Band 7, Sozialpsychologie, 2. Halbband. Hogrefe, Göttingen 1971.

COLEMAN, J. S.: Introduction to mathematical sociology. The Free Press, Glencoe, Ill. 1964.

CRONBACH, L. J., GLESER, G. C.: Psychological tests and personnel decisions. Univ. of Illinois Press, Urbana 1965.

DANN, H. D.: Müssen Aggressionen ausgelebt werden? In: SCHMIDT-MUMMENDEY, A., SCHMIDT, H. D. (Hrsg.): Aggressives Verhalten. Juventa, München 1971.

FESHBACH, S.: The catharsis hypothesis and some consequences of interaction with aggressive and neutral play objects. Journal of Personality *24*, 339, 1956.

FISCHER, G. H.: Neue Entwicklungen in der Psychologischen Testtheorie. In: FISCHER, G. H. (Hrsg.): Psychologische Testtheorie. Huber, Bern 1968.

Einige Gedanken über formalisierte psychologische Theorien. Psychol. Beiträge *13*, 376, 1971.

Conditional maximum-likelihood estimation of item parameters for a linear logistic test-model. Research Bulletin No. 9. Psychol. Inst. der Univ. Wien 1972.

FISCHER, G. H., ALLERUP, P.: Rechentechnische Fragen zu Rasch's eindimensionalem Modell. In: FISCHER, G. H. (Hrsg.): Psychologische Testtheorie. Huber, Bern 1968.

FISCHER, G. H., ROPPERT, J.: Monte-Carlo-Untersuchungen an einem Intelligenzmodell mit vier Faktoren. In: HECKHAUSEN, H. (Hrsg.): Bericht 24. Kongr. d. D. G. f. Psychol. Wien 1964. Hogrefe, Göttingen 1965.

FISCHER, G. H., SCHEIBLECHNER, H. H.: Algorithmen und Programme für das probabilistische Testmodell von Rasch. Psychol. Beiträge *12*, 23, 1970.

GUTTMAN, L.: The basis for Scalogramm-Analysis. In: STAUFFER, S. A. et al. (Eds.): Measurement and Prediction. Princeton Univ. Press, Princeton 1950.

HILKE, R., HÖLLBACHER, M., KEMPF, W. F.: Aggressives Handeln in einer Zwei-Personen-Interaktionssituation. Forschungsbericht 41 des Sonderforschungsbereichs 22. Sozialwissenschaftliches Forschungszentrum, Nürnberg 1973.

HOLZKAMP, K. Theorie und Experiment in der Psychologie. De Gruyter, Berlin 1964.
Zur Problematik der Realitätsverdoppelung in der Psychologie. Psychol. Rundschau *16*, 209, 1965.
Wissenschaft als Handlung. De Gruyter, Berlin 1968.
JÜTTEMANN, G. Was nützen Eigenschaftskonstrukte? Psychol. Rundschau *12*, 91, 1972.
KALVERAM, K. TH.. Über Faktorenanalyse. Archiv f. Psychol. *122*, 92, 1970a.
Probleme der Selektion in der Faktorenanalyse. III. Die „Invarianz" von Faktorenlösungen unter Selektion. Archiv f. Psychol. *122*, 223, 1970b.
KAUFMANN, H.. Definitions and methodology in the study of aggression. Psychol. Bull. *64*, 351, 1965.
KEMPF, W. F. Grundlagenuntersuchungen zum Thematischen Apperzeptionstest. Phil. Diss., Wien 1970.
Probabilistische Modelle experimentalpsychologischer Versuchssituationen. Psychol. Beiträge *14*, 16, 1972a.
Zur Bewertung der Faktorenanalyse als psychologische Methode. Psychol. Beiträge *14*, 610, 1972b.
Basisprobleme der Diagnostik der Aggressivität. In: ECKENSBERGER, L. (Hrsg.): Bericht 28. Kongr. d. D. G. f. Psychol. Saarbrücken 1972. Hogrefe, Göttingen 1974 (im Druck).
KEMPF, W. F., HAMPAPA, P.. The numerical solution of a set of conditional estimation equations arising in a dynamic test model. British Journ. Math. Stat. Psychol. 1974 (in Vorbereitung).
KEMPF, W. F., MACH, G.. Unveröffentlichtes Computerprogramm zur CML-Parameterschätzung in einem dynamischen Testmodell. Institut für die Pädagogik der Naturwissenschaften (IPN), Kiel 1974.
KRISTOF, W. Einige Skalenfragen. In: FISCHER, G. H. (Hrsg.): Psychologische Testtheorie. Huber, Bern 1968.
LAUTMANN, R. Wert und Norm. Begriffsanalysen für die Soziologie. In: Beiträge zur sozialen Forschung *5*. Westdt. Verlag, Opladen 1971.
LUCE, R. D. Individual choice behavior. Wiley, New York 1959.
MERZ, F. Über die Erfassung aggressiver Einstellungen mit Hilfe von Fragebogen. Psychol. Beiträge *5*, 402, 1960.
MICKO, H. C. Eine Verallgemeinerung des Messmodells von Rasch mit einer Anwendung auf die Psychophysik der Reaktionen. Psychol. Beiträge *12*, 4, 1970.
NEYMANN, J., SCOTT, E. L.: Consistent estimates based on partially consistent observations. Econometrica *16*, 1, 1948.
PFANZAGL, J.. Theory of measurement. Physika Verlag, Würzburg 1968.
RASCH, G.. Probabilistic models for some intelligence and attainment test. Nielson & Lydiche, Copenhagen 1960.
On general laws and the meaning of measurement in psychology. Proceedings of the Fourth Berkely Symposion on Mathematical Statistics and Probability *5*, 321, 1961.
Kolloquium über Messmodelle. Hektographiertes Vorlesungsmanuskript, hrsg. von J. STENE, übersetzt von B. REPP, 1965.
– An individualistic approach to item analysis. In: LAZARSFELD, P. F., HENRY, N. W. (Eds.): Readings in mathematical social science. Science Research Associates, Chicago 1966a.

An individual-centered approach to item analysis with two categories of answers. Proceedings of the NUFFIC International Summer Session in Science at „Het Oude Hof", The Hague, 14. - 18. July, 1966b.

An item analysis which takes individual differences into account. British Journ. Math. Stat. Psychol. *19*, 49, 1966c.

ROHRACHER, H.. Einführung in die Psychologie. Urban & Schwarzenberg, Wien 1963 (8. Auflage).

SCANDURA, J. M.. Structural learning I: The theory and empirical research. Gordon & Breach, New York 1973.

SCHEIBLECHNER, H.. The separation of individual- and system-influences on behavior in social contexts. Acta Psychologica *35*, 442, 1971a.

A simple algorithm for CML-parameter-estimation in Rasch's probabilistic measurment model with two or more categories of answers. Research Bulletin No. 5, Psychol. Inst. der Univ. Wien, 1971b.

SELG, H.. Diagnostik der Aggressivität. Hogrefe, Göttingen 1968.

SUPPES, P., ZINNES, J. L.. Basic measurement theory. In: LUCE, R. D., BUSH, R. R., GALANTER, E. (Eds.): Handbook of Mathematical Psychology. Vol. I. Wiley, New York 1963.

TRAXEL, W. Einführung in die Methodik der Psychologie. Huber, Bern 1964.