

## Drei Wege zu mehr Konsistenz in der Selbstbeschreibung: Theoriepräzisierung, Korrespondenzbildung und Datenaggregation<sup>a</sup>

MANFRED SCHMITT, CLAUDIA DALBERT & LEO MONTADA  
Universität Trier

**Zusammenfassung:** Im Rahmen der Konsistenzdebatte in der Differentiellen Psychologie wurde eine Reihe von Vorschlägen gemacht, wie sich die Annahme der Konsistenz interindividueller Unterschiede aufrechterhalten bzw. fairer prüfen lasse. Unter anderem wurde behauptet, die Prüfung der Konsistenzannahme setze die Verwendung (a) inhaltlich korrespondenter und (b) sich theoretisch nahestehender Prädiktoren und Kriterien sowie (c) deren Messung auf Aggregatniveau voraus.

An Datensätzen aus (1) einer Untersuchung interpersonaler Verantwortlichkeit erwachsener Töchter ihren Müttern gegenüber ( $N = 673$ ) und (2) einer Untersuchung existentieller Schuld ( $N = 340$ ) wurde die Wirksamkeit dieser drei Maßnahmen überprüft.

In beiden Untersuchungen wurde Konsistenz als Vorhersagbarkeit einer Kriteriumsvariable aus Prädiktoren, die bezüglich ihrer theoretischen Bezugsnähe und ihrer inhaltlichen Korrespondenz zum Kriterium variiert wurden, konzipiert. Zusätzlich zur theoretischen Bezugsnähe und inhaltlichen Korrespondenz wurde das Aggregatniveau von Kriterium und Prädiktoren variiert.

Alle drei vermuteten Konsistenzquellen konnten nachgewiesen werden. In Untersuchung (1) konnte der stärkste Vorhersagegewinn durch die Herstellung inhaltlicher Korrespondenz erzielt werden, in Untersuchung (2) erwies sich die Datenaggregation (= Erhöhung der Meßwertzuverlässigkeit) als besonders wirkungsvoll.

**Abstract:** During the debate on the consistency of personality and behavior, various suggestions were made how to retain or test more fairly the consistency assumption. Among other means to increase fairness in testing the consistency assumption, it was suggested to use predictor variables (a) that correspond in contents and (b) that are theoretically proximate to the criterion variables. Furthermore, it was suggested to enhance reliability of both predictors and criteria via aggregation.

The effects of these three measures to increase consistency were examined with data from (1) a study on interpersonal responsibility of adult daughters towards their mothers ( $N = 673$ ) and (2) a study on existential guilt ( $N = 340$ ).

In both studies, consistency was defined as the predictability of a criterion by predictor variables that varied with respect to their theoretical proximity and the degree of contents correspondence to the criterion. In addition, criterion and predictors were measured at various levels of aggregation.

Each of the three strategies assumed to enhance predictability was successful. In study (1), the strongest increase in predictability was achieved by having the contents of the predictors and the criterion correspond to each other. In study (2), aggregation had the largest effect on the predictability of the criterion by the predictor variables.

### 1. Einleitung

Die Frage, ob die Annahme der Konsistenz von Eigenschaften und Einstellungen berechtigt ist, hat die persönlichkeitspsychologische Grundsatzdebatte der siebziger Jahre beherrscht und bis heute nichts von ihrer Virulenz verloren. Die pessimistischen Thesen MISCHELS (1968) und die optimistische Sicht von BEM & FUNDER (1978) sind markante Extrempositionen in dieser De-

batte, die differenzierte konzeptuelle und methodische Klärungen brachte. Ihr wichtigstes Ergebnis besteht vermutlich in einer angemessenen Differenzierung der Erwartungen, in welchen Fällen mit mehr, in welchen mit weniger Konsistenz zu rechnen ist. Deshalb geht es heute nicht mehr um die prinzipielle Berechtigung der Konsistenzannahme, sondern um empirisch prüfbare Hypothesen über Bedingungen für Konsistenz und deren angemessene Berücksichtigung in Theoriebildung und Forschungspraxis.

Gemeint ist mit Konsistenz in der Regel die Erhaltung interindividueller Unterschiede be-

<sup>a</sup> Die Untersuchung wurde durch eine Sachbeihilfe der Stiftung Volkswagenwerk unterstützt.

züglich zweier oder mehrerer Variablen. Welche Variablen auf Konsistenz geprüft werden, welchen Status die Variablen haben (endogen, exogen, manifest, latent) und wie sie gemessen werden, hängt von der Fragestellung ab. Entsprechend zahlreich und heterogen sind die Konfigurationen von Variablen, die in Konsistenzprüfungen zueinander in Bezug gesetzt wurden: dieselbe Variable zu zwei oder mehr Zeitpunkten; SelbstbeschreibungsvARIABLEN (Persönlichkeits-, Einstellungs-, Stimmungsinventare etc.) zueinander oder zu analogen Fremdbeschreibungen oder zu offenem Verhalten; offenes Verhalten in einer Situation zu offenem Verhalten in einer anderen Situation; hypothetisches Verhalten (z.B. Reaktion auf Situationsschilderungen in einem experimentellen Fragebogen) zu hypothetischem oder offenem Verhalten in einer ähnlichen oder in einer anderen Situation u. a. m.

Im vorliegenden Beitrag geht es nicht um Konstruktvalidierungen, sondern um einen konsistenztheoretisch begründeten Vergleich verschiedener Prädiktoren hinsichtlich ihrer Vorhersagegüte bezüglich eines Kriteriums. Auf der Basis von SelbstbeschreibungsvARIABLEN wird untersucht, wie sich (1) theoretisch begründete Nähe zwischen Prädiktor(en) und Kriterium, (2) inhaltliche Korrespondenz zwischen Prädiktor(en) und Kriterium und (3) Aggregatniveau der Messung von Prädiktor(en) und Kriterium auf die Präziserbarkeit des Kriteriums auswirken. Bezogen auf dieses Erkenntnisinteresse läßt sich die Wahl der Variablen und ihrer Operationalisierung als Selbstbeschreibungen begründen. Die Daten sind zwei Untersuchungen aus dem Bereich der Moralforschung entnommen, einem Forschungszweig, der seit den Arbeiten von HARTSHORNE & MAY (1928) in der Geschichte der Diskussion um Situationsspezifität von Verhalten bzw. um transsituative Konsistenz von Eigenschaften einen festen Platz hat.

## 2. Perspektiven in der Konsistenzdebatte

Ausgehend von empirischen Befunden, die lediglich geringe intersituative Konsistenz von Verhaltensunterschieden zwischen Personen aufwiesen, hat man sich die Frage gestellt, ob

und wie die Annahme generalisierter Persönlichkeitsmerkmale als Verhaltensursachen begründet sei. Tatsächlich handelt es sich um eine anthropologische Vorannahme, die nicht mehr und nicht weniger an Plausibilität beanspruchen kann als die alternative Erwartung, daß Person- und Situationsmerkmale in Interaktion treten, d.h. daß sich Situationsunterschiede bei unterschiedlichen Personen unterschiedlich im Verhalten auswirken.

Der empirische Nachweis von Interaktionseffekten (BOWERS 1973; ENDLER 1982; ENDLER & MAGNUSSON 1976; MAGNUSSON & ENDLER 1977), der bereits an sich und aus einer Reihe von methodologischen Gründen nicht unproblematisch ist (ALKER 1972; GOLDING 1965; HECKHAUSEN 1980; OLWEUS 1976; SCHMITT 1980), stellt zunächst nur einen deskriptiven Beleg dafür dar, daß nicht alle Personen objektive Situationsunterschiede in gleicher Weise auffassen, verarbeiten bzw. auf objektive Situationsunterschiede in gleicher Weise reagieren. Mit deskriptiven Aussagen über Anteile der Verhaltensvarianz, die mit der Interaktion von Person- und Situationsunterschieden «erklärbar» sind, darf man sich nicht begnügen. Stattdessen müssen theoretisch befriedigende Deutungen der beschriebenen Interaktionen versucht werden. Nur so lassen sich Alternativen zu Konsistenzhypothesen, die auf generalisierten Personeneigenschaften aufbauen, entwickeln.

Dies bedeutet theoriegeleitete Suche nach Konsistenzmoderatoren aus der grundlegenden Annahme «differentieller Reliabilität und Validität von Meßinstrumenten» bzw. «differentieller Diagnostizierbarkeit» und «differentieller Vorhersagbarkeit» von Eigenschaften, Leistungen oder allgemein Verhalten (z.B. AMELANG & BARTUSSEK 1971; AMELANG & BORKENAU 1982, 1984; GHISELLI 1963; JÄGER 1979; ZEDECK 1971). In die Konsistenzdebatte wurde der Moderatorvariablenansatz explizit von ALKER (1972), BEM (1972) und BEM & ALLEN (1974) eingebracht. Als konsistenz erhöhende Moderatoren wurden u.a. vorgeschlagen: selbsteingeschätzte Konsistenz (BEM & ALLEN 1974), ipsative Varianz (individuelle Variabilität über die Indikatoren des interessierenden Merkmals im Vergleich zu den Indikatoren anderer Merkmale; BEM & ALLEN 1974), Self-Monitoring (SNYDER 1982), Self-Consciousness (SCHEIER, BUSS & BUSS 1978),

Identity nach CHEEK & BRIGGS (1982), Empathie nach HOGAN (1969), Beobachtbarkeit des relevanten Merkmals (KENRICK & STRINGFIELD 1980), Ausmaß, zu dem ein Verhalten prototypisch für eine Verhaltensklasse bzw. für ein Persönlichkeitsmerkmal ist (MISCHEL & PEAKE 1982), Zentralität einer Eigenschaft/eines Verhaltens für das Selbstbild einer Person (CHEEK 1982), Selbstaufmerksamkeit (GIBBONS 1978; PRYOR, GIBBONS, WICKLUND, FAZIO & HOOD 1977), Ausmaß, zu dem sich Selbstbeschreibungsmarkmale (z.B. Einstellungen) auf früheres Verhalten stützen (FAZIO & ZANNA 1981). Die Brauchbarkeit dieser Moderatoren zur Differenzierung konsistenter und inkonsistenter Personen wird allerdings unterschiedlich beurteilt; die Befundlage ist uneinheitlich und zum Teil widersprüchlich (vgl. AMELANG & BORKENAU 1982, 1984; CHEEK 1982; KENRICK & STRINGFIELD 1980; SCHLEGEL & DITECCO 1982; TELLEGEN, KAMP & WATSON 1982; ZANNA, OLSON & FAZIO 1980).

Der Suche nach Konsistenzmoderatoren sollte allerdings der Nachweis vorgeschaltet sein, daß eine generalisierte Konsistenzannahme unhaltbar ist. Daß dieser Nachweis nicht immer mit der nötigen Sorgfalt geführt wird und daß die Konsistenzannahme häufig voreilig verworfen wird, legen – mit unterschiedlicher Akzentsetzung – die Arbeiten von BEM & FUNDER (1978), AJZEN & FISHBEIN (1977) und EPSTEIN (z.B. 1979) nahe.

BEM & FUNDER (1978) können mit Hilfe ihrer «template-behavior-matching»-Technik zeigen, daß mit zunehmender funktionaler Entsprechung von Situation und Person die Güte von Verhaltensvorhersagen steigt. Diese Technik von BEM & FUNDER (1978) zur Bestimmung der «Persönlichkeit von Situationen» mittels Q-sort Profil (BLOCK 1961) wurzelt in der Idee der Reaktionskalibrierung von Umwelten (z.B. FREDERIKSEN 1972; vgl. auch ENDLER 1982) und kann als nützlicher Beitrag zu einer theoretisch begründeten Bezugsbildung zwischen Persönlichkeitsmerkmalen und Umweltmerkmalen verstanden werden. Verallgemeinernd läßt sich aus den Befunden von BEM & FUNDER (1978) die Erwartung ableiten, daß theoretische Nähe zwischen Prädiktor und Kriterium Konsistenz (mit)bedingt. Daß diese Erwartung nur auf den ersten Blick trivial ist, läßt sich mit MISCHEL (1983) belegen. Er erinnert an die vielfältigen Versuche der Per-

sonalauslese, der Prognose von Straf- und Rückfälligkeit, der klinischen Prognose mit Persönlichkeits-, Verhaltens- oder biographischen Daten, die in keinem theoretisch explizierten Bezug zum Kriterium stehen: «In its extreme forms, almost any indicator seemed to offer a royal road into the core personality.» (p. 579) In der vorliegenden Arbeit wird *theoretische Nähe zwischen Prädiktor und Kriterium* als einer von drei Konsistenzfaktoren untersucht.

Hier wird deutlich, daß der Begriff Nähe zu spezifizieren ist. Ein Prädiktor kann mit dem vorhergesagten Kriterium «identisch» sein (Merkmalsstabilität), er kann Teil eines Syndroms, Voraussetzung für ein Kriterium usw. sein. Wie theoretisch nahe oder fern ein Prädiktor dem Kriterium ist, läßt sich nur aus einem Zusammenhangsmodell ersehen, das, wenn keine absoluten, so doch relative Distanzen zweier oder mehrerer Prädiktoren zu einem Kriterium expliziert.

Von theoretischer Nähe ist *inhaltliche Korrespondenz von Prädiktor und Kriterium* zu unterscheiden. AJZEN & FISHBEIN (1977) haben mit einer aufwendigen Metaanalyse der Forschungsbefunde zum Einstellungs-Verhaltens-Bezug zeigen können, daß WICKERS (1969) pessimistische Schlußfolgerung, Verhaltensvarianz sei in der Regel nur bis zu 10% mit Einstellungsvarianz zu erklären, in dieser allgemeinen Form nicht aufrechterhalten werden kann. Vielmehr hängt die Enge des Einstellungs-Verhaltens-Bezugs maßgeblich vom Ausmaß ihrer inhaltlichen Korrespondenz ab. Will man beispielsweise Rauschmittelkonsum vorhersagen, ist es weniger sinnvoll, die Einstellung (im Sinne einer evaluativen Stellungnahme oder affektiven Reaktion) gegenüber dem Rauschmittel selbst zu erfragen als gegenüber dem Rauschmittelkonsum. In dieselbe Richtung weisen Vorschläge, zur Verhaltensvorhersage nicht auf globale, sondern auf bereichsspezifische Eigenschaftsmaße zurückzugreifen (in der Angstforschung z.B. ZUCKERMAN 1976). Ganz allgemein scheint zu gelten: «Empirical correlations among measures increase when contents of the measures correspond and overlap as much as possible» (MISCHEL 1983, p. 583). Nach theoretischer Nähe zwischen Prädiktor und Kriterium ist inhaltliche Korrespondenz der zweite Konsistenzfaktor, der in dieser Arbeit untersucht wird.

Neben theoretischer Bezugsnähe und inhaltlicher Korrespondenz von Prädiktor- und Kriteriumsvariablen wirkt sich die Verwendung aggregierter Maße zusammenhangssteigernd aus. Es ist das Verdienst EPSTEINS (1979, 1980, 1983a), die alte Erkenntnis der Testkonstruktion, daß Meßwiederholung/Testverlängerung zuverlässigkeitssteigernd wirkt, in Erinnerung gerufen und in die Konsistenzdebatte eingebracht zu haben. EPSTEIN meint, Voraussetzung zur fairen und gültigen Testung der Konsistenz bzw. Inkonsistenzhypothese sei nicht nur ein zuverlässig gemessener Prädiktor – eine Voraussetzung, die angesichts sorgfältig konstruierter Einstellungs- und Persönlichkeitsskalen häufig erfüllt ist –, sondern auch ein zuverlässig gemessenes Kriterium. Als Mittel zur Gewinnung zuverlässiger Maße schlägt er die Datenaggregation vor. Über welche Einheiten aggregiert werden soll oder muß (Personen, Situationen, Zeitpunkte, Meßmethoden, Untersuchungen etc.; vgl. SCHWENKMEZGER 1984), hängt von der Fragestellung ab. Diese Abhängigkeit des Aggregattyps von der Fragestellung scheint nicht immer hinreichend bedacht worden zu sein; anders lassen sich einige Mißverständnisse und überflüssige Kontroversen kaum erklären (EPSTEIN 1983b; FUNDER 1983; MISCHEL & PEAKE 1982). Varianzquellen, die für eine Fragestellung von höchster Bedeutung sind, mögen in der Untersuchung einer anderen Fragestellung vernachlässigbar sein, als «Fehlervarianz» betrachtet werden, ja ihre Vernachlässigung mag sogar zur Reliabilitätssteigerung angezeigt sein. Das Aggregieren von Daten ist nicht per se richtig oder falsch, sondern je nach Fragestellung sinnvoll oder unsinnig. Inzwischen gibt es eine Reihe von Untersuchungen, die MISCHELS (1968) düsteres Bild von der Angemessenheit der Konsistenzannahme durch die Erhebung sinnvoll aggregierbarer und aggregierter Daten aufhellen (z. B. CHEEK 1982; EPSTEIN 1979, 1980; MISCHEL & PEAKE 1982; RUSHTON, BRAINERD & PRESSLEY 1983). Auch WICKERS (1969) Verneinung der Existenz eines praktisch und nicht nur statistisch bedeutsamen Einstellungs-Verhaltens-Bezuges wird durch die Befunde von FISHBEIN & AJZEN (1974) und WEIGEL & NEWMAN (1976), die mit aggregierten Verhaltenskriterien arbeiteten, korrigiert. Neben theoretischer Nähe und inhaltlicher Korrespondenz wird hier als dritter

Konsistenzfaktor das *Aggregatniveau*, auf dem die Variablen repräsentiert sind, untersucht.

### 3. Konsistenzhypothesen

In allgemeiner Form lauten die drei Hypothesen, die die Existenz der hier untersuchten Konsistenzfaktoren behaupten:

*H1:* Je größer die theoretische Nähe zwischen Prädiktoren und Kriterium, desto enger ist der statistische Zusammenhang zwischen Prädiktor- und Kriteriumsmaßen.

*H2:* Je größer die inhaltliche Korrespondenz zwischen Prädiktoren und Kriterium, desto enger ist der statistische Zusammenhang zwischen Prädiktor- und Kriteriumsmaßen.

*H3:* Je höher das Aggregatniveau von Prädiktor- und Kriteriumsmaßen, desto enger ist der statistische Zusammenhang zwischen ihnen.

Diese Hypothesen sollen an Selbstbeschreibungsdaten aus zwei Untersuchungen zur Thematik «interpersonale Verantwortlichkeit und interpersonale Schuld» (MONTADA 1981) geprüft werden. Zur Analyse der drei Konsistenzquellen muß ihre Konfundierung vermieden werden, z. B. müssen zur Prüfung des Effektes inhaltlicher Korrespondenz das Aggregatniveau und die theoretische Bezugsnähe von Prädiktoren und Kriterium konstant gehalten werden. Diese Voraussetzungen sind durch die Struktur der Daten aus den beiden Untersuchungen relativ gut erfüllbar. In Fällen, in denen die Forderung nach Unabhängigkeit der drei Konsistenzfaktoren verletzt ist, erlauben die Daten einen konservativen Test.

### 4. Untersuchung 1: Verantwortlichkeit erwachsener Töchter ihren Müttern gegenüber

Das erste Datenbeispiel ist einer umfangreichen Untersuchungsreihe über den Umgang erwachsener Töchter mit Wünschen und Bedürfnissen ihrer Mütter entnommen, in der von mutmaßlichen Voraussetzungen über Handlungsentscheidungen bis zu emotionalen Handlungsbewertungen und deren Bewältigung der komplexe Ablauf prosozialen Handelns in mehr als 20 Variablen repräsentiert sein soll. Konzeptbildung und

Hypothesen sind in MONTADA, DALBERT & SCHMITT (1982), Meßinstrumente und ihre statistischen Eigenschaften in SCHMITT, DALBERT & MONTADA (1983) beschrieben. Die folgende Darstellung ist auf die zum Verständnis dieses Beitrages wesentlichen Informationen beschränkt.

#### 4.1 Stichprobe

Eine Zufallsstichprobe von 673 erwachsenen Töchtern, deren Mütter noch leben, wurde aus einer nach drei Variablen stratifizierten Population gewonnen: (a) *geographische Region* (Stadt Trier; ländliche Gemeinden im Umkreis von Trier), (b) *Lebensalter* (Geburtsjahrgänge Kohorte 1: 1957–1962; Kohorte 2: 1947–1952; Kohorte 3: 1929–1934) und (c) *Familienstand* (alleinstehend; verheiratet oder mit Partner zusammenlebend). Die älteste Kohorte ist deutlich unterrepräsentiert ( $n = 74$ ), weil in dieser Alterspopulation die Mütter häufig bereits verstorben sind.

Der Altersmittelwert der Untersuchungsstichprobe beträgt 28,9 Jahre (Streuung: 18 bis 59), die Schulbildung weist einen erhöhten Anteil von Probandinnen mit höherer Bildung auf: Hauptschule oder Hauptschulabschluß ( $n = 252$ ); Realschule oder mittlere Reife ( $n = 183$ ); Abitur, Hochschule oder Hochschulabschluß ( $n = 273$ ); keine Angaben ( $n = 1$ ).

#### 4.2 Erhebungsinstrumente und -modalitäten

Alle in diesem Beitrag erwähnten Variablen beziehen sich inhaltlich auf eine Liste mit 34 Wünschen und Bedürfnissen von Müttern (z. B. nach Kontakt, nach Teilhabe am Leben der Tochter, nach konkreten Hilfen, nach Rücksicht der Tochter in Wertfragen usw.), für deren Erfüllung sich eine Tochter in unterschiedlichem Maße verantwortlich fühlen kann.

Ein Teil der Variablen wurde nicht zu allen 34 Bedürfnissen erhoben. Vielmehr hatte jede Probandin an einem bestimmten Punkt der Untersuchung fünf der Mutter mutmaßlich wichtige Bedürfnisse oder Wünsche aus der Liste auszuwählen. Mit einer Ausnahme (Näheres unter Abschnitt 6) beziehen sich alle hier berichteten Daten nur auf diese fünf, von jeder Probandin individuell ausgewählten Bedürfnisse oder Wünsche ihrer Mutter.

In dieser Arbeit sind nur die folgenden acht Variablen genannt, die am Beispiel des Bedürfnisses Nr. 12 der genannten Liste (= Bedürfnis nach Pflege bei vorübergehender Krankheit) geschildert werden; die Antwortskalen sind jeweils sechsstufige Schätzskalen:

- *EI* (Einstellungen). *Beispielitem:* «Ich finde es richtig, wenn eine erwachsene Tochter ihre vorübergehend kranke Mutter pflegt.» (1 = stimmt genau.../6 = stimmt überhaupt nicht)
- *PN* (Personale Normen). *Beispielitem:* «Prinzipiell fühle ich mich verpflichtet, meine Mutter zu pflegen, wenn sie vorübergehend krank ist.» (1 = stimmt genau.../6 = stimmt überhaupt nicht)

- *VE* (Verhaltenserfahrung). *Beispielitem:* «In der Vergangenheit – das heißt bis heute – habe ich (1 = ständig.../6 = nie) meine Mutter gepflegt, wenn sie vorübergehend krank war.»
- *BM* (Beeinträchtigung der Mutter). *Beispielitem:* «Wie stark leidet Ihre Mutter, wenn ihrem Bedürfnis oder Wunsch Nr. 12 (...) nicht nachgekommen wird (oder nicht nachgekommen würde)?» (1 = sehr stark.../6 = überhaupt nicht)
- *BB* (Berechtigung des Bedürfnisses). *Beispielitem:* «Das Bedürfnis oder den Wunsch Nr. 12 (...) meiner Mutter finde ich (1 = völlig berechtigt.../6 = überhaupt nicht berechtigt).»
- *VN* (Verantwortung für die Notlage). *Beispielitem:* «Meine Mutter hätte sich anders verhalten sollen (oder sollte sich anders verhalten), dann brauchte sie das Bedürfnis oder den Wunsch Nr. 12 (...) nicht zu haben.» (1 = stimmt genau.../6 = stimmt überhaupt nicht)
- *A* (Absicht). *Beispielitem:* «Ich will dem Bedürfnis oder Wunsch Nr. 12 (...) meiner Mutter (1 = voll und ganz.../6 = überhaupt nicht) nachkommen.»
- *H* (Handlungsausführung). *Beispielitem:* «Ich bin dem Bedürfnis oder Wunsch Nr. 12 (...) meiner Mutter (1 = voll und ganz.../6 = überhaupt nicht) nachgekommen.»

Die Erhebung dieser acht Variablen mittels Fragebogen erfolgte postalisch. Die Variablen wurden zu vier Zeitpunkten, zwischen denen jeweils mindestens drei Wochen lagen, erhoben (*EI*, *PN*, *VE* zu  $t_1$ ; *BM*, *BB*, *VN* zu  $t_2$ ; *A* zu  $t_3$ ; *H* zu  $t_4$ ).

### 4.3 Konkretisierung und Testung der Konsistenzhypothesen

#### 4.3.1 Theoretische Nähe

Zur Prüfung des Einflusses der theoretischen Nähe auf Konsistenz können die beiden Prädiktoren Personale Normen (*PN*) und Absicht (*A*) in bezug auf die Genauigkeit der Vorhersage des Kriteriums Handlungsausführung (*H*) miteinander verglichen werden. Die Annahme unterschiedlicher theoretischer Distanz dieser beiden Prädiktoren zum Kriterium ergibt sich aus dem Argument, daß die Personalen Normen generelle, d. h. vom konkreten Handlungskontext unabhängige Sollsetzungen darstellen, während die Absicht zu handeln das Ergebnis einer Synthese aus überdauernden Personmerkmalen (z. B. auch Personalen Normen) und der Analy-

1 Die Probandinnen sollten zwecks Fehlerkontrolle und Fokussierung an dieser Stelle den Wortlaut des Bedürfnisses ihrer Mutter eintragen.

se der spezifischen Situation (hier z. B.: wahrgenommene Beeinträchtigung der Mutter, Einschätzung der Berechtigung des Bedürfnisses der Mutter, wahrgenommene Selbstverschuldung des Bedürfnisses durch die Mutter) darstellt (vgl. AJZEN & FISHBEIN 1980; DALBERT 1980).

Aus diesen Überlegungen und der Forderung nach Entflechtung der drei Konsistenzquellen leitet sich folgende Hypothese ab:

**H1.1:** Bei gleicher inhaltlicher Korrespondenz und gleichem Aggregatniveau ist die Korrelation von *PN* und *H* niedriger als die Korrelation von *A* und *H*.

Die inhaltliche Korrespondenz wird konstant gehalten dadurch, daß sich alle drei Variablen jeweils auf dasselbe ausgewählte Bedürfnis beziehen<sup>2</sup>. Das Aggregatniveau wird konstant gehalten, indem die zu vergleichenden Korrelationskoeffizienten entweder über die Meßwerte zu einem spezifischen Bedürfnis bestimmt werden oder über aggregierte Meßwerte (individuelle Mittelwerte der fünf bedürfnisspezifischen Meßwerte).

Die Daten stehen mit H1.1 im Einklang: Die bedürfnisspezifische Korrelation zwischen *PN* und *H* beträgt im Mittel  $\bar{r} = .30$ , während die mittlere Korrelation zwischen *A* und *H*  $\bar{r} = .51$  ist<sup>3</sup>. Die erwartete Differenz zwischen den Korrelationen zeigt sich auch, wenn sie auf aggregierten Meßwerten beruhen ( $r = .46$  versus  $r = .62$ ). Die Alternativhypothese, daß dieser Befund einen Meßwiederholungseffekt/Meßwiederholungssimplex (*A* wurde zeitlich näher zu *H* erhoben als *PN*) widerspiegelt, kann ausgeschlossen werden: Andere, zum Kriterium theoretisch distale, aber zeitlich nah erfaßte Prädiktoren, über die hier nicht berichtet werden kann, korrelieren mit dem Kriterium erwartungsgemäß niedriger als theoretisch nahe, aber zeitlich distale Prädiktoren.

2 Zwar impliziert die individuelle Auswahl von Bedürfnissen eine Konfundierung von Inhalt und Person. Diese Konfundierung ist für die hier behandelte Fragestellung aber irrelevant, da weder differentielle Konsistenzeffekte für unterschiedliche Inhalte (hier: Bedürfniskategorien) noch differentielle Konsistenzeffekte für unterschiedliche Personen zur Debatte stehen.

3 Alle Mittelungen von Korrelationen erfolgten über FISHERS Z-Transformation (vgl. auch die Anmerkung am Ende des Beitrags).

#### 4.3.2 Inhaltliche Korrespondenz

Die inhaltliche Korrespondenz läßt sich zur Prüfung von H2 variieren, indem die Prädiktoren sich entweder auf dasselbe Bedürfnis beziehen wie das Kriterium (inhaltliche Korrespondenz gegeben) oder aber auf andere Bedürfnisse (inhaltliche Korrespondenz verletzt). Als Kriterium wird wieder die Absicht (*A*), als Prädiktoren die Einstellungen (*EI*), Personalen Normen (*PN*) und Verhaltenserfahrungen (*VE*), die sich als Prädiktoren der Handlungsintention bewährt haben (AJZEN & FISHBEIN 1980; BENTLER & SPECKART 1979; SCHWARTZ 1977), gewählt.

Die Forderung nach Variation inhaltlicher Korrespondenz unabhängig von Aggregatniveau und theoretischer Nähe ist dadurch erfüllt, daß die inhaltliche Korrespondenz derselben Variablen (= konstante theoretische Nähe) bedürfnisspezifisch (= konstantes Aggregatniveau) verändert wird.

**H2.1:** Bei konstantem Aggregatniveau ist die multiple Korrelation zwischen der regressionsanalytisch ermittelten Linearkombination der Prädiktoren *EI*, *PN*, *VE* und dem Kriterium *A* höher, wenn sich Prädiktoren und Kriterium auf dieselben Bedürfnisse beziehen als wenn sie sich auf unterschiedliche Bedürfnisse beziehen.

Auch mit dieser Hypothese stehen die Daten im Einklang: Der mittlere multiple Korrelationskoeffizient bei Verletzung inhaltlicher Korrespondenz beträgt  $\bar{R} = .20$ . Wird dagegen die inhaltliche Korrespondenz von Prädiktoren und Kriterium beachtet, d. h. beziehen sich Prädiktoren und Kriterium auf dasselbe und nicht auf ein anderes Bedürfnis der Mutter, ist der multiple Zusammenhang wesentlich höher und beträgt  $\bar{R} = .60$ .

#### 4.3.3 Aggregatniveau

Das Aggregatniveau läßt sich zur Prüfung von H3 variieren, indem die Variablen bedürfnisspezifisch versus über die fünf bedürfnisspezifischen Werte gemittelt gemessen werden. Inhaltliche Korrespondenz und theoretische Nähe werden nicht variiert. Dieselben Variablen beziehen sich im Fall spezifischer Meßwerte auf dasselbe ausgewählte Bedürfnis, im Fall aggregierter Meßwerte auf die fünf ausgewählten Bedürfnisse. Die Resultate der Prüfung von H1.1 (siehe oben) enthalten diesen Aggregierungseffekt bereits, er soll aber noch an einem weiteren Beispiel demonstriert werden.

Als Kriterium wird wieder die Absicht (*A*), als Prädiktoren die Einstellungen (*EI*), Personalen Normen (*PN*), Verhaltenserfahrungen (*VE*), die wahrgenommene Beeinträchtigung der Mutter (*BM*) gewählt sowie das Ausmaß, in dem das jeweilige Bedürfnis der Mutter als berechtigt beurteilt wird (*BB*) und in dem die Mutter für ihr Bedürfnis/ihre Notlage als selbstverantwortlich wahrgenommen wird (*VN*).

**H3.1:** Das Kriterium *A* wird sich genauer durch eine regressionsanalytisch ermittelte Linearkombination aus den Prädiktoren *EI*, *PN*, *VE*, *BM*, *BB*, *VN* vorhersagen lassen, wenn alle Variablen aggregiert über dieselben fünf ausgewählten Bedürfnisse gemessen werden, als wenn korrespondierende bedürfnisspezifische Meßwerte verwendet werden.

Die Daten stehen mit der Hypothese in Einklang: Der mittlere multiple Korrelationskoeffizient der fünf bedürfnisspezifisch gerechneten Regressionsanalysen beträgt  $\bar{R} = .68$ . Wird dagegen für alle Variablen (Prädiktoren und Kriterium) der individuelle Meßwert als Mittel über die fünf bedürfnisspezifischen Werte gebildet, läßt sich das Kriterium hypothesenkonform besser vorhersagen. Der entsprechende multiple Korrelationskoeffizient beträgt nunmehr  $R = .78$ .

### 5. Untersuchung 2: Umgehen mit eigenen Privilegien und Existentielle Schuld

In einer zweiten Untersuchung wurden Modi der Auseinandersetzung mit eigenen Privilegien, genauer: mit Unterschieden zwischen der eigenen günstigen Lebenslage und der schlechten Lebenslage anderer, analysiert. Angesichts der relativ schlechten Lebenslage anderer kann man mit existentiellen Schuldgefühlen reagieren, aber je nach Fall auch mit Stolz, Freude, Angst um die eigenen Privilegien oder Abwertung der Benachteiligten. In dieser Untersuchung wurde die Art der emotional/kognitiven Reaktion auf die Wahrnehmung relativer Be-

nachteiligungen in Abhängigkeit von Gerechtigkeitsüberzeugungen, sozialen Einstellungen, Tendenzen zur Verantwortlichkeitsabwehr, Kontrollierbarkeitsüberzeugungen u. a. m. analysiert. Näheres zu Konzeptexplikation und erste Ergebnisse der Konstruktvalidierung sind in MONTADA, DALBERT, REICHLE & SCHMITT (1985) und MONTADA, SCHMITT & DALBERT (1986) beschrieben, die statistischen Eigenschaften der Instrumente in DALBERT, MONTADA, SCHMITT & SCHNEIDER (1984). In dieser Arbeit wird nur ein kleiner Ausschnitt aus der Menge der erhobenen Variablen zur Prüfung der Konsistenzhypothesen berichtet.

#### 5.1 Stichprobe

Die Untersuchungsstichprobe setzt sich aus mehreren Teilstichproben zusammen. Neben einer Zufallsstichprobe aus einer altersmäßig (21–70 Jahre Lebensalter) und geographisch (Stadt Trier) begrenzten Population wurden Stichproben aus Gruppen und Vereinigungen gezogen, die durch ihre Ideologien, Aktivitäten und/oder Programme eine Einstellung erwarten lassen, die das Erleben existentieller Schuld unwahrscheinlich oder im Gegenteil wahrscheinlich macht. Die in diesem Beitrag verarbeiteten Analysen beziehen sich auf die Gesamtstichprobe von  $N = 340$ . Der Altersmittelwert beträgt 36.1 Jahre, die Streubreite reicht von 16 bis 72 Jahren. Höhere Bildung ist deutlich überrepräsentiert: kein Schulabschluß ( $n = 6$ ); Hauptschulabschluß ( $n = 68$ ); mittlere Reife ( $n = 69$ ); Abitur bis Hochschulabschluß ( $n = 194$ ); keine Angaben ( $n = 3$ ). Die Stichprobe setzt sich aus 129 Frauen und 210 Männern zusammen (eine Person nannte ihr Geschlecht nicht).

#### 5.2 Erhebungsinstrumente und -modalitäten

Alle Fragebögen zur Erhebung der hier berichteten Daten wurden von den Autoren selbst konstruiert. Mit einem experimentellen Fragebogen (Existentielle Schuld-Inventar) wurden neben existentieller Schuld mehrere kognitive und emotionale Reaktionen auf die Schilderung der relativen Benachteiligung dreier Gruppen unterprivilegierter Menschen – (a) Menschen in der dritten Welt, (b) Körperbehinderte und (c) türkische Gastarbeiter in der BRD – erhoben. Der Fragebogen enthält zu jeder Gruppe von Benachteiligten drei Situationsschilderungen, so daß jede Variable gruppenspezifisch dreimal, insgesamt neunmal gemessen wird. Hier wird nur Bezug genommen auf die beiden Variablen Abwertung der Benachteiligten (*AW*) und Notlagenattribution auf die Benachteiligten (*NA*) im Sinne eines Vorwurfs, die Notlage selbst verschuldet zu haben. Die Itemantwortungen wurden als Zustimmungen auf sechsstufigen Rating-Skalen erhoben.



Neben der Abwertung der Benachteiligten (*AW*) und dem Selbstverschuldungsvorwurf (*NA*) als Reaktionen auf die Schilderung einer konkreten Benachteiligungssituation wird nur noch die Variable Einstellungen gegenüber zwei Gruppen (türkische Gastarbeiter und Körperbehinderte) zur Prüfung der Konsistenzhypothese ausgewählt. Zur Messung von Einstellungen waren in Anlehnung an die Social Distance Scale von BOGARDUS (1925) zwei soziale Distanzskalen konstruiert worden. Den Skalen liegt die Konzeption von Einstellung als Verhaltensschwelle gegenüber dem Einstellungsobjekt zugrunde (vgl. TRIANDIS 1971), wobei die elf Items jeder Skala so formuliert wurden, daß sie eine interindividuell möglichst invariante Schwierigkeitsrangreihe im Sinne des GUTTMAN-Modells bilden. *Beispiel* für ein leichtes Item (niedrige Verhaltensschwelle): «Ich würde mich jederzeit mit einem türkischen Gastarbeiter/einer türkischen Gastarbeiterin unterhalten.» *Beispiel* für ein schweres Item (hohe Verhaltensschwelle): «Ich kann mir vorstellen, einen türkischen Gastarbeiter/eine türkische Gastarbeiterin zu heiraten.» Die Antwortskalen zu den Items sind sechsstufig mit den Polen «stimmt genau» und «stimmt überhaupt nicht»<sup>4</sup>. Die beiden Skalen zur Erfassung der Einstellung gegenüber Körperbehinderten (*EK*) und türkischen Gastarbeitern (*ET*) erzielten in unserer Stichprobe gute statistische Eigenschaften. Die Skalen sind intern konsistent (*ET*: Alpha = .92; *EK*: Alpha = .90) und sind, wenn die Antwortskalen zu den Items an der natürlichen Skalenmitte dichotomisiert werden, mit dem GUTTMAN-Modell kompatibel (*ET*: Reproduzierbarkeit = .92; *EK*: Reproduzierbarkeit = .98).

Die Untersuchung wurde postalisch durchgeführt und war in mehrere Erhebungszeitpunkte unterteilt. Die beiden Einstellungsskalen wurden mehrere Wochen nach dem Existentielle Schuld-Inventar zur Bearbeitung vorgegeben.

### 5.3 Konkretisierung und Testung der Konsistenzhypothesen

#### 5.3.1 Theoretische Nähe

Die Konsistenzhypothese 1 wird über den Vergleich der Vorhersagbarkeit des Kriteriums Abwertung der Benachteiligten (*AW*) durch den theoretisch nahen Prädiktor Notlagenattribution auf die Benachteiligten (*NA*) und den theoretisch distalen Prädiktor Einstellung gegenüber den Benachteiligten (*ET/EK*) geprüft.

Die Annahme, daß *NA* ein theoretisch näherer Prädiktor ist als soziale Einstellung oder so-

ziale Distanz, ist durch das Argument begründet, daß die Variable *NA* wie auch das Kriterium *AW* einen spezifischen Aspekt der Auseinandersetzung mit der Gerechtigkeit von Privilegunterschieden erfassen soll. Wer für die Erklärung der Privilegunterschiede zu Selbstverschuldungsvorwürfen tendiert, wird auch zur Abwertung der Benachteiligten neigen. Demgegenüber ist die Variable Einstellung als generelle soziale Distanz operationalisiert, die keinen direkten Bezug zu Auseinandersetzungen mit Privilegunterschieden hat. Zwar ist anzunehmen, daß eine größere soziale Distanz eine Abwertung der Benachteiligten erleichtert, ob es aber tatsächlich zu einer solchen Abwertung angesichts eines konkreten Privilegunterschiedes kommt, wird neben dem Ausmaß sozialer Distanz von weiteren Variablen, wie z. B. Präferenz von Gerechtigkeitsprinzipien, die eine Abwertung der Benachteiligten erleichtern (z. B. Leistungsprinzip) oder erschweren (z. B. Bedürfnisprinzip) abhängen. Je weniger zwingend aber ein Bezug zwischen Prädiktor und Kriterium ist, je mehr weitere Variablen (Mediatoren, Moderatoren) zur Vorhersage des Kriteriums hinzugezogen werden müssen, desto distaler ist der theoretische Bezug zwischen Prädiktor und Kriterium.

Aus diesen Überlegungen und der Forderung nach Entflechtung der drei Konsistenzquellen leitet sich folgende Hypothese ab:

*H1.2:* Bei gleicher inhaltlicher Korrespondenz und gleichem Aggregatniveau ist die Korrelation zwischen *NA* und *AW* höher als die zwischen *ET/EK* und *AW*.

Inhaltliche Korrespondenz ist als Bezug zweier Variablen auf dieselbe Gruppe von Benachteiligten gegeben (hier: türkische Gastarbeiter und Körperbehinderte). Sie wird für die zu vergleichenden Korrelationen konstant gehalten. Das Aggregatniveau ist ebenfalls so vergleichbar wie möglich gehalten. Die Meßwerte der Probanden auf den Variablen *NA* und *AW* sind Werte für spezifische Gruppen Benachteiligter, genauer: individuelle Mittelwerte über die Meßwerte zu den drei Situationen, in denen eine Benachteiligung der jeweiligen Gruppe geschildert ist. Den Meßwerten der Probanden auf den beiden Einstellungsskalen *ET* und *EK* liegen elf Items zugrunde; die Einstellungen sind also zuverlässiger gemessen als die nur über jeweils

drei Items aggregiert gemessenen Variablen *NA* und *AW*. Dies bedeutet, daß *H1.2* konservativ getestet wird.

Dennoch stützen die Daten die in *H1.2* formulierten Erwartungen. Die beiden gruppenspezifischen Korrelationen zwischen *AW* und *NA* betragen im Mittel  $\bar{r} = .61$ . Mit  $\bar{r} = .41$  sind die beiden Korrelationen zwischen den Einstellungsmaßen (*ET/EK*) und dem Kriterium *AW* im Schnitt deutlich niedriger. Auch hier kann – ebenso wie bei der Überprüfung von *H1.1* – der Verdacht, daß der Befund Ausdruck eines «Meßwiederholungssimplex» ist, mit Daten, die hier nicht referiert werden können, eindeutig ausgeräumt werden.

#### 5.3.2 Inhaltliche Korrespondenz

Die inhaltliche Korrespondenz läßt sich zur Prüfung von *H2* variieren, indem die Prädiktoren sich entweder auf dieselbe Gruppe von Benachteiligten beziehen wie das Kriterium (inhaltliche Korrespondenz gegeben) oder aber auf andere Gruppen (inhaltliche Korrespondenz verletzt). Als Prädiktor werden wieder die Attribution der Notlage auf die Benachteiligten (*NA*) und als Kriterium wieder die Abwertung der Benachteiligten (*AW*) herangezogen.

*H2.2:* Bei konstantem Aggregatniveau wird die Korrelation zwischen dem Prädiktor (*NA*) und dem Kriterium (*AW*) enger sein, wenn sich beide auf dieselbe Gruppe von Benachteiligten beziehen.

Die theoretische Nähe ist konstant gehalten: der Vergleich wird mit denselben beiden Variablen vorgenommen. Auch das Aggregatniveau ist nicht mit der Korrespondenzvariation konfundiert. Der Prädiktor *NA* wird ebenso wie das Kriterium *AW* in allen Fällen über die drei Situationen zu einer Gruppe Benachteiligter aggregiert gemessen.

Die Daten stehen mit der Hypothese in Einklang: Die gruppenspezifisch korrespondente Korrelation zwischen *NA* und *AW* beträgt im Mittel  $\bar{r} = .67$ . Wird dagegen inhaltliche Korrespondenz dadurch verletzt, daß sich *NA* und *AW* auf unterschiedliche Gruppen Benachteiligter beziehen, ist die mittlere Korrelation mit  $\bar{r} = .57$  geringer.

#### 5.3.3 Aggregatniveau

Das Aggregatniveau läßt sich zur Prüfung von *H3* variieren, indem die Variablenzusammenhänge situationsspezifisch, gruppenspezifisch (über jeweils drei Situationen aggregiert) oder unspezifisch (über alle neun Situationen aggregiert) gebildet und miteinander verglichen werden. Die theoretische Nähe wird wieder durch Verwendung derselben Variablen auf allen Aggregatniveaus gewährleistet und die inhaltliche Korrespondenz dadurch, daß sich Prädiktor und Kriterium jeweils auf dieselbe Situation bzw. auf dieselbe Gruppe Benachteiligter beziehen.

*H3.2:* Bei gleicher inhaltlicher Korrespondenz wird die Korrelation zwischen dem Kriterium *AW* und dem Prädiktor *NA* um so höher, über je mehr Situationen/Gruppen Benachteiligter aggregiert die Variablen gemessen werden.

Die Daten entsprechen den Erwartungen: Die mittlere situationsspezifische Korrelation von *NA* und *AW* beträgt  $\bar{r} = .46$ . Werden die drei situationsspezifischen Werte zu jeder Gruppe Benachteiligter aggregiert, beträgt die mittlere gruppenspezifische Korrelation von *NA* und *AW*  $\bar{r} = .67$ . Werden die Meßwerte schließlich über alle neun Situationen aggregiert, steigt die Korrelation von *NA* und *AW* auf  $r = .83$ .

### Diskussion

Alle drei vorgeschlagenen Wege zu mehr Konsistenz erwiesen sich in beiden Untersuchungen als effektiv; ein Effektivitätsvergleich ist nicht leicht und birgt die Gefahr von tautologischen Schlüssen. Es sollen dennoch einige Überlegungen über die Voraussetzungen der Wirksamkeit der drei Strategien zur Konsistenzsteigerung angestellt werden.

Klar – und für den Statistiker trivial – ist freilich die Tatsache, daß die hier berichteten Konsistenzkoeffizienten (einfache und multiple Korrelationskoeffizienten) einzig und allein Funktionen der Varianz-Kovarianzmatrizen der beteiligten Variablen auf höchstem Auflösungs-niveau, d. h. auf Itemebene sind. Der vergleichsweise starke (dämpfende) Einfluß, den die Verletzung inhaltlicher Korrespondenz in Untersuchung 1 (im Vergleich zu Untersu-

<sup>4</sup> Die Einstellung gegenüber den Menschen in der dritten Welt läßt sich nicht über dieselben Verhaltens-schwellen operationalisieren, da diese Überschneidungen des Lebenskontextes von Einstellungsobjekt und Einstellungsobjekt voraussetzen, die hier nicht gegeben sind.

chung 2) auf die Präzifizierbarkeit hat, spiegelt nur wider, daß die Inhalte (= Bedürfnisse) bezogen auf dieselben Variablen heterogen sind und im Vergleich dazu die Variablen bezogen auf dieselben Inhalte eng zusammenhängen. Diese formale «Erklärung» kann aber nur post hoc – d. h. nach der Datenerhebung – erschlossen werden und kann konzeptuelle und theoretische Überlegungen, die zu einer Wahl und Operationalisierung von Variablen führen, nicht ersetzen. Festhalten läßt sich, daß bei konzeptuell, theoretisch oder aus der empirischen Befundlage zu erwartender Heterogenität eines Gegenstandsbereiches inhaltliche Korrespondenz zwischen Prädiktoren und Kriterium Voraussetzung für eine aussagekräftige empirische Überprüfung eines theoretischen Modells der Beschreibung von Variablenzusammenhängen in diesem Bereich ist.

Ebenso klar – und für einen Statistiker ebenso trivial – ist die Feststellung, daß sich die Effektivität der Datenaggregation (z. B. nach der SPEARMAN-BROWN-Formel) berechnen bzw. schätzen läßt (vgl. z. B. DAY, MARSHALL, HAMILTON & CHRISTY 1983). Dennoch wird man auf eine empirische Ermittlung der «oberen Konsistenzgrenze» in vielen Fällen angewiesen bleiben, denn die Annahmen dieser Formel können nicht ungeprüft vorausgesetzt werden und dürften in der Forschungspraxis selten erfüllt sein. Außerdem sind Fälle, in denen Theorien empirisch überprüft werden sollen, zu unterscheiden von Fällen, in denen eine Untersuchung Aufschluß geben soll über die Lösbarkeit von Problemen der Angewandten Psychologie. Während für einen Theorietest etwa Reliabilitätskorrekturen nutzbar gemacht werden können und latente Variablen herangezogen werden können (z. B. via LISREL), werden für praktische Entscheidungen in der Regel zuverlässig meßbare manifeste Variablen benötigt.

Am problematischsten ist die Beurteilung des Einflusses theoretischer Nähe auf Konsistenz. Theoretische Bezugsnähe zwischen Prädiktor und Kriterium ist nur aus einem Modell des Zusammenhangs zu präzisieren. Tatsächlich werden zur Erklärung empirisch gefundener Inkonsistenz häufig exahurierend Argumente angeführt, die die theoretische Nähe problematisieren. Nicht selten gehen in Untersuchungen recht vage Zusammenhangserwartungen ein, die nicht

einmal durch eingehende Inhaltsanalysen von Konzepten und ihren Operationalisierungen gestützt sind, sondern lediglich auf Bedeutungsannahmen beruhen. In der Moralforschung sind beispielsweise mehrfach ohne ausreichend präzise Klärung ihres Inhaltes und ihrer theoretischen Bedeutung Variablen wie Moralisches Urteil sensu KOHLBERG oder Empathie verwendet worden (MONTADA 1983; SCHMITT 1982). Theoretische Nähe ist a priori zu begründen. Eine empirische Korrektur des theoretischen Modells erfordert, daß Inkonsistenzen nicht gleichzeitig oder alternativ auf andere Quellen wie Fehlerbehaftetheit der Meßwerte oder mangelnde inhaltliche Korrespondenz der Konstruktsindikatoren attribuierbar sind.

Welcher der drei hier illustrierten Wege zu mehr Konsistenz beschritten wird, hängt vor allem von der Aufgaben- oder Fragestellung ab. Wenn es um Konstruktvalidierung oder die empirische Überprüfung eines theoretischen Modells geht, wird man den Versuch unternehmen müssen, die Meßwerte der Personen (oder allgemein: Meßwertträger) durch Datenaggregation so zuverlässig wie möglich zu machen, denn Reliabilität ist zwar keine hinreichende Voraussetzung von Validität, aber eine notwendige.

Wenn es um die Vorhersage von Einzelereignissen bzw. ihrer Wahrscheinlichkeit geht, steht die Möglichkeit einer aggregierten Messung des Kriteriums nicht zur Verfügung. Die Vorhersagegenauigkeit hängt aber nicht nur von der Zuverlässigkeit des Kriteriums, sondern auch von der Zuverlässigkeit der Prädiktoren und ihrer inhaltlichen Korrespondenz zum Kriterium ab. Wenn eine Steigerung der Zuverlässigkeit des Prädiktors durch Aggregation paralleler Messungen angestrebt wird, muß meist eine Minderung inhaltlicher Korrespondenz in Kauf genommen werden (da die Zahl inhaltlich korrespondierender Items im gleichen Test ohne Itemwiederholung begrenzt ist). Im Einzelfall ist dann die Effektivität dieser beiden Strategien (Aggregation versus Erhöhung inhaltlicher Korrespondenz) vergleichend abzuwägen. Das kann nur empirisch geschehen. Dazu abschließend noch zwei Beispiele:

In Untersuchung 1 hatte sich inhaltliche Korrespondenz zwischen Prädiktoren und Kriterium als sehr einflußreich auf die Konsistenz er-

wiesen. Um zu prüfen, ob sich die Verletzung inhaltlicher Korrespondenz durch eine zuverlässigere Messung der Prädiktoren kompensieren läßt, wurden die drei Prädiktoren Einstellungen (EI), Personale Normen (PN) und Verhaltenserfahrung (VE) auch über die 34 Bedürfnisse aggregiert gemessen. Die drei Skalen weisen relativ zu ihrer Länge bescheidene, absolut gesehen aber befriedigende interne Konsistenzen auf (EI: Alpha = .92; PN: Alpha = .94; VE: Alpha = .91) und sind entsprechend reliabel. Der mittlere multiple Korrelationskoeffizient zwischen einer reessionsanalytisch ermittelten Linearkombination dieser zuverlässig gemessenen Prädiktoren und dem bedürfnisspezifisch gemessenen Kriterium A beträgt  $\bar{R} = .31$ . Mit  $\bar{R} = .60$  deutlich höher ist dagegen der entsprechende mittlere multiple Korrelationskoeffizient, wenn die Prädiktoren bedürfnisspezifisch gemessen und zum Kriterium inhaltlich korrespondent sind. Inhaltliche Korrespondenz ist also in diesem Datenbeispiel die bessere Strategie zur Vorhersage des Einzelkriteriums als eine zuverlässigere Messung des Prädiktors.

In Untersuchung 2 ist die Befundlage genau entgegengesetzt. Die situationsspezifisch gemessene Abwertung (AW) kann besser durch aggregiert (über die neun Situationen) gemessene Notlagenattribution (NA) vorhergesagt werden (mittleres  $\bar{r} = .57$ ) als durch spezifisch gemessene und damit zwar inhaltlich korrespondente, aber weniger zuverlässig gemessene Variable (mittleres  $\bar{r} = .47$ ).

Dieses Ergebnis bestätigt die weiter oben gegebene Empfehlung, bei heterogenen Untersuchungsgegenständen die Strategie inhaltlicher Korrespondenz als Weg zu mehr Konsistenz zu wählen. Eine Zuverlässigkeitssteigerung der Meßwerte durch parallele Messung ist in diesem Fall nur dann sinnvoll, wenn dadurch die inhaltliche Korrespondenz nicht verletzt wird. Sonst muß auf Meßwiederholung zur Reliabilitätssteigerung zurückgegriffen werden (vgl. auch die Kontroverse zwischen EPSTEIN 1983b und MISCHER & PEAKE 1982; als Beispiel vgl. BUSE & PAWLIK 1984).

Die in den Hypothesen H1, H2 und H3 behaupteten Effekte sind also nicht allgemein nützlich; sie stellen vielmehr Angebote dar, die aufgabenspezifisch zu realisieren sind.

#### Anmerkung

Gegen die Methode des Vergleichs von Korrelationskoeffizienten ließe sich einwenden, sie sei nur sinnvoll oder sogar zulässig, wenn die Verteilungseigenschaften der Variablen (Varianz, Schiefe, Exzeß oder allgemein ihre Verteilungsformen) und die Zuverlässigkeit ihrer Messungen mit berücksichtigt werden. Tatsächlich aber ist die Frage, ob Korrelationskoeffizienten nur bei Gleichheit der Verteilungsformen der beteiligten Variablen verglichen werden dürfen, nicht eindeutig zu beantworten. Zwar dämpfen Unterschiede zwischen den Verteilungsformen zweier Variablen deren maximal mögliche Korrelation, eine solche verteilungsbedingte Minderung ist aber Abbild der Realität, nämlich der eingeschränkten Vorhersagbarkeit der einen Variable aus der anderen. Dagegen würde eine Minderungskorrektur (z. B.  $r/r_{\max}$ ) oder eine nichtlineare Skalentransformation zu einer Variablenpaar-spezifischen Verzerrung führen und die Vergleichbarkeit von Korrelationen einschränken. Auch ist die Varianzabhängigkeit von Korrelationen nur zu bedenken, wenn Gleichheit der Regressionsgewichte vorausgesetzt wird, denn nur bei konstantem Regressionsgewicht ist die Korrelation eine Funktion der Varianzen von Prädiktor und Kriterium. Diese Voraussetzung der Gleichheit der Regressionskoeffizienten ist aber bei psychologischen Skalen, auch wenn sie – wie hier – homolog sind, problematisch. Der Einwand, Korrelationen könne man nur vergleichen, wenn man die Reliabilität der Variablen kenne, ist zutreffend. Hier wurde ja gerade demonstriert, wie die Korrelation mit zunehmender Meßwertreliabilität durch Aggregation steigt. Natürlich wäre es schön, die Zuverlässigkeit der Einzelmessungen unabhängig davon zu kennen. Deren Schätzung über Meßwiederholung oder parallele Items ist aber mit erheblichen Problemen behaftet, denn die Effekte inhaltlicher Unterschiede (parallele Items) bzw. wahrer Veränderungen (Meßwiederholung) lassen sich nicht sicher von Fehlervarianz im eigentlichen Sinn trennen.

#### Literatur

- AJZEN, I. & FISHBEIN, M.: Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 1977, 34, 888–918.
- AJZEN, I. & FISHBEIN, M.: *Understanding attitudes and predicting social behavior*. Englewood Cliffs, N. J.: Prentice Hall, 1980.
- ALKER, H. A.: Is personality situationally specific or intrapsychically consistent? *Journal of Personality*, 1972, 40, 1–16.
- AMELANG, M. & BARTUSSEK, D.: Zur differentiellen Validität von Fragebögen. *Diagnostica*, 1971, 17, 83–84.
- AMELANG, M. & BORKENAU, P.: Über die faktorielle Struktur und externe Validität einiger Fragebogen-Skalen zur Erfassung von Dimensionen der Extraversion und emotionalen Labilität. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1982, 3, 119–146.
- AMELANG, M. & BORKENAU, P.: Versuche einer Differenzierung des Eigenschaftskonzeptes: Aspekte intraindivid-

- vidueller Vorhersagbarkeit. In: AMELANG, M. & AHRENS, H.-J. (Ed.): Brennpunkte der Persönlichkeitsforschung (Band 1). Göttingen: Hogrefe, 1984, p. 89–107.
- BEM, D. J.: Constructing cross-situational consistencies in behavior: Some thoughts on Alker's critique of Mischel. *Journal of Personality*, 1982, 40, 17–26.
- BEM, D. J. & ALLEN, A.: On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review*, 1974, 81, 506–520.
- BEM, D. J. & FUNDER, D. C.: Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review*, 1978, 85, 485–501.
- BENTLER, P. M. & SPECKART, G.: Models of attitude-behavior relations. *Psychological Review*, 1979, 86, 452–464.
- BLOCK, J.: *The Q-sort method in personality assessment and psychiatric research*. Springfield, Ill.: Thomas, 1961.
- BOGARDUS, E. S.: Measuring social distance. *Journal of Applied Sociology*, 1925, 9, 299–308.
- BOWERS, K. S.: Situationism in psychology: An analysis and a critique. *Psychological Review*, 1973, 80, 307–336.
- BUSE, L. & PAWLIK, K.: Inter-Setting-Korrelationen und Setting-Persönlichkeits-Wechselwirkungen: Ergebnisse einer Felduntersuchung zur Konsistenz von Verhalten und Erleben. *Zeitschrift für Sozialpsychologie*, 1984, 15, 44–59.
- CHEEK, J. M.: Aggregation, moderator variables, and the validity of personality tests: A peer-rating study. *Journal of Personality and Social Psychology*, 1982, 43, 1254–1269.
- CHEEK, J. M. & BRIGGS, S. R.: Self-consciousness and aspects of identity. *Journal of Research in Personality*, 1982, 16, 401–408.
- DALBERT, C.: Verantwortlichkeit und Handeln (= Berichte aus der Arbeitsgruppe «Verantwortung, Gerechtigkeit, Moral» Nr. 5). Trier: Universität Trier, Fb I – Psychologie, 1980.
- DALBERT, C., MONTADA, L., SCHMITT, M. & SCHNEIDER, A.: Existentielle Schuld: Ergebnisse der Item- und Skalenanalysen (= Berichte aus der Arbeitsgruppe «Verantwortung, Gerechtigkeit, Moral» Nr. 24). Trier: Universität Trier, Fb I – Psychologie, 1984.
- DAY, H. D., MARSHALL, D., HAMILTON, B. & CHRISTY, J.: Some cautionary notes regarding the use of aggregated scores as a measure of behavioral stability. *Journal of Research in Personality*, 1983, 17, 97–109.
- ENDLER, N. S.: Interactionism comes of age. In: ZANNA, M. P., HIGGINS, E. T. & HERMAN, C. P. (Ed.): *Consistency in social behavior*. Hillsdale: Lawrence Erlbaum, 1982, p. 209–249.
- ENDLER, N. S. & MAGNUSSON, D. (Ed.): *Interactional psychology and personality*. New York: Wiley, 1976.
- EPSTEIN, S.: The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 1979, 37, 1097–1126.
- EPSTEIN, S.: The stability of behavior: II. Implications for psychological research. *American Psychologist*, 1980, 35, 790–806.
- EPSTEIN, S.: Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, 1983, 51, 360–392 (a).
- EPSTEIN, S.: The stability of confusion: A reply to Mischel and Peake. *Psychological Review*, 1983, 90, 179–184 (b).
- FAZIO, R. H. & ZANNA, M. P.: Direct experience and attitude-behavior consistency. In: BERKOWITZ, L. (Ed.): *Advances in Experimental Social Psychology* (Vol. 14). New York: Academic Press, 1981, p. 161–202.
- FISHBEIN, M. & AJZEN, I.: Attitude towards objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 1974, 81, 59–74.
- FREDERIKSEN, N.: Toward a taxonomy of situations. *American Psychologist*, 1972, 27, 114–123.
- FUNDER, D. L.: Three issues in predicting more of the people: A reply to Mischel and Peake. *Psychological Review*, 1983, 90, 283–289.
- GHISELLI, E. E.: Moderating effects and differential reliability and validity. *Journal of Applied Psychology*, 1963, 47, 81–86.
- GIBBONS, F. X.: Sexual standards and reaction to pornography: Enhancing behavioral consistency through self-focused attention. *Journal of Personality and Social Psychology*, 1978, 36, 976–987.
- GOLDING, S. L.: Flies in the ointment: Methodological problems in the analysis of the percentage of variance due to persons and situations. *Psychological Bulletin*, 1965, 82, 278–288.
- HARTSHORNE, H. & MAY, M. A.: *Studies in the nature of character*. Vol. I: Studies in deceit. New York: Macmillan, 1928.
- HECKHAUSEN, H.: *Motivation und Handeln*. Berlin: Springer, 1980.
- HOGAN, R.: Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, 1969, 33, 306–316.
- JÄGER, R. S.: Moderatoren, differentielle Diagnostizierbarkeit und Vorhersage – Methodische Überlegungen und Konsequenzen. *Diagnostica*, 1979, 25, 109–124.
- KENRICK, D. T. & STRINGFIELD, D. O.: Personality traits and the eye of the beholder: Crossing some traditional philosophical boundaries in the search for consistency in all of the people. *Psychological Review*, 1980, 87, 88–104.
- MAGNUSSON, D. & ENDLER, N. S. (Ed.): *Personality at the cross-roads*. New York: Wiley, 1977.
- MISCHEL, W.: *Personality and assessment*. New York: Wiley, 1968.
- MISCHEL, W.: Alternatives in the pursuit of the predictability and consistency of persons: Stable data that yield unstable interpretations. *Journal of Personality*, 1983, 51, 578–604.
- MISCHEL, W. & PEAKE, P. K.: Beyond déjà vu in the search for cross-situational consistency. *Psychological Review*, 1982, 89, 730–755.
- MONTADA, L.: Entwicklung interpersonalen Verantwortlichkeit und interpersonalen Schuld. Projektantrag an die Stiftung Volkswagenwerk (= Berichte aus der Arbeitsgruppe «Verantwortung, Gerechtigkeit, Moral» Nr. 8). Trier: Universität Trier, Fb I – Psychologie, 1981.
- MONTADA, L.: Moralisches Urteil und moralisches Handeln. *Wehrpsychologische Untersuchungen* 18, Heft 2/83, 1983.
- MONTADA, L., DALBERT, C., REICHLE, B. & SCHMITT, M.: Urteile über Gerechtigkeit, «Existentielle Schuld» und Strategien der Schuldabwehr. In: OSER, F., ALTHOF, W. & GARZ, D. (Ed.): *Entstehung moralischer Identität: Soziogenese, moralisches Handeln und Schuld*. München: Peter Kindt Verlag, 1985 (im Druck).
- MONTADA, L., DALBERT, C. & SCHMITT, M.: Interpersonale Verantwortlichkeit erwachsener Töchter ihren Müttern gegenüber: Hypothesen über Zusammenhänge innerhalb der Kernvariablen und zwischen Kernvariablen und Kovariaten (= Berichte aus der Arbeitsgruppe «Verantwortung, Gerechtigkeit, Moral» Nr. 15). Trier: Universität Trier, Fb I – Psychologie, 1982.
- MONTADA, L., SCHMITT, M. & DALBERT, C.: Thinking about justice and dealing with one's own privileges: A study on existential guilt. In: BIERHOFF, H. W., COHEN, R. & GREENBERG, J. (Ed.): *Justice in social relations*. New York: Plenum Press, 1986 (im Druck).
- OLWEUS, D.: Der «moderne» Interaktionismus von Person und Situation und seine varianzanalytische Sackgasse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 1976, 8, 171–185.
- PRYOR, J. B., GIBBONS, F. X., WICKLUND, R. A., FAZIO, R. H. & HOOD, R.: Self-focused attention and self-report validity. *Journal of Personality*, 1977, 45, 513–527.
- RUSHTON, J. P., BRAINERD, C. J. & PRESSLEY, M.: Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, 1983, 94, 18–38.
- SCHIEFER, M. F., BUSS, A. H. & BUSS, D. M.: Self-consciousness, self-report of aggressiveness, and aggression. *Journal of Research in Personality*, 1978, 12, 133–140.
- SCHLEGEL, R. P. & DITECCO, D.: Attitudinal structures and attitude-behavior relation. In: ZANNA, M. P., HIGGINS, E. T. & HERMAN, C. P. (Ed.): *Consistency in social behavior*. Hillsdale, N. J.: Lawrence Erlbaum, 1982, p. 17–49.
- SCHMITT, M.: Person, Situation oder Interaktion? Eine zeitlose Streitfrage diskutiert aus der Sicht der Gerechtigkeitsforschung (= Berichte aus der Arbeitsgruppe «Verantwortung, Gerechtigkeit, Moral» Nr. 6). Trier: Universität Trier, Fb I – Psychologie, 1980.
- SCHMITT, M.: Empathie: Konzepte, Entwicklung, Quantifizierung (= Berichte aus der Arbeitsgruppe «Verantwortung, Gerechtigkeit, Moral» Nr. 9). Trier: Universität Trier, Fb I – Psychologie, 1982.
- SCHMITT, M., DALBERT, C. & MONTADA, L.: Interpersonale Verantwortlichkeit erwachsener Töchter ihren Müttern gegenüber: Ergebnisse der Item- und Skalenanalysen (erster Untersuchungszeitraum) (= Berichte aus der Arbeitsgruppe «Verantwortung, Gerechtigkeit, Moral» Nr. 23). Trier: Universität Trier, Fb I – Psychologie, 1983.
- SCHWARTZ, S. H.: Normative influences on altruism. In: BERKOWITZ, L. (Ed.): *Advances in Experimental Social Psychology* (Vol. 10). New York: Academic Press, 1977, p. 221–279.
- SCHWENKMEZGER, P.: Kann durch das Prinzip der Aggregation von Daten die Konsistenzannahme von Eigenschaften beibehalten werden? *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1984, 4, 251–272.
- SNYDER, M.: When believing means doing: Creating links between attitudes and behavior. In: ZANNA, M. P., HIGGINS, E. T. & HERMAN, C. P. (Ed.): *Consistency in social behavior*. Hillsdale, N. J.: Lawrence Erlbaum, 1982, p. 105–130.
- TELLEGEN, A., KAMP, J. & WATSON, B.: Recognizing individual differences in predictive structures. *Psychological Review*, 1982, 89, 95–105.
- TRIANDIS, H. C.: *Attitude and attitude change*. New York: Wiley, 1971.
- WEIGEL, R. H. & NEWMAN, L. W.: Increasing attitude-behavior correspondence by broadening the scope of behavioral measure. *Journal of Personality and Social Psychology*, 1976, 33, 793–802.
- WICKER, A. W.: Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues*, 1969, 25, 41–78.
- ZANNA, M. P., OLSON, J. M. & FAZIO, R. H.: Attitude-behavior consistency: An individual difference perspective. *Journal of Personality and Social Psychology*, 1980, 38, 432–440.
- ZEDECK, S.: Problems with the use of «moderator» variables. *Psychological Bulletin*, 1971, 76, 295–310.
- ZUCKERMAN, M.: General and situation-specific traits and states: New approaches to assessment of anxiety and other constructs. In: ZUCKERMAN, M. & SPIELBERGER, C. D. (Ed.): *Emotions and anxiety. New concepts, methods, and applications*. Hillsdale, N. J.: Erlbaum, 1976, p. 133–174.

Dipl.-Psych. Manfred Schmitt, Dipl.-Psych. Claudia Dalbert, Prof. Dr. Leo Montada, Fachbereich Psychologie der Universität Trier, Tarforst, 5500 Trier