

Untersuchungen zur faktoriellen Struktur retrospektiv geschätzter und on-line kodierter Verhaltensfrequenzen: Eine Vergleichsstudie¹

PETER BORKENAU und FRITZ OSTENDORF

Universität Bielefeld

Zusammenfassung: Fragestellung der Untersuchung war, ob retrospektive Schätzungen von Verhaltensfrequenzen eine implizite Mehrfachkodierung von Aktivitäten beinhalten. Eine solche Hypothese würde gestützt, wenn sich die Interkorrelationen retrospektiv geschätzter und on-line kodierter Verhaltensfrequenzen dann einander angleichen, wenn das on-line Kodierschema überlappende Verhaltenselemente zwischen Kategorien berücksichtigt. In der vorliegenden Studie waren Faktorenähnlichkeiten das Kriterium einer solchen erwarteten Angleichung. Acht Diskussionsgruppen mit insgesamt 48 Teilnehmern wurden zunächst gefilmt. Die Videobänder wurden sodann mittels dreier Kodierschemata ausgewertet: (a) retrospektiven Schätzungen für Verhaltensfrequenzen, (b) on-line Kodierungen per Zuordnung beobachteten Verhaltens zu einer von 16 Kategorien (forced-choice Format) und (c) on-line Einschätzungen jeder beobachteten Aktivität bezüglich ihrer Prototypikalität für alle 16 Kategorien. Die letzte Methode berücksichtigt Bedeutungsüberlappungen. Entgegen der Erwartung zeigte sich keine maximale faktorielle Korrespondenz zwischen retrospektiven Schätzungen einerseits und multiplen Verhaltenskodierungen andererseits. Vielmehr wiesen die im forced-choice Format kodierten Aktivitäten die höchsten strukturellen Korrespondenzen zu den retrospektiven Ratings auf. Die Generalisierbarkeit des Befundes wird kritisch diskutiert.

Abstract: The present study tested the hypothesis that retrospective estimates of act frequencies reflect an implicit multiple assignment of activities to several behavior categories. This hypothesis would be supported if the structures of retrospectively estimated and on-line recorded act frequencies were to become more similar if the on-line coding scheme took into account overlapping activities among categories. In the present study, factor congruency was the criterion of structural correspondence. Eight discussion groups comprising a total of 48 subjects were videotaped. The tapes were then analyzed by three methods: (a) retrospective estimates, (b) a forced-choice on-line assignment of activities to one of 16 behavior categories, and (c) on-line prototypicality ratings of each activity for all 16 categories. The last method allows for «act overlap». The structural correspondence between retrospective estimates and multiply assigned activities was not the highest one that was obtained. Rather, the highest correspondence was found for retrospective frequency estimates and forced-choice on-line codings. The generalizability of this result is discussed.

1. Problemstellung

Die faktorielle Struktur von Trait-Ratings und Fragebogenscores ist bereits häufig analysiert worden. Zwar führten diese Untersuchungen nicht zu einer Aufgabe der unterschiedlichen Positionen, welche etwa von CATTELL (1972),

EYSENCK (1977; EYSENCK & EYSENCK 1985) und GUILFORD (1975, 1977) bezogen werden. Andererseits jedoch ließ sich ein gleichsam harter Kern gut replizierbarer Faktoren etablieren, deren Zahl gewöhnlich mit fünf (AMELANG & BORKENAU 1982; DIGMAN & INOUE 1986; FISKE 1949; McCRAE & COSTA 1987; NORMAN 1963; TUPES & CHRISTAL 1961) angegeben wird. Diese Faktoren werden mitunter als definitiv in dem Sinne bezeichnet, daß von weiteren faktorenanalytischen Untersuchungen auf der Ebene von Fragebogen- und anderen Einschätzungsdaten keine grundlegend neuen Erkenntnisse

¹ Mit Unterstützung der Deutschen Forschungsgemeinschaft (Az. Bo 774/1–1). Wir danken Ulrich Baumgartner, Kristina Herzig, Corinna Rüting und Angelus Stulik für die Durchführung der mühevollen on-line Beurteilungen.

mehr zu erwarten seien. Dies schließt die weitere Klärung von Detailfragen selbstverständlich nicht aus, wie etwa der nach einer angemessenen Interpretation des Faktors, welcher von NORMAN (1963) als *Culture*, von McCRAE und COSTA (1987) als *Openness to Experience* und schließlich von AMELANG und BORKENAU (1982) in Anlehnung an PAWLIK (1968) als *Unabhängigkeit der Meinungsbildung* interpretiert wird.

Weitgehend ungeklärt ist demgegenüber das Problem, in welcher Beziehung die bei Fragebogen und Trait-Ratings gefundenen Faktoren zur Struktur des eingeschätzten Verhaltens stehen (vgl. BUSE & PAWLIK 1984; MISCHEL & PEAKE 1982; PAWLIK & BUSE 1982). Zum einen nämlich weist die Struktur objektiver Tests sensu CATTELL wenig Korrespondenzen mit der Struktur von Fragebogenscores und globalen Trait-Einschätzungen auf (BLOCK 1977; HÄCKER, SCHWENKMEZGER & UTZ 1982). Weitere Zweifel bezüglich der Berechtigung einer Extrapolation von Fragebogen- und Rating-Faktoren auf Verhaltensstrukturen lassen solche Studien aufkommen, in denen die Korrelationen zwischen retrospektiv geschätzten Verhaltensfrequenzen verglichen werden mit den Interkorrelationen, welche sich für on-line kodierte Frequenzen eben dieser Verhaltensweisen ergeben. Solche Methodenvergleichsstudien zeigen gewöhnlich, daß die Interkorrelationen zwischen den on-line ermittelten Frequenzen geringer ausfallen als die zwischen den retrospektiv geschätzten (D'ANDRADE 1974; NEWCOMB 1929; THORNDIKE 1920). Weiterhin weisen die Interkorrelationen on-line ausgezählter Verhaltensfrequenzen einerseits und retrospektiv geschätzter Frequenzen andererseits nur geringe strukturelle Korrespondenzen auf (D'ANDRADE 1974; NEWCOMB 1929; SHWEDER 1975; SHWEDER & D'ANDRADE 1980).

Schließlich ist noch der Befund zu erwähnen, daß die Interkorrelationen zwischen globalen Persönlichkeitseinschätzungen in aller Regel gut vorhersagbar sind auf Basis der semantischen Ähnlichkeitsbeziehungen zwischen den Persönlichkeits-beschreibenden Begriffen (GARA & ROSENBERG 1981; MULAİK 1964; SHWEDER & D'ANDRADE 1980; siehe jedoch SEMIN & GREENSLADE 1985). Diese Befundkonstellation hat zu einer Kontroverse dahingehend geführt, ob es die Struktur des Verhaltens oder die Struktur der Sprache sei, welche die Interkorrelationen zwi-

schen globalen Persönlichkeitsratings primär bestimme. SHWEDER und D'ANDRADE (D'ANDRADE 1974; SHWEDER 1975, 1982; SHWEDER & D'ANDRADE 1980) haben eine systematische Verzerrungshypothese formuliert, der zufolge es bei retrospektiven Beurteilungen zu systematischen Verzerrungsprozessen komme und zwar in der Weise, daß mit nachlassender Erinnerung mehr und mehr sprachliche Zusammenhänge die Oberhand gegenüber den «wirklichen» Verhaltensstrukturen gewinnen. Entsprechend würden die Interkorrelationen zwischen retrospektiven Einschätzungen primär die semantischen Beziehungen zwischen den Persönlichkeits-beschreibenden Begriffen widerspiegeln und nicht die Struktur des Verhaltens.

Die von SHWEDER und D'ANDRADE präsentierte empirische Evidenz für diese Hypothese ist jedoch in verschiedener Hinsicht problematisiert worden, wobei die Kritik vor allem an der Qualität der on-line Verhaltensauszählungen ansetzte. So wurde etwa von BLOCK, WEISS und THORNE (1979) die geringe Reliabilität dieser Daten bemängelt. Schwerwiegender jedoch erscheint uns der Einwand, daß SHWEDER und D'ANDRADE die Beziehung zwischen Sprache und Verhaltensfrequenzen grundsätzlich falsch konzipierten (BORKENAU 1986; BORKENAU & OSTENDORF 1987; ROMER & REVELLE 1984). Bei den Studien, welche SHWEDER und D'ANDRADE als Beleg für ihre Hypothese heranzogen, war nämlich im Rahmen der on-line Kodierungen Verhalten stets genau einer Verhaltenskategorie zuzuordnen, und zwar völlig unabhängig von der semantischen Ähnlichkeit der Kategorien. Hierbei ergaben sich verblüffende Befunde: So korrelierten etwa die retrospektiven Einschätzungen für die Verhaltenskategorien *Widerspricht* und *Zeigt Antagonismus* erwartungsgemäß hoch ($r = .75$). Demgegenüber ergab sich für diese beiden Kategorien bezüglich der on-line kodierten Verhaltensfrequenzen eine insignifikante Korrelation von $r = .13$ (vgl. D'ANDRADE 1974). SHWEDER und D'ANDRADE erklären solche Befunde durch systematische Verzerrungsprozesse: Bei nachlassender Erinnerung an die exakten Fakten sollen nach und nach die semantischen Beziehungen die Oberhand gegenüber den beobachteten Verhaltensbeziehungen gewinnen.

Die von SHWEDER und D'ANDRADE präsenti-

tierten Befunde lassen sich aber auch damit erklären, daß viele Verhaltensbeispiele etwa für *Widersprechen* ebenfalls gute Beispiele für *Antagonismus* sind. Entsprechende implizite Mehrfachkodierungen könnten sich in den retrospektiven Frequenzeinschätzungen niederschlagen. Die erwähnten verblüffenden Befunde wären dann damit erklärbar, daß sich (a) die beiden Kategorien *Widersprechen* und *Antagonismus* zeigen auf weitgehend überlappende Verhaltensuniversen beziehen, (b) daß deshalb die beiden retrospektiven Frequenzratings hoch positiv korrelieren, und daß (c) bei einer on-line Kodierung im forced-choice Verfahren überlappende Verhaltenselemente per definitionem ausgeschlossen sind, was verminderte Korrelationen zur Folge hat.

Die strikt mengentheoretische Konzeption überlappender Verhaltenselemente ist jedoch aus zwei Gründen unbefriedigend: Zum einen wird sie dem Umstand nicht gerecht, daß die Zuordnung einzelner Aktivitäten zu Trait-Kategorien graduelle Abstufungen kennt. Handlungen unterscheiden sich bezüglich ihrer Prototypikalität für Trait-Kategorien (BUSS & CRAIK 1983; HAMPSON 1982). In dieser Hinsicht ähneln Trait-Kategorien natürlichen Kategorien im Sinne von ELEONOR ROSCH (ROSCH & MERVIS 1975). Hinzu kommt jedoch, daß eine angemessene theoretische Konzeptualisierung von Trait-Kategorien der dimensionalen Betrachtungsweise Rechnung tragen sollte, welche in der Persönlichkeitspsychologie traditionell vorherrscht (BORKENAU 1986; BORKENAU & OSTENDORF 1987; CANTOR & MISCHEL 1979). Die strikt mengentheoretische Konzeption überlappender Verhaltensuniversen ist somit offenbar auch deshalb unangemessen, weil sie nicht zwischen Gegenbeispielen für eine Traitkategorie einerseits und nicht einschlägigen Verhaltensweisen andererseits unterscheidet. Schon rein intuitiv ist ja die Beziehung etwa beleidigenden Verhaltens zur Kategorie *Freundlichkeit* anders gelagert als die Beziehung zwischen «Schreibtisch aufräumen» und der erwähnten Trait-Kategorie. Entsprechend konnte BORKENAU (1986) zeigen, daß bei semantischen Gegensätzen wie *Freundlichkeit* und *Feindseligkeit* gute Verhaltensbeispiele für die eine Kategorie als schlechte Beispiele für die andere Kategorie aufgefaßt werden (negative Korrelation der Prototypikalitätsratings). Hin-

gegen zeigten sich insignifikante Korrelationen, wenn eine Menge von Verhaltensbeispielen einander orthogonalen Traits zuzuordnen war.

Diesen Befunden wurde in der Studie von BORKENAU und OSTENDORF (1987) Rechnung getragen, indem beobachtetes Verhalten auf einer siebenstufigen Ratingskala mit den Endpunkten *krasses Gegenbeispiel* (-3) und *sehr gutes Beispiel* (+3) multipel eingestuft wurde. In einer anderen Hinsicht jedoch sind die bisher vorliegenden Studien zum Strukturvergleich retrospektiv geschätzter und on-line aufgezeichneter Verhaltensfrequenzen sämtlich unbefriedigend: Es wurden keine faktoriellen Strukturen verglichen, sondern Rangreihen von Korrelationskoeffizienten. Die gesamte Forschungsrichtung wurde jedoch motiviert durch die Bedeutung, welche die Faktorenanalyse von Rating-Interkorrelationen im Rahmen der Persönlichkeitspsychologie einnimmt (SHWEDER 1975)². Wenn zentrale Konstrukte der Persönlichkeitspsychologie unangemessen wären, weil sie auf systematische Verzerrungen bei retrospektiven Beurteilungen zurückgingen (D'ANDRADE 1974; MISCHEL 1968; SHWEDER 1975), dann müßten sich on-line ausgezählte und retrospektiv geschätzte Verhaltensfrequenzen auch bezüglich ihrer faktoriellen Struktur unterscheiden. Weiterhin, wenn die strukturellen Diskrepanzen auf der Vernachlässigung von Bedeutungsüberlappungen in den on-line Aufzeichnungen beruhten, so sollte ein Kodierschema, welches Überlappungen berücksichtigt, zu einer Angleichung auch der entsprechenden faktoriellen Strukturen führen. Das Kriterium ähnlicher faktorieller Strukturen unterscheidet sich nämlich von dem ähnlicher Rangreihen von Korrelationskoeffizienten: Sehr ähnliche Rangreihen von Korrelationskoeffizienten können ja in recht unterschiedliche faktorielle Strukturen münden, wenn sich Mittelwerte und Streuungen der Korrelationskoeffizienten für beide Datenmodi deutlich voneinander unterscheiden. So ist es für die faktorielle Struktur von erheblicher Bedeutung, ob die

2 Zwar nimmt die Faktorenanalyse von Einschätzungsdaten in der Persönlichkeitspsychologie eine besonders zentrale Stellung ein. Dieses Verfahren findet aber durchaus auch in anderen Bereichen Anwendung. So wurden etwa die Therapeutenvariablen der Gesprächspsychotherapie mittels der Faktorenanalyse von Einschätzungen erfolgreicher Therapeuten ermittelt.

niedrigsten Korrelationen zwischen den Variablen null sind und somit eine positive Mannigfaltigkeit der Korrelationen vorliegt, oder ob einige Korrelationen stark negative Werte annehmen. Hohe Korrelationen zwischen Korrelationsmatrizen sind somit nur eine notwendige, keineswegs aber eine hinreichende Bedingung für ähnliche Faktorenstrukturen.

Die vorliegende Studie soll diese Lücke füllen. Entsprechend wurden die semantischen Ähnlichkeitsbeziehungen zwischen 16 Verhaltenskategorien verglichen mit den faktoriellen Strukturen retrospektiv geschätzter und on-line kodierter Verhaltensfrequenzen.

2. Methode

2.1 Semantische Ähnlichkeitsurteile

Zwanzig Studenten, je zur Hälfte weiblichen und männlichen Geschlechts, schätzten 16 verhaltensbeschreibende Begriffe bezüglich ihrer Bedeutungsähnlichkeiten ein. Jede der 120 möglichen Kombinationen wurde jedem der 20 Beurteiler per Computer-Bildschirm präsentiert und zwar in jeweils unterschiedlicher, randomisierter Reihenfolge. Die Probanden füllten ihre Ähnlichkeitsurteile auf einer siebenstufigen Schätzskala, deren Pole als *Antonyme* (-3) und *Synonyme* (+3) gekennzeichnet waren. Die entsprechenden Urteile wurden zunächst bezüglich der erzielten Beurteilerübereinstimmung geprüft und sodann über die 20 Beurteiler hinweg gemittelt.

2.2 Bereitstellung einer Beobachtungsbasis

Die retrospektiven Frequenzschätzungen und die on-line Verhaltenskodierungen sollten möglichst auf einer gemeinsamen Beobachtungsbasis beruhen, um die Interpretierbarkeit der Ergebnisse zu steigern. Andernfalls wären Beobachtungsbasis und Auswertungsmethode miteinander konfundiert gewesen. Es fand deshalb folgendes Verfahren Anwendung: Innerhalb der Universität wurde nach männlichen Studenten gesucht, die an Diskussionsgruppen teilnehmen und sich dabei mittels Video filmen lassen wollten. Die Interessenten wurden gebeten, aus einer

Liste von sechs Themen zwei auszuwählen, über die sie gerne diskutieren würden. Weiterhin sollten sie ihre generelle Einstellung («pro» oder «contra») zu diesen Themen angeben, welche zum Zeitpunkt der Untersuchung im Frühjahr 1985 in der Öffentlichkeit kontrovers diskutiert wurden (Beispiel: Tempolimit auf Autobahnen). Der Versuchsleiter stellte auf Basis dieser Angaben sodann acht Gruppen zu je sechs Teilnehmern zusammen, welche: (a) sich untereinander nicht kannten, (b) sich an dem jeweiligen Thema interessiert gezeigt hatten und (c) bezüglich der Problemstellung unterschiedliche Einstellungen angegeben hatten.

Die acht Diskussionsrunden dauerten jeweils 50 Minuten, die vollständig aufgezeichnet wurden. Während der ersten Hälfte sollten die Teilnehmer ihre kontroversen Standpunkte diskutieren, um sodann in der zweiten Hälfte die Punkte zusammenzufassen, über die sie Übereinstimmung erzielt hatten. Nach fünfzig Minuten wurde die Diskussion vom Versuchsleiter unterbrochen. Ziel dieser Vorgaben war, möglichst lebendige und möglichst gleich lange Debatten zu stimulieren. Entsprechend wurde auch die Bezahlung der Diskutanten geregelt: Jeder Teilnehmer erhielt eine Basisprämie von 10,- DM. Zusätzlich wurde eine Prämie von 30,- DM je Teilnehmer für diejenige Diskussionsgruppe ausgelobt, welche die beste Diskussion präsentieren würde.

Die sechs Diskutanten wurden so an zwei Seiten eines quadratischen Tisches plaziert, daß alle sechs während der gesamten Diskussionszeit von einer feststehenden Kamera gefilmt wurden. Vor jedem Teilnehmer stand eine mit einem männlichen Vornamen (Pseudonym) beschriftete Karte, die der Identifikation des Diskutanten diene.

2.3 Kodierung der Verhaltensfrequenzen

2.3.1 Retrospektive Frequenzschätzungen

Fünf Beurteiler (zwei Studentinnen und drei Studenten), welche die Akteure zuvor nicht kannten, sahen sich jeweils eine Diskussionsrunde vollständig an. Die Reihenfolge der Darbietung der acht Gruppen war für die fünf Beurteiler unterschiedlich. Vor dem Betrachten des ersten Bandes wurden sie über die Art ihrer Be-

urteilungsaufgabe und die verwendeten Verhaltenskategorien informiert. Nach dem Betrachten jedes Bandes erhielten sie ein Heft, auf dessen erster Seite folgende Instruktion stand:

Bitte beurteilen Sie nun das Verhalten der einzelnen Diskutanten. Hierzu werden Ihnen einige Begriffe vorgegeben, durch die sich das Verhalten von Diskussionsteilnehmern beschreiben läßt. Ein Beispiel könnte sein «läßt andere Teilnehmer ausreden». Sie sollen sich jeweils überlegen, wie häufig die einzelnen Teilnehmer entsprechendes Verhalten in der Diskussion gezeigt haben. Beziehen Sie sich in Ihren Einschätzungen immer nur auf die vorangegangene Diskussionsrunde.

Jede der 16 Verhaltensklassen wurde sodann auf einer anderen Seite des Heftes präsentiert, um die Unabhängigkeit der einzelnen Urteile zu fördern. Am oberen Rand jeder Seite stand der Kategorienname. Es folgte ein Passus: «Wie häufig haben die einzelnen Teilnehmer dieses Verhalten in der Diskussion gezeigt?» Darunter standen die sechs Pseudonyme. Die Probanden sollten hinter jeden der sechs Namen eine Frequenzschätzung schreiben. Es wurde mithin keine Skala vorgegeben, sondern die Frequenzen wurden frei geschätzt. Die 16 Kategorien wurden dabei den fünf Beurteilern in jeweils unterschiedlicher Reihenfolge präsentiert. Die entsprechenden Schätzungen wurden später auf Beurteilerübereinstimmung geprüft und sodann zwecks Erhöhung der Reliabilität gemittelt.

2.3.2 On-line Kodierungen im forced-choice Verfahren

Zwei Beurteiler (eine Studentin und ein Student) beurteilten die acht Diskussionen in unterschiedlicher Reihenfolge. Ihnen wurden Bänder präsentiert, auf denen die Diskussionen in 15-Sekunden Sequenzen unterteilt worden waren. Jeder 15-Sekunden Sequenz folgte für 10 Sekunden ein Standbild auf dem Monitor. Die Beurteiler wurden instruiert, bei Erscheinen des Standbildes die Pausentaste des Videorecorders zu drücken, sodann eine oder mehrere Entscheidungen zu treffen, und schließlich den Videorecorder wieder zu starten.

Die Beurteiler hatten zu entscheiden, welcher der 16 Verhaltenskategorien die während der letzten 15 Sekunden beobachteten Aktivitäten am besten zugeordnet werden könnten (forced-choice). Deshalb waren pro Sequenz, je nach Anzahl der in den letzten 15 Sekunden verbal

aktiven Diskutanten, ein oder mehrere Urteile abzugeben. Jedoch war pro 15-sek Sequenz für jeden Diskutanten höchstens ein Urteil zu fällen. Entsprechend bezog sich jede Entscheidung auf eine spezifizierte Aktivität eines spezifizierten Akteurs während eines kurzen Zeitintervalls. Die den Ratern vorgegebenen Fragen lauteten z. B.: «Welche der folgenden Kategorien eignet sich am besten, das Verhalten von Frank zu klassifizieren?» oder «Welche der folgenden Kategorien eignet sich am besten, das Verhalten von Peter zu klassifizieren?» Jeder dieser Fragen folgte eine Liste der 16 Verhaltenskategorien sowie eine Residualkategorie «Keine Beurteilung möglich». Diese Restkategorie wurde einbezogen, weil es mitunter unmöglich war, extrem kurze Äußerungen sinnvoll zu klassifizieren. Die Beurteiler markierten ihre Wahl durch ein Kreuz hinter der am besten geeigneten Kategorie.

Die Einteilung der acht Diskussionen in Sequenzen und die Ermittlung der jeweils verbal aktiven Diskutanten hatten einer der Autoren und eine Studentin vorab vorgenommen. Sie hatten insgesamt 3696 Aktivitäten im erwähnten Sinne identifiziert. Entsprechend hatte jeder der beiden on-line Beurteiler 3696 Zuordnungen von Verhalten zu Kategorien vorzunehmen. Aus diesen Einschätzungen wurden Frequenzscores für die 48 Diskutanten ermittelt, indem die Zuordnungen separat für jeden Akteur und jede Verhaltenskategorie ausgezählt wurden. Hierbei wurden zwei unabhängige parallele Berechnungen vorgenommen, um die Replizierbarkeit der Befunde über zwei unabhängig voneinander urteilende Kodierer zu prüfen.

2.3.3 Prototypikalitätsratings

Auch den für diese Aufgabe eingesetzten zwei Studenten (einer weiblich und einer männlich) waren die Diskutanten nicht vorab bekannt. Ihnen wurden die gleichen Videobänder präsentiert wie den Ratern, welche on-line Kodierungen im forced-choice Verfahren vornahmen. Diesmal waren jedoch die Bänder insgesamt 16-mal von jedem Beurteiler anzusehen, jeweils einmal pro Verhaltenskategorie. Die 128 Kombinationen von acht Bändern und 16 Kategorien wurden beiden Beurteilern in unterschiedlicher, randomisierter Reihenfolge vorgegeben.

Die Einschätzungen erfolgten auf Ratingskalen. Die einzelnen Szenen waren nummeriert und der Akteur (die Akteure), dessen (deren) Verhalten einzuschätzen war, war (waren) spezifiziert. Es waren wiederum für 3696 Aktivitäten Einschätzungen vorzunehmen. Die Instruktion lautete jedoch diesmal, «anzugeben, wie gut das gezeigte Verhalten durch die gerade anstehende Kategorie charakterisiert werden» könne. Zu diesem Zweck wurde für jede Aktivität eine siebenstufige Ratingskala mit den Endpunkten krasses Gegenbeispiel (-3) und sehr gutes Beispiel (+3) vorgegeben. Bei jedem Durchlauf eines Bandes waren nur für eine Verhaltenskategorie Einschätzungen vorzunehmen. Beide Beurteiler sahen jedes Band jedoch 16mal und gaben entsprechend $16 \times 3696 = 59.136$ Einschätzungen ab, welche sich über ein halbes Jahr verteilten.

Aus den Prototypikalitätsratings wurden Verhaltensfrequenzen ermittelt, indem jede Aktivität eines Akteurs mit dem Prototypikalitätsrating für die gerade anstehende Verhaltenskategorie gewichtet und danach die gewichtete Summe über alle Aktivitäten des Akteurs gebildet wurde. Entsprechend wurde für alle 48 Akteure und alle 16 Verhaltenskategorien verfahren. Der Berechnungsmodus weist somit Ähnlichkeiten zur Ermittlung von Faktorenwerten auf. Für jeden Diskutanten wurden in dieser Weise zwei parallele Berechnungen vorgenommen: Eine beruhte auf den Prototypikalitätsratings der Beurteilerin, die andere auf denen des Beurteilers.

3. Resultate

Die Darstellung der Resultate beschränkt sich hier auf die Faktorenstrukturen. Die Reliabilität der Maße, die Validitätskoeffizienten, die Interkorrelationen und die Korrelationen zwischen den Interkorrelationsmatrizen sind bereits an anderer Stelle berichtet worden (BORKENAU & OSTENDORF 1987). Es ist hier jedoch festzuhalten, daß sich die 48 Akteure erheblich in ihrer generellen Aktivität unterschieden. Der Mittelwert aller Aktivitäten pro Proband betrug $M = 77$ bei einer Standardabweichung von $s = 37,12$. Der aktivste Proband machte 159 verbale Beiträge, der passivste hingegen nur neun. Dieser

Umstand ist bedeutsam, weil so hauptsächlich positive Interkorrelationen zwischen den Verhaltensfrequenzen für die 16 Kategorien zu erwarten sind. Dies kann gravierende Konsequenzen für die Faktorenstrukturen haben.

Ein Vergleich der verschiedenen Strukturen gewinnt bei einer gleichen Dimensionalität für die untersuchten Datenebenen an Präzision. Hier entschieden wir uns für durchgängig drei Dimensionen. Hierfür sprechen gleich mehrere Überlegungen: Zum einen nämlich sind bei der Analyse von führerlosen Diskussionsgruppen gewöhnlich zwei oder drei Dimensionen identifiziert worden, nämlich eine sozial-emotionale (kooperativ vs. antagonistisch), eine Dominanzdimension und (mitunter) eine Dimension der Aufgabenorientierung. Drei Dimensionen vergleichbarer Bedeutung finden sich etwa im SYMLOG-System von BALES und COHEN (1982). Weiterhin spricht für die Wahl von drei Dimensionen der Umstand, daß dies die Zahl der Faktoren ist, die gemäß dem Kaiser-Guttman Kriterium aus der Korrelationsmatrix zu extrahieren war, die den steilsten Eigenwertverlauf aufwies. Aus den anderen Matrizen wären entsprechend mehr Faktoren zu extrahieren gewesen.

3.1 Multidimensionale Skalierung der Semantischen Ähnlichkeiten

Die Matrix der semantischen Ähnlichkeiten wurde einer Multidimensionalen Skalierung unterzogen (Programm ALSCAL aus dem Programmpaket SPSS-X). Die Zahl der Dimensionen wurde vorab auf drei festgelegt. Die Lösung weist einen Streßwert von .128 auf. Die Koordinaten der 16 verhaltensbeschreibenden Begriffe im dreidimensionalen euklidischen Raum sind in Tabelle 1 aufgeführt. Die erste Dimension kontrastiert kooperative, aufgabenorientierte Aktivitäten einerseits mit antagonistischen und aufgabenirrelevanten andererseits. Dabei weisen die antagonistischen Aktivitäten die ausgeprägteren negativen Skalenwerte auf. Für die zweite Dimension finden sich deutlich negative Werte für die Kategorien, welche aufgabenirrelevante Aktivitäten kennzeichnen. Die dritte Dimension schließlich erweist sich als uninterpretierbar.

Tabelle 1: Ergebnisse der Multidimensionalen Skalierung der Ähnlichkeitsschätzungen für die 120 Kategorienkombinationen.

Kategorie	Dimension 1	Dimension 2	Dimension 3
Unterstützt	1.65	.01	-.54
Greift Beiträge auf	.65	.72	-.38
Scherzt	-.44	-1.48	.54
Vermittelt	1.27	.09	.38
Sucht Ausgleich	1.18	-.20	.52
Stimmt zu	1.79	-.89	-.48
Schlägt vor	.51	.89	-.56
Leitet die Diskussion	-.39	.29	-.70
Kritisiert	-1.57	.61	-.18
Informiert	.75	1.30	-.55
Erklärt	.84	.70	.59
Schweift vom Thema ab	-1.24	-1.80	.06
Fragt nach Meinungen	.65	-.26	1.68
Widerspricht	-2.14	.74	-.13
Lehnt ab	-2.15	.70	.93
Macht lächerlich	-1.37	-1.43	-1.16

3.2 Faktorenanalyse der retrospektiven Frequenzschätzungen

Die 120 Korrelationen zwischen den retrospektiven Frequenzratings streuten von $r = -.07$ für das Kategorienpaar *Sucht Ausgleich* vs. *Lehnt ab* und $r = .93$ für das Kategorienpaar *Informiert* vs. *Erklärt*. Mithin hatten alle statistisch signifikanten Korrelationen ein positives Vorzeichen (vgl. BORKENAU & OSTENDORF 1987, Tab. 2). Es wurde eine Faktorenanalyse mit iterativer Kommunalitätsschätzung berechnet. Die ersten drei Faktoren klärten 84,6% der Gesamtvarianz auf. Diese Faktoren wurden sodann Varimax-rotiert. Es fand somit gemäß den Empfehlungen REVENSTORFS (1978) ein relativ einfaches Verfahren Anwendung. Die Ladungsmatrix der Varimax-rotierten Faktoren ist in Tabelle 2 wiedergegeben. Das auffallendste Merkmal der Ladungsmatrix ist das Fehlen substantieller negativer Ladungen. Die Dimensionen sind somit unipolar zu konzipieren, was für Frequenzschätzungen auch plausibel ist. Der erste Faktor kontrastiert kooperative und aufgabenorientierte Verhaltensweisen mit solchen, welche entweder nicht aufgabenorientiert oder aufgabenorientiert und antagonistisch sind. Demgegenüber kontrastiert der zweite Faktor antagonistische Verhaltensweisen einerseits und entweder kooperative oder aufgabenirrelevante andererseits. Der dritte Faktor schließlich kontrastiert

aufgabenirrelevante Aktivitäten mit aufgabenorientierten, seien sie nun kooperativ oder antagonistisch. Mithin zeigen sich bei den retrospektiven Frequenzschätzungen drei Korrelationscluster: eines für aufgabenorientierte kooperative, eines für aufgabenorientierte antagonistische und eines für aufgabenirrelevante Aktivitäten.

3.3 Faktorenanalyse der on-line Kodierungen im forced-choice Verfahren

Die Interkorrelationen zwischen den Verhaltensauszählungen für verschiedene Kategorien weisen auch in diesem Datenmedium überwiegend positive Vorzeichen auf. Bei Zugrundelegung der Zuordnungen der Kodiererin streuen die Korrelationen zwischen $r = -.18$ für das Kategorienpaar *Vermittelt* und *Erklärt* und $r = .85$ für das Kategorienpaar *Vermittelt* und *Widerspricht*. Die Korrelationen auf Basis der Kodierungen des Kodierers schwanken zwischen $r = -.25$ für das Kategorienpaar *Vermittelt* und *Scherzt* und $r = .65$ für die Kategorien *Schlägt vor* und *Leitet die Diskussion* (vgl. BORKENAU & OSTENDORF 1987, Tab. 8). Die ersten drei Faktoren klären bei iterativer Kommunalitätsschätzung und Zugrundelegung der Urteile der Kodiererin 48,7% der Gesamtvarianz auf. Für den Kodierer beträgt der entsprechende Va-

Tabelle 2: Varimax-rotierte Ladungsmatrix bei Faktorenanalyse der retrospektiven Frequenzeinschätzungen.

Kategorie	Faktor 1	Faktor 2	Faktor 3
Unterstützt	.74	.29	.24
Greift Beiträge auf	.65	.48	.45
Scherzt	.16	-.06	.96
Vermittelt	.94	.12	-.07
Sucht Ausgleich	.90	.08	.02
Stimmt zu	.70	.17	.54
Schlägt vor	.67	.63	.21
Leitet die Diskussion	.76	.48	.16
Kritisiert	.29	.92	.12
Informiert	.56	.62	.35
Erklärt	.55	.70	.27
Schweift vom Thema ab	-.08	.46	.70
Fragt nach Meinungen	.84	.33	-.08
Widerspricht	.26	.88	.10
Lehnt ab	.13	.95	.04
Macht lächerlich	.06	.14	.92

rianzanteil 45,1%. Die Varimax-rotierten Ladungsmatrizen sind in Tabelle 3 wiedergegeben.

Wiederum weisen die meisten Ladungen positive Vorzeichen auf. Dies ist auf die überwiegend positiven Korrelationen zwischen den einzelnen Frequenzen zurückzuführen, welche wiederum die sehr unterschiedliche Gesamtaktivität der einzelnen Diskutanten widerspiegeln. Zunächst zu den Faktoren, welche auf den Ko-

dierungen der Beurteilerin beruhen: Der erste Faktor ist insbesondere wegen der negativen Ladung für *Erklärt* nicht interpretierbar. Auf dem zweiten Faktor weisen die Verhaltenskategorien hohe Ladungen auf, welche sich auf aufgabenirrelevante Aktivitäten (scherzen, vom Thema abschweifen und lächerlich machen) beziehen. Der dritte Faktor schließlich ist wieder schwer interpretierbar: insbesondere stört hier die ge-

Tabelle 3: Varimax-rotierte Ladungsmatrizen bei Faktorenanalyse der on-line Kodierungen im forced-choice Format.

Kategorie	Beurteilerin			Beurteiler		
	F1	F2	F3	F1	F2	F3
Unterstützt	.31	.21	-.11	.51	.48	.07
Greift Beiträge auf	.22	-.02	.72	.28	.59	.38
Scherzt	-.03	.71	-.07	.55	-.15	.06
Vermittelt	.81	.03	.06	-.16	.56	.08
Sucht Ausgleich	.75	.01	.26	.05	.49	.18
Stimmt zu	.13	.49	.26	.44	.35	.21
Schlägt vor	.28	-.02	.66	-.13	.30	.72
Leitet die Diskussion	.66	-.08	.26	.00	.14	.88
Kritisiert	.05	.44	.35	.03	.59	-.09
Informiert	.25	.06	.48	.06	-.06	.62
Erklärt	-.34	.41	.69	-.14	.66	.17
Schweift vom Thema ab	.19	.73	.04	.72	.09	.02
Fragt nach Meinungen	.55	.24	.35	.30	-.02	.39
Widerspricht	.77	.26	.19	.42	.67	.13
Lehnt ab	.26	.31	.15	.15	.48	-.14
Macht lächerlich	.01	.77	-.05	.70	.07	-.08

Anm.: F1 = Faktor 1, F2 = Faktor 2, F3 = Faktor 3.

ringe Ladung für *Leitet die Diskussion*. Auf dem ersten Faktor, welcher auf den Kodierungen des Beurteilers beruht, laden die Variablen am höchsten, welche aufgabenirrelevantes Verhalten kennzeichnen (schweift vom Thema ab, macht lächerlich und scherzt). Allerdings wird dieses Bild durch den Umstand gestört, daß etwa die Variable *Unterstützt* kaum niedriger lädt als *Scherzt*. Für den zweiten Faktor läßt sich keine Interpretation finden, welche dem Ladungsvektor insgesamt gerecht wird. So stören etwa die sehr unterschiedlichen Ladungen für *Informiert* und *Erklärt*. Der dritte Faktor schließlich mit seinen hohen Ladungen für *Leitet die Diskussion*, *Schlägt vor* und *Informiert* läßt sich als dominante und zugleich konstruktive Haltung deuten. Niedrige Ladungen auf diesem Faktor weisen aufgabenirrelevante, antagonistische und unterstützende Aktivitäten auf.

Zusammenfassend läßt sich somit festhalten, daß vor allem ein unipolarer Faktor aufgabenirrelevanten Verhaltens für die on-line Kodierungen im forced-choice Verfahren identifizierbar ist. Im Gegensatz zu den semantischen Ähnlichkeitsrelationen und den retrospektiven Frequenzschätzungen findet sich demgegenüber keine Kontrastierung kooperativer und antagonistischer Verhaltensklassen. Ein gemeinsames Merkmal der retrospektiven Ratings und der on-line Kodierungen im forced-choice Verfahren ist es dabei, daß beide sensibel sind für die unterschiedliche Gesamtaktivität der einzelnen Diskutanten.

3.4 Faktorenanalyse der Prototypikalitätsratings

Zunächst wurden die Prototypikalitätsratings für die 16 Kategorien über alle 3696 Aktivitäten hinweg korreliert und die Korrelationsmatrix einer Faktorenanalyse unterzogen. Eine solche Korrelationsmatrix gibt an, in welchem Ausmaß gute Verhaltensbeispiele für eine Kategorie ebenfalls als gute Beispiele für andere Kategorien angesehen werden. Entsprechend informiert diese Faktorenanalyse über grundlegende Dimensionen bei der Zuordnung von beobachteten Aktivitäten zu Verhaltenskategorien. Entsprechend der systematischen Überlappungshypothese (BORKENAU 1986) sollten diese Dimen-

sionen hohe Ähnlichkeit zu grundlegenden semantischen Dimensionen aufweisen, wie sie etwa aus einer Multidimensionalen Skalierung der semantischen Ähnlichkeitsrelationen resultieren.

Die Interkorrelationen zwischen den Prototypikalitätsratings für verschiedene Kategorien wiesen etwa gleich häufig positive und negative Vorzeichen auf. Dies war insofern zu erwarten, als in diese Korrelationen der unterschiedliche Aktivitätsgrad der einzelnen Diskutanten nicht eingeht. Für die Kodierungen der Beurteilerin streuten die entsprechenden Korrelationen zwischen $r = -.71$ (*Stimmt zu* vs. *Widerspricht*) und $r = .65$ (*Widerspricht* vs. *Lehnt ab*). Für die Prototypikalitätsratings des Beurteilers lauteten die extremsten Korrelationen $r = -.65$ (*Unterstützt* vs. *Widerspricht*) und $r = .65$ (*Unterstützt* vs. *Stimmt zu*). Die Korrelationsmatrizen wurden getrennt für beide Beurteiler einer Faktorenanalyse mit iterativer Kommunalitätsschätzung unterzogen, wobei die Zahl zu extrahierender Faktoren vorab auf drei festgelegt war. Diese drei Faktoren klärten 38,3% der Gesamtvarianz für die Beurteilerin und 41,1% der Gesamtvarianz für den Beurteiler auf. Die Varimax-rotierten Ladungsmatrizen sind in Tabelle 4 wiedergegeben.

Die aus den Einschätzungen der Beurteilerin resultierenden Faktoren lassen sich wie folgt interpretieren: Faktor 1 kontrastiert kooperative (positive Ladungen) mit antagonistischen Verhaltensweisen (negative Ladungen). Faktor 2 kontrastiert organisierende mit desorganisierenden Verhaltensweisen (z.B. *Schweift vom Thema ab*). Faktor 3 ist unipolar; auf ihm laden aufgabenirrelevante Aktivitäten hoch. Die aus den Einschätzungen des Beurteilers resultierenden Faktoren lassen sich vergleichbar konzipieren. Auch hier kontrastiert der erste Faktor kooperative mit antagonistischen Verhaltensweisen. Faktor 2 kontrastiert informierende mit aufgabenirrelevanten Aktivitäten. Faktor 3 schließlich weist ein ähnliches Ladungsmuster auf wie Faktor 2 bei der Beurteilerin und kontrastiert entsprechend organisierende mit desorganisierenden Verhaltensweisen. Die Faktorenanalyse der Prototypikalitätsratings erbringt mithin ähnliche Dimensionen wie die Multidimensionale Skalierung der Semantischen Ähnlichkeitsrelationen.

Tabelle 4: Varimax-rotierte Ladungsmatrix bei Faktorenanalyse der Prototypikalitätsratings.

Kategorie	Beurteilerin			Beurteiler		
	F1	F2	F3	F1	F2	F3
Unterstützt	.74	.02	.09	.78	-.09	.00
Greift Beiträge auf	.22	-.13	-.02	-.01	.15	-.21
Scherzt	.21	.01	.61	.12	-.52	-.03
Vermittelt	.66	.34	.07	.51	-.13	.45
Sucht Ausgleich	.68	.44	.07	.59	-.17	.40
Stimmt zu	.79	-.11	.09	.80	-.05	-.06
Schlägt vor	.23	.50	-.03	.04	.21	.41
Leitet die Diskussion	.02	.56	-.02	-.06	.06	.49
Kritisiert	-.66	.06	-.07	-.71	.25	.02
Informiert	-.12	.21	-.02	-.07	.74	.09
Erklärt	-.12	.24	-.03	-.12	.70	.06
Schweift vom Thema ab	.00	-.34	.39	-.01	-.04	-.35
Fragt nach Meinungen	.29	-.04	-.10	.01	-.44	-.02
Widerspricht	-.78	.08	-.12	-.79	.07	-.03
Lehnt ab	-.81	.09	-.08	-.78	.08	.03
Macht lächerlich	-.06	-.02	.59	.21	-.50	.04

Anm.: F1 = Faktor 1, F2 = Faktor 2, F3 = Faktor 3. Die den Faktoren zugrundeliegenden Korrelationen wurden über 3696 Aktivitäten berechnet.

3.5 Faktorenanalyse der Frequenzscores mit überlappenden Verhaltenselementen

In einem nächsten Schritt wurden auf Basis der Prototypikalitätsratings Frequenzscores für die einzelnen Diskutanten berechnet. Der Berechnungsmodus wurde bereits unter 2.3.3 geschildert. Da hierbei für einzelne Aktivitäten auch negative Gewichte vergeben wurden, resultierten etwa gleich häufig positive wie negative Korrelationen zwischen den entsprechenden Verhaltensfrequenzen. Die ersten drei Faktoren klären hier 70,6 % der Gesamtvarianz auf, wenn die Einschätzungen der Beurteilerin als Gewichte dienen und 66,2 %, wenn hierfür die Ratings des Beurteilers herangezogen werden. Über die Varimax-rotierten Ladungsmatrizen informiert Tabelle 5.

Der erste Faktor kontrastiert kooperative mit antagonistischen Verhaltensweisen. Dies gilt für die Beurteilerin ebenso wie für den Beurteiler. Der zweite Faktor kann für beide Beurteiler als einer organisierenden versus desorganisierenden Verhaltens interpretiert werden, obwohl die entsprechenden Ladungsmuster auch deutliche Diskrepanzen aufweisen. Bezüglich des dritten Faktors weisen bei der Beurteilerin die Variablen *Scherzt* und *Macht lächerlich* hohe Ladun-

gen auf. Man ist deshalb geneigt, ihn als Heiterkeitsfaktor zu deuten. Bei dem Beurteiler hingegen kontrastiert der dritte Faktor die Variablen *Schweift vom Thema ab* und *Fragt nach Meinungen*. Eine Deutung als Aufgabenorientierung bietet sich an.

3.6 Zusammenfassung

Aus den bisherigen intuitiven Deutungen resultieren folgende zusammenfassende Aussagen: Offenbar wegen der erheblichen Differenzen in der Gesamtaktivität der 48 Diskutanten korrelieren die retrospektiven Frequenzschätzungen mit wenigen Ausnahmen positiv. Entsprechend zeigen sich drei unipolare Faktoren, auf denen jeweils (a) aufgabenorientierte und konstruktive, (b) aufgabenorientierte und antagonistische und (c) aufgabenirrelevante Aktivitäten hoch laden. Zentrale Fragestellung der Untersuchung ist, inwiefern sich dieses Faktorenmuster mit dem deckt, welches aus on-line Kodierungen entsprechenden Verhaltens resultiert. Diesbezüglich zeigt sich, daß die on-line Kodierung im forced-choice Verfahren insofern der multiplen Verhaltenskodierung überlegen ist, als das erste Kodierschema ähnlich wie die retrospektiven Ratings zu einer positiven Mannigfaltigkeits-

Tabelle 5: Varimax-rotierte Ladungsmatrix bei Faktorenanalyse der on-line Kodierungen mittels Prototypikalitätsratings.

Kategorie	Beurteilerin			Beurteiler		
	F1	F2	F3	F1	F2	F3
Unterstützt	.90	-.09	-.14	.75	.00	.22
Greift Beiträge auf	.16	.74	.24	-.14	.60	-.14
Scherzt	.45	.36	.62	-.17	-.73	.17
Vermittelt	.70	.38	-.35	.61	.09	.13
Sucht Ausgleich	.71	.46	-.25	.79	-.24	-.01
Stimmt zu	.91	-.17	-.13	.88	.18	-.19
Schlägt vor	.08	.76	-.22	.30	.75	-.10
Leitet die Diskussion	.04	.75	-.02	.32	.52	.32
Kritisiert	-.74	.32	-.14	-.56	.51	.07
Informiert	.05	.84	.13	.19	.92	-.23
Erklärt	-.09	.71	-.34	-.17	.89	-.02
Schweift vom Thema ab	.48	-.70	.31	-.33	.12	-.81
Fragt nach Meinungen	-.09	-.47	-.12	-.27	-.17	.64
Widerspricht	-.87	.22	.10	-.87	-.05	-.04
Lehnt ab	-.94	.13	.01	-.89	-.05	.15
Macht lächerlich	.28	.51	.60	.37	-.77	-.13

Anm.: F1 = Faktor 1, F2 = Faktor 2, F3 = Faktor 3. Die den Faktoren zugrundeliegenden Korrelationen wurden über die N = 48 Diskutanten berechnet.

struktur führt. Diese bringt überwiegend positive Faktorladungen mit sich. Jedoch erweisen sich die drei Faktoren, welche aus der on-line Kodierung im forced-choice Verfahren resultieren, als uninterpretierbar. Demgegenüber ergeben sich bei einer multiplen Gewichtung der 3696 Aktivitäten gemäß ihrer Prototypikalität zahlreiche positive und negative Korrelationen und Faktorladungen. Dieser Umstand unterscheidet die Struktur dieser on-line Kodierungen von der Struktur der retrospektiven Frequenzschätzungen. Hingegen läßt sich die Unterscheidung zwischen kooperativen, antagonistischen und aufgabenirrelevanten Aktivitäten bei Verwendung dieses Kodierschemas auch in den Faktorladungen der on-line Verhaltenskodierungen erkennen. Beide on-line Kodierschemata haben somit offenbar Vor- und Nachteile. Dies läßt eine quantitative Bestimmung von Faktorenähnlichkeiten wünschenswert erscheinen, welche am besten nach einer Zielrotation der einzelnen Faktorenlösungen auf maximale Ähnlichkeit mit einer Vergleichsmatrix erfolgt. Somit stellt sich die Frage, welche Matrix hierfür gewählt werden soll. Zwei Matrizen bieten sich an: die Semantische Ähnlichkeitsmatrix und die Matrix der retrospektiven Frequenzschätzungen.

3.7 Quantitative Ähnlichkeit der Faktorenstrukturen

3.7.1 Ähnlichkeiten mit der semantischen Ähnlichkeitsmatrix

Die systematische Überlappungshypothese (BORKENAU 1986) läßt erwarten, daß sich hohe Entsprechungen zwischen semantischen Ähnlichkeiten einerseits und Interkorrelationen der Prototypikalitätsratings andererseits zeigen, weil beide die rein semantischen Beziehungen zwischen den Kategorien erfassen. In Maße von Verhaltensfrequenzen, welche Bedeutungsüberlappungen berücksichtigen (retrospektive Schätzungen und multiple on-line Kodierungen mittels Prototypikalitätsratings), gehen demgegenüber neben den semantischen Beziehungen auch individuelle Verhaltensstrukturen ein, weshalb hier die Entsprechungen mit der semantischen Ähnlichkeitsmatrix niedriger ausfallen sollten als bezüglich der reinen Prototypikalitätsratings (vgl. GARA & ROSENBERG 1981). Hingegen gehen in die Interkorrelationen zwischen on-line Kodierungen im forced-choice Verfahren die Bedeutungsüberlappungen vermutlich nicht in systematischer Weise ein. Deshalb sollten die Entsprechungen dieser Matrix zur semantischen Ähnlichkeitsmatrix am geringsten ausfallen.

Um die semantische Ähnlichkeitsstruktur mit den verschiedenen Faktorenstrukturen vergleichen zu können, war es erforderlich, die semantische Ähnlichkeitsmatrix zu faktorisieren. In Anwendung eines von ROWE (1982) vorgeschlagenen Verfahrens erfolgte dies in der Weise, daß die semantischen Ähnlichkeitsschätzungen auf der von -3 bis +3 reichenden Skala durch drei dividiert und diese Werte als Korrelationskoeffizienten eingesetzt wurden³. Entsprechend streuten diese Koeffizienten zwischen -.95 und +.77. Diese Koeffizienten wurden sodann einer Faktorenanalyse mit iterativer Kommunalitätenschätzung unterzogen. Die ersten drei Faktoren erklärten 54% der Gesamtvarianz. Der erste Faktor kontrastierte kooperative mit antagonistischen Verhaltensklassen. Der zweite Faktor wies eine Ladung von .91 für die Variable *Leitet*

³ Diese spezielle Transformation semantischer Ähnlichkeitsschätzungen in Korrelationskoeffizienten ist natürlich willkürlich. Deshalb wurde zusätzlich noch eine zweite Transformation durchgeführt, bei der die gemittelten Ähnlichkeitsschätzungen jeweils durch 4,5 dividiert wurden. Entsprechend wurde etwa der höchste Ähnlichkeitswert von 2,85 in einen Korrelationskoeffizienten von $r = .63$ umgewandelt. Bei diesem Vorgehen resultierten etwas geringere Faktorenähnlichkeiten. Ihre Rangordnung blieb aber davon unberührt, so daß identische Schlußfolgerungen nahegelegt wurden. Aus diesem Grund werden die Ergebnisse dieser Analyse nicht gesondert berichtet.

die *Diskussion* auf und ließ sich entsprechend als Dominanzfaktor interpretieren. Der dritte Faktor kontrastierte aufgabenrelevante mit aufgabenirrelevanten Verhaltensweisen.

Die anderen Ladungsmatrizen wurden nunmehr unter Verwendung eines von GEBHARDT (1968) vorgeschlagenen Verfahrens auf maximale Ähnlichkeit mit dieser Zielmatrix rotiert. Diese Prozedur erlaubt nur orthogonale Lösungen. Sodann wurde ein von HARMAN (1975, S. 272) vorgeschlagener Koeffizient für Faktorenähnlichkeiten berechnet. Dieser Index ermöglicht die Ermittlung von Ähnlichkeiten sowohl für die einzelnen Faktoren als auch für die Gesamtlösung. Die entsprechenden Indizes werden in Tabelle 6 berichtet.

Beim Vergleich mit der Semantischen Ähnlichkeitsmatrix bestätigen sich sämtliche Erwartungen: Die Faktorenstruktur der Prototypikalitätsratings weist die höchste Ähnlichkeit auf und die Faktorenstruktur der on-line Kodierungen im forced-choice Verfahren die geringste. Die Befundmuster für parallel arbeitende Kodierer fallen dabei vergleichbar aus. Weiterhin wird deutlich, daß markante Differenzen sich vor allem bezüglich des ersten Faktors ergeben, welcher kooperative mit antagonistischen Verhaltensweisen kontrastiert. Bei den on-line Kodierungen im forced-choice Verfahren – und nur bei dieser Auswertungsmethode – läßt sich diese bei führerlosen Diskussionsgruppen anson-

Tabelle 6: Faktorenähnlichkeiten beim Vergleich der Semantischen Ähnlichkeitsmatrix mit den übrigen Matrizen

Mit der semantischen Ähnlichkeitsstruktur verglichene Matrix	Ähnlichkeitskoeffizienten			
	F1	F2	F3	Gesamtlösung
Retrospektiv geschätzte Frequenzen	.81	.56	.69	.61
On-line Kodierungen im forced-choice Format				
a) Kodiererin	.02	.64	.70	.41
b) Kodierer	.23	.62	.64	.48
Struktur der Prototypikalitätsratings				
a) Kodiererin	.95	.79	.67	.84
b) Kodierer	.93	.74	.75	.83
Gewichtung der Aktivitäten gemäß Prototypik.				
a) Kodiererin	.91	.64	.77	.78
b) Kodierer	.80	.73	.64	.72
Zu erwartender Wert bei Zufallszahlen				.20

Anm.: Die bei Zufallszahlen zu erwartende Übereinstimmung wurde den Angaben von GEBHARDT (1968) entnommen. F1, F2 und F3 bezeichnen die Ähnlichkeiten für die einzelnen Faktoren.

sten gut etablierte Verhaltensdimension nicht identifizieren.

3.7.2 Ähnlichkeiten mit der Struktur der retrospektiven Frequenzschätzungen

Für die Ladungsmatrix der retrospektiven Frequenzschätzungen als Zielmatrix (siehe Tab. 2) lassen sich keine eindeutigen Erwartungen bezüglich der Rangreihe der Ähnlichkeitskoeffizienten formulieren (vgl. 3.6). Über die gefundenen Ähnlichkeitsmaße, welche im übrigen analog zu den in Tabelle 6 berichteten ermittelt wurden, informiert Tabelle 7. Die Strukturähnlichkeit fällt demzufolge für die Interkorrelationsmatrix der on-line Kodierungen im forced-choice Verfahren am höchsten aus. Zudem lassen sich hier bezüglich aller drei Faktoren für beide Beurteiler substantielle Ähnlichkeitskoeffizienten etablieren. Für die anderen Matrizen sind demgegenüber die Ähnlichkeitsindizes geringer.

4. Diskussion

Im vorliegenden Datenmaterial führt der Vergleich von Faktorenstrukturen zu anderen Ergebnissen als die Berechnung von Korrelationen zwischen Korrelationsmatrizen (vgl. BORKENAU & OSTENDORF 1987, Tab. 10). Dies gilt in zwei-

erlei Hinsicht: Zum einen finden sich substantielle Faktorenähnlichkeiten zwischen retrospektiv geschätzten und im forced-choice Verfahren on-line kodierten Verhaltensfrequenzen. Zum anderen führt die Berücksichtigung von Bedeutungsüberlappungen durch das on-line Kodierschema *nicht* zu einer größeren Faktorenähnlichkeit von on-line kodierten und retrospektiv geschätzten Verhaltensfrequenzen. In einer anderen Arbeit (BORKENAU & OSTENDORF 1987) verglichen wir die auch in der vorliegenden Studie analysierten Korrelationsmatrizen mittels der Berechnung von Korrelationen zwischen Korrelationsmatrizen. Dabei ergab sich, daß die Berücksichtigung überlappender Aktivitäten im on-line Kodierschema höhere Entsprechungen sowohl mit der Semantischen Ähnlichkeitsmatrix als auch mit der Matrix der retrospektiven Frequenzschätzungen mit sich bringt.

Wenn bei dem in der vorliegenden Studie vorgenommenen Vergleich von Faktorenähnlichkeiten die Berücksichtigung von Bedeutungsüberlappungen bei der on-line Kodierung zu einer größeren Ähnlichkeit nur mit der Semantischen Ähnlichkeitsmatrix führte, so hat dies seine Ursache vermutlich darin, daß sich in der Faktorenstruktur auch Mittelwert und Streuung der Korrelationskoeffizienten niederschlagen. Retrospektive Frequenzschätzungen und on-line Kodierungen im forced-choice Verfahren teilen in den vorliegenden Daten nämlich das

Tabelle 7: Faktorenähnlichkeiten beim Vergleich der Struktur der retrospektiven Frequenzschätzungen mit den übrigen Matrizen.

Mit der Struktur der retrospektiven Frequenzschätzungen verglichene Matrix	Ähnlichkeitskoeffizienten			
	F1	F2	F3	Gesamtlösung
Semantische Ähnlichkeitsbeziehungen	.76	.57	.32	.61
On-line Kodierungen im forced-choice Format				
a) Kodiererin	.81	.78	.82	.80
b) Kodierer	.85	.73	.82	.79
Struktur der Prototypikalitätsratings				
a) Kodiererin	.68	.51	.80	.64
b) Kodierer	.58	.53	.15	.47
Gewichtung der Aktivitäten gemäß Prototypik.				
a) Kodiererin	.74	.57	.81	.69
b) Kodierer	.62	.63	.35	.57
Zu erwartender Wert bei Zufallszahlen				.20

Anm.: Die bei Zufallszahlen zu erwartende Übereinstimmung wurde den Angaben von GEBHARDT (1968) entnommen. Die Zielmatrix ist mit der in Tabelle 2 berichteten Ladungsmatrix identisch.

Merkmal einer positiven Mannigfaltigkeits-Struktur. Bei der Berücksichtigung von Bedeutungsüberlappungen im on-line Kodierschema finden sich demgegenüber häufig auch negative Interkorrelationen (vgl. BORKENAU & OSTENDORF 1987, Tab. 9) und Faktorenladungen. Dieser Umstand dürfte dafür verantwortlich sein, daß die retrospektiven Frequenzratings die größte Faktorenähnlichkeit zu den on-line Kodierungen mittels forced-choice Verfahren aufweisen. Spricht dieser Umstand nun generell für eine Einfachklassifikation von Verhalten?

Gegen eine solche Schlußfolgerung sprechen mehrere Überlegungen: Erstens ist dagegen einzuwenden, daß bei der Wahl einer anderen Skala für die Prototypikalitätsratings sich eine andere Faktorenstruktur der on-line kodierten Verhaltensfrequenzen, welche Bedeutungsüberlappungen berücksichtigen, ergeben hätte. Die Prototypikalitäten der Aktivitäten bezüglich der 16 Kategorien waren ja auf einer von -3 bis +3 reichenden Skala eingeschätzt worden, wobei eben diese Werte später, bei der Berechnung von Frequenzscores für die einzelnen Diskutanten, als Gewichte dienten. Die Wahl dieser Skala ist aber beliebig. Ebenso gut hätten wir z. B. eine von 1 bis 7 reichende Skala heranziehen können, wie es etwa BUSS & CRAIK (1983) vorschlagen. Wir experimentierten auch mit einer solchen Skala, indem den jeweils abgegebenen Prototypikalitätsratings eine Konstante von 4 hinzuaddiert wurde. Setzten wir die daraus resultierenden Werte als Gewichte bei der Ermittlung von Frequenzscores ein und korrelierten diese sodann über die 48 Diskutanten, so ergaben sich durchgängig Interkorrelationen über .98 zwischen den 16 Kategorien. Die so ermittelten Frequenzscores spiegelten mithin in fast reiner Form die unterschiedliche Gesamtaktivität der Diskutanten wider. Es überrascht deshalb auch nicht weiter, daß eine Faktorenanalyse dieser Interkorrelationsmatrix zu einem überwältigenden Generalfaktor führt, welcher 96,8% (Beurteilerin) bzw. 97,7% (Beurteiler) der Gesamtvarianz aufklärt.

Sicherlich wäre es möglich gewesen, Zwischenlösungen zu erzeugen, indem statt einer 4 eine niedrigere Konstante gewählt worden wäre. Einer solchen Prozedur hätte aber jegliche theoretische Rechtfertigung gefehlt. Hierin liegt denn auch die Crux bezüglich der on-line Kodie-

rung von Verhalten: Ein forced-choice Verfahren ist insofern unangemessen, als es die Bedeutungsüberlappungen zwischen Kategorien systematisch ausblendet und somit der «strukturellen Treue» (LOEVINGER 1957) der Frequenzscores entgegenwirkt. Demgegenüber tritt bei einem Prototypikalitäts-Kodierschema (BORKENAU 1986) bzw. einem Skalierungs-Kodierschema (ROMER & REVELLE 1984) das Problem auf, daß die Befunde in hohem Maße von der gewählten Skala abhängig sind, deren Wahl beim gegenwärtigen Kenntnisstand beliebig ist. Dieses Dilemma wird bei der Berechnung von Korrelationen zwischen Korrelationsmatrizen nicht offenbar. Es zeigt sich aber bei dem hier vorgenommenen Vergleich von Faktorenstrukturen.

Allerdings ist einschränkend zu erwähnen, daß die Bewährungschancen für das on-line Kodierschema, welches Bedeutungsüberlappungen berücksichtigt, in der vorliegenden Studie ausgesprochen ungünstig waren. Dessen Probleme resultieren ja daraus, daß dieses Kodierschema die ganz überwiegend positiven Interkorrelationen zwischen den retrospektiven Ratings nicht abbildet. Eine solche positive Mannigfaltigkeit ist jedoch recht spezifisch für die vorliegende Studie: Erstens resultiert sie aus dem Umstand, daß durchgängig Verbalverhalten in Diskussionsgruppen analysiert wurde. Ähnlich krasse individuelle Unterschiede in der Gesamtaktivität dürften sich bei Einbezug zusätzlicher Verhaltenskategorien (z. B. lesen, um sich schauen, usw.) nicht ergeben. Zwar ergriffen die passiveren Diskutanten in unserer Untersuchung selten das Wort, aber sie waren dennoch nicht gänzlich inaktiv. Lediglich die benutzten Verhaltenskategorien waren unsensibel für ihre Art von Aktivität.

Zweitens resultiert die positive Mannigfaltigkeits-Struktur der retrospektiven Einschätzungen wohl daraus, daß Verhaltensfrequenzen erhoben wurden und keine Eigenschaftsausprägungen. Bei unterschiedlicher Gesamtaktivität der untersuchten Personen ist aber für Verhaltensfrequenzen eher eine positive Mannigfaltigkeits-Struktur zu erwarten, da es für Verhalten keine Gegenbeispiele gibt. Wenn etwa die Anzahl dominanter Handlungen einer Person ausgezählt wird, so ist hierbei die Anzahl submissiver Handlungen unbedeutend. Anders verhält es sich hingegen mit der Eigenschaft *Dominanz*.

Bei deren Messung dürfte es sinnvoll sein, submissive Verhaltensweisen als Gegenbeispiele zu berücksichtigen.

Mithin ist die Generalisierbarkeit des hier erzielten Befundes, daß die Berücksichtigung von Bedeutungsüberlappungen zwischen Verhaltenskategorien nicht zu einer Angleichung der Faktorenstrukturen retrospektiv geschätzter und on-line kodierter Verhaltensfrequenzen führt, fraglich. Verschiedene Überlegungen sprechen vielmehr dafür, daß die positive Mannigfaltigkeits-Struktur der retrospektiven Ratings recht spezifisch für unser setting und die von uns benutzten Kategorien ist. Insbesondere für Trait-Ratings dürften unsere Befunde ausgesprochen unrepräsentativ sein. Diese Einschränkung gilt aber nur für den Vergleichsmaßstab der retrospektiven Ratings. Bezüglich des Vergleichsmaßstabs der semantischen Ähnlichkeitsbeziehungen konnte die vorliegende Studie nämlich zeigen, daß eine Berücksichtigung von Bedeutungsüberlappungen zwischen den Kategorien im on-line Kodierschema zu einer Angleichung der Faktorenstrukturen führt.

Literatur

- 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.
- BALES, R. F. & COHEN, S. P.: *SYMLOG: Ein System für die mehrstufige Beobachtung von Gruppen*. Stuttgart: Klett-Cotta, 1982.
- BLOCK, J.: Advancing the psychology of personality: Paradigmatic shift or improving the quality of research? In: MAGNUSSON, D. & ENDLER, N. S. (Hrsg.): *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N. J.: Erlbaum, 1977.
- BLOCK, J., WEISS, D. S. & THORNE, A.: How relevant is a semantic similarity interpretation of personality ratings? *Journal of Personality and Social Psychology*, 1979, 37, 1055–1074.
- BORKENAU, P.: Toward an understanding of trait-interrelations: Acts as instances for several traits. *Journal of Personality and Social Psychology*, 1986, 51, 371–381.
- BORKENAU, P. & OSTENDORF, F.: Retrospective estimates of act frequencies: How accurately do they reflect reality? *Journal of Personality and Social Psychology*, 1987, 52, 626–638.
- BUSE, L. & PAWLIK, K.: Inter-Setting Korrelationen und Setting-Persönlichkeit-Wechselwirkungen: Ergebnisse einer Felduntersuchung zur Konsistenz von Verhalten und Erleben. *Zeitschrift für Sozialpsychologie*, 1984, 15, 44–59.
- BUSS, D. M. & CRAIK, K. H.: The act frequency approach to personality. *Psychological Review*, 1983, 90, 105–126.
- CANTOR, N. & MISCHEL, W.: Prototypes in person perception. In: BERKOWITZ, L. (Hrsg.): *Advances in Experimental Social Psychology* (Bd. 12, S. 3–52). New York: Academic Press, 1979.
- CATTELL, R. B.: The 16-PF and basic personality structure: A reply to Eysenck. *Journal of Behavioral Science*, 1972, 1, 169–187.
- D'ANDRADE, R. G.: Memory and the assessment of behavior. In: BLALOCK, H. M. (Hrsg.): *Measurement in the social sciences* (pp. 159–186). Chicago: Aldine-Atherton, 1974.
- DIGMAN, J. M. & INOUE, J.: Further specification of the five robust factors of Personality. *Journal of Personality and Social Psychology*, 1986, 50, 116–123.
- EYSENCK, H. J.: Personality and factor analysis: A reply to Guilford. *Psychological Bulletin*, 1977, 84, 405–411.
- EYSENCK, H. J. & EYSENCK, M. W.: *Personality and Individual Differences*. New York: Plenum Press, 1985.
- FISKE, D. W.: Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology*, 1949, 44, 329–344.
- GARA, M. A. & ROSENBERG, S.: Linguistic factors in implicit personality theory. *Journal of Personality and Social Psychology*, 1981, 41, 450–457.
- GEBHARDT, F.: Über die Ähnlichkeit von Faktormatrizen. *Psychologische Beiträge*, 1968, 9, 591–599.
- GUILFORD, J. P.: Factors and factors of personality. *Psychological Bulletin*, 1975, 82, 802–814.
- GUILFORD, J. P.: Will the real factor of Extraversion-Introversion please stand up? A reply to Eysenck. *Psychological Bulletin*, 1977, 84, 412–416.
- HAMPSON, S. E.: Person memory: A semantic category model of personality traits. *British Journal of Psychology*, 1982, 73, 1–11.
- HARMAN, H. H.: *Modern factor analysis*. Chicago: University of Chicago Press, 1975.
- HÄCKER, H., SCHWENKMETZGER, P. & URTZ, H.: Empirische Untersuchungen zur Bedeutung objektiver Testfaktoren: Eine kritische Auseinandersetzung mit Cattells Faktorenkonzeption. *Diagnostica*, 1982, 28, 117–130.
- LOEVINGER, J.: Objective tests as instruments of psychological theory. *Psychological Reports*, 1957, 3, 635–694.
- MCCRAE, R. R. & COSTA, P. T.: Validation of the five factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 1987, 52, 81–91.
- MISCHEL, W.: *Personality and assessment*. New York: Wiley, 1968.
- MISCHEL, W. & PEAKE, P. K.: Beyond déjà vu in the search for cross-situational consistency. *Psychological Review*, 1982, 89, 730–755.
- MULAİK, S. A.: Are personality factors raters' conceptual factors? *Journal of Consulting Psychology*, 1964, 28, 506–511.

- NEWCOMB, T. M.: *Consistency of certain extrovert-introvert behavior patterns in 51 problem boys*. New York: Columbia University. Teachers College, Bureau of Publications, 1929.
- NORMAN, W. T.: Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 1963, 66, 574–583.
- PAWLIK, K.: *Dimensionen des Verhaltens*. Bern: Huber, 1968.
- PAWLIK, K. & BUSE, L.: Rechnergestützte Verhaltensregistrierung im Feld: Beschreibung und erste psychometrische Überprüfung einer neuen Erhebungsmethode. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1982, 3, 101–118.
- REVENSTORF, D.: Vom unsinnigen Aufwand. *Archiv für Psychologie*, 1978, 130, 1–36.
- ROMER, D. & REVELLE, W.: Personality traits: Fact or fiction? A critique of the Shweder and D'Andrade systematic distortion hypothesis. *Journal of Personality and Social Psychology*, 1984, 47, 1028–1042.
- ROSCH, E. & MERVIS, C. B.: Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, 7, 573–605.
- ROWE, D. C.: Monozygotic twin cross-correlation as a validation of personality structure: A test of the semantic bias hypothesis. *Journal of Personality and Social Psychology*, 1982, 43, 420–428.
- SEMIN, G. R. & GREENSLADE, L.: Differential contributions of linguistic factors to memory-based ratings: Systematizing the systematic distortion hypothesis. *Journal of Personality and Social Psychology*, 1985, 49, 1713–1723.
- SHWEDER, R. A.: How relevant is an individual difference theory of personality? *Journal of Personality*, 1975, 43, 455–484.
- SHWEDER, R. A.: Fact and artifact in trait perception: The systematic distortion hypothesis. *Progress in Experimental Personality Research*, 1982, 11, 65–100.
- SHWEDER, R. A. & D'ANDRADE, R. G.: The systematic distortion hypothesis. In: SHWEDER, R. A. (Hrsg.): *New directions for methodology of social and behavioral science* (Bd. 4, S. 37–58). San Francisco: Jossey-Bass, 1980.
- THORNDIKE, E. L.: A constant error in psychological ratings. *Journal of Applied Psychology*, 1920, 4, 25–29.
- TUPES, E. C. & CHRISTAL, R. E.: *Recurrent Personality Factors Based on Trait Ratings*. USAF ASD Technical Report, 1961, No. 61–97.

Anschrift der Verfasser: PD Dr. Peter Borkenau, Dipl.-Psych. Fritz Ostendorf, Abteilung für Psychologie der Universität Bielefeld, Postfach 8640, D-4800 Bielefeld.