

B I E L E F E L D E R A R B E I T E N Z U R S O Z I A L -
P S Y C H O L O G I E

Nr. 19

(April 1977)

Hans Dieter Mummendey,
Bernhard Hesener und
Margret Isermann:

Selbstwahrnehmung als interper-
sonelle Wahrnehmung: Methoden
und Ergebnisse

Zusammenfassung:

In zwei Experimenten wurde die Annahme überprüft, daß Personen auch bei der Selbstbeurteilung differenzierter und extremer urteilen, wenn sie über mehr Informationen über den Urteilsgegenstand verfügen. Ein Experiment mit studentischen Untersuchern und Teilnehmern führte zu keiner Bestätigung. Dagegen fand sich in einem Experiment mit Wirtschaftsfachhochschulern folgendes: Personen, mit denen ein längeres Gespräch über ihre Persönlichkeit geführt wurde, geben häufiger polarisierte und seltener mittlere Selbsteinschätzungen ab. Personen, mit denen ein solches Gespräch geführt wurde und Personen, denen man ihre Ergebnisse in Intelligenz- und Persönlichkeitstests mitteilte, neigen dazu, sich etwas differenzierter zu beurteilen als Personen einer Kontrollgruppe. Bei verschiedenen Gesprächsleitern zeigen die Selbstbeurteiler unterschiedliches Urteilsverhalten.

Problemstellung und Hypothesen

Ausgehend von der begründeten Auffassung, das intrapersonelle Urteilsverhalten sei dem interpersonellen im wesentlichen funktional äquivalent, wurde im ersten Teil der vorliegenden Arbeit (SCHMIDT 1976) argumentiert, wichtige und empirisch hinreichend belegte Erkenntnisse über Bedingungen der Genauigkeit interpersonellen Urteilens müßten in analoger Weise bei intrapersoneller Beurteilung, also bei der Selbstwahrnehmung bzw. Selbstbeurteilung wirksam sein.

Als Bedingungsvariable der Genauigkeit interpersoneller Urteile wurde der Grad der Vertrautheit zwischen Urteilssubjekt und -objekt, also zwischen Beurteiler und Beurteiltem diskutiert.

Als gut gesichertes Resultat der Personwahrnehmungsforschung erscheint, daß Personen, die einander gut kennen bzw. über mehr Informationen über den jeweils anderen verfügen, einander differenzierter beurteilen. Mangelnde Urteilsdifferenziertheit drückt sich unmittelbar im Ausmaß des Halo-Effekts, des prominentesten Beurteilungsfehlers bei interpersonellen Urteilsprozessen aus.

Hier soll eine längere, zweifellos berechtigte Diskussion darüber, inwieweit ein Halo-Effekt stets als "Fehler" aufzufassen sei, da er doch gerade über wesentliche Stile des Kognizierens sozialer Objekte Auskunft gebe, vermieden werden. Halo-Effekte sind Ausdruck relativ einseitiger, undifferenzierter Beurteilung und lassen sich durch geeignete, immunisierende Instruktionen nachhaltig reduzieren (SCHMIDT 1966). Insofern die Präzision bzw. Zuverlässigkeit subjektiver Urteile - in der Psychologie am häufigsten durch Rating-Skalen erfaßt - mangels eines angemesseneren urteilstheoretischen Modells als Grad der relativen Abwesenheit typischer Urteilstendenzen wie des Halo-Effekts definiert werden kann, Urteilsgenauigkeit also u.a. durch Urteilsdifferenziertheit ausgedrückt werden kann, wirken Halo-Effekte reliabilitätsmindernd und tragen zur Fehlervarianz der Urteile bei.

Personen, die über eine größere Menge von Information über eigene Persönlichkeitseigenschaften verfügen, müßten daher bei der Selbstbeurteilung in geringerem Maße einem Halo-Effekt unterliegen und sich somit in dem genannten Sinne zuverlässiger beurteilen.

Entsprechendes kann für die Extremität von Urteilen gelten. Personwahrnehmungs-Urteile werden - auch dies erscheint als ein empirisch gut gesicherter Befund - mit anwachsender, die Stimulusperson beschreibender Informationsmenge extremer. Dies bedeutet im Zusammenhang der Erfassung von Persönlichkeitseigenschaften mit Rating-Skalen, daß mittlere bzw. neutrale Skalenpositionen eher vermieden und stärker extreme, polarisierte Urteile abgegeben werden. Auch diese Urteilstendenz könnte man im Sinne größerer Rating-Zuverlässigkeit interpretieren, da ihr Gegenteil einem weiteren, traditionell als Fehler bezeichneten Urteilseffekt, nämlich der Tendenz zur Mitte (error of central tendency) entspricht.

Personen, denen eine größere Menge von Information über eigene Persönlichkeitseigenschaften zur Verfügung steht, müßten sich daher eindeutiger, d.h. extremer beurteilen.

Ungeklärt erscheint bei der hier vorgenommenen Übertragung von Zuverlässigkeitskonzeptionen der Fremdbeurteilung auf die Selbsteinschätzung, inwiefern die Vermeidung eines Halo-Effekts von der Vermeidung des Fehlers der zentralen Tendenz theoretisch und empirisch unabhängig ist. Eine konsequent eindeutige, extreme Selbsteinschätzung wäre nämlich zugleich relativ undifferenziert. Je nachdem welche Methode zur Erfassung des Halo-Effekts angewendet würde, schlossen Urteilsgenauigkeit im Sinne von Urteilsextremität und Urteilsgenauigkeit im Sinne von Halo-Vermeidung einander teilweise aus. Man könnte sich, um diese Schwierigkeit zu umgehen, für nur eines dieser Fehlerkonzepte entscheiden. Eine zweite Möglichkeit bestünde darin, experimentelle Annahmen getrennt zu formulieren. Bei der Formulierung der Hypothesen wurde die zweite Lösung vorgezogen.

Da man sich unter der einer Person (über andere Personen oder über sich selbst) zur Verfügung gestellten Information Unter-

schiedliches vorstellen kann, sollten schließlich die der Untersuchung zugrundeliegenden Annahmen bezüglich der Art der Information differenziert werden. Einmal erscheint es möglich, jemandem eine größere Menge gleichsam abstrakt vorgetragener Information, z.B. Testdaten, zugänglich zu machen. Zum anderen wäre es möglich, spezifischere, aber für die Person vermutlich bedeutsamere Informationen zu vermitteln. "Viel Information" hätte in diesem Sinne einen "qualitativen" Klang. Daher müßten sich Personen umso differenzierter und umso extremer beurteilen können, je mehr Informationen sie über einen Verhaltenssektor ihrer Persönlichkeit erhielten, der aus persönlich bedeutsamen Tätigkeiten (beruflicher, familiärer Art) besteht.

Demgemäß wurden folgende Hypothesen formuliert:

- 1) Personen, denen eine größere Menge von Informationen über eigene Persönlichkeitseigenschaften zur Verfügung gestellt wird, werden sich in einem mehrdimensionalen Selbsteinschätzungsverfahren a) differenzierter (im Sinne eines geringeren Halo-Effekts) b) eindeutiger (im Sinne größerer Urteilsextremität) beurteilen als solche Personen, denen eine entsprechende Menge von Informationen über sich selbst nicht zur Verfügung steht.
- 2) Personen, denen eine größere Menge von Informationen über eigenes Verhalten und eigene, hierfür als relevant angesehene Persönlichkeitsmerkmale in gut vertrauten und für die Person bedeutsamen Alltagssituationen zur Verfügung gestellt wird, werden sich in einem mehrdimensionalen Selbsteinschätzungsverfahren a) differenzierter b) eindeutiger beurteilen als solche Personen, denen Informationen dieser Art nicht zur Verfügung stehen.
- 3) Personen, denen Informationen über eigenes Verhalten und eigene Persönlichkeitsmerkmale in gut vertrauten, alltagsrelevanten Situationen zur Verfügung stehen, werden sich a) differenzierter b) eindeutiger beurteilen als solche Personen, denen lediglich eine große Zahl von Informationen über eigene Persönlichkeitseigenschaften (i.S.d.1.Hypothese) zur Verfügung stehen.

Methoden und Ergebnisse

Zur Prüfung der Hypothesen wurden zwei Experimente ausgeführt. Experiment I hatte den Charakter eines Vorversuchs, da es von einer Praktikumsgruppe von Studenten ausgeführt wurde. Die Un-

tersucher waren also einerseits relativ unerfahren im Umgang mit den Methoden und den Versuchspersonen. Andererseits konnte in Experiment I eine größere Gruppe von Untersuchern und Versuchspersonen gewonnen werden. Experiment II wurde von den Autoren ausgeführt. Die Zahl der Versuchspersonen war hier geringer, doch konnte in diesem Falle der gesamte Versuchsplan realisiert werden. Experiment I zielte nur auf Hypothese 2.

EXPERIMENT I

In diesem Experiment hatten Gruppen von zwei oder drei Studenten die Aufgabe, jeweils zwei einander in mehrfacher Hinsicht ähnliche, männliche Studenten auszusuchen ("experimentelle Zwillinge"). Sie sollten aus nicht-sozialwissenschaftlichen Fachrichtungen stammen, den Untersuchern nicht näher bekannt sein, gleichen Studiengang, gleiches Semester und gleiches Alter (nicht über 30) aufweisen und - soweit erkennbar - vergleichbare weltanschauliche/hochschulpolitische Einstellungen aufweisen (z.B. beide aus RCDS, KBW oder Evang.Studentengemeinde etc.). Auf diese Weise wurden 40 Personen, d.h. 20 Paare gewonnen.

Je eine Person eines Paares wurde sodann der Bedingung "Selbstbezogene Information" (i.S.d.Hypothese 2) und der Bedingung "Keine Information" unterworfen. Selbstbezogene Information sollte in einem maximal einstündigen Gespräch, das die Studenten mit der Vp über deren Charaktereigenschaften führten, gegeben werden (bzw. aktualisiert werden). Der Befragte sollte nach seinen allgemeinen Persönlichkeitsmerkmalen gefragt werden, ferner nach Eigenschaften, die sich im beruflichen und im privaten Bereich für den Betreffenden seiner Meinung nach als eher vorteilhaft oder weniger vorteilhaft auswirkten. Den groben Aufbau der Gesprächsführung diskutierten die Untersucher zuvor in einem Praktikum anhand zweier praktischer Fälle. Alle Vpn füllten im Anschluß an das Gespräch den von JOHN & KEIL (1972) adaptierten Selbsteinschätzungsbogen aus, der aus 18 bipolar verbal verankerten Rating-Skalen besteht (vgl. MUMMENDEY, MIELKE, MAUS & HESENER 1977): Führungsrolle, Person und Fähigkeiten, Geselligkeit, Ausgeglichenheit im Umgang, Selbstsicherheit,

Probleme und Beschwerden, Verantwortungsbewußtsein, Gut in Gesellschaft hineingewachsen, Selbstbeherrschung, Toleranz, Um guten Eindruck bemüht, Schwierigkeiten in Gemeinschaft, Leistungsstärke und Anpassungsfähigkeit, Selbständigkeit und Leistungsstärke, Geistige Fähigkeiten, Verständnis für Andere, Beweglichkeit und Anpassungsfähigkeit, Weibliche Interessen. Der Versuchsplan zu Experiment I stellte sich also schematisch wie folgt dar:

<u>Gruppe 1:</u> (Gesprächsgruppe)	}	(paralleliert nach persönlichen Merkmalen)	}	<u>Gespräch</u>	<u>Selbst-Rating</u>
<u>Gruppe 2:</u> (Kontrollgruppe)				<u>Kein Gespräch</u>	<u>Selbst-Rating</u>

Von den 20 Paaren von Vpn waren die meisten zwischen 20 und 23 Jahre alt, hauptsächlich Mathematiker und Juristen im dritten bis fünften Semester, die überwiegend liberalen und sozialdemokratischen Richtungen nahestanden. Die Gespräche dauerten (nach Angaben der Gesprächsleiter) zur Hälfte etwas weniger als eine Stunde, zur Hälfte eine Stunde und länger.

EXPERIMENT II

In diesem Versuch sollten alle drei bei der Hypothesenbildung unterschiedenen Bedingungen realisiert werden. 46 Fachhochschüler (3.Semester) der Fachrichtung Wirtschaft (im Alter zwischen 20 und 26; Durchschnittsalter 23 Jahre), die von einem ihrer Dozenten zur Teilnahme an einem psychologischen Versuch angeworben worden waren, wurden im Anschluß an eine Vorlesung über Investitionsrechnung nach Zufall in drei Gruppen aufgeteilt und den Bedingungen "Selbstbezogene Information: Gespräch", "Selbstbezogene Information: Testdaten" und "Keine Information" unterworfen.

Bedingung "Selbstbezogene Information: Gespräch": Wie in Experiment I wurde mit der Person ein Gespräch über allgemeine Charakteristika ihres Verhaltens bzw. allgemeine Persönlichkeitseigenschaften sowie über die für den beruflichen und pri-

vaten Bereich mehr oder weniger förderlichen Eigenschaften gesprochen. Dabei sollte herausgearbeitet werden, welches die für die Person besonders bedeutsamen Situationen zu sein schienen, ferner, welches für sie wichtige Bezugsgruppen oder -personen sind. Indem bewußt zwischen Eigenschaften derselben Person in unterschiedlichen Situationen unterschieden wurde, sollte nicht nur auf Unterschiede zu anderen Menschen, sondern auch auf Unterschiede von Eigenschaften innerhalb derselben Person abgehoben werden. Den Befragten gegenüber wurde das Gespräch damit begründet, daß herauszufinden sei, "in welchem Maße Persönlichkeitseigenschaften für bestimmte, für eine Person besonders relevante Situationen nützlich oder hinderlich sein können". Wegen des Zeitaufwandes bei diesem Experiment wurde das Gespräch auf eine halbe Stunde beschränkt, doch herrschte bei den Gesprächsleitern der Eindruck vor, daß es sich bei den Befragten überwiegend um sehr interessierte und intensiv mitarbeitende Personen handelte, so daß die Zeit nicht wesentlich zu kurz angesetzt schien. Anschließend wurden die Personen gebeten, das 18 teilige Selbst-Rating vorzunehmen.

Bedingung "Selbstbezogene Information: Testdaten": Mit den Vpn dieser Gruppe wurden zwei Tests ausgeführt, und es wurde ihnen angekündigt, daß die Testergebnisse anschließend mitgeteilt würden. Zunächst wurde eine Kurzform des "Intelligenz-Struktur-Test" (IST, Form A) von AMTHAUER (1953) mit den Untertests "Satzergänzung" (SE), "Analogien" (AN), "Gemeinsamkeiten" (GE) und "Rechenaufgaben" (RA) angewendet (Dauer: 30 Minuten). Sodann wurde die Halbform A des "Freiburger Persönlichkeits-Inventar" (FPI) von FAHRENBERG & SELG (1970) gegeben (Dauer: ca. 30 Minuten), so daß Aussagen über "Nervosität", "Aggressivität", "Depressivität", "Erregbarkeit", "Geselligkeit", "Gelassenheit", "Dominanzstreben", "Gehemmtheit", "Offenheit", "Extraversion", "Emotionale Labilität" und "Maskulinität" möglich wurden. Die Ergebnisse des Intelligenz- und des Persönlichkeitstests wurden anschließend der Vpn-Gruppe wie folgt erläutert: Es wurden Durchschnittsnormen für männliche Personen im Alter zwischen 15 und 30 Jahren angegeben (ungefähre Mittelwerte, ferner der Bereich einer Standardabweichung unter und

über diesen Werten), so daß jede Vp für sich feststellen konnte, ob sie im Mittelbereich, eher unter oder über diesem Bereich lag. Die Intelligenz-Untertests wurden dabei weitgehend vulgärpsychologisch gedeutet (z.B. in Bezug auf "eher praktische Begabung", "eher theoretische Begabung" usw; etwa so wie bei AMTHAUER, a.a.O., 19 und 25). Unmittelbar im Anschluß an diese "Erläuterung der Testergebnisse" wurden die Vpn gebeten, das 18 teilige Rating ihrer eigenen Person vorzunehmen.

Bedingung "Keine Information": Die unter dieser Bedingung untersuchten Personen nahmen lediglich die Selbstbeurteilung auf den 18 Rating-Skalen vor.

Somit stellte sich der Versuchsplan in Experiment II wie folgt dar:

<u>Gruppe 1:</u> (Gespr.Gr.)] (randomisiert)	<u>Gespräch</u>	<u>Selbst-Rating</u>
<u>Gruppe 2:</u> (Test-Gr.)		<u>Tests</u>	<u>Selbst-Rating</u>
<u>Gruppe 3:</u> (Kontr.Gr.)		-----	<u>Selbst-Rating</u>

Es gelang in der Praxis der Untersuchung nicht, gleich viele Personen auf alle drei Teilnehmergruppen zu vereinigen. Die Gesprächs-Gruppe umfaßte schließlich 16, die Test-Gruppe 10 und die Kontrollgruppe 20 Personen.

Abhängige Variablen in Experiment I und II

Urteilsextremität. Da es sich bei allen Ratings um siebenstufige numerische Ratings handelte (Extreme "3", Skalenmitte "0"), erschienen Häufigkeitsvergleiche der Verwendung dieser Skalenpositionen als einfacher Indikator der Urteilsextremität angebracht.

Urteilsdifferenziertheit. Zur Erfassung der Urteilsdifferenzierung im Sinne einer Vermeidung des Halo-Effekts jeder Person (bzw. der jeweiligen Versuchs- oder Kontrollgruppe) in Bezug auf die 18

Selbstbeurteilungsskalen erschienen verschiedene Verfahren angebracht, da es unterschiedliche Ansätze zur Erfassung von Halo-Effekten gibt.

a) Als grobes und oberflächliches Verfahren zur Schätzung der Urteilsdifferenzierung bzw. eines Halo-Effekts einer Gruppe von Beurteilern kann die durchschnittliche Höhe der Interkorrelationen zwischen den beurteilten Eigenschaften gelten, da Halo-Effekte zur Erhöhung der Interkorrelationen von Beurteilungen führen (THORNDIKE 1920). Daher soll das arithmetrische Mittel der 153 Korrelationskoeffizienten, die den Zusammenhang der 18 zu beurteilenden Eigenschaften beschreiben, zur ersten Schätzung des Halo-Effekts herangezogen werden.

b) Urteilsdifferenzierung in diesem Sinne könnte auch aus dem Eigenwertabfall einer Faktorenanalyse nach dem Hauptkomponentenmodell der 18 Variablen pro Gruppe erschlossen werden. Damit würde die Zahl interpretierbarer Faktoren und das Maß der von ihnen aufgeklärten Varianz zu einer weiteren Beschreibungsgröße für Urteilsdifferenzierung bei einer Beurteilerguppe.

c) Erwünscht wäre ferner ein Verfahren, das die Größe des Halo-Effekts oder seiner Vermeidung möglichst präzise angibt. Das Verfahren von GUILFORD (1954, 280 ff.), angewendet bei SCHMIDT (1966), schätzt die Größe des Halo-Effekts aus dem Varianzanteil der Wechselwirkung zwischen Beurteilern und Beurteilten, wenn eine Mehrzahl von Eigenschaften beurteilt wird. Es kommt hier nicht in Frage, da eine dieser drei Varianzquellen fehlt. RUDINGER & FEGER (1970) haben zudem darauf hingewiesen, daß man wegen wiederholter Messungen in zwei der drei Faktoren ein erhöhtes Risiko zweiter Art eingeht. Das von ihnen bevorzugte Verfahren von GLESER, CRONBACH & RAJARATNAM (1965) scheidet ebenfalls aus, da auch hier für mehrere Personen von mehreren Beurteilern Beurteilungen in mehreren Eigenschaften vorzuliegen haben. Stattdessen kann die von WINER (1971, 288) beschriebene varianzanalytische Methode zur Reliabilitätsbestimmung Anwendung finden, bei welcher die Differenz der Varianzausdrücke "Zwischen Beurteilten" und "Innerhalb Beurteilten" ins Verhältnis zur Varianz innerhalb der Beurteilten gesetzt wird. WINER diskutiert diesen Fall für die Beurteilung einer einzigen Eigenschaft bei mehreren (n) Beurteilten (wiederholte Messungen) durch mehrere (k) Beurteiler (unabhängige Stichproben). Dieser Fall der Fremdbeurteilung wird für unser Problem adaptiert, wenn bei der Selbstbeurteilung mehrere (n) eigene Eigenschaften (wiederholte Messungen) von mehreren (k) Beurteilern (unabhängige Stichproben) beurteilt werden. Aus der für den Fehler des Verhältnisses zweier Varianzkomponenten korrigierten Schätzgröße

$$\Theta' = (\text{MQUS Zw.Eig.} - m \cdot \text{MQUS In.Eig.}) / k \cdot m \cdot \text{MQUS In.Eigensch.},$$

wobei $m = (n(k-1)) / (n(k-1)-2)$,

läßt sich eine Schätzung der Reliabilität des Durchschnitts von k Beurteilungen errechnen, die dem Intraclass-Koeffizienten von EBEL (1951) entspricht:

$$r_k' = k \cdot \Theta' / (1 + k \cdot \Theta').$$

Die gute Eignung varianzanalytischer Maße wie des vorliegenden zur Reliabilitätsbestimmung von Ratings wurde bereits von mehre-

ren Autoren für den Fall von Fremdbeurteilungen bestätigt (WERNER 1976, BINTIG 1976). Für einen präzisen Vergleich der Urteilsdifferenziertheit der beiden Gruppen in Experiment I und der drei Gruppen in Experiment II kommt demnach schließlich der Vergleich der Reliabilitätskoeffizienten r_k in Frage.

Ergebnisse

EXPERIMENT I

Urteilsextremität. Da Versuchs- und Kontrollgruppe gleiche Größe haben, kann die Nutzung der extremen Urteilkategorien "3" bei beiden Gruppen direkt auf Häufigkeitsunterschiede geprüft werden. Die Differenz der Häufigkeiten der Gesprächsgruppe (74) und der Kontrollgruppe (60) ist nicht interpretierbar ($\text{Chi}^2 = 1.46$; $\text{df} = 1$; $p > 0,05$). Gleiches gilt für die Häufigkeiten, mit denen die neutrale Urteilkategorie "0" bei der Gesprächsgruppe (127) und der Kontrollgruppe (140) auftrat ($\text{Chi}^2 = 0.63$; $\text{df} = 1$; $p > 0.05$).

Urteilsdifferenzierung. (a) Der Durchschnitt der Interkorrelationen beträgt in beiden Gruppen gleichermaßen 0.23. Eine zufallskritische Differenzprüfung erübrigt sich daher.

(b) Die Faktorenanalyse der 18 Variablen erbringt bei beiden Gruppen sechs Faktoren mit Eigenwerten über 1.0, die bei der Gesprächsgruppe 74%, bei der Kontrollgruppe 77% der Varianz aufklären. Die Faktormatrizen bieten ein sehr einheitliches Bild, so daß sich keine Unterschiede zwischen beiden Gruppen interpretieren lassen.

(c) Bei der Reliabilitätsbestimmung der Selbsteinschätzungen beider Gruppen nach WINER ergeben sich folgende Werte:

$$r_{40}' \text{ (Gespr. Gruppe)} = 0.974$$

$$r_{40}' \text{ (Kontr. Gruppe)} = 0.954.$$

Auch hier erscheint der Schluß gerechtfertigt, daß sich beide Gruppen nicht unterscheiden.

Zusammenfassung der Ergebnisse von Experiment I: Die beiden von studentischen Versuchsleitern untersuchten Gruppen parallelisierter Personen, die entweder ein längeres Gespräch über die eigene Persönlichkeit geführt hatten, oder nicht, unterscheiden

sich weder hinsichtlich Urteilsextremität noch in Bezug auf Urteilsdifferenziertheit.

EXPERIMENT II

Urteilsextremität. Die Häufigkeit, mit der die Personen der drei untersuchten Gruppen die extremen Urteilskategorien "3" und die mittlere Kategorie "0" ankreuzten, muß wegen ungleich großer Beurteilerstichproben als Proportion ausgedrückt werden. In der folgenden Tabelle sind (1) die rohen Häufigkeiten, (2) die Stichprobengrößen, (3) die Gesamtzahlen abgegebener Selbsturteile, (4) die proportionalen Anteile extremer (bzw. mittlerer) Urteile, und (5) die Ergebnisse der zufallskritischen Prüfung der Unterschiede je zweier Proportionen p , sofern die Differenzen das 5%-Niveau erreichen, aufgeführt. Die Prüfung erfolgt mit $t = (p_1 - p_2) / \sigma_{p_1 - p_2}$ bei $df = n_1 + n_2 - 2$, wobei $\sigma_{p_1 - p_2} = \sqrt{p_1 q_1 / n_1 + p_2 q_2 / n_2}$. Wir setzen hierbei jedoch α auf 1 Promille fest, da von den vorliegenden Proportionen nicht unbedingt auf eine Normalverteilung von Proportionen geschlossen werden kann.

	<u>Gruppe 1</u> (Gespr.Gr.)	<u>Gruppe 2</u> (Test-Gr.)	<u>Gruppe 3</u> (Kontr.Gr.)
(1)	57	23	35
(2)	16	10	20
(3)	288	180	360

(Verwendung extremer Urteilskategorien:)

(4)	0.20	0.13	0.10
(5)			

(Verwendung der mittleren Urteilskategorie:)

(4)	0.08	0.16	0.18
(5)			

Demnach bestehen interpretierbare Unterschiede zwischen den Gruppen 1 und 3: Die Personen der Gesprächsgruppe verwenden extreme Selbstbeurteilungskategorien häufiger, und die mittlere Kategorie seltener als die Personen der Kontrollgruppe. Es zeigen sich auch Tendenzen derart, daß sich das Urteilsverhalten der Gesprächsgruppe von demjenigen der Testgruppe in ähnlicher Weise unterscheidet, während zwischen den Urteilen der Test-Gruppe und der Kontrollgruppe keinerlei Differenzen zu entdecken sind.

Urteilsdifferenzierung. (a) Die mittleren Interkorrelationen der 18 Eigenschaften bei den drei untersuchten Gruppen betragen:

<u>Gruppe 1</u> (Gespr.Gr.)	<u>Gruppe 2</u> (Test-Gr.)	<u>Gruppe 3</u> (Kontr.Gr.)
r = 0.26	r = 0.36	r = 0.33
(n = 16)	(n = 10)	(n = 17)
(s = .20)	(s = .22)	(s = .21)

Danach hat es den Anschein, als urteilten die Personen der Gesprächs-Gruppe insgesamt differenzierter. Allerdings dürfte die Höhe der Korrelationen von der Stichprobengröße abhängen (Gruppe 3 ist hier wegen missing data reduziert). Dies bestätigt sich, wenn die Versuchsgruppen 1 und 2 zusammengefaßt und ihre mittlere Interkorrelation derjenigen der Kontrollgruppe gegenübergestellt wird:

<u>Gruppen 1 u. 2</u> (Gespr., Test)	<u>Gruppe 3</u> (Kontr.Gr.)
r = 0.23 (n=26)	r = 0.33 (n=17)

Prüfbar erscheint also nur der Unterschied zwischen den Gruppen 1 (n = 16) und 3 (n = 17). Die zufallskritische Prüfung ergibt $t = 0.18$ bei $df = 29$; der Unterschied ist daher als noch zufällig zu bezeichnen. Zieht man aus den Gruppen 1 und 3 Zufallsstichproben der Größe 10 und vergleicht nunmehr die drei gleich großen Gruppen in Bezug auf die Höhe der mittleren Rating-Interkorrelation, so ergibt sich:

Gr.1: r = 0.35 Gr.2: r = 0.36 Gr.3: r = 0.41

Tendenziell urteilen demnach die Personen der Gesprächs- und der Test-Gruppe differenzierter als diejenigen der Kontroll-Gruppe. Die zufallskritische Prüfung der Korrelationsdifferenzen ergibt jedoch keine signifikanten Unterschiede.

(b) Ein Vergleich der Faktorenstrukturen erscheint in diesem Falle als problematisch, weil zumindest im Falle der Gruppe 2 die Zahl der Personen diejenige der Variablen unterschreitet. Vergleicht man lediglich Gruppe 1 mit Gruppe 3, so ergibt sich ein Hinweis auf eine differenziertere Faktorenstruktur bei der ersteren: Bei der Gesprächsgruppe weisen sechs, bei der Kontrollgruppe fünf Faktoren Eigenwerte über 1.0 auf; sie erfassen in beiden Fällen 82% der Varianz. Der Unterschied erscheint jedoch als wenig aussagekräftig. Faßt man einmal sämtliche Versuchsgruppen aus den Experimenten I und II ($n = 66$) und beide Kontrollgruppen ($n = 57$) zusammen, so ergibt sich kein sichtbarer Unterschied mehr.

c) Aufgrund der Berechnung der Reliabilitätskoeffizienten (als Ausdruck der Urteilsdifferenzierung der Gruppen) ergeben sich folgende Resultate:

<u>Gruppe 1:</u> (Gespr.Gr.) ($k=16, n=18$)	QUS Zw.Eig. = 394.13	df = 17	MQUS = 23.18
	QUS In.Eig. = 619.37	df = 270	MQUS = 2.29
	Theta' = 0.57		$r_{16}' = 0.90$ =====
<u>Gruppe 2:</u> (Test-Gr.) ($k=10, n=18$)	QUS Zw.Eig. = 349.74	df = 17	MQUS = 20.57
	QUS In.Eig. = 367.90	df = 162	MQUS = 1.65
	Theta' = 1.15		$r_{10}' = 0.92$ =====
<u>Gruppe 3:</u> (Kontr.Gr.) ($k=20, n=18$)	QUS Zw.Eig. = 215.22	df = 17	MQUS = 12.66
	QUS In.Eig. = 619.75	df = 342	MQUS = 1.81
	Theta' = 0.30		$r_{20}' = 0.86$ =====

Die r_k -Koeffizienten scheinen sich in der erwarteten Art und Weise zu unterscheiden: Die Rating-Reliabilitäten der Gesprächs- und Test-Gruppen sind höher als die der Kontrollgruppe. Die Unterschiede sind jedoch nicht prüfbar, da für r_k keine Fehlerverteilungen vorliegen. r_k wächst mit zunehmender Stichprobengröße; für die Versuchsgruppen 1 und 2 gemeinsam liegt r_{26}' bei 0.96, so daß ein der Kontrollgruppen-

größe 20 entsprechender Koeffizient der Versuchsgruppen etwa zwischen 0.90 und 0.96 zu liegen käme (gegenüber 0.86 der Kontrollgruppe).

Unterschiede zwischen Gesprächsleitern. Da sich die Absicht bei der Planung des Experiments, den Gesprächsleiter in Gruppe 1 konstant zu halten, als praktisch nicht möglich erwies, wurden die Gespräche von den drei Autoren, im folgenden in Zufallsfolge als A (6 Vpn), B (5 Vpn) und C (5Vpn) bezeichnet, geführt. Die Autoren A und B konnten gegenüber C als relativ erfahrener im Führen von Interviewgesprächen gelten. Um daher mögliche Unterschiede der Wirksamkeit der drei Gesprächsführungen aufzudecken, wurde innerhalb der Gruppe 1 nochmals nach den drei Bedingungen A,B und C unterschieden.

In Bezug auf Urteilsextremität unterscheiden sich die Daten der Personen von Gesprächsleiter A und C sehr deutlich (0.001-Niveau), diejenigen von B und C tendenziell (0.05-Niveau). Bei den relativ erfahreneren Gesprächsleitern kam es demnach - entgegen der Erwartung - zu weniger extremen Selbstbeurteilungen der Vpn. In Bezug auf mittlere Urteile ergaben sich keine Unterschiede.

Unterschiede der mittleren Interkorrelationen der 18 Selbstbeurteilungen als Ausdruck unterschiedlicher Urteilsdifferenzierung gab es nicht; die Koeffizienten waren 0.41, 0,40 und 0,45 für die Gesprächsleiter A, B und C. Faktorenanalysen unterblieben wegen der geringen Personenzahlen. Differentielle Effekte ergaben sich jedoch offensichtlich in Bezug auf die Höhe der Reliabilitätskoeffizienten nach WINER, wenn man zwischen den von verschiedenen Gesprächsleitern geführten Gesprächsgruppen unterschied:

<u>Gespr.leiter A:</u> (k = 6, n = 18)	QUS Zw.Eig. = 158.80	df = 17	MQUS = 9.34
	QUS In.Eig. = 128.20	df = 90	MQUS = 1.42
	Theta' = 0.96		$r_6' = 0.85$ =====

<u>Gespr.leiter B:</u> (k = 5, n = 18)	QUS Zw.Eig. = 195.30	df = 17	MQUS = 11.49
	QUS In.Eig. = 137.20	df = 72	MQUS = 1.91
	Theta' = 0.97		$r_5' = 0.83$ =====

<u>Gespr.leiter C:</u> (k = 5, n = 18)	QUS Zw.Eig. = 150.00	df = 17	MQUS = 8.82
	QUS In.Eig. = 238.40	df = 72	MQUS = 3.31
	Theta' = 0.32		$r_5' = 0.61$ =====

Zusammenfassung der Ergebnisse von Experiment II: Die Personen der Gesprächs-Gruppe verwenden extreme Selbstbeurteilungen häufiger und mittlere Selbstbeurteilungen seltener. Die Personen der Gesprächs- und der Testgruppe neigen auch zu differenzierteren Selbsturteilen als diejenigen der Kontrollgruppe. Bei verschiedenen Gesprächsleitern zeigen die Selbstbeurteiler unterschiedliches Urteilsverhalten.

Diskussion

Die Annahmen dieser Untersuchung konnten für Experiment I in keinem Falle, für Experiment II dagegen in mehrfacher Hinsicht bestätigt werden.

Wichtige Unterschiede zwischen Experiment I und II aufzudecken fällt leicht, da sie teilweise bereits in der Konzeption beider Versuche enthalten waren. Experiment II sollte als kritisches Experiment dienen, Experiment I nur als Vorversuch; in letzterem Falle standen Studenten als Gesprächsleiter und als Versuchspersonen zur Verfügung, so daß weder bei der Auswahl der Versuchspersonen noch bei der Ausführung des Experiments (oft diente die Cafeteria der Universität als Gesprächsraum) große Sorgfalt erwartet werden konnte. Das unmißverständliche Ergebnis von Experiment I überrascht daher wenig.

Experiment II wies demgegenüber alle Vorteile bis auf einen, nämlich die Zahl der Versuchsteilnehmer, auf. Die Versuchspersonen waren gut motivierte, an psychologischen Ergebnissen offensichtlich interessierte Personen. Sie unterschieden sich im Gesamtverhalten sehr stark von den "sophisticated subjects" aus der Universität. Aufgrund von Erfahrungen mit Experiment I konnte bei der Applikation der Versuchsbedingungen "Gespräch" und "Tests" mehr Sorgfalt aufgewendet werden. Der nicht zu unterschätzende Mangel von Experiment II ist die Zahl der Versuchspersonen. Sie wirkt sich besonders beim hier nicht immer möglichen Vergleich der Faktorenstrukturen der 18 Selbst-Ratings (Schätzung der Urteilsdifferenzierung nach Methode b) aus.

Dennoch zeigen sich in Experiment II einige deutliche Resultate. Personen, für die eine größere Menge von Informationen über die eigene Persönlichkeit (in Form der Problematisierung eigener Charaktereigenschaften und Verhaltensmerkmale in bestimmten, bedeutsamen Situationen) verfügbar - oder durch die entsprechende Versuchsbedingung aktualisiert - ist, geben eindeutiger bzw. extremere Selbsturteile ab, und ihre Selbst-Ratings weisen insgesamt einen höheren Differenzierungsgrad auf. In Bezug auf Urteilsdifferenzierung (bzw. Vermeidung von Halo-Effekten) bei der Selbsteinschätzung gilt dies auch für jene Personen, denen man lediglich einen Katalog von Testresultaten aus einem zuvor ausgeführten Intelligenz- und Persönlichkeitstest zur Verfügung stellt. Zwar sind die genannten Unterschiede hier nicht als einschneidend zu bezeichnen und oft nur tendenziell vorhanden, doch ist es auffallend, daß sich entsprechende Tendenzen gleichsinnig in mehreren von uns herangezogenen Indikatoren zur Erfassung von Urteilsdifferenzierung zeigen. Da sich sowohl nach Methode (a) als auch nach Methode (c) die Mitglieder der Versuchsbedingungen "Selbstbezogene Information: Gespräch" und "Selbstbezogene Information: Testdaten" von denjenigen der Kontrollgruppe unterscheiden, läßt sich für die abhängige Variable "Urteilsdifferenzierung" sagen, daß die Hypothesen 1 und 2 (vgl.S.4) in der Tendenz bestätigt wurden, nicht jedoch Hypothese 3. In Bezug auf die abhängige Variable "Urteilsextremität" wurden die den Hypothesen 1 und 2 entsprechenden Nullhypothesen zurückgewiesen, nicht aber die der Hypothese 3 entsprechende Nullhypothese.

Ungeklärt bleibt, inwiefern die aufgetretenen Veränderungen im Urteilsverhalten tatsächlich das Ergebnis der durch das selbstbezogene Gespräch oder die durch die Tests und deren Rückmeldung veränderten Selbstwahrnehmung sind. Es wäre beispielsweise auch denkbar, daß durch die experimentelle Behandlung der Versuchsgruppen ein allgemeines Klima der Bereitschaft zu psychologischer Selbstreflexion geschaffen wurde, das - unabhängig von vermittelter konkreter Information - zu teils polarisierteren, teils differenzierteren Selbstbeurteilungen geführt hat. Die Frage ist aber unter pragmatischem Gesichtspunkt wenig erheblich, da sich die Prozesse

der Informationsverarbeitung "innerhalb" des Selbstbeurteilers ohnehin nicht durch Experimente wie das vorliegende aufklären lassen. Die vorliegende Arbeit reiht sich also in Untersuchungen zur Wirkungsforschung (zur Wirkung von Instruktionen etc.) ein, die den der erzielten Wirkung zugrundeliegenden inneren Prozeß nicht aufklären.

Daneben muß angemerkt werden, daß im vorliegenden Experiment versucht werden sollte, Urteilsänderungen auf möglichst sparsame und gefährlich simple Art und Weise hervorzurufen. Die gesprächsweise Beschäftigung mit dem eigenen Selbst kann ebenso wie die Information über testmäßig erfaßte und formulierte Persönlichkeitsmerkmale quantitativ und qualitativ wesentlich voluminöser gestaltet werden. Man sollte dabei jedoch überlegen, ob gleichsam therapeutische Sitzungen mit Versuchspersonen nicht zu weiteren, unerwünschten Nebeneffekten führen können.

Wer die Diskussion über Persönlichkeitspsychologie in den 70er Jahren verfolgt (vgl. GRAUMANN 1975), wird das Vorgehen im vorliegenden Experiment, das durch die Aktualisierung von Kognitionen über starre Persönlichkeitseigenschaften charakterisiert ist, ein wenig antiquiert finden. Die Praxis des Experiments zeigt jedoch, daß es den meisten Personen eher leicht fällt, sich in Begriffen starrer Persönlichkeitseigenschaften zu beurteilen. Die Konzepte der überkommenen Charakterologie scheinen das Denken und Urteilen der meisten Personen zu bestimmen. Es wäre daher vermutlich eine Überforderung der untersuchten Personen, auf Eigenschaftskonzepte und die ihnen entsprechenden Verfahren der Eigenschaftsbeurteilung zu verzichten.

In der vorliegenden Arbeit konnte gezeigt werden, daß Selbstwahrnehmung bzw. Selbstbeurteilung in gewissem Maße ähnlichen Prinzipien folgt wie Fremdwahrnehmung bzw. Fremdbeurteilung. Entsprechende Untersuchungen müßten jedoch in jedem Falle mit größerem Aufwand wiederholt und modifiziert werden.

Literatur

- AMTHAUER, R. 1953. Intelligenz-Struktur-Test. Göttingen: Hogrefe.
- BINTIG, A. 1976. Über die Zuverlässigkeit sozialwissenschaftlicher Beurteilungsskalen bei der Wahrnehmung und Beurteilung von Personen. Unveröff. Diss., Fak.f. Soziologie, Univ. Bielefeld.
- EBEL, R.L. 1951. Estimation of the reliability of ratings. Psychometrika, 16, 407-424.
- FAHRENBERG, J. & SELG, H. 1970. Das Freiburger Persönlichkeitsinventar. Göttingen: Hogrefe.
- GLESER, G.C., CRONBACH, L.J. & RAJARATNAM, N. 1965. Generalizability of scores influenced by multiple sources of variance. Psychometrika, 30, 395-418.
- GRAUMANN, C.F. 1975. Person und Situation. In: Lehr, U.M. & Weinert, F.E. (Hrsg.) Entwicklung und Persönlichkeit. Stuttgart: Kohlhammer, 15-24.
- GUILFORD, J.P. 1954. Psychometric methods. New York: McGraw-Hill.
- JOHN, D. & KEIL, W. 1972. Selbsteinschätzung und Verhaltensbeurteilung. Psychologische Rundschau, 23, 10-29.
- MUMMENDEY, H.D., MIELKE, R., MAUS, G. & HESENER, B. 1977. Untersuchungen mit einem mehrdimensionalen Selbsteinschätzungsverfahren. Bielefelder Arbeiten zur Sozialpsychologie, Nr.14 (Januar 1977).
- RUDINGER, G. & FEGER, H. 1970. Die Beurteilung formaler Verhaltensmerkmale durch Rating-Skalen: Eine Generalisierbarkeitsstudie. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 2, 96-112.
- SCHMIDT, H.D. 1966. Über die Zuverlässigkeit von Verhaltensbeurteilungen durch Rating-Skalen. Archiv für die gesamte Psychologie, 118, 47-72.
- SCHMIDT, H.D. 1976. Selbstwahrnehmung als interpersonelle Wahrnehmung: Hypothesen. Bielefelder Arbeiten zur Sozialpsychologie, Nr.11 (September 1976).
- THORNDIKE, E.L. 1920. A constant error in psychological ratings. Journal of Applied Psychology, 4, 25-29.
- WERNER, J. 1976. Varianzanalytische Maße zur Reliabilitätsbestimmung von ratings. Zeitschrift für experimentelle und angewandte Psychologie, 23, 490-500.
- WINER, B.J. 1971. Statistical principles in experimental design. New York: Wiley, 2nd ed.

BIELEFELDER ARBEITEN ZUR SOZIALPSYCHOLOGIE

1976 erschienen:

- Nr.1 (2/76): Hans Dieter Schmidt: Die Förderung nationalistischer Einstellungen durch internationale Sportwettkämpfe in sozial-psychologischer Sicht
- Nr.2 (2/76): Hans Dieter Schmidt, Amélie Schmidt-Mummendey, Friedrich-Wilhelm Schröer und Rüdiger Stallberg: Der Einfluß aggressiven Modellverhaltens, Status des Modells und Selbsteinschätzung auf offen aggressives Verhalten
- Nr.3 (3/76): Hans Dieter Schmidt, Amélie Schmidt-Mummendey, Ulrich Goertz, Günter Maus, Friedrich-Wilhelm Schröer und Uwe Troske: Der Einfluß von Modell-Aggression und Modell-Status auf das Erteilen von Geldstrafen als instrumentell-aggressives Verhalten
- Nr.4 (4/76): Manfred Bornewasser: Naturwissenschaftliche und verhaltenstheoretische Orientierungen in der Sozialpsychologie (I)
- Nr.5 (4/76): Rosemarie Mielke und Hans Dieter Schmidt: Zur Psychologie politischen Verhaltens - eine empirische Analyse von Verhaltensmustern beim Kampf um Gleichberechtigung der Frau
- Nr.6 (4/76): Hans Dieter Schmidt: Selbstwahrnehmung und Beeinflussung
- Nr.7 (4/76): Lothar P. Schardt: Wem nützt die Eignungsdiagnostik? Zum konservativen Charakter der klassischen Testtheorie und ihrer Anwendung in der betrieblichen Praxis
- Nr.8 (5/76): Lothar P. Schardt: Ansätze zu einer arbeitsorientierten Eignungsdiagnostik
- Nr.9 (7/76): Manfred Bornewasser, Fritz Köster, Rosemarie Mielke und Cornelia Pawelcik: Erstellung und erste Überprüfung eines Fragebogens zur instrumentellen Arbeitsorientierung bei Jugendlichen
- Nr.10 (8/76): Hans Dieter Schmidt, Bernd Schiebel und Uwe Troske: Untersuchung der Stabilität von Modell-Effekten aggressiven Verhaltens gegenüber Populations- und Situationsunterschieden
- Nr.11 (9/76): Hans Dieter Schmidt: Selbstwahrnehmung als interpersonelle Wahrnehmung: Hypothesen
- Nr.12 (10/76): Hans Dieter Schmidt, Lothar P. Schardt und Rosemarie Mielke: Ein- und mehrdimensionale Skalierung der gewerkschaftlichen Kampfbereitschaft von Belegschaften
- Nr.13 (Herbst 1976): Hans Dieter Schmidt: Auf dem Wege zu einer sozialen Psychophysiologie des akuten Vampirismus (vergriffen)

BIELEFELDER ARBEITEN ZUR SOZIALPSYCHOLOGIE

Bis April 1977 erschienen:

- Nr.14 (1/77): Hans Dieter Mummendey, Rosemarie Mielke, Günter Maus und Bernhard Hesener: Untersuchungen mit einem mehrdimensionalen Selbsteinschätzungsverfahren
- Nr.15 (2/77): Rosemarie Mielke: Zur Erfassung von Umweltmerkmalen im Ausbildungsbereich
- Nr.16 (2/77): Manfred Bornwasser: Die Konsistenzbeziehung zwischen Einstellungen und offenem Verhalten
- Nr.17 (3/77): Hans Dieter Mummendey: Einstellungen (sets) bei der Erforschung der Beziehung zwischen Einstellungen (attitudes) und offenem Verhalten
- Nr.18 (4/77): Hans Dieter Mummendey, Bernd Schiebel und Uwe Troske: Experimentelle Untersuchung der Stabilität instrumentell-aggressiven Verhaltens bei Meßwiederholungen in verschiedenen Zeitabständen
- Nr.19 (4/77): Hans Dieter Mummendey, Bernhard Hesener und Margret Isermann: Selbstwahrnehmung als interpersonelle Wahrnehmung: Methoden und Ergebnisse

Forschungsberichte, herausgegeben von H.D.Mummendey, Universität Bielefeld
Fakultät für Soziologie, Postfach 8640, 4800 Bielefeld 1, pro Heft DM 1,50
(Selbstkostenpreis)