

---

## SOZIALE ERWÜNSCHTHEIT ALS PROBLEM PSYCHOLOGISCHER FORSCHUNG

Hans Dieter Mummendey

---

Summary (Social desirability as a problem of psychological research): Social desirability is regarded as a psychological construct that has to be differentiated into several partial constructs. A review of methods of control for social desirability response set is given, stressing the bogus pipeline paradigm as an efficient strategy to modify the total response situation. Since social desirability reactions and all methods to control them make use of deception, several ethical implications of controlling social desirability are discussed.

### 1. Das Konzept sozial erwünschter Reaktionen

Seit langem ist bekannt (Cronbach, 1946, 1950; Edwards, 1953, 1957), daß Reaktionen auf verbale Stimuli hinsichtlich ihrer Bejahungswahrscheinlichkeit u.a. vom Grad der sozialen Erwünschtheit der Bejahung abhängen. Damit wird ein psychologisches Konstrukt bzw. eine Variable

etabliert, die als Fehlerquelle für alle subjektiven psychologischen Meßmethoden fungiert. Das Schwellen-Modell von Voyce & Jackson (1977) mag zeigen, wie man sich die beiden prominentesten Reaktionstendenzen (Erwünschtheits- und Jasage-Tendenz) als Spielarten der Tendenz zur sozial erwünschten Reaktion vorstellen kann. Die Autoren beschreiben Reaktionen auf Persönlichkeits-Items als abhängig von zwei Parametern: der Schwelle, d.h. jenem kritischen Grad an Erwünschtheit, der den Übergang von einer "Falsch"- zu einer "Richtig"-Reaktion markiert, und einem Salience-Parameter, der eine Schätzung der Sensibilität der Vp gegenüber der Erwünschtheitsdimension darstellt. Die Vp reagiert so, als "kenne" sie den Erwünschtheitsgrad jedes Items. Liegt ein Item unterhalb ihres eigenen Schwellenwertes, so wird sie es nicht bejahen. Eine Vp mit einer niedrigen Schwelle wird eine größere Zahl von "Stimmt"-Reaktionen, eine Vp mit einem hohen Schwellenwert eher "Stimmt-nicht"-Antworten geben, so daß sich im neutralen Bereich der Erwünschtheit individuelle Differenzen in der Jasagetendenz ergeben. Da es in Differentieller und Diagnostischer Psychologie um individuell zurechenbare Reaktionen - unbeeinflußt von dem, was sozial erwünscht sein mag - geht, kommt der Variablen "Social Desirability" (SD) der Status einer die Validität psychologischer Messungen mindernden Suppressorvariablen zu (Wiggins, 1973).

Während sich Strategien zur Kontrolle der allgemeinen Jasagetendenz aufgrund des "formalen" Charakters dieses Konzepts besser realisieren lassen, da es in Grenzen die Möglichkeit gibt, Statements hinsichtlich ihrer positiven/negativen Formulierung abzustufen bzw. auszubalancieren (vgl. Jackson, 1967; Messick, 1966; Keil, 1968, 1973; Bentler, Jackson, & Messick, 1971; Ehlers, 1973), scheint das Problem der Nicht-Separierbarkeit von SD und Inhalten schwieriger zu sein, weil es sich stärker um eine "inhaltliche" Störvariable zu handeln scheint. "Positivität" kann variiert werden, ohne den skalierbaren Gehalt eines Items stark zu verändern; mit der Manipulation des Erwünschtheitsgrades wird dagegen eine stärkere Veränderung des Item-Inhaltes vorgenommen.

Janke (1973) wies darauf hin, daß neutrale Items, die sich mittlerer Wertschätzung in einer Gesellschaft erfreuen, in Fragebogen schwach repräsentiert sind. Der MMPI als Vater (bzw. Großvater) der meisten gegenwärtigen Persönlichkeitsinventare enthält kaum ein als neutral in bezug auf seine Wertung zu beurteilendes Item. Auch die vielen Versuche, Adjektive zur Persönlichkeitsbeschreibung hinsichtlich SD beurteilen zu lassen (z.B. Anderson, 1968; Cowen & Budin, 1964; deutschsprachig z.B. Schönbach, 1972; Busz, Cohen, Poser, Schümer, Schümer, & Sonnenfeld, 1972; Klapprott, 1972; Buxbaum, Anm. 1 u. 2) ergeben zweigipflige Häufigkeitsverteilungen. Diese Erwünschtheits-Ratings von Eigenschaften korrelieren, wie Kirby & Gardner (1971) zeigten, 0.98 mit den Edwards'schen Social Desirability Scale Values (SDSV). Betrachtet man die SDSV von fast 3000 Statements aus Persönlichkeitsfragebogen, so verteilen sich auch diese klar zweigipflig (Edwards, 1970). Es ist Wiggins (1979) zuzustimmen, der Persönlichkeitsvariablen, die nicht hinsichtlich Erwünschtheit stark variieren, als armselige Repräsentationen von Kategorien des wirklichen Lebens ansehen möchte.

Der SD-Suppressoreffekt ist selten empirisch genau bestimmt worden. Die Untersuchung von Dicken (1963) mit dem CPI hatte zwar nur in vier von 24 bzw. in zwei von 36 verglichenen Fällen und diejenige von Goldberg, Rorer, & Greene (zit.n.Wiggins, 1973) in 30 von 169 Fällen einen Suppressoreffekt ergeben. Andererseits zeigen Untersuchungen mit Instruktionsmanipulation enorme Beschönigungseffekte noch im Vergleich zu unbehandelten Kontrollgruppen, bei denen bereits SD-Tendenzen zu erwarten sind. So fanden u.a. Häcker, Schwenkmezger, & Utz (1977) bei ca. der Hälfte der Skalen verschiedener Cattell'scher Fragebogentests signifikante Verfälschungs-

effekte. Untersuchungen von Diagnostikern, die Tests vom Q-Typ in ausleseähnlichen Situationen (beim TÜV, bei Strafgefangenen, usw.), also in Ernstsituationen angewendet haben (z.B. Böcher, 1967; Schmidt & Vorthmann, 1971; Reitzner, zit.n.Fahrenberg, Selg, & Hampel, 1978; Häcker et al., 1977), zeigen erhebliche Beschönigungseffekte in den durch erhöhten SD-Druck zu charakterisierenden Situationen.

Es wird deutlich, daß das psychologische Konzept sozial erwünschter Reaktionen prinzipiell Täuschungsverhalten seitens der untersuchten Personen impliziert, denn je nach Situation, Thema und Persönlichkeit der Vp kommt es zu mehr oder weniger starken Substitutionen von als "echt" aufgefaßten durch im Sinne sozialer Normen korrigierte bzw. im herkömmlichen Sinne von Diagnostischer und Differentieller Psychologie verfälschte Reaktionen. Weiter unten wird zu diskutieren sein, wie der Psychologe mit eigenen Mitteln der Täuschung versuchen kann, das aus seiner Perspektive gestörte Gleichgewicht zwischen V1 und Vp wiederherzustellen.

## 2. Differenzierungen des Erwünschtheits-Konzepts

Zunächst ist auf die schon früh (vgl. Jackson & Messick, 1958) erfolgte Unterscheidung zwischen Response-Set und -Style kurz einzugehen, obgleich sie nicht immer einheitlich ist (vgl. Rorer, 1965). Sie bezieht sich auf den Grad der Situationsspezifität bzw. -generalität der SD-tendenz und faßt SD als "Stil" im Sinne eines unterschiedliche Antwortsituationen übersteigenden Persönlichkeitsmerkmals auf (vgl. Bentler, Jackson, & Messick, 1971). Wenn gelegentlich gezeigt wurde, daß Personen mit hohen SD-Scores z.B. besser konditionierbar sind (Crowne & Marlowe, 1964; Epstein, 1964), diverse Lernaufgaben besser erledigen (Brown, 1960; Norman, 1963; Stephens, zit.n.Allaman, Joyce, & Crandall, 1972; Katz, Henchy, & Allen, 1968), motorische Aufgaben schneller ausführen können (Ford & Rubin, 1970), aber auch beeinflubarer bzw. konformer sind (Crowne & Marlowe, 1964), weniger Self-Esteem besitzen und unsicherer sind (Crowne & Marlowe, 1964), Mütter mit relativ höheren Ausprägungsgraden an Feindseligkeit, Kritizismus, Restriktivität, Zwang und Bestrafungstendenz bei Mädchen und Jungen, bei Jungen zusätzlich mit weniger Ermutigung zur Entwicklung von Fertigkeiten als Erziehungsverhalten gehabt zu haben (Allaman et al., 1972), so wird hieraus die radikale Konzeption von SD als Persönlichkeits-eigenschaft deutlich. Allerdings hat sich oft gezeigt, daß die Generalität der SD-Tendenz bereits über verschiedene Meßmethoden hinweg begrenzt ist (vgl. Fiske & Pearson, 1970; Edwards & Abbott, 1973). Es mag daher eine Frage der Toleranz des Forschers gegenüber den "neuen" Erkenntnissen der in den 70er Jahren wiederbelebten Person-Situation-Interaktions-Diskussion sein (vgl. Mischel, 1968; Magnusson & Endler, 1977; Herrmann, 1980), ob SD als Stil bzw. Quasi-Persönlichkeitsmerkmal formaler Art aufgefaßt wird.

Die meisten Response-Set-Forscher nehmen in relativ undifferenzierter Weise ein allgemeines Konstrukt "soziale Erwünschtheit" an. Dies ist scheinbar dadurch gerechtfertigt, daß die SDSV von Items gewöhnlich über verschiedene Gruppen von Beurteilern hinweg positiv miteinander korrelieren (Edwards, 1970). Andererseits variieren Verhaltensbewertungen in Abhängigkeit von Alter, Geschlecht und sozialer Schicht, und es kommt zu Wechselwirkungen bezüglich SD zwischen diesen Merkmalen (vgl. z.B. Crott & Rossrucker, 1974). SD-Ratings fallen auch unterschiedlich aus je nachdem, ob Selbsteinschätzungen oder Fremd-Selbsteinschätzungen abgegeben werden (vgl. Crott, Prüfer, & Wolfshörndl, 1977). Altersunterschiede in SD-Ratings wurden ferner von Milturn & Lewis (1868), Ahammer (1971) sowie Ahammer & Baltés (1972) gefunden; sie scheinen besser gesichert als z.B. Geschlechts- und Bildungsunterschiede (vgl. z.B. Klett, 1957; Edwards, 1957; Marlowe & Crowne, 1962; Allaman et al., 1972; Klein, 1974). Mit großer Konsistenz wurden dage-

gen interkulturelle Übereinstimmungen - zwischen norwegischen, arabischen, nizäischen, griechischen und amerikanischen Vpn - gefunden (Edwards, 1957; Lovaas, 1958; Klett & Yaukey, 1959; Triandis, 1964); in anderen Fällen (vgl. z.B. Klein, 1974) ergaben sich jedoch Nationalitätenunterschiede, z.B. zwischen Deutschen und Franzosen.

Insgesamt gesehen erscheint es notwendig, die Annahme eines allgemeinen, gruppen-unspezifischen SD-Konstrukts fallenzulassen. Ergänzt wird dieses Urteil durch die gelegentlich gefundene niedrigen Korrelationen zwischen verschiedenen SD-Kontrollskalen (vgl. Ehlers, 1973). Eine generelle Konzeption sozialer Erwünschtheit, und sei es auch nur für die Fragebogenbeantwortung als spezifischer Methode, die zugleich eine spezifische Situation konstituiert, erscheint als problematisch, so daß es ratsam erscheint, auf spezifische Stichproben von Methoden und Personen, Situationen und Themen bezogene SD-Konzepte zu formulieren. Solche Teilkonzepte sozialer Erwünschtheit wären im äußersten Falle durch das Konzept einer persönlichen Erwünschtheit zu ersetzen. So hielt Scott (1963) aufgrund von Untersuchungen mit dem EPPS SD nicht für ein Item-Charakteristikum, sondern für das Ergebnis einer Item-Person-Interaktion. Er gab zugleich den lernpsychologisch interessanten Hinweis, daß die Erwünschtheit eines Items auch als Konsequenz von Bejahung aufgefaßt werden könne, und nicht nur umgekehrt. Diese Interpretation des Verhältnisses von Inhalt und Set bzw. Style ist - vielleicht mangels lerntheoretischer Orientierung der meisten differentiell- und sozialpsychologischen Autoren - nicht weiter beachtet worden. Nancy Wiggis (1966) zeigte nach Anwendung des Verfahrens von Tucker & Messick (1963), daß die MMPI-Items sechs funktional unabhängigen Ideal-Personen zuzuordnen sind, die sehr unterschiedliche Itemcluster mit jeweils hohen SDSV reflektieren; die Autorin postuliert die Multidimensionalität des SD-Konzepts und folgert daraus differentielle SD-Strukturen für Items und Individuen.

In der Praxis der psychologischen Forschung haben jedoch die notwendigen Arbeiten an Differenzierungen bzw. Teilkonstrukten sozialer Erwünschtheit in den 70er Jahren keineswegs stattgefunden; vielmehr ist ein relativer Rückgang an Publikationen über das SD-Konzept festzustellen, wenn man die absolute Häufigkeit der Veröffentlichungen zum Stichwort "Social Desirability" in den "Psychological Abstracts" auf die Gesamtmasse an Publikationen bezieht. Stattdessen hat man den Eindruck, daß man auf das SD-Konzept als Allerweltskonstrukt mit Vorliebe dann zurückgreift, wenn es an die Interpretation hypothesennonkonformer Ergebnisse empirischer Untersuchungen geht. Den SD-Tendenzen geht es damit ähnlich wie den mit ihnen verwandten Vpn-Effekten der "Sozialpsychologie des Experiments" (vgl. z.B. Orne, 1962; Page, 1973) - sie sind oft für globale Vermutungen und schnelle Erklärungen gut, werden aber selten in ihrer spezifischen Wirkweise nachgewiesen.

### 3. Die Kontrolle sozialer Erwünschtheits-Tendenzen

#### 3.1 Kontrolle durch Itemkonstruktion

Es kann versucht werden, bei der Itemselektion solche Items zu präferieren, die hinsichtlich des zu messenden Merkmals möglichst wenig transparent sind. Solche Kontrollversuche stossen jedoch bei Selbstberichten über Verhalten ebenso wie bei Einstellungs-Urteilen auf die Schwierigkeit, daß sich Inhalt und Set bzw. Style nicht dadurch trennen lassen, daß man gewissermaßen den Inhalt auszuschalten sucht. Andererseits gibt es durchaus Statements, die trotz mangelnder SD-Anfälligkeit ausreichende Trennschärfe besitzen. Der Versuch einer SD-Kontrolle auf Itemebene wurde in bislang systematischster Weise bei der Konstruktion der PRF von Jackson

(1974) und der entsprechenden deutschen Version von Angleitner, Stumpf, & Wieck (Anm.3)vorge-  
nommen. Um Statements mit hoher SD-Varianz bei der Selektion zu eliminieren, wurden Korrelati-  
onen mit einem gleichzeitig ermittelten SD-Score berechnet und zur Item-Trennschärfe in Be-  
ziehung gesetzt. Der so gewonnene "Differentielle Reliabilitäts-Index" diente als Kriterium  
der Itemselektion. Prinzipiell handelt es sich bei dieser Kontrollstrategie um die Anwendung  
spezieller SD-Kontrollskalen bereits auf Itemebene. Eine mögliche Schwierigkeit dieses Verfah-  
rens könnte darin bestehen, daß sich in einer Menge derart selektierter Items neue Erwünsch-  
theits-Hierarchien herausbilden, da Inhalt und Set bezüglich ihres relativen Varianzbeitrages  
nur relativ, nicht aber absolut getrennt werden können.

### 3.2 Kontrolle durch Antwortkombination

Da Bejahung und SD direkt miteinander zusammenhängen, können die Antwortalternativen hin-  
sichtlich ihres SD-Grades ausbalanciert werden. Systematisch geschieht dies z.B. in der auf  
entsprechenden Paarvergleichsurteilen basierenden Forced-Choice-Technik wie sie z.B. im EPPS  
(Edwards, 1954) realisiert ist. Neben der relativen Unökonomie der Methode läßt sich gegen sie  
u.U. auch einwenden, daß es nach vollzogener Selektion der Antwortkombinationen zu einer Re-  
organisation der SD-Hierarchie der Items kommen kann (vgl. auch Cowen et al., 1964). Gegen  
den Paarvergleich von SD-balancierten Items spricht jedoch auch grundsätzlich die mögliche  
Mehrdimensionalität des SD-Konzepts (vgl. Scott, 1963; N.Wiggins, 1966), die die Ausbalancie-  
rung der Items gemäß einem modalen SD-Wert als relativ sinnlos erscheinen ließe; in letzter  
Konsequenz wären entsprechende Paarungen stets nur für eine bestimmte Person ausbalanciert.

### 3.3 Kontrolle durch spezielle Kontrollskalen

Die seit den 50er Jahren existierenden SD-Skalen (z.B. Edwards, 1957; Crowne & Marlowe,  
1960; deutschsprachig z.B. Lück & Timaeus, 1969; Schmidt & Vorthmann, 1971; Dickenberger,  
Holtz, & Gniech, 1978) dienen zumeist der statistischen Post-hoc-Kontrolle durch Auspartiali-  
sierung der SD-Variablen. Sie können nach den Regeln der klassischen Testkonstruktion erstellt  
werden, nachdem die SDSV von Statements durch Rater bestimmt oder aufgrund von Untersuchungen  
mit Instruktionsmanipulation besonders SD-anfällige Statements selektiert worden sind. Das in  
die diversen Kontrollskalen gesetzte Vertrauen muß aber schwinden, wenn sich herausstellt,  
daß sie untereinander oft nur in geringem Maße (vgl. Ehlers, 1973; Angleitner, 1976) und -  
wie erwähnt - mit einer ganzen Reihe von Persönlichkeitsvariablen kovariieren. Müßte auch  
hier in der Konsequenz eine auf die jeweilige Stichprobe abgestellte Itemanalyse vorgenommen  
werden, so erwiesen sich die SD-Kontrollskalen nur scheinbar als ökonomischstes Mittel der  
SD-Kontrolle (vgl. Ehlers, 1973).

### 3.4 Kontrolle durch Instruktion

Viele Autoren versuchen, dem SD-Response-Set mittels Instruktion einen alternativen Set  
entgegenzusetzen, z.B. in Gestalt der Aufforderung, ganz besonders ehrlich zu antworten und  
sich nicht von sozialen Normen leiten zu lassen. Solche immunisierenden Instruktionen mögen  
Wirkungen zeitigen, solange sie nicht bereits als Routine-Zusätze zu Habituationseffekten füh-  
ren. Als effizient für die Zuschreibung von Erwünschtheits-Indizes zu Statements oder Fragen  
haben sich Faking-Instruktionen, d.h. Aufforderungen zur absichtlichen Verstellung erwiesen  
(vgl. z.B. Wiggins, 1959; Hoeth, Kucklick, & Simmat, 1965; Gorman, 1968; Lennertz, 1969; Eg-  
gert, 1971; Angleitner, 1976; Häcker et al., 1977). Mit dieser Methode ließ sich z.B. auch  
zeigen, daß mittels Forced-Choice-Technik kontrollierte diagnostische Verfahren gegenüber Fa-  
king-Instruktionen nach wie vor anfällig sind (Stollak, 1965), daß Neurotizismus-Messungen  
verfälschbarer als Extraversions-Messungen sind (Farley & Goh, 1976; Angleitner, 1976), usw.

Daß zur Herstellung eines "guten Eindrucks" die kürzeste, zur Herstellung eines "schlechten Eindrucks" dagegen die meiste Zeit aufgewendet wird, zeigten Schneider & Hübner (Anm.4); hieraus ließe sich schließen, daß die üblichen Tempoinstruktionen ("nicht lange nachdenken", "zügig antworten") ihr Ziel, SD-Effekte zu verringern, verfehlen. Daß Faking-Instruktionen situativ unterschiedlich wirken, zeigten wir kürzlich in einem unveröffentlichten Experiment: Verschiedene ethnische Gruppen wurden signifikant unterschiedlich beurteilt je nachdem, ob es sich um Faking-Instruktionen in einer Gruppe, in einer "minimal social situation" oder um eine Faking-Instruktion vom Typ "Bewerbungssituation" handelte. SD-Kontrolle durch Instruktion scheint also zwar (vgl. auch Ehlers, 1973) für manche Zwecke sinnvoll, weil sie die komplizierten Prozesse der Item- und Antwortselektion umgeht und nur auf die allgemeine Einstellung (set) der Vp abhebt; gruppen-, themen- und situationsspezifische Differenzierungen erscheinen aber am Platze, wenn sich nicht auch diese Kontrollmethode abnutzen soll.

#### 4. Experimentelle SD-Kontrolle durch Manipulation der Rahmensituation

Auf dem Gebiet der sozialpsychologischen Einstellungsmessung wurden Überlegungen und Experimente zur SD-Kontrolle unternommen, die einen Sonderfall der Kontrolle durch Instruktionsmanipulation darstellen. Sie zielen ab auf die Modifikation der gesamten Rahmensituation der Messung, d.h. auf die Herstellung einer Untersuchungssituation, in der Reaktionstendenzen durch Anstrengung der Vp selbst minimiert werden. Die Vp wird dabei unter den Laboratoriumsbedingungen scheinbarer physiologischer Messung in einen Zustand erhöhter Selbstaufmerksamkeit versetzt, der es für sie vordringlich macht, bei Selbstberichten über eigenes Verhalten zutreffender, also ehrlicher zu antworten und bei Urteilen über Einstellungsobjekte stärker individuellen als allgemeinen Erwünschtheitsnormen zu gehorchen. Das Bemühen des Vp, Täuschungs- bzw. Verfälschungstendenzen seitens der Vp durch eigene Täuschungsstrategien entgegenzuwirken, wird bei dieser Methode der SD-Kontrolle besonders deutlich.

Das als Bogus-Pipeline (BPL-) -Paradigma bezeichnete Verfahren wurde von E.E.Jones und Mitarbeitern entwickelt (vgl. Jones & Sigall, 1971). Die Vp lernt im ersten Teil des Versuchs, daß ein "Elektromyograph" die unwillkürlichen Muskelschwingungen in beiden Unterarmen registriert, während die Vp sich auf eine bipolare Urteilstkala konzentriert und ein spezielles Steuerrad festhält. Durch eine einfache, aber überzeugende Täuschungsprozedur gibt ein mit dem Gerät gekoppelter Computer Registrierungen, d.h. Urteils-Werte wieder, die den vermeintlich einigermaßen echten Werten der Vp ziemlich genau entsprechen (vgl. Sigall & Page, 1971; Mummendey, Schiebel, Troske, Hesener, & Bolten, (Anm.5)). Im zweiten Versuchsabschnitt geht es dann angeblich darum, zu untersuchen, in welchem Maße die Vp ihre wirklichen Einstellungen bzw. Bewertungen kenne bzw. ihr eigenes Verhalten richtig wahrnehme; dazu soll die Vp bei einer Reihe folgender Stimuli die "Computerwerte" vorhersagen. Diese Schätzungen werden als relativ SD-freie Urteile angesehen.

Einen Nachweis der Überlegenheit des BPL-Paradigmas über Papier-Bleistift-Verfahren bezüglich der Verringerung von SD-Tendenzen brachten Sigall & Page (1971). Eine Reihe von Replikationen und weiteren Untersuchungen über Urteilsobjekte, die SD-anfällig sind, hat im Laufe der Zeit bestätigt, daß die Herstellung einer verschärften Urteils-Rahmensituation im BPL-Paradigma zur Verringerung von SD-Tendenzen geeignet ist (vgl. zusammenfassend Brackwede, 1980). Über die im wesentlichen zufriedenstellenden Ergebnisse einer deutschen Replikation des Experiments von Sigall & Page (1971) wurde kürzlich in Tübingen berichtet (Mummendey, Anm. 6).

Geht man davon aus, daß durch die Instruktion an die Vp, so zu antworten, daß dies mit ihren "wirklichen Einstellungen" übereinstimmt, sie solle sich bemühen, "ihr eigenes Verhalten richtig wahrzunehmen", eine intensive Lenkung der Aufmerksamkeit auf die eigene Person gelingt, so kann man den Zustand der Vp als einen solchen der selbstzentrierten Aufmerksamkeit oder - in der Terminologie von Duval & Wicklund (1972) - der objektiven Selbstaufmerksamkeit (OSA) bezeichnen. Insofern wäre das Anschließen der Vp an den "Elektromyographen" funktional äquivalent dem Aufstellen eines Spiegels oder einer Kamera, der Wiedergabe der eigenen Stimme vom Tonband oder der Konfrontation mit einem Publikum, also jenen Hilfsmitteln, denen man sich bei Experimenten im Rahmen der OSA-Theorie bedient, um eine Person besonders selbstaufmerksam zu machen (vgl. Wicklund, 1975, 1979a, 1980). Da sich zeigen ließ, daß Personen im Zustande erhöhter OSA eher in Übereinstimmung mit ihren zuvor gemessenen Einstellungen urteilen, so daß es bei ihnen auch zu besserer Einstellungs-Verhaltens-Konsistenz kommt, scheint das OSA-Konzept ein geeigneter Ansatz zur Erklärung der SD-Kontrolle im BPL-Experiment. Berücksichtigt man ferner, daß das Individuum zur Vermeidung von Inkonsistenz im Zustande erhöhter OSA dem stärksten in der Situation aktuellen Druck folgt, so kann nach Betrachtung der spezifischen Charakteristika der BPL-Situation gut begründet werden, warum hier interne, personale Standards stärker bzw. salienter als Gruppenstandards werden: Die Vp wird vom V1 nachdrücklich auf eben diese "wirklichen eigenen" Wahrnehmungen bzw. Urteile hingewiesen, deren Messung dient offensichtlich und explizit rein wissenschaftlichen bzw. naturwissenschaftlichen Zwecken, usw. Experimentelle Belege für die Angemessenheit eben dieser theoretischen Interpretation stehen allerdings noch aus. Inzwischen haben wir lediglich in einer ganz direkten, auch von den amerikanischen Vorbildern nicht realisierten Weise gezeigt, daß in der BPL-Situation Erwünschtheit im üblichen (Fragebogen-) Sinne reduzierbar ist (Mummendey & Bolten, Anm. 7).

Die Konstruktion einer experimentellen Rahmensituation, in der die Vp lernt, sich beim Reagieren an relativ stärker individuell als gruppenspezifisch bestimmten Normen zu orientieren, erscheint für die Zukunft aussichtsreich, wenn die Praktikabilität des Verfahrens erhöht wird und ihm zugleich das Odium des Spektakulären genommen werden kann. Beide Probleme reichen zumindest teilweise bereits in forschungsethische Fragen der Anwendung und Verbreitung von Täuschungsverfahren hinein.

#### 5. Ethische Probleme der Kontrolle sozialer Erwünschtheit

Kurioserweise scheint sozial Erwünschtes unter Gesichtspunkten psychologischer Forschung nicht immer erwünscht zu sein. Zur Abwendung von Täuschung auf seiten der Vp setzt der Forscher Mittel der Täuschung (deception) ein - am deutlichsten im BPL-Paradigma, aber in abgeschwächter Form auch bei der Anwendung von Papier-Bleistift-Verfahren. Daher muß zumindest er sich die Fragen nach der Ethik gefallen lassen. Diese Fragen werden konventionellerweise von interessierten Psychologiekritikern, neuerdings jedoch auch von vielen experimentell arbeitenden Psychologen selbst aufgeworfen (vgl. z.B. Kelman, 1967, 1968; Schlenker & Forsyth, 1977; Schwitzgebel & Kolb, 1978; Stolz & Associates, 1978; West & Gunn, 1978; Diener & Crandall, 1978; American Psychological Association, 1979; Irle, 1979; Schuler, 1980; Kumpf, im Druck).

Einige Gegenargumente zur Anwendung von Täuschung lauten: Unter Ausnutzung sozialer Macht erschleicht sich der Forscher Erkenntnisse auf betrügerische Weise (vgl. Kelman, 1972); durch falsche Rückmeldung an die Vp wird diese u.U. nachhaltig geschädigt (vgl. Kumpf, im Druck); Täuschung im psychologischen Experiment kann im Sinne einer Modellierung von Unaufrichtigkeit

wirken (Baumrind, 1964); bei getäuschten Vpn kann Mißtrauen entstehen, das letztlich für die weitere Forschung schädlich sein kann (vgl. Kumpf, im Druck); das Wissen um Täuschung hat sich in einigen Untersuchungen in geringerem Maße als erwartet ausgewirkt (vgl. Schuler, 1980). Als Argumente für die Anwendung von Täuschung können u.a. dienen: Täuschung ist aus methodologischen Gründen, z.B. zur Erhaltung von Nonreaktivität, Spontaneität etc. geboten (vgl. Métraux, im Druck); Täuschung ist schon deshalb wenig problematisch, da sie in der psychologischen Forschung bekanntlich weithin üblich ist (vgl. Schuler, 1980); Täuschung geschieht nur vorübergehend, da postexperimentelle Aufklärung erfolgt; viele Vpn beurteilen Täuschung nachträglich als keineswegs ethisch anfechtbar - sie bewerten sie oft weniger negativ als Psychologen (vgl. Kumpf, im Druck); Täuschung könnte Teil einer impliziten oder sogar expliziten Vereinbarung zwischen Vp und Vl sein, wenn sich die Vp zu einem Experiment oder Test bereit erklärt (vgl. Schwitzgebel & Kolb, 1978); vollständige Offenheit und Ehrlichkeit ist eigentlich nirgendwo im öffentlichen Leben üblich (vgl. Schuler, im Druck); es ist bislang nichts über tatsächlich eingetretene Schädigungen durch Täuschung bekanntgeworden (vgl. Kumpf, im Druck).

Angesichts der begründet erscheinenden relativen Notwendigkeit der sparsamen Anwendung von Täuschung (vgl. Irle, 1979) wird die Einhaltung der einschlägigen Richtlinie aus dem Verhaltenskodex von Diener & Crandall (1978) in Verbindung mit den Subprinzipien (d) und (h) der APA-Richtlinien in der Version von 1979 (American Psychological Association, 1979) empfohlen: Täuschung nur nach einer Analyse der Forschungssituation unter ethischen Gesichtspunkten anwenden und die Vpn zum frühestmöglichen Zeitpunkt über die Notwendigkeit vorübergehender Täuschung informieren. Im Falle des BPL-Paradigmas wäre es sicherlich ethisch bedenklich (im Sinne des Prinzips 8a der APA-Richtlinien), wenn zur SD-Kontrolle Vpn mit einem funktionierenden Gerät, also einem effizienten Lügendetektor heimlich untersucht würden. Es wird jedoch in genau umgekehrter Weise verfahren: Mit einem nicht funktionierenden Gerät wird eine exakte Kontrolle vorgetäuscht. Ethisch bedenklich wäre es vielleicht auch, in Differentieller und Diagnostischer Psychologie SD-Kontrolle mittels Item- und Antwortformen sowie Kontrollskalen auszuüben, die tatsächlich effizient wären. Zum gegenwärtigen Zeitpunkt stellen diese Verfahren jedoch Hilfsmittel dar, deren Anwender einstweilen von allzu starkem Gewissensdruck entlastbar erscheinen.

#### 6. Anerkennung sozialer Erwünschtheit in Teilbereichen psychologischer Forschung?

Die Frage, ob Forschungsfragestellungen sinnvoll sind, die ohne die Auffassung von SD-Varianz als Fehlervarianz auskommen, so daß sich Täuschungsstrategien seitens der untersuchten Person wie auch des Untersuchers erübrigen, könnte für den Fall bejaht werden, daß (a) kein Wahrheitskriterium im strengen Sinne existiert (z.B. bei Einstellungs-Urteilen) und daß (b) die Kontamination von individuellem Urteil und sozialer Norm tolerierbar bzw. willkommen ist. Vermutlich treffen beide Voraussetzungen gegenwärtig weitgehend erst auf Teile der Selbstkonzeptforschung zu. Bei der Produktion von Selbstbildern und ihrer Messung (vgl. Mummendey, 1979) interessieren oft gerade die internalisierten oder auch bei psychologischer Untersuchung aktualisierten Orientierungen an Gruppennormen als selbstverständliche Bestandteile des Selbstkonzepts. Da letzterem allgemein eine verhaltensbeeinflussende Funktion zugeschrieben wird (vgl. z.B. Wicklund, 1979b; Schuler, 1980) erscheinen Selbstkonzeptmessungen auch dann als valide, wenn sie mehr oder weniger "realistisch" sind.

Anm. 1: Buxbaum, O. Die soziale Erwünschtheit von Persönlichkeitseigenschaften. Berichte aus dem Institut für Psychologie der Universität Graz, Juni 1976.

- Anm. 2: Buxbaum, O. Die soziale Erwünschtheit von Persönlichkeitseigenschaften (1979). Unveröffentlichtes Manuskript, Universität Graz, Institut für Psychologie, 1979.
- Anm. 3: Angleitner, A., Stumpf, H., & Wieck, T. Die "Personality Research Form" von Jackson: Konstruktion, bisheriger Forschungsstand und vorläufige Ergebnisse zur Äquivalenzprüfung einer deutschen Übersetzung. Wehrpsychologische Untersuchungen, 1976, 3.
- Anm. 4: Schneider, J.F., & Hübner, R. Einfluß von Verfälschungsinstruktionen auf die Bearbeitungszeit von Persönlichkeitsfragebogen. Arbeiten der Fachrichtung Psychologie der Universität des Saarlandes, Nr. 68, Universität Saarbrücken, 1980.
- Anm. 5: Mummendey, H.D., Schiebel, B., Troske, U., Hesener, B., & Bolten, H.-G. Experimentelle Replikation des Bogus-Pipeline-Effekts für ethnische Stereotype. Bielefelder Arbeiten zur Sozialpsychologie, Nr. 55, Universität Bielefeld, Dezember 1979.
- Anm. 6: Mummendey, H.D. Das Bogus-Pipeline-Paradigma für ethnische Stereotype: Eine experimentelle Replikation mit vereinfachter Versuchsanordnung. Referat auf der 22. Tagung Experimentell Arbeitender Psychologen in Tübingen, 1980.
- Anm. 7: Mummendey, H.D., & Bolten, H.-G. Die Veränderung von Social-Desirability-Antworten im Bogus-Pipeline-Experiment. Bielefelder Arbeiten zur Sozialpsychologie, Nr. 60, Universität Bielefeld, April 1980.