

Descriptive consistency and social desirability in self- and peer reports

PETER BORKENAU and FRITZ OSTENDORF
University of Bielefeld, FRG

Abstract

In the present study, 300 subjects were administered 20 sets of four trait-descriptive terms where aspects of content and evaluation were unconfounded (e.g. firm, severe, lenient, and lax). Each subject was also evaluated by three peers using the same sets of four trait terms. Moreover, the subjects responded to several personality inventories and rating scales, and they were also described on these rating scales by their peers. The results showed that the subjects frequently ascribed to themselves or to their peers two favourable trait terms that were descriptively inconsistent (e.g. firm, lenient). A measure of individual differences in socially desirable responding was constructed by summing all desirable responses. Subjects who described themselves in a socially desirable manner were less neurotic and more conscientious according to self-reports as well as peer reports. Several implications of the findings are discussed, and the present SD measure is compared with several well-known desirability scales.

INTRODUCTION

The social desirability (SD) of traits and behaviours exerts a major influence on personality trait ratings and questionnaire item endorsements. The endorsement frequency of items is almost perfectly predictable from the SD of the item content (Edwards, 1953; Edwards and Diers, 1962; Norman, 1967). Moreover, the first principal component in the Minnesota Multiphasic Personality Inventory (MMPI) distinguishes among item responses that are socially desirable or undesirable (Jackson and Messick, 1958). Similar findings have been obtained for Wiggins's (1979) circumplex model of interpersonal behaviour (Jackson and Helmes, 1979). Accordingly, SD may be a major source of the covariation among personality measures too.

Several factors have been suggested which may contribute to these phenomena. The relationship between endorsement frequency and SD may reflect that people

Correspondence concerning this article should be addressed to Peter Borkenau, Abteilung Psychologie der Universität Bielefeld, Postfach 8640, D-4800 Bielefeld 1, F.R.G.

learn to behave in socially desirable ways in their everyday life, and that therefore desirable activities occur more frequently than undesirable activities. Furthermore, it is reasonable to suppose that at least some people tend to describe themselves more favourably than is appropriate. This may happen because their self-images are more favourable than is justified (self-deception), or because they deliberately fake their responses in order to create a favourable impression that others should hold of them (other-deception). The same explanations may account for positive correlations, across subjects, among traits and behaviours that are evaluated in a similar way. Thus, there may be either a strong factor of good versus bad character, or evaluation may account for an important proportion of the variance in person perception, or the subjects' tendency to fake may be a highly general one. The latter two sources may contribute to what is usually called the *halo effect* (Cooper, 1981; Thorndike, 1920).

The most popular approach to measuring socially desirable responding is that suggested by Crowne and Marlowe (1960). In their scale, the subjects are asked if they performed either behaviours that are highly desirable but very rare (e.g. '*Are you always courteous, even to people who are disagreeable?*'), or if they performed behaviours that are undesirable but very common within our culture. Endorsement of the first type of items and denial of the second type of items are thereby assumed to indicate a tendency to respond more favourably than is justified.

Another approach to contrast effects of evaluation with those of descriptive aspects has been suggested by Peabody (1967) in the context of efforts to investigate the importance of evaluation in trait inferences. Peabody formed sets of four trait-descriptive terms such that either similar dispositions were described by trait-descriptive terms with an opposite evaluation, such as *firm* and *severe*, or opposite dispositions were described by trait-descriptive terms with a similar evaluation, such as *firm* and *lenient*. One set of four trait terms that exemplifies Peabody's approach is depicted in Table 1.

Table 1. Scheme for unconfounding content and desirability

	Evaluative contrast	
	Favourable term	Unfavourable term
Descriptive	Firm	Severe
Contrast	Lenient	Lax

Peabody asked his subjects to estimate whether, for example, a *firm* person was more likely to be *severe* or to be *lenient*. For 70 appropriate comparisons, it turned out that the descriptively similar and evaluatively opposite term was uniformly preferred to the evaluatively similar and descriptively opposite term. Thus, for example, a *firm* person was deemed more likely also to be *severe* than to be *lenient*. Accordingly, when evaluatively consistent trait inferences resulted in clear contradictions at the descriptive level, the subjects preferred descriptive consistency over evaluative consistency in trait inferences for fictitious targets. But it is still to be shown that the higher importance of descriptive aspects, than of evaluative aspects, holds also for direct ratings of self and peers.

There are two reasons for doubting that the finding of a higher importance of descriptive than of evaluative aspects, which was obtained by Peabody (1967) for fictitious targets, should also be obtained for self- and peer ratings. First, negative

evaluations may more readily be made for fictitious stimulus persons than for the self or for acquaintances. Second, descriptive inconsistencies at the trait level may be felt less pressing if actual persons are described. If fictitious stimulus persons are introduced as *firm*, for example, in a trait inference task, this may be understood to imply that they are prototypes of firm persons. It may then appear obvious that these same targets cannot appropriately be described as lenient. By contrast, if a subject prefers one pole of a rating scale to the other pole in a self-report, this does *not* imply that the preferred term perfectly matches the subject's personality. The subject may sometimes even behave in ways that justify the ascription of the opposite trait. Thus, for example, persons may be very firm in general, but lenient towards their own children. It follows from this reasoning that, whenever subjects start from the assumption that the more positive of two traits matches their personality better than the more negative trait, some confirming behavioural evidence can be found that justifies the ascription of the positive trait. Accordingly, descriptive inconsistencies may occur more frequently in self-reports than in trait inferences for fictitious stimulus persons.

A study by Norman (1988) supports this hypothesis. He used two of Peabody's (1967) 15 sets of four trait terms in a self-report task. Moreover, he added a second (descriptively and evaluatively) equivalent term to each of the eight (2×4) cells. Thus, there were 16 traits under study. Ratings on each item were made on a nine-point scale ranging from 'extremely inaccurate as a self-description' (1) to 'extremely accurate as a self-description' (9). The total variance in the subjects' responses was then decomposed, and the effect of item differences in social desirability accounted for about 25 per cent of the total variance, implying that the subjects generally preferred desirable traits to undesirable traits. Moreover, the Desirability × Subjects one-way interaction accounted for another 10 per cent of the total variance, implying that some subjects described themselves more favourably than did others.

The present study differs from Norman's (1988) in two respects. First, more trait words were used, as self-ratings were collected for 20 sets of four trait-descriptive terms that resembled Peabody's (1967) terms. Second, the subjects had to choose between two of the traits within a set of four, because, for each set of four, the two descriptive *and* evaluative opposites were combined into one item. The subjects were then asked which of the two terms described them more appropriately. For example, the subjects got the item-pair *I am more firm than lax* and *I am more lenient than severe*. Three different response patterns may occur for these items: (a) subjects may describe themselves as *firm* and *lenient*, thus choosing two opposite and favourable traits; (b) subjects may describe themselves either as *firm* and *severe* or as *lenient* and *lax*, thus responding consistently at the descriptive level at the cost of some desirability; or (c) subjects may describe themselves as *severe* and *lax*. The last response pattern is inconsistent and extremely undesirable, and it can thus be expected to be very rare. This response pattern may indicate that the subjects did not even read the items or answered randomly.

However, a problem with such item-pairs is that some subjects may appropriately endorse both items: (a) *I am more firm than lax*, and (b) *I am more lenient than severe*. Such a response pattern may be appropriate for subjects with an intermediate position on the respective trait dimension, because negatively evaluated traits tend to describe more extreme personal dispositions (Borkenau

and Ostendorf, 1987; Goldberg, 1982; Peabody, 1967). Thus, if evaluation is ignored, it is possible to project the four above-mentioned traits onto a single dimension of tough versus loose control, as indicated in Figure 1. Individuals who hold the position *X* on the underlying trait dimension, can appropriately be described as both: more *firm* than *lax* and more *lenient* than *severe*. This problem may be circumvented, however, by letting subjects fill out rating scales instead of instructing them to make binary choices. Whereas it is credible that the same persons are somewhat more firm than lax as well as being somewhat more lenient than severe (see Figure 1), it is highly unlikely that persons are extremely more firm than lax and, simultaneously, extremely more lenient than severe. Accordingly, it can be concluded from the latter response pattern that the subjects, in giving maximally socially desirable responses, allowed for pronounced descriptive inconsistencies.



Figure 1. Projection of a set of four trait terms on an underlying dimension of tough versus loose control

In the present study, we investigated (a) whether Peabody's finding of a higher importance of descriptive than of evaluative aspects may be generalized from trait inferences for fictitious stimulus persons to self- and peer reports, (b) whether the lower extremity of favourable terms accounts for the socially desirable response pattern, (c) which self-reported personality variables covary with this form of socially desirable responding, and (d) whether subjects who respond desirably are also more favourably evaluated by their peers.

METHOD

Subjects

The subjects were 300 German adults (129 male, 171 female) with a mean age of 26.4 years and an age standard deviation of 9.84 years. The subjects were recruited by an article in the local newspaper. They pursued various professions and were paid for their cooperation. A precondition for their participation was that each subject had to be accompanied by three relatives and/or acquaintances who had to provide peer ratings. In this way, the independence of self- and peer reports was secured. But the subjects could also participate in this study as groups of four mutual acquaintances, and they had then to report about each other reciprocally. Almost all subjects made use of this provision.

Material

Measures of socially desirable responding

Two versions of this measure were used. A typical item-pair of the first version was *I am more firm than lax* and *I am more severe than lenient*. These items could be

either endorsed or denied. The 40 items of this version were intermingled among 531 dichotomously answered questionnaire items. The other version consisted of pairs of six-point rating scales (i.e. without a neutral point), for example:

lax	-3	-2	-1	+1	+2	+3	firm
severe	-3	-2	-1	+1	+2	+3	lenient

The 40 scales of the second version were interspersed among 102 six-point rating scales. Whereas the dichotomous items were only filled out by the subjects in their self-reports, the rating scales were administered as self-ratings and as peer ratings.¹

The development of 20 German sets of four trait terms that meet Peabody's (1967) criteria has been described in more detail by Borkenau and Ostendorf (1987).² Put in a nutshell, three criteria were employed: First, undesirable traits had a mean social desirability rating of below -1 and desirable traits of above $+1$ on a seven-point rating scale with the endpoints *extremely undesirable* (-3) and *extremely desirable* ($+3$). Second, within each set of four trait terms, the judged similarity of a given trait with its descriptive counterpart (which was opposite in evaluation), was higher than the similarity with its descriptive opposite (which was similar in evaluation). For example, *firm* was judged as more similar to *severe* than to *lenient*. Finally, another group of judges who completed a trait-inference task, preferred the descriptively similar term to the evaluatively similar term. For example, when asked whether a *firm* person was more likely to be either *severe* or *lenient*, the subjects indicated that this person was more likely to be *severe*. Only those sets of four trait terms where each single trait met all three criteria were retained. Thus, our measure of socially desirable responding is a conservative one as it favours the impact of descriptive consistency at the expense of evaluative consistency.

Personality scales and lie scales

The subjects were administered German versions of: A short version, with 14 scales, of the Personality Research Form (Jackson, 1967); Costa and McCrae's (1985) NEO Personality Inventory; Form A of the Eysenck Personality Inventory

¹ To unconfound socially desirable responding and acquiescence, the items were administered to the subjects such that desirable responses were associated with positive digits for 20 items and with negative digits for the remaining 20 items. The latter responses, however, were afterwards recoded such that positive digits indicate a socially desirable response.

² If readers intend to use SD scales as here suggested in English, it is recommended to use Peabody's (1967) terms. The German terms that were used were: (a) beständig, änderungsfähig, unveränderlich, wechselhaft; (b) standhaft, beweglich, unflexibel, unstetig; (c) konsequent, flexibel, starr, unschlüssig; (d) zielstrebig, genügsam, gierig, faul; (e) selbstbewußt, bescheiden, eingebildet, schüchtern; (f) gesellig, unaufdringlich, zudringlich, verschlossen; (g) offen, diskret, aufdringlich, zugeknöpft; (h) freimütig, diplomatisch, taktlos, unehrlich; (i) kultiviert, natürlich, gekünstelt, naiv; (j) locker, zuverlässig, unzuverlässig, pedantisch; (k) aktiv, entspannt, hektisch, lahm; (l) lebhaft, gemütlich, ruhelos, träge; (m) munter, ruhig, exaltiert, schlaff; (n) agil, gelassen, überdreht, apathisch; (o) energievoll, behutsam, explosiv, energielos; (p) ernsthaft, humorvoll, verbissen, kindisch; (q) sachlich, fröhlich, humorlos, albern; (r) sparsam, grosszügig, geizig, verschwenderisch; (s) gefestigt, anpassungsfähig, eingefahren, labil; (t) kritisch, begeisterungsfähig, mäkelig, unkritisch. For each set of four trait terms, the first two are desirable and the latter two are undesirable.

(Eysenck and Eysenck, 1964); and the Freiburger-Persönlichkeits-Inventar (Fahrenberg, Hampel and Selg, 1984), a personality inventory that is widely used in Germany. Moreover, the subjects filled out German versions of the Marlowe-Crowne Social Desirability Scale (Crowne and Marlowe, 1960) and of Edwards' Social Desirability Scale (Edwards, 1957). Furthermore, each subject was independently described by three relatives and/or acquaintances with the use of 102 rating scales: three rating scales for each of the 14 PRF scales were administered which measured different facets of each of the 14 constructs. Forty more scales, that is 20 scale-pairs, were used as our measure of social desirability in peer ratings. Moreover, Norman's (1963) 20 marker scales for the five major factors of personality were administered. The peers filled out the rating scales under the experimenter's supervision. Finally, all rating scales that were administered in peer ratings were also administered in self-reports.

RESULTS

Descriptively consistent versus socially desirable responding

The responses to the two items that referred to a common set of four trait terms were first cross-tabulated, and it turned out that inconsistent response patterns occurred more frequently than consistent response patterns. When the respective items had to be either endorsed or denied, 59.2 per cent of the response patterns were of the socially desirable variety. This implies that the subjects reported, for example, that they were both, more *firm* than *lax* and more *lenient* than *severe*. Another 3.5 per cent of the patterns were inconsistent and extremely undesirable. The remaining 37.3 per cent were consistent, that is the subjects chose one desirable term and one undesirable term.

However, because the first seemingly inconsistent response pattern may be appropriate for some subjects if the items have to be either endorsed or denied (see Figure 1), the cross-tables for the rating scale version of our SD measure are more revealing. Table 2 reports the entries (percentages) in each of the 36 cells for the self-ratings, whereas Table 3 reports these percentages for the peer ratings.

Tables 2 and 3 are extremely similar: the majority of the response patterns (i.e. 66 per cent for the self-ratings and 61 per cent for the peer ratings) cluster in the lower right quadrant. This implies that the judgments are on the positively evaluated side of the rating scale for both scales of a pair. In contrast, very few response patterns (2 per cent for the self-ratings and 4 per cent for the peer ratings) are found in the upper left quadrant (which indicates an undesirable and inconsistent choice). The remaining third of the response patterns is clearly consistent. The high similarity of Tables 2 and 3 is remarkable and shows that the effects of social desirability (for self-ratings) and of leniency (for peer ratings) are highly similar. It should be considered in this context, however, that the subjects themselves nominated the peers. The relationship between subjects and peers can therefore be assumed to be favourable.

Can the high proportion of socially desirable response patterns be explained by the higher extremity of negatively evaluated terms? Tables 2 and 3 do not support that explanation. In both tables, for example, the highest number of entries is

Table 2. Cross-tabulation of the rating categories applied in self-reports

		First scale						Desirable trait (e.g. lenient) +3 Sum	
		Undesirable trait (e.g. severe)							
		-3	-2	-1	+1	+2	+3		
Second scale	Undesirable trait (e.g. lax)	-3	0.02	0.03	0.03	0.10	0.35	0.57	1.11
		-2	0.05	0.20	0.39	1.06	2.17	0.76	4.63
		-1	0.10	0.34	0.89	2.70	3.79	1.23	9.05
	Desirable trait (e.g. firm)	+1	0.15	1.31	2.91	8.27	7.78	1.94	22.38
		+2	0.40	2.93	5.88	10.45	17.74	4.79	42.19
		+3	0.72	1.90	2.81	4.28	6.76	4.16	20.64
	Sum		1.45	6.72	12.92	26.10	38.60	13.45	100.00

Note: The entries in the body of the table are percentages that refer to 6000 entries altogether (i.e. 300 subjects \times 20 scale-pairs).

Table 3. Cross-tabulation of the rating categories applied in peer-reports

		First scale						Desirable trait (e.g. lenient) +3 Sum	
		Undesirable trait (e.g. severe)							
		-3	-2	-1	+1	+2	+3		
Second scale	Undesirable trait (e.g. lax)	-3	0.16	0.16	0.13	0.24	0.39	0.44	1.51
		-2	0.16	0.41	0.64	1.32	2.09	1.05	5.66
		-1	0.13	0.62	1.35	3.70	4.66	1.63	12.08
	Desirable trait (e.g. firm)	+1	0.29	1.37	3.70	8.91	7.95	2.42	24.63
		+2	0.60	2.53	5.26	10.50	12.68	4.51	36.09
		+3	0.73	1.92	3.02	4.41	5.75	4.19	20.02
	Sum		2.06	7.02	14.09	29.07	33.52	14.24	100.00

Note: The entries in the body of the table are percentages that refer to 18 000 entries altogether (i.e. 300 subjects \times 3 peer ratings per subject \times 20 scale-pairs).

found in the +2/+2 cell. This indicates that the ratings on the two scales of a pair tend to be both quite favourable and quite extreme. Furthermore, if we focus on those cases where one scale has received an extremely favourable rating (i.e. +3), we find that about 75 per cent of the responses are on the favourable side for the other scale too. Such a response pattern, however, is descriptively inconsistent even in view of the higher extremity of unfavourable traits. For example, if persons describe themselves as extremely *firm* as opposed to *lax*, it is inappropriate that they also describe themselves as more *lenient* than *severe*. Accordingly, subjects who infer traits of fictitious targets appropriately infer *severe* (rather than lenient) from *firm* (Borkenau and Ostendorf, 1987; Peabody, 1967).

These findings suggest that what Peabody (1967) found for trait inferences in fictitious targets, namely that subjects prefer descriptive consistency to evaluative consistency in case of clear descriptive relations among trait terms, is not found for self- and peer reports. Pronounced descriptive inconsistencies occur in judgments

about actual people, and the great majority of these inconsistencies is of the socially desirable variety.

Individual differences in socially desirable responding

Internal structure of the SD measure

For each of the 20 pairs of self-report rating scales, the responses to the two scales were correlated across subjects. For example, the *severe-lenient* scale was correlated with the *lax-firm* scale. The appropriate correlations for the 20 scale-pairs range from -0.36 to 0.28 , and their average is -0.04 . This implies that the responses to the two scales of a pair are nearly uncorrelated. Whereas a strong negative correlation indicated the predominance of content in the subjects' responses, a strong positive correlation indicated the predominance of evaluation. Thus, the present zero correlation indicates that the effects of content and desirability are of similar importance and cancel each other out.

Next, the responses to the two scales of each pair were added. Given that high scores indicate preference for the socially desirable trait (see Table 2), this composite indicates the extent of socially desirable responding and, because the two scales of a pair are counter-balanced for content, this measure of SD is unconfounded with content. The 20 composite scores were then correlated and factor-analysed. A positive manifold structure was obtained; all correlations were positive in sign, but there were three factors with eigenvalues $\lambda > 1.00$ that explained 35 per cent of the total variance. Moreover, the three Varimax-rotated factors were clearly interpretable in terms of content. Whereas the scales with high loadings on the first factor refer to the activity domain, those with high loadings on the second factor refer to perseverance, and those with high loadings on the third factor emphasize facets of sociability. Thus, some subjects respond more descriptively consistent in case of specific content domains. However, the positive manifold structure is also remarkable, and it indicates that individuals differ consistently in their degree of socially desirable responding.

Covariates of socially desirable responding

The number of desirable responses to the 40 items (i.e. 20 item-pairs) was used as a measure of the subjects' overall SD tendency. In the questionnaire version of our SD measure, choices of the favourable term were scored 1 and choices of the unfavourable term were scored 0. The mean for this measure was 31.12, its standard deviation 5.35, and its reliability $\alpha = 0.83$. In the rating scale version of this measure, the response to the 40 scales (which had been recoded in the direction of SD) were added. The mean of this measure was 51.72, its standard deviation 22.22, and its reliability 0.90. The correlation between the two versions was 0.80 and thus approximated the highest covariation that can be expected, given the limited reliability of the measures.

Relationship with lie scales and SD scales

As the subjects had been administered several SD scales, namely the Marlowe-Crowne SDS, Edwards' SDS, and the EPI Lie Scale, these scales were correlated and factor-analysed together with the questionnaire version of the present SD

measure. Two factors with eigenvalues $\lambda > 1.00$ were obtained which together accounted for 78 per cent of the total variance. The correlations are reported in Table 4, and the Varimax-rotated factor pattern is reported in Table 5.

Table 4. Correlations among measures of socially desirable responding

	1	2	3	4
1. Present SD measure (Questionnaire version)	—	0.45	0.23	0.52
2. Marlowe–Crowne Scale		—	0.53	0.37
3. EPI Lie Scale			—	0.19
4. Edwards' Scale				—

Table 5. Varimax-rotated factor pattern of the SD scales

Scale	Factor 1	Factor 2
1. Present SD measure (Questionnaire version)	0.84	0.22
2. Marlowe–Crowne SDS	0.40	0.77
3. EPI Lie Scale	0.03	0.92
4. Edwards' SDS	0.87	0.10

Note: Principal components analysis.

The present SD measure and Edwards' Scale have high loadings on the first factor, whereas the Marlowe–Crowne Scale and the EPI Lie Scale have high loadings on the second factor. However, the Marlowe–Crowne Scale also has a substantial loading on the first factor. Given that Edwards' Scale has been demonstrated to be a marker variable for *Self-Deception*, whereas the EPI Lie Scale is a marker for *Impression Management*, and the Marlowe–Crowne Scale has substantial loadings on both factors (Paulhus, 1986), the first factor in Table 5 may be interpreted as *Self-Deception* and the second factor as *Impression Management*. Accordingly, the present SD measure is highly saturated with *Self-Deception*.

Personality correlates of socially desirable responding

An overview of the personality correlates of socially desirable versus consistent responding is conveyed by a common factor analysis of all personality scales and all SD scales under study. Five factors were retained which together accounted for 64 per cent of the total variance. The Varimax-rotated factor pattern is reported in Table 6. It replicates the finding that the five-factor model of personality is not only supported by factor-analytic studies on peer ratings, but that highly similar factors are also obtained when a heterogeneous sample of personality scales is submitted to a common factor analysis (Amelang and Borkenau, 1982; McCrae and Costa, 1987). The first factor may be interpreted as *Neuroticism*, the second factor as *Extraversion*, the third factor as *Conscientiousness*, the fourth factor as *Agreeableness*, and the fifth factor as *Openness to Experience* or *Intellect*. The respective scales of Costa and McCrae's (1985) NEO Personality Inventory are marker variables for these factors. Of the various SD scales, the EPI Lie Scale has the lowest loadings on these factors. This, however, may be due to its low reliability in

Table 6. Varimax-rotated factor pattern of all the questionnaire scales

Scale	Factors				
	1	2	3	4	5
Personality Research Form (PRF)					
Achievement	0.06	0.13	0.76	0.15	-0.24
Affiliation	-0.06	0.64	-0.04	0.48	0.27
Aggression	0.34	0.48	-0.11	-0.58	0.01
Dominance	-0.16	0.63	0.36	-0.36	-0.12
Endurance	-0.25	-0.06	0.75	0.03	-0.16
Exhibition	0.02	0.79	-0.08	-0.13	-0.15
Harm-avoidance	0.20	-0.48	0.13	0.19	0.42
Impulsivity	0.29	0.39	-0.61	0.01	-0.28
Nurturance	0.17	0.20	0.21	0.73	0.02
Order	-0.10	-0.01	0.60	0.10	0.47
Play	-0.01	0.69	-0.44	-0.01	0.06
Social recognition	0.31	0.36	0.10	0.24	0.51
Succorance	0.50	0.14	-0.37	0.38	0.29
Understanding	0.00	0.01	0.19	0.04	-0.71
Freiburger Persönlichkeitsinventar (FPI)					
Contentedness	-0.64	0.21	0.22	0.11	0.26
Social orientation	0.23	0.03	0.11	0.69	-0.31
Achievement orientation	-0.15	0.49	0.59	-0.10	-0.07
Restraint	0.39	-0.64	-0.19	0.06	0.19
Irritability	0.69	0.17	-0.13	-0.06	0.17
Aggression	0.29	0.42	-0.05	-0.56	0.04
Proneness to stress	0.71	-0.02	0.23	0.09	-0.03
Somatic complaints	0.70	-0.11	0.01	0.12	0.07
Health concerns	0.23	-0.11	0.30	-0.03	0.39
Extraversion	-0.04	0.89	0.09	-0.07	-0.06
Emotionality	0.91	-0.00	-0.04	-0.03	-0.06
NEO Personality Inventory (NEO-PI)					
Extraversion	-0.02	0.84	0.03	0.14	0.00
Neuroticism	0.80	-0.03	-0.32	0.08	-0.05
Openness	0.18	0.24	-0.12	0.09	-0.72
Agreeableness	-0.01	-0.14	-0.00	0.81	0.14
Conscientiousness	-0.17	0.02	0.75	0.07	0.36
Eysenck Personality Inventory (EPI)					
Extraversion	0.03	0.86	-0.19	-0.03	0.01
Neuroticism	0.89	-0.09	-0.15	0.07	0.01
Lie Scale	-0.22	-0.21	0.35	0.22	0.12
Marlowe-Crowne SDS					
Marlowe-Crowne SDS	-0.43	-0.28	0.49	0.29	0.14
Edwards' SDS					
Edwards' SDS	-0.82	0.18	0.24	0.03	-0.07
Present SD measure					
Present SD measure	-0.56	-0.09	0.42	0.23	0.03

Note: Principal components analysis.

our study ($\alpha=0.31$). In contrast, Edwards' SD Scale is a marker for *Emotional Stability-Neuroticism*. Thus, the classic critique of this scale that it is a measure of psychological health is supported by the present study. Surprisingly, the pattern of factor loadings for the Marlowe-Crowne Scale and for the present SD measure are quite similar to each other; subjects who score high on these scales describe themselves as less neurotic and as more conscientious than subjects who score low.

Table 7. Correlations between the present SD measure and peer ratings

Peer rating	SD scale version	
	Questionnaire	Rating scales
PRF constructs		
Achievement	0.19**	0.09
Affiliation	-0.09	-0.02
Aggression	-0.10	-0.07
Dominance	0.00	0.02
Endurance	0.27**	0.19**
Exhibition	-0.15*	-0.10
Harm-avoidance	0.06	0.06
Impulsivity	-0.26**	-0.21**
Nurturance	0.10	0.05
Order	0.25**	0.24**
Play	-0.25**	-0.14*
Social recognition	-0.16*	-0.12
Succorance	-0.25**	-0.21**
Understanding	0.14*	0.07
Norman factors		
Neuroticism	-0.20**	-0.23**
Extraversion	-0.12	0.00
Conscientiousness	0.36**	0.32**
Agreeableness	0.02	0.02
Culture	-0.05	-0.05

Note: $N = 300$; * $p < 0.05$; ** $p < 0.01$.

This implies that a high score in these measures covaries with socially desirable responding in the diagnostic scales. Comparable results are also obtained when the rating scale version of the present SD measure is factor-analysed together with all other self-ratings. Similar factors are identified and the SD measure has loadings of -0.71 on *Neuroticism*, of 0.09 on *Extraversion*, of 0.20 on *Conscientiousness*, of 0.43 on *Agreeableness*, and of 0.27 on *Openness/Culture*. Thus, the highest (negative) loading is again found for *Neuroticism*.

These findings are somewhat ambiguous, however, because it is dubious whether the subjects who score high on SD are indeed less neurotic and thus appropriately described by more favourable trait terms, or if they only appear as less neurotic in questionnaires because they are more inclined to camouflage their personal weaknesses. This problem can be clarified by correlating the SD scores of the subjects with their descriptions by the peers. Accordingly, the composite score of the three peer ratings per scale was calculated. Moreover, because three scales had been administered as criterion variables for each PRF scale, these three scales were also aggregated. The same procedure was followed for the four scales that are marker variables for each of the five Norman factors. Accordingly, there are 14 scores that measure the PRF constructs and five scores that measure the Norman factors. These scores were correlated with the two versions of the present SD measure. The results are reported in Table 7 and show that subjects who respond more desirably are less neurotic and more conscientious according to self-report as well as according to peer report. Subjects who are more favourably evaluated by their peers tend to describe themselves in a favourable way at the cost of descriptive consistency.

DISCUSSION

The results of the present study may be summarized as follows: First, inconsistent response patterns occur about twice as frequently as consistent response patterns, and nearly all inconsistent response patterns originate from two socially desirable responses. This finding cannot be explained by the higher extremity of undesirable traits. Second, the socially desirable response patterns for the 20 sets of four trait terms have a positive manifold structure; they correlate positively across subjects in their entirety, but more highly so if they refer to similar content. Finally, the overall tendency to respond in a socially desirable way covaries directly with self-reported and peer-reported conscientiousness, and it covaries inversely with self-reported and peer-reported neuroticism.

Regarding the importance of descriptive and evaluative aspects, the present findings differ from those for trait inferences in fictitious stimulus persons (Peabody, 1967; Borkenau and Ostendorf, 1987). For example, Peabody asked whether a *firm* person was more likely to be *severe* or *lenient*, and his subjects responded that such a person was more likely to be *severe*. In contrast, the subjects of the present study were asked to rate themselves (or their acquaintances) on a *lax-firm* as well as on a *severe-lenient* scale, and even those subjects who described themselves (or their peers) as extremely *firm*, described themselves (or their peers) predominantly as more *lenient* than *severe*. Accordingly, in order that subjects prefer descriptive consistency to evaluative consistency, it is not sufficient that unambiguous descriptive relations can be identified among the attributes. This condition held in both studies, Peabody's and the present one. The source of the discrepant findings must therefore be located in the remaining differences between the studies. One difference is the higher ego involvement of subjects in the present study who evaluated themselves or their acquaintances instead of fictitious stimulus persons. Another difference is the higher salience of descriptive inconsistencies in Peabody's study in cases where the subjects opted for the evaluatively similar trait. In Peabody's (1967) study, the subjects compared three trait terms which were presented to them simultaneously in a single trait-inference task. In contrast, in the present study, the subjects compared pairs of trait terms for their appropriateness to describe actual persons.

It may be argued that, in the present study, subjects may have been completely accurate if they described themselves as, for example, extremely *firm* and extremely *lenient*, because behaviour depends on the situation and the same person may therefore be sometimes *firm* and sometimes *lenient*. It is highly likely indeed that some confirming evidence may be found for both of these traits in the same person. However, situational specificity of behaviour does not justify this response pattern because such persons could then also describe themselves as extremely *severe* and extremely *lax*. The most reasonable explanation of the present findings is therefore that most subjects hold the expectation that favourable traits describe their personality better than unfavourable traits. But there are also pronounced individual differences in this respect. Moreover, these differences covary with the favourability of peer evaluations. Thus, a self-esteem variable may be involved which reflects individual differences in observable conduct and different peer evaluations to some extent. Subjects with a lower self-esteem and a less favourable

evaluation by their peers are more willing to concede socially undesirable traits and respond more consistently to the present SD scale, accordingly.

A parallel may be seen here to findings in research on depression that depressive subjects sometimes have more accurate cognitions than non-depressives (Abramson, Alloy and Rosoff, 1981; Kuiper and McDonald, 1982; Lewinsohn, Mischel, Chaplin and Barton, 1980). Although no measure of depression was administered in the present study, such a measure would presumably correlate substantially with the measures of neuroticism. It may therefore be expected that depressive subjects would have more descriptively consistent response patterns than would non-depressives, if they were administered the present SD measure.

ACKNOWLEDGEMENT

The research reported in this paper was supported by a grant from the Deutsche Forschungsgemeinschaft to the first author (Az. Bo 774/2-1).

This paper was written while the first author was a visitor at the Institute of Personality Assessment and Research at the University of California, Berkeley, USA.

REFERENCES

- Abramson, L. Y., Alloy, L. B. and Rosoff, R. (1981). 'Depression and the generation of complex hypotheses in the judgment of contingency', *Behavior Research and Therapy*, **19**: 75-86.
- Amelang, M. and Borkenau, P. (1982). 'Über die faktorielle Struktur und externe Validität einiger Fragebogen-Skalen zur Erfassung von Dimensionen der Extraversion und emotionalen Labilität' [On the factor structure and external validity of some questionnaire scales measuring dimensions of extraversion and neuroticism], *Zeitschrift für Differentielle und Diagnostische Psychologie*, **3**: 119-146.
- Borkenau, P. and Ostendorf, F. (1987). 'Ein Versuch der Trennung deskriptiver und evaluativer Bedeutungskomponenten in Eigenschafts-beschreibenden Begriffen' [An approach to unconfound descriptive and evaluative meaning components in trait-descriptive terms], *Archiv für Psychologie*, **139**: 189-207.
- Cooper, W. H. (1981). 'Ubiquitous halo', *Psychological Bulletin*, **90**: 218-244.
- Costa, P. T. and McCrae, R. R. (1985). *The NEO Personality Inventory Manual*, Psychological Assessment Resources, Odessa, FL.
- Crowne, D. P. and Marlowe, D. (1960). 'A new scale of social desirability independent of psychopathology', *Journal of Consulting Psychology*, **24**: 349-354.
- Edwards, A. L. (1953). 'The relationship between the judged desirability of a trait and the probability that the trait will be endorsed', *Journal of Applied Psychology*, **37**: 90-93.
- Edwards, A. L. (1957). *The Social Desirability Variable in Personality Assessment and Research*, Dryden, New York.
- Edwards, A. L. and Diers, C. J. (1962). 'Social desirability and the factorial interpretation of the MMPI', *Educational and Psychological Measurement*, **22**: 501-509.
- Eysenck, H. J. and Eysenck, S. B. G. (1964). *Manual of the Eysenck Personality Inventory*, University of London Press, London.
- Fahrenberg, J., Hampel, R. and Selg, H. (1984). *Das Freiburger Persönlichkeitsinventar*, The Freiburg Personality Inventory, Hogrefe, Göttingen.
- Goldberg, L. R. (1982). 'From Ace to Zimbie: Some explorations in the language of personality'. In: Spielberger, C. D. and Butcher, J. N. (Eds), *Advances in Personality Assessment*, Vol. 1, pp. 203-234, Erlbaum, Hillsdale, NJ.
- Jackson, D. N. (1967). *Personality Research Form Manual*, Research Psychologists Press, Goshen, NY.

- Jackson, D. N. and Helmes, E. (1979). 'Personality structure and the circumplex', *Journal of Personality and Social Psychology*, **37**: 2278–2285.
- Jackson, D. N. and Messick, S. (1958). 'Content and style in personality assessment', *Psychological Bulletin*, **4**: 243–252.
- Kuiper, N. A. and McDonald, M. R. (1982). 'Self and other perception in mild depressives', *Social Cognition*, **3**: 223–239.
- Lewinsohn, P. M., Mischel, W., Chaplin, W. and Barton, R. (1980). 'Social competence and depression: The role of illusionary self-perceptions?', *Journal of Abnormal Psychology*, **89**: 203–212.
- McCrae, R. R. and Costa, P. T. (1987). 'Validation of the five-factor model of personality across instruments and observers', *Journal of Personality and Social Psychology*, **52**: 81–90.
- Norman, W. T. (1963). 'Toward an adequate taxonomy of personality attributes. Replicated factor structure in peer nomination personality ratings', *Journal of Abnormal and Social Psychology*, **66**: 574–583.
- Norman, W. T. (1967). 'On estimating psychological relationships. Social desirability and self-report', *Psychological Bulletin*, **67**: 273–293.
- Norman, W. T. (1988). 'On separating substantive, stylistic, and evaluative components in personality measurements: A cross-national comparison', (Manuscript submitted for publication).
- Paulhus, D. L. (1986). 'Self-deception and impression management in test responses'. In: Angleitner, A. and Wiggins, J. S. (Eds), *Personality Assessment via Questionnaires: Current Issues in Theory and Measurement*, pp. 143–165, Springer, Berlin.
- Peabody, D. (1967). 'Trait inferences: Evaluative and descriptive aspects', *Journal of Personality and Social Psychology*, **7** (Whole No. 644).
- Thorndike, E. L. (1920). 'A constant error in psychological ratings', *Journal of Applied Psychology*, **4**: 25–29.
- Wiggins, J. S. (1979). 'A psychological taxonomy of trait-descriptive terms: The interpersonal domain', *Journal of Personality and Social Psychology*, **37**: 395–412.

RÉSUMÉ

Dans cette recherche, 300 sujets ont reçu 20 groupes de quatre termes descriptifs de traits de personnalité. Dans ces groupes, le contenu et l'évaluation étaient séparés par des adjectifs (par exemple: ferme, sévère, indulgent et laxiste). Chaque sujet était aussi évalué par trois pairs du même âge au moyen des mêmes groupes de quatre termes qualifiant des traits de personnalité. De plus, les sujets remplissaient différents questionnaires sur la personnalité et des échelles d'évaluation. Ils étaient également décrits par leurs pairs à l'aide de ces échelles d'évaluation. Les résultats montrèrent que les sujets s'attribuaient et attribuaient souvent aux autres deux traits souhaitables inconsistants au niveau descriptif (par exemple: ferme, indulgent). Une mesure des différences individuelles dans les réponses socialement souhaitables a été construite en totalisant toutes les réponses souhaitables. Les sujets qui se décrivaient d'une manière socialement souhaitable étaient moins névrotiques et plus consciencieux tant d'après leurs descriptions personnelles que d'après les évaluations des pairs. Les différentes implications de ces résultats sont discutées et la nouvelle mesure SD est comparée aux diverses échelles connues de SD.

ZUSAMMENFASSUNG

In der vorliegenden Studie bearbeiteten 300 Probanden 20 Quartette Eigenschaftsbeschreibender Begriffe, in denen Inhalts- und Erwünschtheitsaspekte getrennt werden können (z.B. *gefestigt, eingefahren, anpassungsfähig, und labil*). Jeder Proband wurde auch durch drei Bekannte bezüglich des gleichen Satzes von Adjektivquartetten beurteilt.

Weiterhin bearbeiteten die Probanden mehrere Persönlichkeitsinventare und Ratingskalen und wurden ebenfalls seitens der Bekannten auf diesen Ratingskalen eingestuft. Es zeigte sich, daß die Probanden sich selbst und ihren Bekannten häufig zwei erwünschte Traitbegriffe zuschreiben, welche deskriptiv inkonsistent sind. Ein Maß individueller Differenzen in der sozial erwünschten Antworttendenz wurde durch Aggregation aller sozial erwünschten Antworten gebildet. Probanden, welche sich in sozial erwünschter Weise darstellen, haben Selbst- und Peerberichten zufolge geringere Neurotizismus- und erhöhte Gewissenhaftigkeits-Scores. Verschiedene Implikationen der Ergebnisse werden diskutiert, und das hier vorgestellte Maß wird mit traditionellen Erwünschtheitskalen verglichen.