

In H.I. Braun, D.N.
Jackson, & D.E. Wiley
(Eds.), *The role of constructs
in psychological and
educational measurement*
(pp. 49-69). Mahwah NJ:
Erlbaum ©

2002

4

Socially Desirable Responding: The Evolution of a Construct

Delroy L. Paulhus
University of British Columbia

OVERVIEW

Socially desirable responding (SDR) is typically defined as the tendency to give positive self-descriptions. Its status as a response style rests on the clarification of an underlying psychological construct. A brief history of such attempts is provided. Despite the growing consensus that there are two dimensions of SDR, their interpretation has varied over the years from minimalist operationalizations to elaborate construct validation. I argue for the necessity of demonstrating departure-from-reality in the self-reports of high SDR scorers: This criterion is critical for distinguishing SDR from related constructs. An appropriate methodology that operationalizes SDR directly in terms of self-criterion discrepancy is described. My recent work on this topic has evolved into a two-tiered taxonomy that crosses degree of awareness (conscious vs. unconscious) with content (agentic vs. communal qualities). Sufficient research on SDR constructs has accumulated to propose a broad reconciliation and integration.

INTRODUCTION

I define *response biases* as any systematic tendency to answer questionnaire items on some basis that interferes with accurate self-reports. Examples are tendencies to choose the desirable response or the most moderate response, or to agree with statements independent of their content (for a review, see Paulhus, 1991). Following Jackson and Messick (1958), I distinguish *response styles*—biases that are consistent across time and questionnaires—from *response sets*—short-lived response biases attributable to some temporary distraction or motivation.

The topic of this essay is restricted to one response bias—*socially desirable responding* (SDR)—defined here as the tendency to give overly positive self-descriptions. Note that my qualification—*overly*—is seldom included in definitions of SDR, but is of central importance in this essay. Indeed, I will argue that no SDR measure should be used without sufficient evidence that high scores indicate a departure from reality.

This essay begins with a selective review of the wide variety of constructs held to underlie SDR scores. Coverage of the early developments is particularly selective because that history has already been reviewed elsewhere (Messick, 1991; Paulhus, 1986). The latter part of the essay emphasizes the recent developments with which I have been associated. Although my approach departs from theirs in some respects, my understanding of the topic of SDR draws liberally from the substantial empirical and theoretical contributions of the team of Sam Messick and Doug Jackson (e.g., Jackson & Messick, 1962; Messick & Jackson, 1961). And specific to this volume, my depiction of the interplay between response styles and personality can be traced to Messick's insightful analyses (Damarin & Messick, 1965; Messick, 1991).

A PLETHORA OF OPERATIONALIZATIONS

Assessment psychologists have agreed, for the most part, that the tendency to give socially desirable responses is a meaningful construct. In developing measures of SDR, however, they have used a diversity of operationalizations. A singular lack of empirical convergence was the unfortunate result. Commentators who were already wary of the very concept of SDR have exploited this disagreement to buttress their skepticism (e.g., Block, 1965; Kozma & Stones, 1988; Nevid, 1983). And the skeptics have a point in that the allegation that SDR contaminates personality measures is difficult to substantiate without a clarification of the SDR construct itself. This chapter aims to provide such a clarification. I argue that the attention given to SDR research cannot be dismissed as a red herring (Ones, Viswesvaran, & Reissa, 1996), but represents a process of construct validation that has now accumulated to the point where a coherent integration is possible. Accordingly, my review of the literature begins by laying out the three approaches that require integration.

1. Minimalist Constructs. A number of contributors have erred on the cautious side by using a straightforward operationalization of SDR with minimal theoretical elaboration. One standard approach entails (a) collecting social desirability ratings of a large variety of items, and (b) assembling an SDR

measure consisting of those items with the most extreme desirability ratings. (e.g., Edwards, 1953; Jackson & Messick, 1961; Saucier, 1994). The rationale is that individuals who claim the high-desirability items and disclaim the low-desirability items are likely to be responding on the basis of an item's desirability rather than its accuracy.

The validity of such SDR measures has been supported by demonstrations of consistency across diverse judges in the desirability ratings of those items (Edwards, 1970; Jackson & Messick, 1962¹). Moreover, scores on SDR scales developed from two different item domains (e.g., clinical problems, personality) were shown to be highly intercorrelated (Edwards, 1970). In short, the same set of respondents was claiming to possess a variety of desirable traits.

Exemplary of the minimalist approach was the psychometrically rigorous but theoretically austere creation of the SD scale by Allen Edwards (1957, 1970). Throughout his career, Edwards remained cautious in representing SD scores as "individual differences in rates of SD responding" (Edwards, 1990, p. 287).² At the same time, the prominence of his work derived undoubtedly from the implication that (a) high SD scores indicate misrepresentation and that (b) personality measures correlating highly with his SD-scale were contaminated to the point of futility (Edwards & Walker, 1961). Such inferences were easily drawn from his frequent warnings about the utmost necessity that personality measures be uncorrelated with SD (Edwards, 1970, p. 232; Edwards, 1957, p. 91).

An important alternative operationalization of SDR has been labeled *role-playing* (e.g., Cofer, Chance, & Judson, 1949; Wiggins, 1959). Here, one group of participants are asked to "fake-good", that is, respond to a wide array of items as if they were trying to appear socially desirable. The control group does a "straight-take": That is, they are asked simply to describe themselves as accurately as possible. The items that best discriminate the two groups' responses are selected for the SDR measure. This approach led to the construction of the MMPI Malingering scale and Wiggins's Sd scale, which is still proving useful after 30 years (see Baer, Wetter, & Berry, 1992).

Both of the above operationalizations seemed reasonable yet the popular measures ensuing from the two approaches (e.g., Edwards's SD-scale and Wiggins's Sd-scale) showed notoriously low intercorrelations (e.g., Holden &

1 Nonetheless, both authors noted elsewhere that multiple points of views SD must be recognized to understand the role of SD in personality (Jackson & Singer, 1967; Messick, 1960).

2 On the few occasions where he lost his equanimity, his opinion was clear: "Faking good on personality inventories, without special instructions to do so, I would consider equivalent to the tendency to give socially desirable responses in self-description" (Edwards, 1957, p.57).

Fekken, 1989; Jackson & Messick, 1962; Paulhus, 1984; Wiggins, 1959). Although both measures comprise items with high desirability ratings, a critical difference is that the endorsement rate of SD items (their communalities) were relatively high (e.g., "I am not afraid to handle money") whereas the endorsement rate for Sd items (e.g., "I never worry about my looks") was relatively low. To obtain a high score on the Sd scale, one must claim many rare but desirable traits. Thus the Sd scale (and similarly-derived measures) incorporated the notion of exaggerated positivity.

2. *Elaborate Constructs.* Some attempts to develop SDR measures have involved more theoretical investment at the operationalization stage and, in varying degrees, have provided a detailed construct elaboration. Here, item composition involved specific hypotheses regarding the underlying construct (e.g., Crowne & Marlowe, 1964; Eysenck & Eysenck, 1964; Hartshorne & May, 1930; Sackeim & Gur, 1978). The items were designed to trigger different responses in honest responders than in respondents motivated to appear socially desirable. In short, these measures incorporated the notion of exaggerated positivity.

In the earliest example, Hartshorne and May's (1930) monumental program of research on deceit included the development of a lie scale. The items asked about behaviors that "have rather widespread social approval but . . . are rarely done" (p. 98). High scores on the lie scale were assumed to mark a dishonest character. A more influential lie scale, the MMPI Lie, scale was written with a similar rationale to identify individuals deliberately dissembling their clinical symptoms (Hathaway & McKinley, 1951). Eysenck and Eysenck (1964) followed a similar rational procedure in developing the Lie scale of the Eysenck Personality Inventory.

The most comprehensive program of construct validity was that carried out by Crowne and Marlowe (1964) in developing their social desirability scale. Like the above, they assembled items claiming improbable virtues and denying common human frailties. In contrast to the purely empirical methods, high scores were accumulated by self-descriptions that were not just positive, but improbably positive.

Crowne and Marlowe (1964) fleshed out the character of high scorers by studying their behavioral correlates in great detail. The authors concluded that a need for approval was the motivational force behind both (a) high scores on the Marlowe-Crowne scale and (b) public-behavior that was both conforming and socially harmonious. Further resolution of this character was provided by Crowne (1979). Thus the construct had evolved appropriately in response to accumulating data.

3. *Accuracy Constructs.* Serious consideration must be given to the theorists who argue that those scoring high on SDR instruments should be taken at their word: That is, they actually do possess an abundance of desirable traits (Block, 1965; McCrae & Costa, 1983; Milholland, 1964). To support the accuracy position, these researchers showed that the self-reports on SDR instruments correlated with reports by knowledgeable observers. More recent analyses, however, have revealed that the evidence regarding the accuracy of the claims made by high SDR respondents is mixed, at best (Paulhus & John, 1998).

The most prominent example of the accuracy position is Block's (1965) book, the *Challenge of Response Sets*. His view was that high scores on Edwards's SD scale (as well as the first factor of the MMPI) represented a desirable personality syndrome called *ego-resiliency*. His evidence included the confirmation by knowledgeable observers (e.g., spouses) of many of the desirable qualities that were self-ascribed on the SD scale. No doubt there is some degree of accuracy in SD scores, but my recent analysis of Block's Ego Resiliency measure confirmed that it also includes a demonstrable degree of distortion (Paulhus, 1998a).

McCrae and Costa (1983) articulated a similar argument for the accuracy of self-descriptions on the Marlowe-Crowne (MC) scale. They showed that spouses sustained many of the claims by high scorers that they possessed a variety of desirable traits. In apparent contradiction, a series of studies by Millham and Jacobson (1978) showed that high-MCs would lie and cheat to impress experimenters with their character. These conflicting depictions can be reconciled within the construct of need for approval. High scorers on MC may realize that socially conventional behavior is usually the best way to gain approval yet believe that deceit works better in a number of situations where detection is very unlikely. In short, the data do not support the naive claim that high MCs (or high-SDs) are simply those with desirable character.³

In sum, the two most popular measures of SDR (SD and MC) appear to tap both reality and distortion. Confirmation of the distortion component makes it easier to understand why some respondents describe themselves in consistently positive terms across a variety of trait dimensions.

TWO-FACTOR MODELS OF SDR

The notion that SDR appears in two distinct forms was recognized by a number of early researchers (Cattell & Scheier, 1961; Edwards, Diers, &

³ See Paulhus and John (1998) for other reasons why the claims of high MC scores cannot be taken at face value.

Walker, 1962; Jackson & Messick, 1962; Messick, 1962; Wiggins, 1959). Factor analyses revealed two relatively independent clusters of measures non-committally labeled *Alpha* and *Gamma* by Wiggins (1964). The Alpha factor was clearly marked by Edwards SD-scale and the Gamma factor by Wiggins's Sd scale. Subsequent research positioned other measures on the first factor including the MMPI K-scale, Byrne's (1961) Repression-Sensitization scale and Sackeim and Gur's (1978) Self-Deception Questionnaire. Measures falling on the second factor included Eysenck's Lie scale and Sackeim and Gur's Other-Deception Questionnaire. A third set of measures loading largely, but not exclusively, on the second factor included the Marlowe-Crowne scale, the Good Impression scale (Gough, 1957), and the MMPI Lie scale (Hathaway & McKinley, 1951). With a growing consensus regarding two empirical factors, the conceptual task was now doubly challenging—what are the psychological constructs underlying these two SDR factors?

Damarin and Messick

It was not until the review by Damarin and Messick (1965)⁴ that a detailed theoretical interpretation of the two factors was offered (see Figure 4.1a). Factor 1 was said to involve the defensive distortion of one's private self-image to be consistent with a global evaluative bias. As a substantive label for this factor, they proposed *autistic bias in self-regard*. Associated personality traits included self-esteem and ego-resiliency. Factor 2 was labeled *propagandistic bias* to indicate a naive tendency to promote a desirable public reputation. Here, the underlying motivation was linked to factors varying from social approval to habitual lying. For the first time, a detailed characterological analysis had been provided for both factors.

Sackeim and Gur

Perhaps the most clear-cut example of the rational approach to SDR scale development was the work by Sackeim and Gur (1978; Gur & Sackeim, 1979; Sackeim, 1983). They applied to the process of questionnaire responding a distinction between the constructs of *self-deception* and *other-deception*. Some respondents report unrealistically positive self-depictions about which they appear to be convinced; other respondents consciously and deliberately distort their self-descriptions to fool an audience (See Figure 4.1b).

⁴ This was a technical report with limited circulation but much of the material was reviewed in the subsequent chapter by Messick (1991).

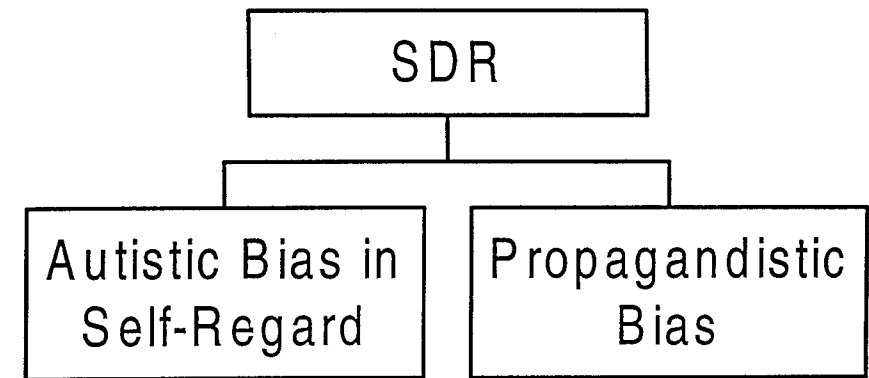


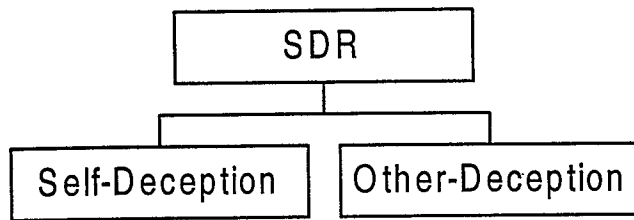
FIG. 4.1a. Two Constructs proposed by Damarin-Messick.

To compose a set of items indicating self-deception, the authors drew on the psycho-dynamic notion that sexual and aggressive thoughts are universally experienced yet often denied. If respondents overreact to questions with offensive content (e.g., "Have you ever thought about killing someone?"), then they are assumed to have self-deceptive tendencies. To measure other-deception, the authors wrote items describing desirable behaviors that are so public and blatant that they are not subject to self-deception (e.g., "I always pick up my litter"). According to the authors' reasoning, then, excessive claims of such commendable behaviors must involve conscious dissimulation.

The result of Sackeim and Gur's rational item composition was the "dynamic" duo of measures labeled the Self-Deception Questionnaire and the Other-Deception Questionnaire. Use of the word "deception" in both labels made it clear that exaggeration was an integral part of both conceptions. But to ensure that this exaggeration tendency was captured by the items, they recommended a scoring procedure that gave credit only for exaggeratedly positive item responses: Specifically, only responses of '6' or '7' on a 7-point scale were counted.

Early Paulhus

My early work was essentially an attempt to link and integrate the provocative concepts and instruments developed by Sackeim and Gur with the integrative structure provided by Damarin and Messick (see Paulhus, 1984; 1986). My



Unconscious

Conscious

FIG. 4.1b. Sackeim and Gur's two deception constructs.

and Gur scales was revealing: Those two scales clearly marked the two factors suggesting a theoretical interpretation of the factors that was consistent with, but more theoretically trenchant than, the labels provided by Damarin and Messick. I settled on the labels, *self-deception* and *impression management* (see Figure 4.1c). The term, *other-deception*, was replaced because its implication of deliberate lying seemed presumptuous. Instead I argued, following Damarin and Messick, that habitual presentation of a specific positive public impression could be construed as an aspect of personality, rather than a deception (see also Hogan, 1983). Hence, the term, *impression management*, was judged to be more apt.

I also devoted some effort to evaluating the psychometric properties of Sackeim and Gur's Self- and Other-Deception Questionnaires, with some dismaying conclusions. To begin with, all the items on the former measure were negatively keyed and all items on the latter measure, positively keyed. Because the measures were thus confounded in opposite directions with acquiescence, their observed intercorrelation of .30 was likely to have

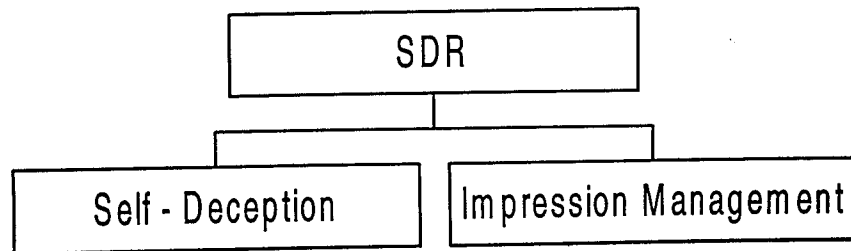


FIG. 4.1c. Two constructs proposed by Paulhus (1984).

underestimated the true value. As feared, when reversals were added to each scale, the intercorrelation exceeded .50. Although the balanced versions of these measures still loaded on their original factors, the high intercorrelation negated their advantage over single-factor measures. Moreover, some of the items on the Self-Deception Questionnaire were blatantly confounded with adjustment. To say the least, this state of affairs was discouraging for the two-factor conception.

A New Look at Socially Desirable Responding

Instead of conceding to the one-factor conception, my research group embarked on a new phase of item-writing. An extensive range of items were rationally composed to tap every conceivable form of self-deception and impression management (Paulhus, Reid, & Murphy, 1987). A swarm of factor analyses consistently revealed one factor of impression management and two factors of self-deception. The impression management items that cohered were largely the same items from earlier versions of the measure going back to Sackeim and Gur (1978). The two clusters of self-deception items appeared to involve *enhancement* (promoting positive qualities) and *denial* (disavowing negative qualities; Paulhus & Reid, 1991). Figure 4.2 shows the resulting subscales labeled Impression Management (IM), Self-Deceptive Enhancement (SDE), and Self-Deceptive Denial (SDD): They were incorporated into Version 6 of the *Balanced Inventory of Desirable Responding (BIDR)*, which I began distributing in 1988. Table 4.1 provides examples of the three types of items.

Construct Validity of the BIDR. The SDE and IM scales, in particular, form a useful combination of response style measures because they are relatively uncorrelated but capture the two major SDR dimensions (Paulhus, 1988, 1991). Their utility was demonstrated recently in a study of self-presentation during a job application situation (Paulhus, Bruce, & Trapnell, 1995). The IM scale, but not SDE, was extremely sensitive to faking instructions requesting various degrees of self-presentation. The sensitivity of the IM scale also far exceeded that of any of the NEO-FFI measures of the Big Five personality traits (Costa & McCrae, 1989). A similar pattern was observed in a study of job applicants vs. incumbents (Rosse et al., 1998)

In other studies, the SDE scale, but not the IM, predicted various kinds of self-deceptive distortions, for example, hindsight bias (Hoorens, 1995; Paulhus, 1988). More than 40 other studies, most outside of our laboratory, have added to the construct validity. For a more extensive review, see Paulhus (1998b).

TABLE 4.1
Sample items from the Balanced Inventory of
Desirable Responding
Version 6

Subscale	Items
Self-Deceptive Enhancement	My first impressions always turn out to be right.
Self-Deceptive Denial	I have never thought about killing someone.
Impression Management	I always pick up my litter on the street.

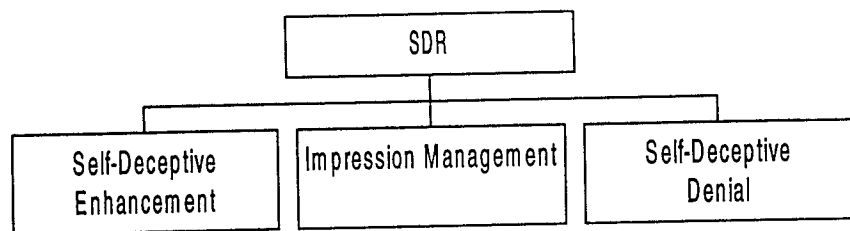


FIG. 4.2. Refined constructs proposed by Paulhus (1988).

Personality and Adjustment Correlates of the BIDR. One argument against interpreting SDR factors as personality constructs is that they rarely appear as independent factors in comprehensive factor analyses of personality. One possible explanation is that the BIDR response styles are simply disguised measures of normal personality. To clarify the interrelationships, my colleagues and I administered both kinds of measures to the same student samples under relatively anonymous conditions (Meston, Heiman, Trapnell, & Paulhus, 1998). Table 4.2 provides the correlations of the three response styles with Costa and McCrae's (1989) measures of the Big Five personality traits.

Although the response styles do not line up directly with any of the Big Five, the two self-deception subscales, SDE and SDD, do seem to pervade all five personality factors. Given the anonymous conditions of administration, the

TABLE 4.2
Correlations of BIDR Subscales with the Big Five personality factors.

	Self-Deception		Impression Management
	Enhancement	Denial	
Extraversion	.40	.16	.10
Openness	.48	.10	.11
Stability	.29	.32	.13
Conscientiousness	.19	.41	.29
Agreeableness	-.10	.50	.25

overlap suggests that self-deceptive bias plays a role in all personality dimensions. Correlations with the Impression Management scale are weaker but the fact that they are non-zero is noted here and discussed later.

The adjustment correlates of these response style measures have also been examined. In general, SDE, but not IM, is positively related to self-perceptions of mental health (e.g., Bonanno et al., in press; Brown, 1998; Nichols & Greene, 1997; Paulhus & Reid, 1991; Paulhus, 1998b). High SDE can also have a positive impact on performance in certain circumstances (Johnson, 1995). In a recent study of interpersonal adjustment, however, high SDE scorers were perceived negatively after 7 weeks of interaction. Moreover, high-SDE but not high-IM or high-SDD participants exhibited a discordance with reality as indicated by an inflation in self-ratings relative to ratings by fellow group members (Paulhus, 1998a).

That research bears directly on the debate about whether positive illusions are adaptive (Taylor & Brown, 1988; Yik, Bond, & Paulhus, 1998) or maladaptive (Colvin, Block, & Funder, 1995; John & Robins, 1994). The SDE scale (along with measures of narcissism) represents a trait operationalization of positive illusions, that is, trait self-enhancement. The two studies by Paulhus (1998a) indicated, in short, that trait self-enhancement was adaptive in promoting high self-esteem and positive first impressions, but had negative interpersonal consequences (see also, Bonanno et al., in press).

THE STRUCTURE OF SELF-FAVORING BIAS

Disentangling these sources of variance will be no easy task, but a splendid beginning might be made by providing a right-answer key for each subject's answer to each item. (Damarin & Messick, 1965, p. 63).

The implication of this understated comment was that researchers making allegations about response bias must do the work of demonstrating departures from reality: This task requires the collection of credible measures of personality to be partialled from self-reports. Damarin and Messick (1965) went on to lay out the statistical partitioning necessary to isolate the residual bias component (p. 21).

This recommendation proved invaluable in my work with Oliver John on determining the structure of self-favoring bias (John & Paulhus, 2000; Paulhus & John, 1998). We needed a unit of bias to represent each part of the personality space.⁵ For each personality variable, we collected self-ratings to compare with a more objective criterion, namely, ratings by knowledgeable peers (i.e., friends, family). In the case of intelligence, we also used IQ scores as a criterion. Each self-rating was regressed on its corresponding criterion to create a residual score representing the departure of the self-rating from reality. Factor analysis of a comprehensive set of such residuals should uncover the structure of self-favoring bias.

Using the Big Five dimensions of personality plus intelligence to represent personality space, our factor analyses of residuals revealed a smaller space than the 5-space of either self- or peer-ratings. The first two major dimensions appeared as in Figure 4.3. Factor 1 was marked by the Extraversion and Openness residuals whereas Factor 2 was marked by the Agreeableness and Conscientiousness residuals. Apparently, the structure of bias bears little resemblance to the standard Big Five structure. If anything, these factors look more like agency and communion (see Bakan, 1966; Wiggins, 1991).⁶

A replication study helped to clarify the meaning of the bias factors through the addition of a wide variety of self-report measures. These included traditional measures of SDR (Marlowe-Crowne scale) as well as related measures of self-enhancement (e.g., Narcissistic Personality Inventory). The additions allowed us to project a variety of bias and personality measures onto the two bias factors. The resulting projections (correlations with the factors) are depicted in Figure 4.4.

⁵ The last 15 years of work on the Five Factor Model suggests that it captures the 5 most important dimensions of personality (Wiggins, 1996). There is some dispute, however, about which rotation is optimal.

⁶ The results were more clear when we separated Conscientiousness into Dutifulness and Ambition following Paunonen and Jackson (1996). It is the Dutifulness measure that is most faithful, conceptually and empirically, to the Gamma factor.

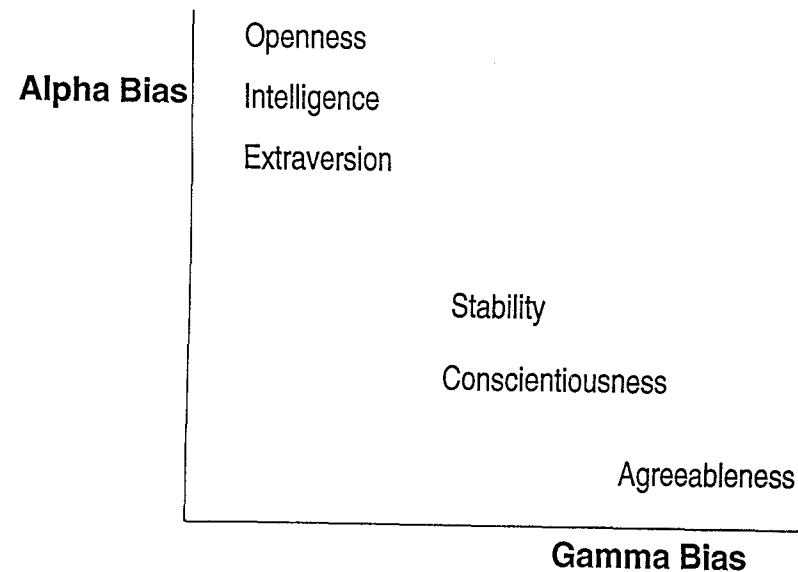


FIG. 4.3. Structure of Big Five Residuals.

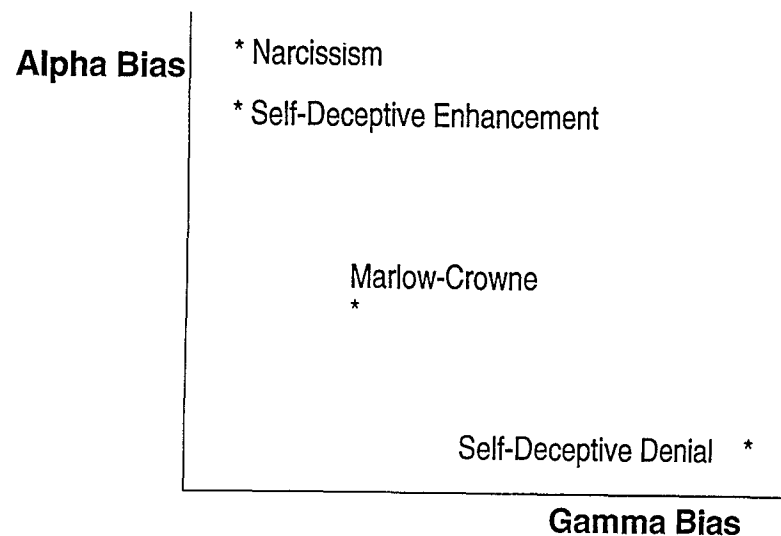


FIG. 4.4. Response style correlates of Alpha and Gamma.

Note first the striking match of the two BIDR subscales (SDE and SDD) to the two bias dimensions. Immediately, we have reason to believe that the factors represent Alpha and Gamma, the bias factors named by Wiggins (1964) and explicated by Damarin and Messick (1965). Note further that narcissism, as measured by the Narcissistic Personality Inventory (Raskin & Hall, 1981), marks Factor 1 along with SDE⁷. Factor 2 resembles earlier studies in being well-marked by the IM and SDD scales and less well by Eysenck's Lie scale, the MMPI Lie scale and the Marlowe-Crowne scale.

Remarkably, the venerable Alpha and Gamma SDR factors (noted above) have been re-generated via a novel technique requiring only personality content measures. The convergence of results across the two techniques adds substantial credibility to the Alpha and Gamma factors. In particular, the new technique provides evidence that both Alpha and Gamma assess departure-from-reality. That is, high scores on both SDR factors involve overly positive self-descriptions (Q.E.D.).

One remaining puzzle is the fact that two conceptually different response style measures, Self-Deceptive Denial and Impression Management, fall on the same SDR factor, Gamma. How can the response styles previously held to tap conscious and unconscious distortions (Paulhus, 1986; Sackeim & Gur, 1978) now coalesce at this point? In anonymous student samples, where pressure for self-presentation is minimal, SDD and IM appear to be capturing similar personality content. Yet IM is more responsive to instructional manipulations. In short, Gamma subsumes both conscious and unconscious aspects of common content. Apparently, I have to question my previous contention that level of consciousness is the core difference between Alpha and Gamma factors of SDR (e.g., Paulhus, 1986). This theoretical revision makes it easier to explain why the Gamma loading of an allegedly conscious deception measure (IM) does not disappear entirely in anonymous responses. With Gamma as a content factor, it is now quite understandable that IM should appear even when there is no audience to motivate impression management.

CONTENT VERSUS STYLE REDUX

In a more recent set of experiments, we sought to clarify the Alpha and Gamma factors via a series of studies varying self-presentation instructions (Paulhus & Notareschi, 1993). First we wondered why Gamma measures were

7 A number of recent reports on narcissism and self-deceptive enhancement suggests substantial overlap in both the constructs and the primary measures, NPI and SDE (McHoskey, Wortzel, & Szyarto, 1998; Paulhus, 1998a; Raskin, Novacek, & Hogan, 1991).

so sensitive to instructional manipulations (e.g., Paulhus, 1984). When given standard instructions to respond in a socially desirable fashion, respondents reported that they interpreted the instruction to mean that they should respond like a "nice person" or "good citizen." It struck us that this interpretation of social desirability was rather narrow, focusing on content related to agreeableness and dutifulness, i.e., communal traits. Accordingly we tried a more agentic form of instruction to respondents: "Respond to the questionnaire in a way to impress an experimenter with how strong and competent you are." Lo and behold, the SDE was more sensitive than the IM scale⁸ to these instructions (Paulhus, Tanchuk, & Wehr, 1999). In retrospect, these findings seem embarrassingly obvious; yet they have dramatic implications for previous research on SDR.

First, it is now apparent why the items on Wiggins's Sd and other Gamma factor scales contained those socially desirable but distinctively conventional items.⁹ Recall that these measures were developed using role-playing instructions that emphasized communion-related desirability. Second, it now seems obvious why Gamma-related scales were so responsive to instructions. They contain the very content that is implied by the instructions. Third, Alpha-related measures may be no more unconscious (and therefore self-deceptive) than Gamma measures.

Then what, after all, can we make of these two factors? Both appear under anonymous conditions. Both respond to faking instructions. Both have conscious and unconscious aspects to them. At least we don't have to withdraw the (thankfully noncommittal) labels, *Alpha* and *Gamma*.

The "final" two-tier conception suggests that (a) Alpha and Gamma be distinguished in terms of personality content, and that (b) each involves a self-deceptive style and an impression management style (see Figure 4.5). Alpha and Gamma are held to be two constellations of traits and biases that have their origins in two fundamental values, agency and communion (Paulhus & John, 1998). Excessive adherence to these values results in self-deceptive tendencies, which we label *egoistic bias* and *moralistic bias*.

Associated with Alpha is an *egoistic bias*, a self-deceptive tendency to exaggerate one's social and intellectual status. This tendency leads to unrealistically positive self-perceptions on such agentic traits as dominance, fearlessness, emotional stability, intellect, and creativity. Self-perceptions of

8 When subjects were notified that their answers could land them a summer job (emphasizing competence), then both scales showed significant increases (Paulhus, Lysy, & Yik, 1994).

9 The items were also low in communality (Wiggins, 1964).

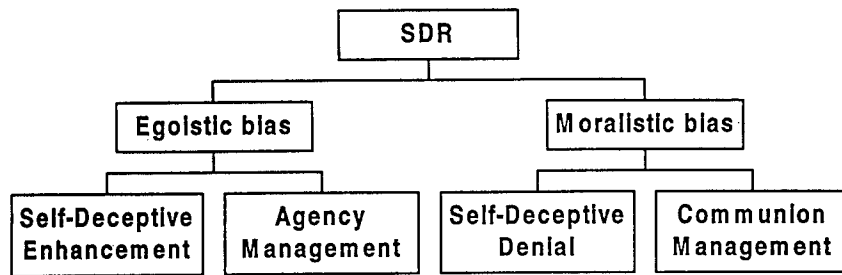


FIG. 4.5. Proposed two tier system.

high scorers have a narcissistic, “superhero” quality. Associated with Gamma is the *moralistic bias*, a self-deceptive tendency to deny socially-deviant impulses and claim sanctimonious, “saint-like” attributes. This tendency is played out in overly positive self-perceptions on such traits as agreeableness, dutifulness, and restraint.

At the impression management level, people are often motivated to deliberately exaggerate their attainment of agency and communion values. Thus the same two clusters of traits are involved but the exaggeration is more conscious. At this level, Alpha involves *Agency Management*, that is, asset-promotion or bragging. Such deliberate promotion of competence, fearlessness, physical prowess, etc. are most commonly seen in job applicants or in males attempting to impress a dating partner. Deliberate exaggeration of Gamma is termed *Communion Management* and involves excuse-making and damage control of various sorts. Such deliberate minimization of faults might also be seen in religious settings, or in employees who are trying to hold on to the status quo, or legal defendants trying to avoid punishment.

To fully assess the two-tiered system of SDR constructs, four types of measures are needed. Fortunately, three out of the four have been available for some time. Self-deceptive enhancement can be measured with its namesake (SDE) or the Narcissistic Personality Inventory (see Paulhus, 1998a). Self-deceptive denial can also be measured by its namesake scale (SDD). Communion management may be assessed using the traditional Impression Management scale, which has varied little since Sackeim and Gur (1978). Tentatively, it is renamed *Communion Management*.

The fourth type of desirable responding construct, agency management, required the development of a new instrument by the same name (AM). It consists of items related to agency content but with low endorsement rates in straight-take administrations. The low communalities permit room for manipulators to deliberately enhance impressions of their agency. Examples are

“I am very brave” and “I am exceptionally talented”. Such items tend not to be claimed, even by narcissists, under anonymous conditions. But the endorsement rate is higher under agency-motivated conditions than under anonymous conditions.

In recent studies, we have found that the impression management scales, AM and CM, are more useful as response sets than response styles. They were designed for the purpose of capturing instructional sets to appear agentic or communal and they perform that task very well. These measures do not perform very well as styles, presumably because impression management has so many sources and is so sensitive to situational demands.¹⁰ To the extent that a response bias is self-deceptive, the motivation for bias is more trait-like and therefore consistent with the definition of response style.

CONCLUSIONS

The fact that theories evolve is not a deficiency of science. Indeed, its responsiveness to new data can be seen as science’s greatest asset. In this light, the evolution of constructs underlying SDR should be viewed as inevitable rather than distressing. At the same time, science should exhibit net progress rather than veer haphazardly. The ideas about SDR presented here are the result of such progress: They were founded on and developed from earlier work. In particular, the earlier writings by Messick (1991; Damarin & Messick, 1965) were a necessary precursor for many of the ideas presented here.

For example, Messick’s writings emphasized the necessity of demonstrating departure-from-reality in assessing SDR. To this end, he suggested the statistical analysis of partial correlations. That notion and that method proved to be central to the development of our residuals method of determining the structure of bias in self-reports (Paulhus & John, 1998).

Yet those earlier ideas could not account for all the newly-collected data. In particular, the new data required a more elaborate structural model of SDR. This final two-tiered system incorporates a content-level (agency vs. communion) as well as a process level (conscious vs. unconscious). All four types of SDR were shown to involve the departure from reality that distinguishes response biases from content dimensions of personality. And they reaffirm the continuing challenge of response biases to valid assessment.

¹⁰ Promising new methods for measuring individual differences in impression management include the overclaiming technique (Paulhus & Bruce, 1990) and response latencies (Holden & Fekken, 1993).

REFERENCES

- Baer, R. A., Wetter, M. W., & Berry, D. T. (1992). Detection of underreporting of psychopathology on the MMPI: A meta-analysis. *Clinical Psychology Review, 12*, 509-525.
- Bakan, D. (1966). *The duality of human existence: Isolation and communion in Western man*. Boston: Beacon.
- Block, J. (1965). *The challenge of response sets*. New York: Century.
- Bonanno, G. A., Field, N. P., Kovacevic, A., & Kaltman, S. (in press). Self-enhancement as a buffer against extreme adversity. *Journal of Personality and Social Psychology*.
- Brown, J. D. (1998). *The self*. Boston: McGraw-Hill.
- Byrne, D. (1961). The repression-sensitization scale: Rationale, reliability, and validity. *Journal of Personality, 29*, 334-349.
- Cattell, R. B., & Scheier, I. H. (1961). Extension of meaning of objective test personality factors: Especially into anxiety, neuroticism, questionnaire, and physical factors. *Journal of General Psychology, 61*, 287-315.
- Cofer, C. N., Chance, J., & Judson, A. J. (1949). A study of malingering on the Minnesota Multiphasic Personality Inventory. *Journal of Psychology, 27*, 491-199.
- Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly-positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology, 68*, 1152-1162.
- Costa, P. T., & McCrae, R. R. (1989). *Manual for the NEO Personality Inventory: Five Factor Inventory/NEO-FFI*. Odessa, FL: PAR.
- Crowne, D. P. (1979). *The experimental study of personality*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York: Wiley.
- Damarin, F., & Messick, S. (1965). *Response styles as personality variables: A theoretical integration* (ETS RB 65-10). Princeton, NJ: Educational Testing Service.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- Edwards, A. L. (1970). *The measurement of personality traits by scales and inventories*. New York: Holt-Rinehart-Winston.
- Edwards, A. L. (1990). Construct validity and social desirability. *American Psychologist, 45*, p. 287-289.
- Edwards, A. L., Diers, C. J., & Walker, J. N. (1962). Response sets and factor loadings on sixty-one personality scales. *Journal of Applied Psychology, 46*, 220-225.
- Edwards, A. L., & Walker, J. N. (1961). A short form of the MMPI: The SD scale. *Psychological Reports, 8*, 485-486.
- Eysenck, H. J., & Eysenck, S. B. G. (1964). *The manual of the Eysenck Personality inventory*. London: U of London Press.
- Gough, H. C. (1957). *Manual for the California Psychological Inventory*. Palo Alto: Consulting Psychologists Press.
- Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology, 37*, 147-169.
- Hartshorne, H., & May, H. A. (1930). *Studies in the nature of character*. New York: MacMillan.
- Hathaway, S. R., & McKinley, J. C. (1951). *MMPI manual*. New York: Psychological Corporation.
- Hogan, R. (1983). A socioanalytic theory of personality. In M. M. Page (Ed.), *Nebraska Symposium on Motivation* (pp. 336-55), Lincoln, NE: University of Nebraska Press.
- Hogan, R., & Nicholson, R. A. (1988). The meaning of personality test scores. *American Psychologist, 43*, 621-626.
- Holden, R. R., & Fekken, G. C. (1989). Three common social desirability scales: Friends, acquaintances, or strangers? *Journal of Research in Personality, 23*, 180-191.
- Holden, R. R., & Fekken, G. C. (1993). Can personality test item responses have construct validity? Issue of reliability and convergent and discriminant validity. *Personality and Individual Differences, 15*, 243-248.
- Hoorens, V. (1995). Self-favoring biases, self-presentation, and the self-other asymmetry in social comparison. *Journal of Personality, 63*, 793-818.
- Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin, 55*, 243-252.
- Jackson, D. N., & Messick, S. (1961). Acquiescence and desirability as response determinants on the MMPI. *Educational and Psychological Measurement, 21*, 771-792.
- Jackson, D. N., & Messick, S. (1962). Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology, 65*, 285-299.
- Jackson, D. N., & Singer, J. E. (1967). Judgements, items, and personality. *Journal of Experimental Research in Personality, 2*, 70-79.
- John, O. P., & Paulhus, D. L. (2000). *The structure of self enhancement*. Unpublished ms., University of California, Berkeley.
- John, O. P., & Robins, R. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*, 206-219.
- Johnson, E. A. (1995). Self-deceptive coping: Adaptive only in ambiguous contexts. *Journal of Personality, 63*, 759-792.
- Kozma, A., & Stones, M. J. (1988). Social desirability and measures of subjective well-being: Age comparisons. *Social Indicators Research, 20*, 1-14.
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882-888.
- McHoskey, J. W., Worzel, W., & Szyarto, C. (1998). Machiavellianism and psychopathy. *Journal of Personality and Social Psychology, 74*, 192-210.
- Messick, S. (1960). Dimensions of social desirability. *Journal of Consulting Psychology, 24*, 279-287.
- Messick, S. (1962). Response style and content measures from personality inventories. *Educational and Psychological Measurement, 22*, 41-56.
- Messick, S. (1991). Psychology and methodology of response styles. In R. E. Snow & D. E. Wiley (Eds.), *Improving inquiry in social science*. (pp. 161-200). Hillsdale, NJ: Erlbaum.
- Messick, S., & Jackson, D. N. (1961). Desirability scale values and dispersions for MMPI items. *Psychological Reports, 8*, 409-414.
- Meston, C. M., Heiman, J. R., Trapnell, P. D., & Paulhus, D. L. (1998). Socially desirable responding and sexuality self-reports. *Journal of Sex Research, 35*, 148-157.
- Milholland, J. E. (1964). Theory and techniques of assessment. *Annual Review of Psychology, 15*, 311-346.
- Millham, J., & Jacobson, L. I. (1978). The need for approval. In H. London & J. E. Exner (Eds.), *Dimensions of Personality* (pp. 365-390). New York: Wiley.
- Nevid, J. S. (1983). Hopelessness, social desirability, and construct validity. *Journal of Consulting and Clinical Psychology, 51*, 139-140.
- Nichols, D. S., & Greene, R. L. (1997). Dimensions of deception in personality assessment: The example of the MMPI-2. *Journal of Personality Assessment, 68*, 251-266.

- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660–679.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598–609.
- Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaire* (pp. 143–165). New York: Springer-Verlag.
- Paulhus, D. L. (1988). *Manual for the Balanced Inventory of Desirable Responding* (BIDR-6). Unpublished manual, University of British Columbia.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). New York: Academic Press.
- Paulhus, D. L. (1998a). Intrapsychic and interpersonal adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology, 74*, 812–820.
- Paulhus, D. L. (1998b). *Manual for the Balanced Inventory of Desirable Responding* (BIDR-7). Toronto/Buffalo: Multi-Health Systems.
- Paulhus, D. L., & Bruce, M. N. (1990, June). *The Over-Claiming Questionnaire (OCQ)*. Presented at the meeting of the Canadian Psychological Association, Ottawa, Canada.
- Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and structure. *Personality and Social Psychology Bulletin, 21*, 100–108.
- Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic bias in self-perceptions: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*, 1024–1060.
- Paulhus, D. L., Lysy, D., & Yik, M. (1994). *Self-presentation on a real-world job application*. Unpublished manuscript, University of British Columbia.
- Paulhus, D. L., & Notareschi, R. F. (1993). *Varieties of faking manipulations*. Unpublished data, University of British Columbia.
- Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*, 307–317.
- Paulhus, D. L., Reid, D. B., & Murphy, G. (1987). *The Omnibus Study of Desirable responding*. Unpublished data, University of British Columbia.
- Paulhus, D. L., Tanchuk, T., & Wehr, P. (1999, August). *Value-based faking on personality questionnaires: Agency and communion rule*. Presented at the meeting of the American Psychological Association, Boston.
- Paunonen, S., & Jackson, D. N. (1996) The Jackson Personality Inventory and the five-factor model of personality. *Journal of Research in Personality, 30*, 42–59.
- Raskin, R. N., & Hall, C. S. (1981). The narcissistic personality inventory: Alternative form reliability and further evidence of construct validity. *Journal of Personality Assessment, 45*, 159–160.
- Raskin, R. N., Novacek, J., & Hogan, R. T. (1991). Narcissism, self-esteem and defensive self-enhancement. *Journal of Personality, 59*, 19–38.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on pre-employment testing and hiring decisions. *Journal of Applied Psychology, 83*, 634–644.
- Sackeim, H. A. (1983). Self-deception, self-esteem, and depression: The adaptive value of lying to oneself. In J. Masling (Ed.), *Empirical studies of psychoanalytic theories* (pp. 101–157). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sackeim, H. A., & Gur, R. C. (1978). Self-deception, other-deception and consciousness. In G.E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research* (Vol. 2; pp. 139–197). New York: Plenum Press.
- Saucier, G. (1994). Separating description and evaluation in the structure of personality attributes. *Journal of Personality and Social Psychology, 66*, 141–154.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social-psychological perspective on mental health. *Psychological Bulletin, 103*, 193–210.
- Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23*, 419–427.
- Wiggins, J. S. (1964). Convergences among stylistic response measures from objective personality tests. *Educational and Psychological Measurement, 24*, 551–562.
- Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In W. Grove & D. Cicchetti (Eds.), *Thinking clearly about Psychology: Essays in honor of Paul Meehl* (Vol.2; pp. 89–113). Minneapolis: University of Minnesota Press.
- Wiggins, J. S. (Ed.). (1996). *The Five Factor Model*. New York: Guilford.
- Yik, M. S. M., Bond, M. H., & Paulhus, D. L. (1998). Do Chinese self-enhance or self-efface? It's a matter of domain. *Journal of Research in Personality, 24*, 399–406.