

# Issues in socially desirable responding and personality research

**Jan-Erik Lönnqvist**



Department of Psychology  
University of Helsinki, Finland

Academic dissertation to be publicly discussed,  
by due permission of the Faculty of Behavioural Sciences  
at the University of Helsinki in Auditorium XII, Unioninkatu 34,  
on the 28th of November, 2008, at 12 o'clock.

UNIVERSITY OF HELSINKI  
Department of Psychology  
Studies 52: 2008

- Supervisors     Markku Verkasalo, PhD  
                     Department of Psychology  
                     University of Helsinki  
                     Finland
- Professor Sampo Paunonen  
                     Department of Psychology  
                     University of Western Ontario  
                     Canada
- Reviewers        Professor Katariina Salmela-Aro  
                     Department of Psychology  
                     University of Jyväskylä  
                     Finland
- Professor Jüri Allik  
                     Department of Psychology  
                     University of Tartu  
                     Estonia
- Opponent         Professor Rainer Riemann  
                     Department of Psychology  
                     Bielefeld University  
                     Germany

ISSN 0781-8254

ISBN 978-952-10-4989-7 (pbk.)

ISBN 978-952-10-4990-3 (PDF)

<http://www.thesis.helsinki.fi>

Helsinki University Printing House  
Helsinki 2008

# CONTENTS

<b>ABSTRACT</b> .....	<b>5</b>
<b>TIIVISTELMÄ</b> .....	<b>6</b>
<b>LIST OF ORIGINAL PUBLICATIONS</b> .....	<b>7</b>
<b>ABBREVIATIONS</b> .....	<b>8</b>
<b>1 INTRODUCTION</b> .....	<b>9</b>
1.1. Conceptualizing and measuring SDR	10
1.1.1. Response set vs. response style vs. response substance	10
1.1.2. Communal bias vs. agentic bias	12
1.1.2.1. The Big Five	14
1.1.2.2. Personal values	15
1.2. SDR as a moderator or suppressor variable	15
1.3. Effects of SDR in research on personality stability	17
1.3.1. Inflated test-retest correlations?	17
1.3.2. The maturity-stability hypothesis	18
1.4. Effects of SDR on self-selection	19
<b>2 RESEARCH QUESTIONS AND HYPOTHESES</b> .....	<b>21</b>
<b>3 METHODS AND RESULTS</b> .....	<b>23</b>
3.1. Study I	23
3.1.1. Participants and procedure	23
3.1.2. Measures	24
3.1.2.1. Socially desirable responding	24
3.1.2.2. Personality	24
3.1.3. Results	25
3.1.3.1. Response set vs. response style vs. response substance	25
3.1.3.2. Communal bias vs. agentic bias	26
3.1.3.3. Moderator variables vs. suppressor variables	28
3.1.3.4. Effects of SDR on stability estimation	29
3.2. Study II	30
3.2.1. Participants and procedure	30
3.2.2. Measures	31
3.2.2.1. Socially desirable responding	31
3.2.2.2. Personal values	31

3.2.3. Results	32
3.2.3.1. Response set vs. response style vs. response substance	32
3.2.3.2. Communal bias vs. agentic bias	33
3.2.3.3. Moderator variables vs. suppressor variables	34
3.3. Study III	35
3.3.1. Participants and procedure	35
3.3.2. Measures	36
3.3.2.1. Socially desirable responding	36
3.3.2.2. Personality	36
3.3.2.3. Psychiatric evaluations	36
3.3.2.4. Competence indices	37
3.3.3. Results	37
3.4. Study IV	39
3.4.1. Participants and procedure	39
3.4.2. Measures	40
3.4.2.1. Socially desirable responding	40
3.4.2.2. Personality	40
3.4.3. Results	41
3.4.3.1. SDR and self-selection	41
3.4.3.2. Moderator variables vs. suppressor variables	42
<b>4 GENERAL DISCUSSION.....</b>	<b>43</b>
4.1. Communion management	43
4.2. Self-deceptive enhancement	45
4.3. SDR and personality stability	47
4.4. SDR and self-selection	48
4.5. Limitations	49
4.6. Conclusions	50
4.6.1. The BIDR scales	50
4.6.2. Controlling for SDR	51
4.6.3. The influence of SDR on personality research	52
4.6.4. Future directions	53
<b>5 ACKNOWLEDGEMENTS.....</b>	<b>55</b>
<b>6 REFERENCES.....</b>	<b>57</b>
<b>APPENDIX.....</b>	<b>65</b>

## **ABSTRACT**

The four studies presented in this dissertation were designed to examine the influence of socially desirable responding (SDR) on personality research outcomes. The assessment of personality relies heavily on the use of self-report questionnaires. Their validity could be threatened by people being dishonest in their self-descriptions and ascribing more desirable traits to themselves than would be warranted by their behaviour. Scales designed to detect SDR have been around for half a century, but their status continues to be debated. Paulhus' (1991) Balanced Inventory of Desirable Responding (BIDR) is perhaps the most prominent of the scales developed to distinguish between those individuals who have distorted their responses and those who have not. The first two studies included in this dissertation mostly deal with the properties of the BIDR. The other two studies are less focused on SDR scales and investigate, more generally, the potential effects of SDR on two phenomena that are of central interest to the general personality discourse – personality stability over time and volunteering as participants in psychological research.

The data of Studies I and II showed that Paulhus' BIDR scales, designed to be indicators of SDR, are not pure measures – both the communion management and self-deceptive enhancement scales are, at once, measures of response bias and measures of more substantive individual differences in behaviour. The data further suggested that the communion management and self-deceptive enhancement scales of the BIDR are somewhat accurate measures of communal and agentic bias, respectively. No evidence for a suppressor model of SDR, and only weak evidence for a moderator model, was found in those studies. Concerning research on personality stability, some data in Study I suggested that SDR may add reliable and common variance to a personality questionnaire administered at two different points in time, thus artificially inflating the test-retest correlation of that questionnaire. Furthermore, Study III demonstrated that the maturity-stability hypothesis may be in part, but not entirely, a product of SDR. Study IV suggested that some of the observed personality differences between research volunteers and nonvolunteers may be due to heightened SDR of volunteers. However, those personality differences were by no means exclusively attributable to differences in SDR.

In sum, the work presented in this thesis reveals some ambiguity regarding the effects of SDR on personality research, as is true of much of the previous research on SDR. Clear-cut conclusions are difficult to reach, as the data were neither fully consistent with the view that SDR can be ignored, nor with the view that SDR needs to be controlled in some way. The struggle to understand the influence of SDR on personality research continues.

## TIIVISTELMÄ

Tähän väitöskirjaan kuuluvat neljä tutkimusta suunniteltiin selvittämään miten sosiaalisesti suotuisa vastaamistyyli vaikuttaa persoonallisuuspsykologisen tutkimuksen tuloksiin. Persoonallisuutta mitatessa luotetaan useimmiten itsearviointimittareihin. Näiden mittareiden validiteettia voi uhata itseään epärehellisesti kuvaavat henkilöt, jotka väittävät omaavansa enemmän suotavia piirteitä kuin mihin heidän käyttäytymisensä oikeuttaisi. Sosiaalisesti suotuisan vastaamisen havaitsemiseen suunniteltuja mittareita on ollut saatavilla jo puoli vuosisataa, mutta niiden hyödyllisyydestä kiistellään yhä. Todennäköisesti paras vastauksiaan vääristelevien ja rehellisesti vastaavien erottamiseen kehitetty mittari on Paulhusin (1991) sosiaalisen suotavuuden mittari. Kaksi ensimmäistä tähän väitöskirjaan kuuluvaa tutkimusta käsittelevät tämän mittarin ominaisuuksia. Toiset kaksi tutkimusta keskittyvät niihin vaikutuksiin jota sosiaalisesti suotavalla vastaamisella voi olla kahteen persoonallisuuspsykologisessa tutkimuksessa keskeiseen ilmiöön – persoonallisuuden stabiliteettiin ajan yli ja halukkuuteen osallistua koehenkilönä psykologiseen tutkimukseen.

Tutkimusten I ja II tulokset osoittivat, että Paulhusin (1991) mittarin asteikot eivät ole puhtaita mittareita – sekä vaikutelmanluonti- että itsepetosasteikko mittaavat samanaikaisesti vastaustaipumusta ja todellisia yksilöiden välisiä eroja. Tulokset viittasivat myös siihen, että vaikutelmanluonti- ja itsepetosasteikot mittaavat yhteisöllistä ja agenttista vastaustaipumusta, tässä järjestyksessä. Tukea sosiaalisesti suotuisan vastaamisen tilastolliselle kontrolloinnille ei juuri löytynyt. Mitä tulee persoonallisuuden stabiliteetin tutkimukseen, Tutkimus I osoitti, että sosiaalisesti suotuisa vastaaminen voi lisätä luotettavaa ja yhteistä varianssia persoonallisuusmittariin joka annetaan kahtena eri ajankohtana, vääristäen näin arvioita persoonallisuuden stabiliteetista. Tämän lisäksi Tutkimus III osoitti, että kypsyyss-stabiliteetti hypoteesi voi olla osittain, mutta ei täysin, sosiaalisesti suotavan vastaamisen tuottama. Tutkimus IV osoitti, että mitatut persoonallisuuserot tutkimuksiin vapaaehtoisesti osallistuvien ja ei-osallistuvien välillä voivat olla osittain, mutta eivät täysin, sosiaalisesti suotavan vastaamisen tulosta.

Tässä väitöskirjassa esitellyn tutkimuksen tulokset ovat monitulkintaisia sen suhteen miten sosiaalisesti suotava vastaamistyyli vaikuttaa persoonallisuuspsykologian tutkimustuloksiin. Toisaalta tulokset eivät tue näkemystä, että sosiaalisesti suotavan vastaamisen voi jättää huomiotta, mutta eivät toisaalta myöskään sitä näkemystä että sosiaalisesti suotavaa vastaamista pitäisi jollain lailla kontrolloida. Ponnistelut sosiaalisesti suotavan vastaamisen ymmärtämiseksi jatkuvat.

## LIST OF ORIGINAL PUBLICATIONS

I Lönnqvist, J.-E., Paunonen, S. V., Tuulio-Henriksson, A., Lönnqvist, J., & Verkasalo, M. (2007). Substance and style in socially desirable responding. *Journal of Personality*, 75, 291-322.

II Lönnqvist, J.-E., Verkasalo, M., & Bezmenova, I. (2007). Agentic and communal bias in socially desirable responding. *European Journal of Personality*, 21, 853-868.

III Lönnqvist, J.-E., Mäkinen, S., Paunonen, S. V., Henriksson, M., & Verkasalo, M. (2008). Psychosocial functioning in young men predicts their personality stability over 15 years. *Journal of Research in Personality*, 42, 599-621.

IV Lönnqvist, J.-E., Paunonen, S. V., Verkasalo, M, Leikas, S., Tuulio-Henriksson, A., Lönnqvist, J. (2007). Personality characteristics of research volunteers. *European Journal of Personality*, 21, 1017-1030.

## ABBREVIATIONS

SDR	Socially Desirable Responding
BIDR	Balanced Inventory of Desirable Responding
IM	Impression Management
CM	Communion Management
SDE	Self-Deceptive Enhancement
$N$	Number of participants
$t$	Value on t-test
$r$	Pearson's $r$ (measure of effect size)
$d$	Cohens's $d$ (measure of effect size)
$\eta_p^2$	Partial Eta squared (measure of effect size)
$\Delta R^2$	Increase in variance explained
$\beta$	Regression coefficient



# 1 INTRODUCTION

One of the concerns of personality psychology is the assessment of individual differences in personality traits. The measurement of personality attributes most frequently relies on the use of self-report paper-and-pencil or computerized personality questionnaires. These questionnaires typically present respondents with a series of items and ask them to indicate whether or to what extent each item accurately describes their personality. The method of self-report thus assumes that respondents are willing and able to provide reasonably accurate responses. However, this might not always be the case.

It is clear that questionnaire responses may be susceptible to a number of different distortions. One potential threat to personality assessment validity is socially desirable responding (SDR). In one of the earliest works on SDR, Edwards (1953) noted that the more desirable a trait was judged, the more people ascribed that trait to themselves. He surmised that people in general either had more desirable than undesirable traits or they were being dishonest in their self-descriptions.

If response distortion is widespread and common, as it may be, for instance, in organizational or clinical settings, a personality questionnaire may fail the usual tests of psychometric accuracy. This may have severe consequences, such as not hiring the best job applicant. But even if the scale seems to work well in general, the scores of some respondents may be severely distorted, yielding inaccurate and misleading personality profiles for those people.

One of the more popular defenses against SDR response distortion is the use of scales designed to verify the creditability of a person's responses to the items of the questionnaire. SDR scales typically contain very desirable items, and a respondent's score on the scale is the number of such items he or she endorses. High scores on these scales are typically used either to identify suspicious protocols that may be discarded, or sometimes to adjust scores on the substantive personality scales to account for a desirability bias.

Although formal scales designed to detect SDR have been around for half a century or more (Edwards, 1957), their status continues to be debated. Some researchers have suggested that the empirical evidence does not justify the use of SDR scales to assess the validity of substantive personality scales (e.g., McCrae & Costa, 1983; Ones, Viswesvaran, & Reiss, 1996). However, in their 1996 *Annual Review of Psychology* article, Butcher and Rouse (1996) recommended that all personality questionnaires include validity scales that

can appraise the subject's willingness to share personal information and their tendency for response exaggeration. Consistent with this mandate, SDR scales continue to appear, for instance, as control variables (Finkel, Burnette, & Scissors, 2007), as indicators of scale validity (Starzyk, Holden, Fabrigar, & MacDonald, 2006), and even as primary variables of interest (e.g., Lalwani, Shavitt, & Johnson, 2006).

The main controversy over SDR is whether SDR is or is not a major problem in personality assessment that needs to be controlled. The present thesis contributes to this discussion with four empirical studies. These studies pertain to SDR as it affects the outcomes of empirical studies in the personality area. The first two studies mostly dealt with the properties of the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991), perhaps the most prominent of the questionnaires designed to measure SDR. These studies investigated whether the BIDR measures response bias, or some substantive individual differences in personality. Related to this is the question of whether SDR can be controlled for in some way. The third and fourth studies were less focused on SDR scales, and more generally investigate the potential effects of SDR on two phenomena that are of central interest to the general personality literature – personality stability across time and volunteering as a participant for psychological research.

## **1.1. Conceptualizing and measuring SDR**

Perhaps the most frequent criticism of SDR scales has been that some individuals actually possess more desirable traits than do others, and it is not possible to distinguish, among individuals with high SDR scale scores, those who have distorted their responses from those who have not (e.g., McCrae & Costa, 1983). Paulhus' (1991) BIDR, however, was developed to make such a distinction possible. The BIDR consists of items that describe very undesirable but common attributes or behaviours (e.g., "I have said something bad about a friend behind his or her back"), as well as items that describe very desirable but uncommon attributes or behaviours (e.g., "I never cover up my mistakes").

### **1.1.1. Response set vs. response style vs. response substance**

Following Jackson and Messick (1958), the BIDR was developed to distinguish between response sets and response styles. According to Paulhus (2002), response sets refer to

short-lived response biases attributable to the situation or to motivation in a particular context, whereas unconscious response styles are consistent across time and questionnaires. The BIDR was developed to measure these two types of response bias with, respectively, a scale called Impression Management (IM), and a scale called Self-Deceptive Enhancement (SDE). High scorers in IM are thought to consciously and deliberately describe themselves to others as highly desirable, whereas those scoring high in SDE are thought to report unrealistic but honestly held positive self-descriptions.

There are several types of evidence one could consider in determining whether an SDR measure represents primarily an index of response set, response style, or some substantive personality characteristic. When the demand for SDR increases, such as in an applicant context or when participants are explicitly instructed to fake good, scores on a response set measure should increase, whereas scores on a response style measure should be relatively stable. Consistent with the interpretation of the IM scale as a measure of response set, scores on that scale (but not the SDE scale) change markedly with changes in the demand to present a good impression (Paulhus, Bruce, & Trapnell, 1995). Another type of evidence one could consider is the correlations between SDR scales and other desirable measures. If an SDR scale measures response set, then the correlations between that scale and other desirable measures should be stronger in assessment conditions with a high demand for SDR, as more variance in scale scores should be attributable to SDR in those conditions.

Different evidence is needed to conclude that a measure is an index of response style. Because response styles are proposed to be general and consistent across conditions, unlike response sets, scores on measures of the former should be largely invariant across assessment contexts differing in SDR demand. Thus, the test-retest stability of such measures across, for instance, applicant and incumbent assessments should be relatively high. In line with this, SDE scale scores (but not IM scales scores) have been shown to be relatively invariant across assessment contexts differing in their demand for SDR (Paulhus, Bruce, & Trapnell, 1995). Like response set measures, response styles should also correlate with self-reports on other desirable and undesirable personality traits. Unlike response set measures, such correlations should be found regardless of the SDR demand extant at the time of the assessments.

Several previous SDR scales, such as Edwards' Social Desirability Scale (Edwards, 1957), the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960), and the Multidimensional Personality Questionnaire Unlikely Virtues and Desirable Response Inconsistency scales (Tellegen & Waller, 1992) have been suggested to be measures of

substantive traits rather than indicators of response bias (Borkenau & Ostendorf, 1992; McCrae & Costa, 1983; Piedmont, McCrae, Riemann, & Angleitner, 2000). How is it possible to determine if a putative measure of SDR is actually measuring real behaviours rather than simply a bias for presenting a favorable impression? For this, it is necessary to show that the measure of SDR is actually not measuring a departure from reality – i.e., a criterion of behaviour is needed that is independent of self-report and, thus, unlikely to be affected by respondent's SDR tendencies.

In arguing that the BIDR measures response bias and not real behaviour tendencies, Paulhus (2002) used ratings by knowledgeable others as an external criterion of reality, and showed that desirable responses to the BIDR items reflected a departure from reality in self-descriptions. McCrae and Costa (1983) came to the opposite conclusion. If responses to SDR items are indeed accurate, they argued, then SDR scores could be expected to correlate with others' reports of more general personality traits. This is what McCrae and Costa (1983) found in their study on the Marlowe-Crowne SDS, and which led them to conclude that the scale measures more substantive individual differences in personality than response bias.

Studies I and II of the present thesis were designed to evaluate whether, as Paulhus (2002) has claimed, the BIDR CM and SDE scales measure response set and response style, respectively, or whether those scales, like many other purported measure of SDR, should rather be interpreted as measures of substantive individual differences in personality (Borkenau & Ostendorf, 1992; McCrae & Costa, 1983; Piedmont, McCrae, Riemann, & Angleitner, 2000). Related to this issue, these studies also investigated whether the influence of SDR can somehow be controlled using the BIDR scales.

### **1.1.2. Communal bias vs. agentic bias**

The BIDR was originally designed to measure SDR regardless of item content (Paulhus, 1991). However, Paulhus (2002) has more recently classified response biases as communal or agentic, according to the content of the items endorsed, both of which exist on the conscious and unconscious levels. Communion and agency are terms developed by Bakan (1966) to reflect two general patterns of human existence. Communion refers to a focus on others and forming connections with them. This relates to the participation of the individual in a larger organism of which he or she is a part and includes co-operation, group participation, and attachment. Agency refers to a focus on self and differentiating oneself from others, which reflects one's existence as an individual and includes self-direction,

self-assertion, and self-expansion. Two orthogonal personality descriptive dimensions corresponding to agency and communion have consistently been implicated in the personality literature (for reviews, see Digman, 1997; Paulhus & John, 1998). However, the names given to these dimensions have varied. For instance, Hogan (1983) labelled these dimensions as ‘getting ahead’ and ‘getting along,’ Deci and Ryan (2000) referred to the need for competence and the need for relatedness, and DeYoung (2006) called these dimensions stability and plasticity.

On the basis of his two-stage classification (communal vs. agentic and conscious vs. unconscious), Paulhus (2002) has renamed his Impression Management (IM) scale the Communion Management (CM) scale. It supposedly measures conscious communal bias and is thought to indicate how strongly the person wants to appear a “saint” – i.e., as someone who can control his or her earthly impulses to a greater degree than can others. This type of self-deception pertains to overly positive self-evaluations regarding, for example, loyalty, dutifulness, honesty, and law-abidingness. Paulhus’ (2002) Self-Deceptive Enhancement Scale (SDE) fits into this two-stage classification as an index of unconscious agentic bias. Paulhus maintained that SDE measures the extent to which a person overestimates his or her competence and strives to give an impression of being a “superhero” – i.e., as someone who can accomplish brave and heroic feats that are unattainable by others. This type of self-deception pertains to overestimation of own skill level or abilities, and pertains to overly positive self-evaluations regarding, for example, power, achievements, individual talents, and intelligence.

The correlations between SDR scales, whether they are interpreted as response sets or response styles, and their relations with other individual differences measures are relevant in determining the type of response bias the SDR scales represent. In addressing this issue, the present research locates agentic and communal bias within the frameworks provided by the Big Five model of personality structure and by Schwartz’ (1992) values theory. More specifically, the present research examined to what extent the BIDR CM and SDE scales are measures of communal and agentic bias, respectively. For this end, those scales were linked to both the Big Five personality factors (Study I), and to personal values (Study II). In the following sections, the Big Five model of personality structure and values theory are briefly introduced.

### 1.1.2.1. The Big Five

Currently the most popular model among psychologists studying personality structure is the Big Five model of personality. The Big Five model of personality structure emerged within the psycholexical research tradition. As early as the 1930s, psychologists in search of a descriptive taxonomy of human dispositions turned to everyday language. The first step was to select from the lexicon all words that could be used to distinguish between people (Allport & Odbert, 1936/1970). This approach was inspired by the lexical hypothesis, according to which “Those personality traits that are most salient and socially relevant in people’s lives have become encoded into their language; the more important such a trait, the more likely is it to become expressed as a single word” (Goldberg, 1982, p. 204).

Allport and Odbert (1936/1970) found 18000 words that could be used to describe individual differences. However, their list was of little value as it did not organize individual differences in behaviour and experience. The first to attempt such an organization was Cattell (1943). Based on theoretical considerations and empirical work, Cattell ended up with 35 personality traits, which he by means of factor analysis further reduced to 12 personality factors. However, based on the intercorrelations between the 35 personality traits identified by Cattell, Tupes and Christal (1961/1992) established that five broad factors, but not much else, could be found in self-ratings, peer ratings, and clinician’s ratings of personality. The five factors were later labeled by Goldberg (1982, 1990) as the Big Five, and are now known as Neuroticism, Extraversion (or Surgency), Openness to Experience (or Intellect), Agreeableness, and Conscientiousness.

The popularity of the five factor approach grew when Costa and McCrae (1988, 1992; McCrae & Costa, 1990) showed that most, if not all, comprehensive personality questionnaires measured the same five factors as identified by the psycholexical approach. There is now widespread agreement among personality psychologists that those five factors form an appropriate taxonomy for the description of individual differences (for a review, see John; 1990; cf. Block, 1995; McAdams & Pals, 2006). Within this framework, communal bias is most clearly revealed in overestimation of one’s Agreeableness and Conscientiousness, whereas agentic bias is evident in overestimation of Extraversion and Openness to Experience (Paulhus, 2002). Exaggeration of own Emotional Stability (the opposite of Neuroticism) involves aspects of both communal and agentic bias (Paulhus, 2002).

### 1.1.2.2. Personal values

Many researchers have suggested that values can be defined as trans-situational goals that serve as guiding principles in the life of a person or group (e.g., Allport, 1961; Feather, 2002; Kluckhohn, 1951; Rohan, 2000; Rokeach, 1973; Schwartz, 1992). Schwartz and Bilsky (1987) identified five formal features that recur in conceptual definitions of values: Values (a) are concepts or beliefs, (b) guide selection or evaluation of behaviour, policies, people, and events, (c) transcend specific actions and situations, (d) are less numerous and more central to personality than are attitudes, and (e) are ordered by importance relative to one another.

One of the most influential models of values is the two-dimensional taxonomy of personal values advanced by Shalom Schwartz (1992). Schwartz and his colleagues (1992, 2005a, 2005b; Schwartz et al., 2001; Schwartz & Sagiv, 1995) have conducted research in almost 50 languages, revealing a close to universal set of values differentiated by motivational content. Within the framework offered by Schwartz' values theory, communal bias is correlated with adherence to values such as Obedience, Honoring of Parent and Elders, and Honesty, whereas agentic bias is correlated with adherence to values such as Authority, Social Power, and Influence (Verkasalo & Lindeman, 1994). However, exaggeration of value endorsement has not previously been linked to response bias.

## 1.2. SDR as a moderator or suppressor variable

There are two primary ways in which SDR scales can be used to improve personality assessment validity. They can be used as moderator or as suppressor variables. Moderator variables are variables that affect the direction and/or strength of the relation between dependent and independent variables (Baron & Kenny, 1986). For instance, within a correlational analysis framework, SDR can be thought of as a third variable that affects the zero-order correlation between two other variables, such as self-reports and other-reports of personality. Only people low in SDR may show the expected positive connection between self-reports and other-reports. People high in SDR may be more likely to substitute their true answers with socially desirable answers, and variation in their responses may thus reflect mostly understanding of social norms and less a description of own behaviour. This means that statistical manipulations that strive to construct valid test scores are not likely to succeed (Comrey & Backer, 1975), because there is little or no true score variance in the

item responses of people high in SDR. In this moderator model of SDR, the substantive scales of individuals high in SDR have to be viewed with caution.

Suppressor variables are variables that lower (suppress) the correlations between two other variables. They do not measure variance in the dependent measure, but do measure some of the variance in the independent measures which is not found in the dependent measure. By eliminating this error variance from the independent variable, the suppressor variable allows the independent variable to explain more of the variance of the dependent variable. SDR can be interpreted as a suppressor variable, in which case it represents correctible bias that results from simply shifting one's answers from the true response toward the desirable response by some amount. In this suppressor model, SDR, be it a response set or a response style, will contribute variance to self-reports of desirable personality traits. When correlating those desirable personality traits with criteria not influenced by SDR (e.g., objective indicators of the trait), SDR will be treated as error variance in terms of the correlation, with the result that the coefficient of correlation will normally be attenuated. Consequently, partialling SDR from the self-ratings should show evidence for a suppressor effect, whereby the value of the correlation increases when SDR is controlled for. SDR can thus be conceptualized as a suppressor variable that attenuates the real relation between a personality test and a criterion of interest (see Conger & Jackson, 1972; Paulhus, Robins, Trzesniewski, & Tracy, 2004).

A recent meta-review by Ones, Viswesvaran, and Reiss (1996) of research conducted in real world settings failed to find any suppressor effects on validity for SDR (see also Piedmont, McCrae, Riemann, & Angleitner, 2000). That is, partialling SDR out of test-criterion correlations did not result in any improvements to test validity estimates. However, that meta-analysis suffered from summarizing a wide variety of personality measures and an equally wide variety of SDR measures. For instance, all SDR scales were treated as measuring the same underlying construct, which is hardly warranted given Paulhus' (2002) two-stage classification of SDR measures. The different SDR scales were thus prone to overlap haphazardly with different personality measures. Rosse, Stecher, Miller, and Levin (1998) detailed other problems with Ones et al.'s (1996) meta-analysis.

In contrast to the meta-analytic finding reviewed above, some simulation studies have found evidence that response distortion on personality scales hurts the criterion-related validity of those measures. Schmit, Ryan, Stierwalt, and Powell (1995) reported that instructions to respond as if applying for college admission destroyed the validity of a measure of conscientiousness used to predict students' college grade point average. In a



similar vein, Douglas, McDaniel, and Snell (1996) witnessed the correlations between personality measures and supervisory job performance ratings dissipate when the respondents were instructed to fake good. However, whether this means that test protocols of those engaged in SDR need to be discarded (moderator model), or that those test protocols can be adjusted for SDR (suppressor model), is unclear. The moderator and suppressor models of SDR were compared and evaluated in Studies I, II, and IV.

### **1.3. Effects of SDR in research on personality stability**

The degree of consistency in personality over time has been a matter of some debate in the psychological literature (Block & Block, 2006). Recent research on personality stability shows evidence of both stability and change in personality during the transition from childhood to adulthood. It is important to note that consistency of personality across time can be measured both at the level of the group and at the level of the individual. Both group-level and individual-level consistency estimates could be affected by SDR.

#### **1.3.1. Inflated test-retest correlations?**

At the group level, one can assess absolute stability and differential stability. Absolute stability refers to the consistency in a group of people in their average level of a trait over time. It is typically estimated by the difference between average trait scores for a group assessed on two different occasions. However, as will be seen, the present research focuses only on differential or rank-order stability. Differential stability refers to the rank-order consistency of individual differences within the group over time. It is usually indexed by the test-retest correlation coefficient computed for a specific personality trait measure administered to the same people at two different points in time.

Because the test-retest correlations on personality dimensions are typically computed from self-ratings, SDR can inflate the size of these correlations, especially on highly evaluative dimensions. If SDR is contributing variance to both variables being correlated (e.g., self-ratings of the same personality trait at two different points in time), then the response bias will spuriously increase the value of that consistency correlation. Partialling SDR out of the correlation will consequently lower the estimate of consistency. Note also that if a partial test-retest correlation, controlling for desirability, were found to be less than the zero-order correlation, this would sustain a response style or substantive interpretation of the desirability measure rather than a response set interpretation. This is because the bias

would need to be stable across the two assessments to result in the enhanced zero-order correlation. Studies I and III investigated how SDR influences test-retest correlations and thereby estimates of personality stability.

### **1.3.2. The maturity-stability hypothesis**

Test-retest correlations reflect group-level continuity. But there is no reason to presume that every individual in a group will show the same degree of continuity. Ipsative stability refers to the stability of the configuration of an individual's personality profile across time. Ipsative stability is a summary index of changes in a person's personality trait levels compared to some previous point in time. Studies concerned with ipsative stability have assessed characteristics of persons who differ in this variable. Roberts, Caspi, and Moffitt (2001) have argued that individuals with more mature personalities should experience less personality change over time than individuals with less mature personalities. This maturity-stability hypothesis has been confirmed in a number of studies that have found predictors of ipsative stability, including psychological factors related to the concepts of ego resiliency (Asendorpf & van Aken, 1991), planful competence (Clausen, 1991), constraint, positive emotionality, and low negative emotionality (Donnellan, Conger, & Burzette, 2007; Roberts, Caspi, & Moffitt, 2001).

Positive connections between maturity and ipsative stability have been explained in terms of, for example, personal capital in the form of increased resiliency (Roberts, Caspi, & Moffitt, 2001), or better compatibility with the scripts society and culture provide (Clausen, 1991). However, the indicators of maturity typically used to evaluate those relations have been self-reports on standard questionnaires. Thus, social desirability response bias could add reliable variance to personality scale scores that is not there for people who are low in desirability response tendencies. It is possible, therefore, that those who consistently engage in socially desirable responding will produce comparatively high ipsative stability (profile-profile) correlations compared to those who are more truthful in their responses. It is conceivable, therefore, that the maturity-stability hypothesis is based, in part, on an artifact produced by individual differences in SDR. This possibility was examined in Study III.

## 1.4. Effects of SDR on self-selection

Self-selection takes place when individuals volunteer to join a group. Self-selection can be considered a major predicament in research in the behavioural and medical sciences, because studies in those fields are largely based on volunteer participants. The difficulty that occurs is that the research may therefore be dealing with the psychology of the volunteer and, thus, shaped by a selection bias. That is, perhaps the persons who are giving rise to much of the empirical data are not comparable to the people who choose not to participate in such research. This, of course, represents a challenge to the generalizability and external behavioural and medical validity of research results.

Although there is some research on personality traits that distinguish research volunteers from nonvolunteers, most of that research was conducted in the 1960s and 70s. And, perhaps in part because no general framework within which to assess personality was available at that time, the results are somewhat patchwork. In 1975, Rosenthal and Rosnow reviewed the accumulated evidence on the personality traits of research participants and reached the following conclusions. Volunteers compared with nonvolunteers were likely to be, approximately in order of importance, more intelligent, higher in the need for social approval, more sociable, more arousal-seeking, less conventional, lower in authoritarianism, and to some extent more altruistic. Notice that some of these characteristics, such as intelligence, sociability, and altruism, are very highly evaluative and socially desirable characteristics.

One of the variables that Rosenthal and Rosnow (1975) found to be among the most important predictors of research participation is something they called need for social approval. This might explain some of the results implying that research volunteers usually have more desirable personality traits than do nonvolunteers. Individuals high in need for social approval are likely to volunteer for a research study to earn the high opinion and high regard of some unfamiliar researcher. The same individuals are also more likely to respond in a desirable direction on personality measures for the same reason. The insinuation of this is that volunteers may have higher scores on desirable traits than do nonvolunteers partly because of an SDR response bias.

Note that this reasoning presupposes an interpretation of SDR as reflecting substantive individual differences in personality and not just a response style or response set. That is, those scoring high in SDR are presumed to actually behave in a socially desirable manner – i.e., to volunteer as research participants. But it is important to determine the extent to

which observed differences in personality traits between research volunteers and nonvolunteers, such as the greater intelligence and sociability of the former, can be explained in terms of SDR. A related question is whether these differences in personality traits reflect differences in broader personality traits, such as the Big Five personality factors. Study IV of this dissertation was designed to address these issues.

## 2 RESEARCH QUESTIONS AND HYPOTHESES

This thesis describes the results of four studies designed to contribute to the discussion on the effects of SDR on research outcomes in the personality area. A list of the research questions is given below. Table 1 provides an overview of the present research, and shows which research questions are pursued in each study.

a) Do the CM and SDE scales of the BIDR measure response set, response style, or substantive individual differences in personality? Based on the work of Paulhus (2002), the CM scale would be expected to be a response set measure and the SDE scale a response style measure. Based on other findings (e.g., McCrae & Costa, 1983), there could be a substantive component to those scales. Because of the generally inconsistent findings in the literature, as described in the introduction, no hypotheses were formulated.

b) Do the CM and SDE scales measure social desirability bias regardless of content, agentic bias, or communal bias? Based on the work of Paulhus (2002; & John, 1998), the CM scale was expected to measure communal bias and the SDE scale to measure agentic bias.

c) Should the test protocols of individuals scoring high in SDR be discarded, or should these protocols be adjusted for SDR? That is, should the CM or SDE scales be interpreted as moderator variables or as suppressor variables? Or are both interpretations false? Because these questions are intertwined with the question of what the BIDR scales actually measure, and previous results are inconclusive, no hypotheses were formulated.

d) In research on personality stability, is differential stability inflated by SDR? Furthermore, can SDR, in part, explain some of the evidence for the maturity-stability hypothesis in research on ipsative stability? Differential stability was expected to be influenced by SDR, and part of the evidence for the maturity-stability hypothesis was also expected to be explainable by SDR.

e) Can individual differences in SDR explain some of the observed personality differences between research volunteers and non-volunteers? And, do these groups differ on broader personality traits as well? SDR was expected to explain some of the observed personality differences, but no hypotheses were formulated concerning broader personality differences.

Table 1. Overview of the present studies.

	Study I	Study II	Study III	Study IV
Research questions	a, b, c, d	a, b, c	d	c, e
Participants	Military cadets (Samples 1 and 2); Married couples (Sample 3)	Finnish humanities students, military cadets, and business and technology students; Russian business and technology students	Finnish conscripts	Finnish military cadets (Sample 1); Adult siblings from large families (Sample 2)
Procedure	Participants responded as applicants and as incumbent (Samples 1 and 2); Self- and spouse ratings (Sample 3)	Participants responded with standard instructions and with instructions to fake good	15 year follow-up study on personality stability	Cadets were asked to volunteer (Sample 1); Self- and sibling ratings of personality (Sample 2)
Measures	BIDR; Personality measures	BIDR; Values measure	MMPI Lie Scale; Personality measure	BIDR; Personality measures
Dependent variables	Responses as incumbents (Samples 1 and 2); Spouse-ratings of personality (Sample 3)	Responses under standard instructions	Personality stability	Volunteering vs. not-volunteering (both Samples 1 and 2)

*Note.* For research questions a, b, c, d, & e, see text. Different samples of military cadets participated in Studies I, II, and IV.

## 3 METHODS AND RESULTS

Regarding the statistical tests used in this research, for differences in group means, effect sizes are reported as Cohen's  $d$ . Differences in means between dependent groups (Study I; Study II; Study IV, Sample 2) were tested with paired samples  $t$ -tests, and differences between independent groups (Study IV, Sample 1) with independent samples  $t$ -tests. All estimates of Cohen's  $d$  were computed using the means and original standard deviations rather than the  $t$ -test values (see Dunlop, Cortina, Vaslow, & Burke, 1996). For all General Linear Model analyses, effect sizes are given as  $\eta_p^2$ . For regression analyses, increases in explained variance ( $\Delta R^2$ ) are reported. Studies I, II, III, and IV refer in order to the four original publications listed above.

### 3.1. Study I

Study I was designed to evaluate whether Paulhus' (1991) SDE and CM SDR scales should be interpreted as response set measures, response style measures, or measures of substantive individual differences in personality. Furthermore, those scales were evaluated as measures of communal and agentic bias, and as suppressor and moderator variables. In Samples 1 ( $N = 57$ ) and 2 ( $N = 62$ ), army officer trainees were tested as applicants to their program and retested as incumbents three years later. In Sample 3 ( $N = 70$ ), self-reports on SDR scales and other personality measures were compared with spouse-reports.

#### 3.1.1. Participants and procedure

Three different samples were used in Study I. The first two were samples of male applicants who were accepted to the Military Officer training program at the National Defense College in Lappeenranta, Finland, in 1992 and 1993. The 1992 applicant sample consisted of 57 men (mean age = 20.5 years), and the 1993 applicant sample of 62 men (mean age = 21.0 years). The third sample consisted of 70 participants (35 married couples) who were randomly selected from the Helsinki phonebook. The mean age of that sample was 51.5 years and the couples had been married for an average of 24.6 years.

In the first two military samples, all applicants completed the questionnaires as part of their application process. Furthermore, the accepted applicants completed the questionnaires three (Sample 1) or two (Sample 2) years later as incumbents. The applicant

condition was reasoned to be a situation that would be likely to elicit socially desirable responding in self-reports relative to the incumbent condition.

Participants in Sample 3 were given the opportunity to participate in the study by completing both a self-report and a spouse report questionnaire, which would be mailed to them. In return, they were promised short feedback on their personality profiles. In all three samples, participants' responses were recorded non-anonymously.

### **3.1.2. Measures**

#### **3.1.2.1. Socially desirable responding**

In all three samples, the Finnish translation (Lindeman & Verkasalo, 1995a) of Paulhus' (1991) Balanced Inventory of Desirable Responding (BIDR) version 6 (presented in Appendix A), which consists of the 20-item Communion Management (CM) and 20-item Self-Deceptive Enhancement (SDE) subscales, was administered. Each item was rated on a 7-point scale. The sample of spouses completed the BIDR in self-rating and spouse rating format. In the latter case, those scales were rewritten in the third person form, so as to allow for the rating of one's spouse's behaviour. For instance, "I never swear" was rewritten as "She/He never swears." Across the three samples, the internal consistency reliability (Cronbach's  $\alpha$ ) of the self-report CM scale varied from .79 to .85, and the reliability of the self-report SDE scale varied from .67 to .78. The reliability of the spouse rating CM scale was .82, and the reliability of the spouse rating SDE scale was .44.

#### **3.1.2.2. Personality**

One sample of military cadets (Sample 2) were asked to rate the self-descriptiveness of 30 personality adjectives (in Finnish; see Rauste-von Wright & von Wright, 1984) using a rating scale ranging from 1 (*not at all like me*) to 7 (*very much like me*). These adjectives were then scored on the Big Five personality dimensions. For each dimension, the scores on the 30 adjectives were added to form a weighted composite variable. The adjectives' weights on the five factors were based on the factor loadings derived from an independent sample of university students responding anonymously ( $N = 884$ ). To assess the criterion validity of the 30-adjective form, it was administered along with the NEO-FFI (Costa & McCrae, 1992) to another independent sample of students ( $N = 206$ ). The correlations across measures were .68 (Neuroticism), .71 (Extraversion), .26 (Openness), .62 (Agreeableness), and .65 (Conscientiousness). The reliabilities (Cronbach's  $\alpha$ ) of these five dimensions ranged from .46 to .84.



The sample of spouses completed a Big Five personality measure in self-rating and spouse rating format. The personality factors, each of which is constituted of six lower-level facet scales, were measured by the scales of the NEO-PI-R (Costa & McCrae, 1992) form S (self-rating) and form R (spouse rating). Each of the 30 facet measures contains 8 items, and each item is responded to on a 5-point rating scale. The reliabilities of the Big Five self-report factor scales ranged from .80 to .93, and the reliabilities of the spouse reports from .86 to .92.

### **3.1.3. Results**

#### **3.1.3.1. Response set vs. response style vs. response substance**

As argued in the Introduction, a measure of response set should change level with changes in the demand to present a good impression. Consistent with a response set interpretation of the BIDR CM scale, CM scores were elevated in the applicant condition as compared to the incumbent condition ( $d = 0.59$ ,  $p < .01$ , and  $d = 1.07$ ,  $p < .01$ , in Samples 1 and 2, respectively).

In contrast to response sets, response styles would not be expected to vary across conditions. Scores on the SDE scale were, therefore, not expected to vary across conditions or instructional sets. The results were mixed in this regard. In the first sample, applicant scores were similar to incumbent scores ( $d = 0.02$ ,  $p = ns$ ). However, in the second sample, applicant scores were higher than incumbent scores ( $d = 0.36$ ,  $p < .01$ ). This latter result can thus be interpreted as suggesting that the SDE scale, at least to some extent, measures response set.

The effects of the situation on SDR were expected to depend on the BIDR scale evaluated. More specifically, the CM scale was expected to react more strongly to the contextual demand for SDR than the SDE scale. Consistent with this, the effect of the situation, where completing an application causes one to present a good impression, was noticeably stronger for the CM scale than the SDE scale ( $\eta_p^2 = .35$  and  $\eta_p^2 = .54$ , both  $p < .01$  in Samples 1 and 2, respectively).

Differential stability can also be considered when evaluating whether the CM and SDE scales are measures of response set or response style. The differential stability of the CM scale was predicted to be relatively low, being a response set measure. The test-retest correlations between applicant and incumbent CM three years later was quite high at .68 ( $p < .01$ ) in both samples. The differential stability of the SDE scale was similar to that of the CM scale. The test-retest correlations were .44 ( $p < .01$ ) and .71 ( $p < .01$ ) in Samples 1 and

2, respectively. Thus, individual differences in the relative strength of CM and SDE tendencies were remarkably stable over the applicant and incumbent conditions. These results suggest a response style interpretation of the CM and SDE scales.

At this stage, it is important to note that the type of evidence presented so far cannot distinguish between response style and some other substantive individual differences. In the former case, respondents misrepresent themselves as being good. In the latter case, they are good in some normative sense. To distinguish between these two interpretations, indicators that are not subject to response bias are needed. In Sample 3, spouses corroborated, to some extent, claims made on the CM and SDE scales. Correlations between the self- and spouse reports of CM and SDE were positive and statistically significant, .35 ( $p < .01$ ) for CM and .33 ( $p < .05$ ) for SDE. Furthermore, self-reports on the SDE scale were correlated with spouse reports of low Neuroticism ( $r = -.24$ ,  $p < .05$ ), and high Extraversion ( $r = .33$ ,  $p < .01$ ). These correlations, although not high, suggest that both BIDR scales measure some substantive individual differences, and not merely response bias. However, concerning the CM scale, these substantive individual differences appear not to be covered by the Big Five factor structure.

### 3.1.3.2. Communal bias vs. agentic bias

One type of evidence that could be considered in determining whether an SDR scale measures response set or response style is how the correlations with other desirable personality scales change as a function of the situational demand for SDE. With regard to response set measures, associations between response sets and self-reports on desirable personality scales should be higher when an incentive to present a positive impression exists. When such an incentive is present, more of the variance of response set scales and desirable personality scales should be attributable to the motivation to respond desirably. In contexts with little motivation to respond desirably, response set scores as well as desirable personality scores should be much less extreme, resulting in attenuated correlations between response sets and desirable personality scales. With regard to response style measures, generally the same correlations with desirable personality traits should be found regardless of the demand for SDR.

When looking at the correlations between SDR measures and other desirable personality scales, questions concerning the content of the SDR scales are intertwined with question concerning the type of response bias measured. Measures of communal and agentic bias should be correlated with exaggeration of communal and agentic traits,

respectively. For measures of response set, these correlations should be higher when an incentive to present a good impression is present. For measures of response style, these correlations should not vary as a function of the incentive to present a good impression.

In the applicant context, the CM scale was correlated with high scores on Big Five Emotional Stability, Agreeableness, and Conscientiousness ( $r = .38$  to  $.44$ , all  $p < .01$ ). Consistent with a response set interpretation, those correlations did not persist into the incumbent condition. In the sample of spouses, CM was correlated  $.33$  ( $p < .01$ ) with self-ratings of Agreeableness and the more interpersonal facets of Conscientiousness; that is, Dutifulness ( $r = .34$ ,  $p < .01$ ) and Deliberation ( $r = .31$ ,  $p < .01$ ). These correlations are consistent with the view that the CM scale measures communal bias.

The SDE scale was, in the applicant context, positively correlated with Emotional Stability, Extraversion, Openness, Agreeableness, and Conscientiousness ( $r = .37$  to  $.72$ , all  $p < .01$ ). The correlations obtained in the applicant context can be interpreted as suggesting that the SDE scale measures SDR regardless of content. Consistent with a response set interpretation, the correlations of SDE with Emotional Stability and Conscientiousness were attenuated in the incumbent condition, whereas the correlations with Openness and Agreeableness disappeared. Only the correlation between SDE and Extraversion persisted over the two conditions. In the sample of spouses, SDE was positively correlated with self-ratings of Emotional Stability, Extraversion, and Conscientiousness ( $r = .34$  to  $.66$ , all  $p < .01$ ). Closer analyses of the Conscientiousness scale revealed that SDE was positively correlated with Competence, Self-Direction, and Achievement ( $r = .32$  to  $.40$ , all  $p < .01$ ), but not with the more interpersonal facets of Conscientiousness. These correlations, obtained in circumstances with little pressure to respond desirably, are consistent with Paulhus' (2002) notion of the SDE scale as a measure of agentic bias.

Important in determining whether the CM and SDE scales measure bias or real individual differences is to look at whether the scales predict a departure from reality. A criterion of real behavior independent from self-descriptions is, of course, necessary for such a determination. The spouse ratings given by the third sample provide such a criterion. To measure the discrepancy between self- and spouse ratings, self-ratings of personality were regressed on spouse ratings, and the residuals were saved. (The residual analysis presented here was not reported in the original publication.) The residuals indicate how much of the personality self-ratings cannot be explained by spouse ratings (see Paulhus, 2002). These residuals were then correlated with CM and SDE scores. As already noted, the CM scale was correlated with self-ratings of Agreeableness, as well as

Dutifulness and Deliberation, but not with the spouse ratings on these traits. Consistent with this, CM correlated with the Agreeableness residual, as well as with the Dutifulness and Deliberation residuals ( $r = .28$  to  $.34$ , all  $p < .05$ ), but with no other Big Five residuals. As for the SDE scale, it was, as noted above, correlated with self-ratings of Emotional Stability, Extraversion, and the Conscientiousness facets Competence, Self-Direction, and Achievement. The correlations with spouse ratings were much weaker. Consistent with this, SDE correlated with the Emotional Stability residual ( $r = .63$ ,  $p < .01$ ), the Extraversion residual ( $r = .23$ ,  $p < .05$ ), and the residuals of the three Conscientiousness facet ( $r = .31$  to  $.38$ , all  $p < .01$ ). These correlations suggest that the CM and the SDE scales are correlated with exaggeration of communal and agentic characteristics, respectively.

### 3.1.3.3. Moderator variables vs. suppressor variables

In the following analyses, possible moderator and suppressor effects of the CM and SDE scales were examined (the moderator analyses were not included in the original publication). Hierarchical multiple regression analyses (see Aiken & West, 1991) were performed to examine the moderating effects of the BIDR scales on the predictive power of (a) applicant personality trait scores in the prediction of incumbent personality trait scores (Sample 2), and (b) self-ratings of personality in the prediction of spouse ratings (Sample 3). Before the interaction terms were calculated, the criterion variables and the predictor variables were standardized to reduce possible multicollinearity among the independent and interaction terms as well as to facilitate interpretation of the interaction effects (see Aiken & West, 1991).

To first evaluate the moderator model of SDR, incumbent ratings on each of the Big Five traits (one at a time) were predicted with applicant ratings on the same trait, plus applicant CM or SDE scores (one at a time), plus the interaction between the Big Five trait and the BIDR scale. For this purpose, 10 regression analyses were run, of which none showed a statistically significant effect for the interaction term (all  $t(57) < 1.34$ ,  $p = ns$ ), nor a statistically significant increase in variance explained when the interaction term was added to the model (all  $\Delta R^2 < .02$ ,  $p = ns$ ). Spouse ratings on each of the Big Five traits were then predicted with self-rating on the same trait, the CM or SDE score (one at a time), and their interaction. For this purpose, 10 regression analyses were run, of which none showed a statistically significant effect for the interaction term (all  $t(66) < 1.88$ , all  $p = ns$ ) nor an increase in variance explained (all  $\Delta R^2 < .03$ , all  $p = ns$ ).

Next, possible suppressor effects were considered. Partialling out CM from the correlations between applicant and incumbent personality scores did not affect the correlations (difference in zero-order and part correlations ranged from -.02 to .01). However, partialling out SDE actually decreased the correlations in some cases. This result will be returned to in the chapter on personality stability. In the sample of spouses, partialling out CM or SDE from the correlations between self-reports and spouse reports of the Big Five had very little effect on the correlations (difference in zero-order and part correlations ranged from -.09 to .01). Again, there was no evidence for the existence of a suppressor effect.

#### 3.1.3.4. Effects of SDR on stability estimation

As noted in the introduction, SDR can influence personality stability estimates both at the level of the group (differential stability) and at the level of the individual (ipsative stability). Concerning differential stability, desirability response bias can inflate the size of test-retest correlations on evaluative dimensions because those correlations are typically computed from self-ratings. If SDR is contributing variance to both variables being correlated (e.g., self-ratings on a desirable personality trait measured at two points in time), then the response bias will increase the value of the test-retest correlation. Partialling SDR out of the correlation will thus lower it.

In Sample 2, high correlations were found between subjects' Big Five personality scores obtained as applicants and those obtained as incumbents, averaging .61 across the five factors. When the subjects' CM scores were partialled out of the applicant-incumbent personality factor correlations, very little change in those correlations was observed. The range in differences between the zero-order and partial correlations was -.02 to .01. But this might be anticipated for the CM scale. It is assumed to be primarily a measure of a response set related to impression management, and that response bias should not consistently affect both the applicant scores and the incumbent scores of the military cadets.

The situation with the partial correlations was rather different when the SDE scale was used as the covariate applied to the applicant-incumbent personality correlations. Two of the Big Five personality factors showed clear declines in correlations when self-deceptive enhancement was controlled: the .71 correlation for Neuroticism declined to .53 ( $p < .05$ ), and the .57 correlation for Conscientiousness declined to .41 (this change narrowly failed to reach statistical significance, probably due to the relatively small sample size of  $N = 62$ ). These results, by themselves, point towards a response style or substantive interpretation of

the SDE scale, as SDE contributed variance to the personality ratings across two different contexts. If the SDE scale is a response style measure, then partialling SDE from the test-retest correlations would be to remove only unwanted or nuisance variance. However, if the SDE scale measures substantive individual differences, then to partial SDE from the test-retest correlations would be removing wanted variance from both the predictor and criterion variables.

## **3.2. Study II**

The purpose of Study II was to evaluate Paulhus' (1991) CM and SDE scales as measures of response set and response style, and also as measures of communal and agentic bias. The CM and SDE scales, as well as the Schwartz values survey (SVS), were administered to four samples drawn from different cultures and subcultures ( $N = 900$ ). Participants completed the questionnaires in honest and fake good conditions. Study II also investigated moderator and suppressor models of SDR.

### **3.2.1. Participants and procedure**

Data from four samples, three Finnish and one Russian, were used for Study II. The first sample consisted of 232 humanities students (55 men, mean age 24.4 years). The second sample consisted of 140 business and technology students (80 men, mean age 22.6 years). The third sample was an all male sample of military cadets ( $N = 264$ , mean age 22.5 years). The fourth sample was a sample of 264 Russian business and technology students (135 men, mean age = 19.2 years).

In all samples, the participants were first given an SDR measure and a personal values questionnaire with standard instructions (Honest condition). Situational pressure to respond desirably was assumed to be low, as participants responded anonymously (Lindeman & Verkasalo, 1995b). About one week later, participants were given the same questionnaires with the instructions to "fake good." In that condition (Fake condition), roughly half of the participants were given the instructions to "respond as you would respond if you were trying to make as positive an impression as possible on your classmates," whereas the other half were instructed to make a positive impression on a professor. Different targets of the good impression were used because they may influence

self-presentation; people tend to tailor their public images to the perceived values and preferences of significant others (for a review, see Leary & Kowalski, 1990).

### **3.2.2. Measures**

#### **3.2.2.1. Socially desirable responding**

As in Study I, Paulhus' (1991) Balanced Inventory of Desirable Responding (BIDR) was used to measure SDR. In the Honest condition, across all four samples, the reliability of the CM scale ranged from .66 to .81, and the reliability of the SDE scale from .60 to .74. In the Fake condition, across all samples, the reliability of the CM scale ranged from .66 to .81, and the reliability of the SDE scale from .79 to .91.

#### **3.2.2.2. Personal values**

Personal values were measured using the Schwartz Value Survey (SVS; Schwartz, 1992). This measure includes 56 single values each assessed using a nine-point scale, ranging from  $-1$  (*opposed to my values*) to  $7$  (*of supreme importance*). The SVS was administered in order to assess the content of the response bias measured by the BIDR scales. In factor analyses of different personality measures, the first extracted factor is often associated with response tendencies (e.g., Davison, 1985), as well as the general desirability of responses (e.g., Boies, Lee, Ashton, Pascal, & Nicol, 2001; Saucier, 2003). Therefore, factor analysis was expected to yield a first factor representative of SDR regardless of content. Furthermore, if communion and agency are fundamental modalities of human existence (Bakan, 1966; Paulhus, 2002), then one would expect the SVS, as a measure of universal human values, to reveal a communal values factor and an agentic values factor. To determine these three values factors, maximum likelihood factor analysis was conducted on the 44 universal values (see Schwartz, 1992) included in the SVS (only responses in the Honest condition were used, and all groups were combined). The first extracted factor correlated .95 with the total sum score of the 56 variables from the SVS. The first factor was thus interpreted as measuring general SDR (recall that all values are scored in the desirable direction) and was named the Elevation factor, referring to the general elevation of the values profile. The second and third factors were rotated graphically using the Survo MM version 2.40 statistical package (Mustonen, 1992) to correspond with communion and agency. The second values factor was rotated to have high loadings of the values Obedient, Honoring of Parent and Elders, and Honest, and low loadings of Independent, Choosing Own Goals, and An Exciting Life. The third values factor was simultaneously rotated to

have high loadings of Authority, Social Power, and Influence, and low loadings of Social Justice, Equality, and Protecting the Environment. The second and third values factors thus corresponded to communion and agency, respectively. The item weights obtained in the Honest condition were used to construct the Elevation, Communal, and Agency values factors in the Fake condition. In the Honest condition, the average internal consistency reliabilities (computed using the General Reliability Coefficient; Tarkkonen & Vehkalahti, 2005) of the three values factors were .92 (Elevation), .85 (Communion), and .79 (Agency). These reliabilities were somewhat higher when participants were instructed to fake.

### 3.2.3. Results

#### 3.2.3.1. Response set vs. response style vs. response substance

Consistent with a response set interpretation of the BIDR CM scale, CM scores were, across all four samples, elevated in the two Fake conditions as compared to the Honest condition ( $d = 0.45$  to  $1.29$ , all  $p < .01$ ). In contrast to response sets, response styles would not be expected to vary across conditions. However, across all four samples, SDE scores were elevated in the Fake conditions as compared to the Honest condition ( $d = 0.45$  to  $1.29$ , all  $p < .01$ ). These results can thus be interpreted as suggesting that the SDE scale, at least to some extent, measures response set.

The effects of the situation on SDR were expected to depend on the BIDR scale being evaluated. More specifically, the CM scale was expected to react more strongly to the contextual demand for SDR than was the SDE scale. The effects of the instructions to fake were different for the two different BIDR scales, for the four different samples, and for the two different targets of the good impression (the four-way interaction term was significant,  $\eta_p^2 = .05$ ,  $p < .01$ ). Follow-up analyses revealed that the instructions to fake had a stronger effect on CM scores than on SDE scores in the sample of military cadets ( $\eta_p^2 = .05$ ,  $p < .01$  and  $\eta_p^2 = .13$ ,  $p < .01$ , for peer and professor as targets, respectively), and in the sample of business students instructed to make a good impression on their professor ( $\eta_p^2 = .29$ ,  $p < .01$ ). In contrast, instructions to fake had a greater effect on SDE scores than on CM scores in the sample of humanities students ( $\eta_p^2 = .05$ ,  $p < .05$  and  $\eta_p^2 = .15$ ,  $p < .01$ , for peer and professor as targets, respectively). The results thus suggest that some characteristics of the respondents, as well as the target of the good impression, influence which of the two BIDR scales reacts more strongly to the situational demand for SDR.

Concerning differential stability, the differential stability of the CM scale was predicted to be relatively low, being a response set measure. The correlations between honest and



fake CM ranged from .04 to .54 across samples, with a weighted average test-retest correlation of .25 ( $p < .01$ ). The differential stability of the SDE scale was predicted to be relatively high, as a response style measure. However, the differential stability of the SDE scale was similar to that of the CM scale; the test-retest correlations varied from -.10 to .51 with a weighted average test-retest correlation of .28 ( $p < .01$ ). Thus, individual differences in the relative strength of CM and SDE tendencies were somewhat stable over the Honest and Fake conditions.

### 3.2.3.2. Communal bias vs. agentic bias

Concerning the BIDR CM scale, in the Honest condition it was positively correlated with communal values in all four samples ( $r = .14$  to  $.25$ , all  $p < .05$ ). Furthermore, consistent with a response set interpretation, these correlations increased in the samples of military cadets and Russian business students when participants were instructed to fake good (all  $p < .05$  for the increases in correlation). Correlated change analyses confirmed that increases in CM were correlated with increases in the Communal values factor in these two samples ( $r = .18$  to  $.28$ ,  $p < .05$ ). CM was also related to the Elevation factor, interpreted as indicative of general SDR. In the honest condition, CM was related to general SDR in the samples of humanities students ( $r = .25$ ,  $p < .01$ ) and military cadets ( $r = .13$ ,  $p < .01$ ). Furthermore, increases in CM were, in all four samples and in at least one of the two Fake conditions, correlated with increases in the elevation values factor ( $r = .18 - .38$ , all  $p < .05$ ). In general the results thus suggest that the CM scale measures both general SDR regardless of content and communal bias.

Concerning the BIDR SDE scale, in the Honest condition it was positively correlated with agentic bias only in the sample of Finnish business students ( $r = .23$ ,  $p < .05$ ). However, SDE was correlated with general SDR in the samples of humanities students ( $r = .20$ ,  $p < .01$ ) and military cadets ( $r = .31$ ,  $p < .01$ ). Furthermore, in the Fake condition, SDE was, for at least one of the two instructional sets, correlated with general SDR in all four samples ( $r = .20$  to  $.35$ , all  $p < .05$ ). Analyses of correlated change confirmed that when SDE increased, general SDR also increased across all four samples ( $r = .19$  to  $.38$ , all  $p < .05$ ). However, increases in SDE were related to increases in agentic values in only two of the samples: military cadets with peer as target ( $r = .19$ ,  $p < .05$ ), and Finnish business students with professor as target ( $r = .24$ ,  $p < .05$ ). The above correlations generally suggest that the SDE scale measures both agentic response bias and general SDR regardless of content.

### 3.2.3.3. Moderator variables vs. suppressor variables

As in Study I, hierarchical multiple regression analyses (see Aiken & West, 1991) were performed to examine the moderating effects of the BIDR scales on the predictive power of values measured in the Fake condition in the prediction of values measured in the Honest condition. (Note that the moderator analyses below differ from those reported in the original publication, where Honest condition scores were used to predict Fake condition scores. Either type of analysis is possible, but the original publication was deviated from in the present analyses in order to keep the moderator analyses of Study II consistent with those of Studies I and IV. The below suppressor analyses were also not included in the original publication.)

Each of the three values factors measured in the Honest condition were predicted with the same values factor measured in the Fake condition, with the CM or SDE (one at a time) scale measured in the Fake condition, and with their interaction. This resulted in 48 regression analyses, 11 of which showed a statistically significant interaction term at  $p < .05$ . Due to the large number of analyses and risk for Type 1 error, only the results that were consistent across two or more samples are presented. The interaction term between CM and communal values was significant ( $\beta = -.17$  to  $-.34$ ,  $t = -2.25$  to  $-3.05$ , all  $p < .05$ ), and the variance explained by the model increased significantly ( $\Delta R^2 = .03$  to  $.11$ , all  $p < .05$ ), in the samples of military cadets (either target), the sample of Finnish business students (professor as target), and the sample of Russian business students (professor as target).

Tests of simple slopes at plus and minus one standard deviation of the CM mean revealed that the interaction effect was similar across samples: the association between Communal values in the Fake condition and Communal values in the Honest condition was strong when CM was low ( $\beta = .62$  to  $.69$ ,  $t = 4.61$  to  $7.05$ , all  $p < .01$ ), but much weaker when CM was high ( $\beta = .06$  to  $.35$ ,  $t = 0.44$  to  $3.15$ ).

The interaction term between SDE and agentic values was significant ( $\beta = -.17$  to  $-.23$ ,  $t = -2.07$  to  $-2.72$ , all  $p < .05$ ), and the variance explained by the model increased significantly ( $\Delta R^2 = .03$  to  $.05$ , all  $p < .05$ ), in the samples of Finnish humanities students (peer as target), Finnish business students (peer as target) and Russian business students (peer as target). Tests of simple slopes at plus and minus one standard deviation of the SDE mean revealed that the interaction was similar across samples: the association between Agentic values in the Fake condition and Agentic values in the Honest condition was strong when SDE was low ( $\beta = .53$  to  $.74$ ,  $t = 4.57$  to  $5.31$ , all  $p < .01$ ), but much weaker when SDE was high ( $\beta = .13$  to  $.30$ ,  $t = 1.19$  to  $2.25$ ).

Possible suppressor effects were examined next. Partialling out CM or SDE scores from values in the Fake condition did not affect the correlations between values in the Fake and Honest conditions (difference in zero-order and part correlations ranged from -.06 to .02). Thus, no reliable suppressor effects were found.

### **3.3. Study III**

Study III examined the influence of SDR on personality stability research. More specifically, the focus was on differential stability estimates and on the maturity-stability hypothesis. A sample of male military conscripts in the Finnish Defence Forces ( $N = 74$ ) completed the Sixteen Personality Factor Questionnaire (16PF; Cattell, Eber, & Tatsuoka, 1970) and the MMPI Lie scale at age 20 and the 16PF again as civilians at age 35. The original publication had a focus slightly different from that of the present dissertation, and it did not include any of the analyses below that involve the MMPI Lie Scale.

#### **3.3.1. Participants and procedure**

This third study is part of a bigger research project examining the influences of father presence or absence on male military conscripts' behaviour (Mäkinen, 1991). Possible participants were all the 626 men registered at the Uusimaa Light Infantry Battalion in October 1987 or February 1988. Of these, 600 volunteered in writing for the first phase of the research, referred to as Time 1. Of these 600, 138 were removed because (a) their 20th birthdays did not occur in the course of their service ( $N = 108$ ), or (b) they were categorized as having a medical condition that restricted their duty assignments ( $N = 30$ ). Next, the remaining 462 conscripts were separated into those from father-absent families ( $N = 124$ ), father-present families ( $N = 298$ ), or mixed families ( $N = 40$ ). Finally, 90 conscripts were selected randomly (see Mäkinen, 1991) from the father-absent group and 62 from the father-present group, resulting in 152 participants assessed at Time 1. In the second phase of this research, called Time 2, 147 of the participants at Time 1 (5 were deceased) were contacted by mail and by telephone 15 years later as civilians; that is, in the year of their 35th birthdays. In total 74 former conscripts (50.3%) agreed to take part in a follow-up study.

Psychiatric interviews were carried out with each conscript at Time 1 and at Time 2. The interviews were carried out by a psychiatrist specialized in both adolescent and adult psychiatry. At the time of the interviews, the psychiatrist was blind to both the hypotheses

of the present research and the respondents' scores on all administered measures. The Finnish Defence Forces provided data from which seven measures of conscript competence were derived.

### **3.3.2. Measures**

#### **3.3.2.1 Socially desirable responding**

A 25-item version of the MMPI Lie scale (Hathaway & McKinley, 1946) developed by the Finnish Defence Forces (Nyman, 2007) was used as the measure of SDR in this study. It was administered along with other MMPI scales and other questionnaires to new recruits during the first two weeks of training. Unfortunately, only total scores on the MMPI were provided by the Finnish Defence Forces, and the internal consistency reliability of the measures could thus not be estimated. The Lie Scale's correlations with communal and agentic bias resemble those of the IM and SDE scales (Paulhus, 1984).

#### **3.3.2.2. Personality**

Personality was measured in the sample of men two times with Cattell's Sixteen Personality Factor Questionnaire (16PF; Cattell, Eber, & Tatsuoka, 1970; Cattell, 1981), once at Time 1 and once at Time 2. The 16 PF (Fourth Edition) consists of fifteen personality scales, and a Reasoning scale. The personality scales are labeled Warmth, Emotional Stability, Dominance, Liveliness, Rule-Consciousness, Social Boldness, Sensitivity, Vigilance, Abstractedness, Privatness, Apprehension, Openness to Change, Self-Reliance, Perfectionism, and Tension. Each scale is comprised of either 10 or 13 items. Two of the scales, Privatness and Openness to Change, had slightly negative mean interitem correlations in this study. Consequently, those two scales were not considered further. The reliabilities of the remaining scales ranged at the initial assessment from .26 to .86 and at the follow-up from .16 to .83.

#### **3.3.2.3. Psychiatric evaluations**

Psychiatric assessments were conducted on each participating volunteer both at Time 1 and at Time 2. The assessments consisted of semi-structured, comprehensive, interviews lasting approximately two hours and covering a variety of themes (e.g., family relations, sexual maturity, substance abuse). The psychiatrist rated the participants immediately after on (a) level of psychological adjustment, and (b) level of self-esteem, both on a scale from

1 (*poor*) to 4 (*excellent*). The Time 1 ratings were used as indicators of maturity, and the differences between Time 1 and Time 2 as criteria of person change over 15 years.

#### 3.3.2.4. Competence indices

The Finnish Defence Forces provided seven different indices conscript competence at Time 1. These competence indices were self-evaluated competence (average score on a seven-item scale with coefficient alpha reliability of .77), superior-evaluated competence (average score on a ten-item scale collected at three separate occasions with coefficient alpha reliability of .92 and test-retest correlation of .59), general and field evaluation (these are rated in the military passport that conscripts receive after fulfillment of military service), rank (whether or not the conscript was selected for leadership training), sick days (the number of days the participant was exempt from duty due to some physical or mental ailment), and disciplinary actions (the number of days a conscript was subjected to formal military punishment).

#### 3.3.3. Results

Concerning the influence of SDR on differential stability estimates, contrary to expectations, partialling out the MMPI Lie scale from the personality test-retest correlations did not have a statistically significant effect on those correlations. The differences between zero-order and part correlations ranged from -.08 to .01.

The above statistics reflect group-level changes. However, SDR may also influence ipsative stability. In particular, as argued in the introduction, SDR may, in part, underlie the maturity-stability hypothesis. To assess ipsative stability (the stability of the configuration of a person's personality profile across time), profile correlations were used (e.g., Robins, Fraley, Roberts, & Trzesniewski, 2001). Each of the trait scales was standardized across subjects before correlating the personality profiles, and to deal with the arbitrary direction of keying we used Cohen's (1969) coefficient of profile similarity,  $r_c$ . Consistent with the maturity-stability hypothesis, high Emotional Stability, Rule-Conscientiousness, and Perfectionism and low Apprehension, Tension, and Vigilance, measured when participants were 20, all predicted ipsative stability ( $r = .26$  to  $.46$ , all  $p < .05$ ). Besides being indicators of greater maturity, these personality traits are also possible indicators of desirability responding. Most pertinent to the present discussion, the MMPI Lie Scale also predicted greater ipsative stability ( $r = .25$ ,  $p < .05$ ). Consistent with this, correlations between the above personality traits and ipsative stability consistently became

weaker, albeit by non-significant amounts, when the Lie scale was controlled for (absolute differences in zero-order and part correlations ranged from .02 to .08). This pattern of correlations suggests that individual differences in SDR may, in part, contribute to the maturity-stability phenomenon.

At least four aspects of the data, described above, disagree with a pure desirability interpretation of the stability prediction results. First, other personality trait scales that could also be expected to incorporate a strong evaluative component, and thus be likely to be influenced by SDR, did not predict person stability. Those scales include Warmth, Self-Reliance, Sensitivity, and Dominance. Second, the relatively low test-retest correlations for some 16PF scales, despite an apparent desirability component (e.g., Apprehension showed a test-retest correlation of .17,  $p = ns$ ), indicates that the rank orderings of respondents changed substantially over the 15-year interval on those measures. Such low retest correspondence disagrees with the interpretation that desirability bias perseveres from age 20 to age 35, which in turn disagrees with the interpretation that the desirability bias caused the ipsative stability prediction results. Third, many of the indicators of greater maturity consisted of other-ratings, and objective indices such as military rank and number of sick days. These measures also predicted greater stability; for example, a composite index based on the first unrotated component of the psychiatrist evaluation and the competence as conscript variables correlated .34 ( $p < .01$ ) with ipsative stability. These measures are not likely to have been distorted by SDR. Fourth, a desirability interpretation of the stability results supposes that those who were recognized as being less temporally stable had relatively low ipsative stability because they changed their stylistic response tendencies on the personality scales from age 20 to age 35, most probably in the direction of greater desirability (at the group level, mean-level changes in the direction of greater desirability were found). However, the results show that such changes in the subjects' self-reports in the direction of greater desirability generally correlated with changes in the psychiatrist's ratings of them in the direction of greater adjustment (e.g., changes in self-reports of three Neuroticism related traits correlated on average .56 ( $p < .01$ ) with changes in psychiatrist's ratings of adjustment). In case the psychiatrist was not simply misled by the respondents' impression management tendencies, this finding is coherent with true personality changes, in the direction of greater desirability, in those participants over 15 years.

### **3.4. Study IV**

Study IV evaluated differences in SDR and in the Big Five personality factors between research volunteers and nonvolunteers. In the first sample, 158 military officers were asked to participate in a mail survey. The personality scores of the volunteers ( $N = 61$ ) and nonvolunteers ( $N = 97$ ) were available from an archival data set. In our second study, adult siblings from large families were invited to participate in extensive clinical epidemiological evaluations. The personality scores of volunteers ( $N = 55$ ) and nonvolunteers from the same families ( $N = 29$ ) were estimated from sibling ratings made by those who participated in the study.

#### **3.4.1. Participants and procedure**

Data from two samples were used for Study IV. In the first sample, the starting point was archival data on 199 male military officers in training in the Finnish Defence Forces. The archival data included an SDR measure as well as a Big Five personality questionnaire. Three years after the archival data had been collected at the National Defense College, 158 of the initial 199 cadets (mean age 25.0 years) were found to have pursued their military careers as paid officers of the Finnish Defence Forces. These officers were approached by mail and asked to take part in a follow up study to the original. Of the 158 contacted officers, 61 (39%) agreed to participate.

The data for the second sample were taken from a broader epidemiological project in which siblings from numerous large families were asked to go through a wide range of assessments connected to psychological and physical health. These assessments consisted of paper-and-pencil questionnaires, physical measurements, neuropsychological tests, and psychiatric interviews. Most important to the present study was the fact that the assessment battery also included sibling ratings of Big Five personality traits. A total of 118 siblings from 28 families took part in the epidemiological study. From those participants 15 families were identified in which (a) at least three siblings agreed to volunteer for the study (mean = 4.2 respondents per family) and (b) one or more siblings refused to volunteer (mean = 1.9 nonrespondents per family). This gave the chance to compare the personality traits of participants, as rated by their siblings, with the personality traits of nonrespondents, as rated by the same siblings. Thus, each of these 15 families would be contributing sibling ratings to both groups. The group of volunteers consisted of 55 targets who were rated by one or more siblings, having a mean age of 42.0 years, and 34 of whom

were women. The group of nonrespondents consisted of 29 targets who were also rated by one or more siblings, having a mean age of 43.1 years, and 13 of whom were women. In the investigation of these personality trait ratings, any manifold ratings of a target (by two or more siblings) were first averaged, and all self-ratings were excluded.

### **3.4.2. Measures**

#### **3.4.2.1. Socially desirable responding**

As in Studies I and II, socially desirable responding was assessed using the BIDR (Paulhus, 1991). This measure was administered to the military cadets that constituted Sample 1. The reliability of the CM scale was .82 and the reliability of the SDE scale was .61.

#### **3.4.2.2. Personality**

In the cadet sample, the Big Five personality factors were measured by the scales of the NEO Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992). Each scale contains 12 items, and each item is responded to on a five-point rating scale. The reliabilities of the scales ranged from .67 to .80.

In the second sample, participants were asked to rate their siblings on 15 bipolar adjective scales, each chosen to correspond to a dimension underlying one of the Big Five personality factors. Responses to the adjective scales were recorded on visual analogue scales. Each scale consisted of a 100mm line, with an adjective anchor at each end (e.g., introverted, extraverted) and with the word “average” placed near the midpoint. In making ratings of a sibling, participants were instructed to put a slash through the 100mm line indicating his or her judgment of the sibling’s position on the continuum. Furthermore, participants were instructed to put all of his or her siblings on the same line, without allowing for any ties (i.e., overlapping slashes). To minimize the tendency to give overly positive ratings, participants were told that not everyone can reasonably be rated above (or below) the scale midpoint on all of the dimensions.

Each of the Big Five personality factors was measured in Sample 2 by three bipolar adjective pairs (adapted from Goldberg, 1992) that were translated into Finnish. The number of raters per sibling varied from 1 to 6, with an average of 2.34 raters per nonrespondent, and 3.32 raters per respondent. The average pair-wise cross-observer correlations, computed between two siblings rating the same target, were .38 (Neuroticism), .43 (Extraversion), .56 (Openness), .58 (Agreeableness), and .58 (Conscientiousness). The



reliabilities of the five personality factor scales were .73 (Neuroticism), .86 (Extraversion), .20 (Openness), .84 (Agreeableness), and .89 (Conscientiousness).

### **3.4.3. Results**

#### **3.4.3.1. SDR and self-selection**

As noted in the introduction section, prior research has shown that people who are high in need for social approval are more likely to volunteer as participants for research. It was argued that the same people who volunteer in order to earn the appreciations of some unknown experimenter may, for the same reason, tend to describe themselves in a desirable manner on personality questionnaires. This could contribute to the observed differences in desirable personality traits between research volunteers and nonvolunteers.

In Sample 1, no differences were found on the BIDR SDE scale between respondents and nonvolunteers. However, a marginal difference ( $d = .31, p < .07$ ) was found on the CM scale – volunteers tended to score slightly higher in communion management tendencies than did nonvolunteers. Of the two BIDR constructs, communion management is arguably the closest to what would normally be defined as the need for social approval (Paulhus, 1984, 1991, 2002). These results suggest that social desirability may play some part in the explanation of personality differences between volunteers and non-volunteers. That is, those inclined to answer questionnaire items desirably may also be somewhat more likely to comply with requests for socially desirable behaviour, such as a request to participate in experimental research. Note that this reasoning supports a substantive interpretation of social desirability (see McCrae & Costa, 1983), as the “good” personality shows up in actual behaviour (volunteering) and not just as a response bias on paper-and-pencil questionnaires.

In the second sample, where sibling ratings of personality were used to assess personality differences between respondents and nonrespondents, respondents were rated as higher in Extraversion, Agreeableness, and Conscientiousness, but lower in Neuroticism ( $d = .18$  to  $.58$ , all  $p < .05$ ). The use of sibling ratings arguably precludes the effects of social desirability response bias as a determinant of differences between volunteers and nonvolunteers – that bias generally pertains to self-ratings and less so to other ratings (but see Murray, Holmes, Dolderman, & Griffin, 2000; see also Konstabel, Aavik, & Allik, 2006).

#### 3.4.3.2. Moderator variables vs. suppressor variables

The question of whether somehow controlling for the BIDR scales could improve the predictive power of the personality scales was examined next. Volunteering for the follow-up study was used as a binary criterion variable (Sample 1). (The analyses below were not presented in the original publication.) Hierarchical multiple regression analyses (see Aiken & West, 1991) were performed to examine the moderating effects of the BIDR scales on the predictive power of self-ratings of personality in the prediction of volunteering behaviour. That is, volunteering behaviour was predicted with Big Five personality traits (one at a time), CM or SDE scores (one at a time), and their interaction. None of the 10 binary logistic regression analyses that were run showed a statistically significant effect for the interaction term (all Wald statistics  $< 2.89$ , all  $p = ns$ ).

Finally, partialling out the BIDR scales from the correlations between Big Five traits and volunteering in research did not improve the predictive power of the Big Five traits (difference in zero-order and partial correlations ranged from  $-.02$  to  $.00$ ). The data of Study IV thus supported neither a moderator nor suppressor model of SDR.

## **4 GENERAL DISCUSSION**

The data presented in this dissertation suggest that measures of SDR can represent multiple sources of variance. Both the CM and the SDE scales of the BIDR, for example, can be interpreted as measures of response set, response style, and substantive individual differences in personality. The results further suggest that SDR can influence research on personality stability, but that a maturity-stability link is not purely an artifact of SDR. Finally, the results show that SDR cannot by itself explain the observed differences in personality between research volunteers and non-volunteers.

The general discussion below is organized into six sections. In the first section, results pertaining to the CM scale are evaluated, whereas the second section turns its attention to the SDE scale. The third section discusses the effects of SDR on research outcomes in the area of personality stability research. The fourth section focuses on the effects of SDR on volunteering as a participant for psychological research and on observed personality differences between respondents and nonrespondents. The fifth section discusses some of the limitations of the studies presented in this thesis. Finally, the sixth section presents the conclusions that the data presented in this thesis allow.

### **4.1. Communion management**

Paulhus (2002) has suggested that the BIDR CM scale is, on the one hand, a measure of response set, which is a short-lived response bias attributable to a contextual motivation to present a favorable impression. On the other hand, regarding the content of the bias, he has suggested that the CM scale is a measure of communal bias. In support of the response set interpretation, in both Studies I and II, CM scores were significantly higher in situations with high demand for SDR (applicant and Fake conditions) than in situations with low demand for SDR (incumbent and Honest conditions). Furthermore, in both Studies I and II, significant correlations were found between CM and desirable traits and values, especially in situations where the demand for SDR was high. Those correlations tended to be substantially reduced for the same subjects in a low demand situation. In further support of the response set interpretation, self-report CM scores were not related to spouse ratings of personality (Study I), which would suggest that the scale measures a response bias that is largely independent of other personality traits (at least as far as the Big Five factors are concerned).

In contrast to the results referred to above, some of the other results are inconsistent with CM as purely affected by response set. Specifically, applicant and incumbent CM scores were strongly correlated (Study I), and Honest and Fake CM scores moderately correlated (Study II), suggesting stability in communion management across different assessment contexts differing in their demand for SDR. These results suggest the existence of stable individual differences in communion management. With regard to the causes of this stability, some of the results suggest that the CM could be interpreted as a measure of substantive individual differences. In Study I, spouses corroborated, to some extent, the claims made by respondents on the CM scale. Furthermore, in Study IV, those scoring high in CM were more likely to comply with a request for socially desirable behaviour (i.e., to volunteer for research). These results support a substantive interpretation of the CM scale, because communion management tendencies are reflected in spouse ratings of personality and in actual socially desirable behaviour (volunteering), and not merely in self-ratings of personality.

Concerning the content of bias measured by the CM scale, the results were generally consistent with expectations, and support Paulhus' (2002) interpretation of the scale as a measure of communal bias. In the applicant context of Study I, CM was correlated with low Neuroticism, high Agreeableness, and high Conscientiousness. Furthermore, in the sample of spouses, CM was correlated with self-ratings of own Agreeableness and the more interpersonal facets of Conscientiousness. CM was also correlated with the exaggeration of these same traits in comparison to spouse ratings. In Study II, CM was correlated with endorsement of communal values both in the Honest and in the Fake conditions. Correlations with the general elevation of the values profile (SDR regardless of content) were fewer and weaker. Furthermore, increases in CM between the Honest and Fake conditions were correlated with increases in endorsement of communal values, although they were also related to increases in the general elevation of the values profile.

Assuming that variance in the CM scale is mostly due to bias, two different recommendations can be given on how to treat high scorers. In the moderator variable model, the test protocols of high scorers need to be discarded. In the suppressor variable model, the test protocols of high scorers need to be adjusted to take into account CM tendencies. No support for the latter, suppressor variable model of CM was found. However, the data of Study II supported somewhat the moderator variable model of CM. The results suggested that Communal values in the Fake condition predicted Communal values in the Honest condition, but only if CM was low. This means that, as far as

communal values are concerned, the test protocols of participants scoring high in CM in the Fake condition should be discarded. This result also supports Paulhus' (2002) interpretation of the CM scale as a response set measure of communal bias. However, it should also be noted that the interpretation of CM as a moderator variable did not receive any support in the ecologically more valid Studies I or IV.

## **4.2. Self-deceptive enhancement**

Paulhus (2002) has interpreted the SDE scale as, on the one hand, a measure of stable response style, and on the other hand, a content measure of agentic bias. In Studies I and II, individual differences in SDE scores persisted over applicant and incumbent conditions, as well as over Honest and Fake conditions. This result suggests that the SDE scale measures either stable response style, or some other stable individual differences in personality. The sample of spouses in Study I provided some evidence supporting the latter substantive interpretation. First, spouses partly corroborated the claims made by respondents in their responses to the SDE scale items. Second, self-rated SDE was associated with spouse ratings of low Neuroticism and high Extraversion. These results suggest that overestimation of one's competence and one's strivings to appear a "superhero," which is what the SDE scale was designed to measure (Paulhus, 2002), are, to some extent, transparent to others, or at least to spouses, and characterize a certain personality type. That SDE correlated significantly in the desirable direction for some personality traits and some personal values, even in conditions with low demand for SDR (incumbents and Honest conditions), could also be accounted for by a substantive interpretation of the scale

Some evidence for a response set interpretation of the SDE scale was also found. Studies I and II showed that SDE scores were consistently elevated in the applicant and Fake conditions as compared to the incumbent and Honest conditions. However, it seems likely that some degree of dissimulation is possible on all self-report personality measures, and it might therefore be more instructive to compare changes on the SDE scale with changes on the CM scale. In the groups of military cadets (Studies I and II) and business students instructed to make a good impression on their professor (Study II), the applicant context (Study I) and the faking instructions (Study II) affected CM scores more strongly than SDE scores. In contrast, in the sample of Finnish humanities students, instructions to fake influenced SDE scores more strongly than CM scores. Thus, both characteristics of the responders and the target of the good impression appear to determine which BIDR scale

is more strongly influenced by situational demands. Another piece of evidence in support of a response set interpretation is that the correlations between SDE scores and self-ratings of some personality traits (i.e., Neuroticism and Conscientiousness) attenuated or dissipated by nontrivial amounts when going from the applicant condition to the incumbent condition.

Concerning the content of the SDE scale, the results suggest that, besides measuring agentic bias, as was expected based on Paulhus' (2002) work, the SDE scale may also measure SDR regardless of content. In the applicant context of Study I, the SDE scale was correlated in the desirable direction with all Big Five personality traits. In Study II, increases in SDE across the Honest and Fake good conditions were, in all samples, correlated with increases in the general elevation of the values profile. In contrast, the correlations between increases in SDE and increases in agentic values were fewer and weaker. Interestingly, in situations with low demand for SDR, the SDE scale was associated with more desirable self-reports on agentic, but not communal, traits. In incumbent military cadets' ratings, SDE was correlated with low Neuroticism, high Extraversion, and high Conscientiousness. These correlations were replicated in the sample of spouses, where SDE was further determined to correlate with the Achievement, Self-Discipline, and Competence facets of Conscientiousness. Although SDE was also correlated with spouse-ratings on these traits, SDE was more strongly related to self-ratings of personality than to spouse ratings. Consistent with this observation, SDE correlated with overestimation, as compared to spouse ratings, of own Emotional Stability, Extraversion, and the Competence, Achievement, and Self-Direction facets of Conscientiousness.

The suppressor and moderator variable models of SDE were compared. No support was found for the suppressor model; that is, adjusting scores for SDE is not likely to improve validity of the scores. However, in Study II, some evidence for a moderator variable model of SDE was found. The results showed that Agentic values in the Fake condition predicted Agentic values in the Honest condition, but only if SDE was low. This means that the Agentic values factor of participants scoring high in SDE in the Fake condition could be discarded, having little or no useful true score variance. This result supports an interpretation of the SDE scale as a response set measure of agentic bias. However, no support for the moderator variable interpretation of SDE was found in Studies I or IV.

### **4.3. SDR and personality stability**

As argued in the introduction section of this dissertation, SDR could have several consequences for research on personality stability. Study I showed, as expected, some evidence that differential or rank-order stability may be inflated by SDR – controlling for SDE attenuated the test-retest correlations of Neuroticism and Conscientiousness. This means that SDE may contribute variance to the self-ratings at both points in time, thus artificially inflating the test-retest correlations. However, this result must be interpreted with caution, as the SDE scale appears to measure, to some extent, substantive individual differences. This means that controlling for SDE may be controlling for wanted rather than unwanted variance. Contrary to expectations, in Study III, no evidence was found for the attenuating effects of SDR on test-retest correlations over a 15-year time span. However, it should be noted that in that study the measure of SDR was the MMPI Lie Scale, and this may explain the unexpected result.

In Study III, the maturity-stability hypothesis was considered from the perspective of SDR. Over a 15-year time span, in which personality was measured at age 20 and again at age 35, desirable traits such as high Emotional Stability, Rule-Conscientiousness, and Perfectionism, and low Apprehension, Tension, and Vigilance, predicted greater ipsative stability. SDR was reasoned to possibly contribute reliable variance to personality scale scores that is not there for people who are low in desirability response tendencies. Therefore, those who consistently engage in SDR should yield relatively high ipsative stability (profile-profile) correlations compared to those who are more honest in their self-descriptions. Consistent with expectations, the MMPI Lie scale predicted greater stability over the 15-year time span.

Several aspects of the data in Study III contradicted a pure desirability interpretation of the results. Most important, several of the indicators of greater maturity consisted of other-ratings and objective indices, such as military rank and number of sick days. These measures also predicted greater stability, and are unlikely to have been interfered with by response bias. Another important aspect that contradicted a desirability interpretation was that this view presupposes that those who were identified as being less temporally stable had relatively low ipsative stability because they changed their stylistic response tendencies on the personality scales from age 20 to age 35. However, such changes in subjects' self-reports in the direction of desirability were found to be correlated with changes in psychiatrist's ratings of them in the direction of greater adjustment. Assuming the

psychiatrist was not totally deceived by the respondents' impression management tendencies, this finding is consistent with true personality changes in those participants over 15 years.

#### **4.4. SDR and self-selection**

Some of the results of Study I suggested that the CM and SDE scales of the BIDR may measure more than response bias. More specifically, these scales may to some extent also measure substantive individual difference in personality. This means that people scoring high on these scales, besides responding in a more desirable manner, may also behave differently than people scoring low on these scales. Based on their desire for social approval, they may, for instance, be more strongly inclined to perform socially desirable behaviours, such as volunteering for research. If people scoring high in CM or SDE are more likely to participate in research, then the personality scores of volunteers are prone to be more desirable than those of nonvolunteers, as some studies have suggested (Dollinger & Leong, 1993; Rosenthal & Rosnow, 1975).

In Study IV, the personality traits of respondents were compared with those of nonrespondents. As expected, some of the evidence suggested that observed personality differences between these groups, may, in part, stem from differences in SDR. In the first sample, respondents, who had more desirable Big Five personality traits than nonrespondents, also had marginally higher CM scores. This means that the differences in Big Five personality scores could be due to heightened SDR among respondents. To further evaluate this possibility, personality was evaluated with sibling ratings rather than self-ratings in the second sample. In that sample, Big Five personality differences between respondents and nonrespondents were also found. This means that respondents tend to differ from nonrespondents on the broad Big Five personality factors, and not merely in their need for social approval.

Studies I (but only Sample 3) and III of the present research relied on volunteers, and the results of Study IV may thus have some implications for the interpretation of the results of those studies. In our sample of spouses in Study I, only 35 of the contacted 200 couples volunteered to participate. The strong self-selection implies that those who volunteered were likely to have more desirable personality traits than those who did not volunteer. Such a restriction in the range of the personality and SDR variables may have attenuated the correlations we found between self- and spouse ratings, and also those between our



measures of SDR and discrepancies between self- and spouse ratings. In Study III, those participants who volunteered for the follow-up were likely to have more desirable personality profiles than those who did not. As there were no personality differences at age 20 between those who volunteered at age 35 and those who did not, the existence of such personality differences at age 35 would imply more variable trajectories of personality development, which would in turn mean more variance in estimates of personality stability. Thus, as in Study I, volunteer bias may have restricted the range of some variables, and thus attenuated correlations, such as those between the MMPI Lie scale and personality stability.

#### **4.5. Limitations**

Some limitations of the present research need to be acknowledged. One important limitation was that the experimental situations that were used to elicit socially desirable responding from some respondents were not properly justified theoretically. In particular, it may be that crude instructions, such as the instructions to fake good, cannot distinguish response style measures from response set measures. It seems possible that both types of measures, as well as all other personality measures, will react strongly to such instructions. In contrast, situations where a demand to appear honest is also present, such as the applicant context of Study I, may influence responses differently. In future research, more effort should be put into explicating those properties of the situation that influence response bias.

Some limitations concerning our participants and measures should also be noted. Concerning our participants, as discussed above, volunteer bias may have influenced the results of two of our studies. Furthermore, several of the samples consisted of male military cadets (Samples 1 and 2 in Study I, one of the samples in Study II, and Sample 1 in Study IV), and the one study that did not include military cadets was also conducted with an all male sample (Study III). This severely limits the generalizability of the results. However, it should be noted that sex differences in SDR have not generally been found, suggesting that the results pertaining to men may be generalizable to women.

Another methodological limitation concerns some of the measures that were used. As noted above, Study III did not include the BIDR as a measure of SDR, but rather the MMPI Lie scale, which does not distinguish between response set and response style. This distinction would have been relevant in the prediction of personality stability (only

response style would be expected to predict stability of self-reports of personality). Another measurement-related limitation was that some of the personality measures that were used were not standard measures and showed very low reliabilities (e.g., the different measures of trait Openness had reliabilities of .26 and .20, in Studies I (Sample 2) and IV (Sample 2)). The low reliabilities are likely to have attenuated any correlations.

Another important limitation of the present series of studies was that the sample sizes were sometimes quite small. Study II, conducted with the largest samples, was the only one to find moderator effects for the BIDR scales. The largest effect accounted for an increase of .11 in variance explained. The probabilities of correctly detecting an effect of such power in the Study I samples of military cadets ( $N = 62$ ) and spouses ( $N = 70$ ) were .51 and .65, respectively (with  $\alpha = .05$ ). Furthermore, as moderator effects typically are much smaller (e.g., Chaplin, 1991, suggests that they will seldom account for more than a .01 increase in variance explained), the small sample sizes seriously limit the ability of our studies to detect them.

## **4.6. Conclusions**

### **4.6.1. The BIDR scales**

The data of Studies I and II show that the BIDR scales, designed to be indicators of SDR, are not pure measures – both the CM and SDE scales are, at once, measures of response bias and measures of more substantive individual differences in behaviour. Although such an ambiguous conclusion can be considered unsatisfying, it is perhaps the only reasonable outcome. After all, the same situation can be thought to apply to more typical personality measures. That is, although some personality measures can be considered first and foremost indicators of some substantive personality traits, they too might only seldom be pure measures, being influenced in varying degrees by SDR bias or other response determinants.

The relative amount by which either the CM or the SDE scale was affected by bias versus substance is not possible to determine with the data of the present thesis. However, relevant to such an evaluation is the following observation. The most important source of evidence for a substantive interpretation of the CM and SDE desirability measures was based on significant positive correlations in Study I between self-ratings and spouse ratings on those scales. That result meant that the self-descriptions on the scales' items showed some accuracy relative to spouse-descriptions and, as such, could not be entirely distortions

or misrepresentations. But those correlations,  $r = .35$  for CM and  $r = .33$  for SDE, were much lower than the self-spouse correlations for the personality trait scales. For instance, Big Five factor measures yielded a mean self-spouse correlation of .52. It is thus clear that the substantive individual differences represented by the CM and SDE items are not rated as accurately as are the individual differences measured by the items of more conventional personality trait measures.

#### **4.6.2. Controlling for SDR**

Another clear conclusion of the present research was that no evidence in support of a suppressor model of SDR was found. Although SDR increased respondents' obtained scores on desirable personality traits in high demand as compared to low demand situations, there was substantial stability to the effect across respondents. Thus, the rank-ordering of the respondents was noticeably invariant across high demand and low demand assessment situations (e.g., the test-retests correlations of the Big Five averaged .61 across the incumbent and applicant contexts in Study I), and controlling for SDR did not improve the rank-ordering of respondents. The results of the present research thus suggest that there is no need to statistically control for SDR, as least as measured by the CM and SDE scales.

The results were more ambiguous concerning possible moderator effects. The results of Study I did not support the use of SDR scales as moderator variables. In contrast, some of the results of Study II, conducted with larger sample sizes, suggested that CM and SDE scores in a Fake good condition indicate the extent to which participants substitute their true answers with socially desirable answers, and that the test protocols of those scoring high in CM or SDE should be discarded. The proportions of variance that the CM and SDE scales could account for as moderator variables were reasonable, reaching as high as .11 and .05, respectively. Moderator effects of such size are larger than the moderator effects usually found in personality psychology (see Chaplin, 1991). However, if such measurement error is tolerable, then there is no need to consider the CM or SDE variables as moderator variables. It should also be noted that, whereas Study I used data collected in a real applicant context, Study II used data that was collected under instructions to fake. Thus, the results of Study I, which did not support the use of the BIDR scales as moderator variables, have higher ecological validity than the results of Study II.

### **4.6.3. The influence of SDR on personality research**

In Studies III and IV of the present research, the influence of SDR on two phenomena central to personality research was examined. The purpose of these studies was to evaluate the consequences SDR may have for different fields of research within personality psychology. More specifically, Study III investigated the possible influences of response bias on research on personality stability. Study IV investigated whether differences in SDR explain observed personality differences between people who volunteer for research and people who do not.

The results of Study I suggested that SDR may distort personality stability research in the sense that SDR may add reliable and common variance to personality questionnaires administered at two different points in time, thus artificially inflating the test-retest correlation of the questionnaire. However, no such effect was found in the 15-year follow up study (Study III). That study further showed that the maturity-stability hypothesis may, in part be influenced by SDR, but that it is at least not entirely an artifact of SDR.

Research volunteers generally have more desirable traits than do nonvolunteers. Study IV suggested that some of the observed personality differences between research volunteers and nonvolunteers may be due to heightened SDR in volunteers. However, the personality differences were by no means exclusively attributable to differences in SDR, as volunteers were also described in more desirable terms in siblings' ratings of the Big Five personality traits.

It seems likely that response bias does affect personality measurement. For instance, it is a common finding that applicant personality scores differ from incumbent personality scores, such as in Study I of the present research. Furthermore, personality questionnaires administered in an applicant, as compared to an incumbent context, have been found to show differential item functioning (Griffin, Hesketh, & Grayson, 2004) and to reveal a different personality structure (e.g., Schmit & Ryan, 1993). Even in contexts with little pressure to fake, perceptions of the social desirability of items are positively correlated with the likelihood of endorsing those items (Konstabel et al., 2006). An important contribution of Studies III and IV was that they illustrate the importance of relying on evidence other than self-reports to rule out the possible effects of SDR. Personality stability and self-selection are only two phenomena in the personality literature that SDR may influence. It is important that other phenomena within personality psychology also be studied with an eye on the possible effects of response distortion. For this, methods other than self-reports are necessary.

The effects of social desirability do not confine themselves to the field of personality psychology. For instance, in the field of social cognition research, information access and motivation are the two broad psychological processes thought to bring about asymmetries in which the self is favored over others. The motivational process of impression management influences, for instance, actor-observer asymmetries in explanations of behaviour, where the self is seen as more rational (Malle, Knobe, & Nelson, 2007) and less conforming than are others (Pronin, Berger, & Molouki, 2007).

#### **4.6.4. Future directions**

The present study can offer some suggestions for future directions for research on SDR. The purpose of a measure of SDR is to distinguish those individuals who have distorted their responses from those who have not. The present data suggest that neither the CM nor the SDE scale fulfill that purpose very well. This means that alternative ways of operationalizing SDR should be pursued. In the 1950s, Edwards first obtained social desirability ratings of personality questionnaire items, and then computed respondents' social desirability scores based on how they responded to those items. Konstabel et al. (2006) showed that controlling for SDR, as conceptualized with a similar method to the one proposed by Edwards, improved the agreement between self- and peer ratings. This method can thus be viewed as a promising prospect for future SDR research.

Another important issue concerns the indicators of dishonest responding. For instance, in the present research, discrepancy between self- and spouse ratings was thought to indicate SDR. Consistent with our expectations, discrepancies on some personality traits were correlated with our measures of SDR. However, these correlations cannot tell whether the spouse is underestimating the target, or whether the target is overestimating him or herself. To resolve this issue, and more generally the issue of how to detect dishonesty in responses, more objective measures (e.g. behavioural measures) of traits are needed. However, this might also raise other problems, as specific instances of behaviour are psychometrically treacherous (e.g. low reliabilities; Epstein, 1979), and their relevance for traits may be questionable (e.g., Goldberg, 1972). Paulhus and colleagues have suggested that at least exaggeration of intelligence can be measured with the over-claiming technique, which uses respondents' ratings of their knowledge of various events, persons etc. to measure self-enhancement (Paulhus, Harms, Bruce, & Lysy, 2003). Because some of the items do not exist, responses can be analyzed with signal detection formulas to index response bias. This method thus gives the researcher an objective indicator of external

reality against which to measure respondents' departure from reality. Similar measures for domains other than general knowledge would constitute an important advance in SDR research.

Kwan and colleagues have recently reviewed studies concerned with self-enhancement, with a focus on the definition and measurement of self-enhancement (Kwan, John, Kenny, Bond, & Robins, 2004). They observed that about half the reviewed articles used a definition originating in Festinger's (1954) social comparison theory, whereas the other half used a definition originating in Allport's (1937) notion of self-insight. The former articles derived an index of self-enhancement from the discrepancy between self-perceptions and the way the individual perceives others. The discrepancy, of course, is generally in the direction that puts the self in a positive light. This better-than-average effect is a well-established finding in social psychology (e.g., Alicke, 1985; Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995). The response bias measured by the BIDR scales – discrepancy between the individual's self-ratings and the way the individual is rated by knowledgeable and unbiased others – has its roots in the definition of self-enhancement originating in Allport's (1937) notion of self-insight. For instance, Study I used spouse ratings as criteria for actual trait level. Future work on SDR could benefit from also considering response bias in the context offered by work on the better-than-average effect.

In sum, the work presented in this thesis raises, perhaps, more questions than it answers, as is true of much of the previous research on SDR. Clear-cut conclusions are difficult to reach, as the data were neither consistent with the view that research on SDR is like beating a dead horse and should be dropped (Piedmont et al., 2000), nor with the view that SDR needs to be constantly monitored and controlled for in some way (Butcher & Rouse, 1996). There is no question, however, about the need to continue with research in this important area of personality assessment.

## 5 ACKNOWLEDGEMENTS

My greatest debt is to my excellent supervisor, Dr. Markku Verkasalo. This dissertation is an attempt to publish some of the ideas that he has developed and data that he has collected for the past fifteen years. Thank you Markku for giving me this possibility, without you this thesis would not have been written. As an instructor, although you have always treated me with the utmost respect, as you would treat a colleague, you have at the same time managed to pave my way throughout the process. This combination of freedom and protection has made my work easy and highly enjoyable.

Without the involvement of my other supervisor, Professor Sampo Paunonen, this thesis might have been started upon, but it would have remained in a rather awkward and primitive state. Thank you Sampo for the extreme levels of patience, encouragement, persistence, and friendliness that you have shown towards me when teaching me how to write academic papers. No-one could have been a better teacher.

The third defining person in the history of this thesis was Professor Klaus Helkama, who employed me for three years with his Academy of Finland research grant. Thank you Klaus for giving me this opportunity, for organizing the intellectually stimulating biweekly values seminar, and, most importantly, for setting an example of how enthusiastic one should be about one's work.

I am also grateful for the positive reviews of the two pre-examiners of this dissertation, Professor Katariina Salmela-Aro and Professor Jüri Allik.

Two of the original articles presented in this thesis present, in part, data I collected when employed at the National Health Institute. Thank you Dr. Annamari Tuulio-Henriksson and Professor Jouko Lönnqvist for employing me, for providing a nice working atmosphere, and for giving me the freedom to pursue my personal research interests besides my actual duties.

One of the original articles presented in this thesis presents data that was laboriously collected by Dr. Seppo Mäkinen. Thank you Seppo for allowing me to use your data for my thesis. I also want to thank my coauthors Dr. Irina Bezmenova and Dr. Markus Henriksson for valuable comments on the original papers, and Jari Lipsanen and Pekka Lahti-Nuutila for providing statistical expertise. I also express my gratitude to the Emil Aaltonen foundation for funding the last year of my work, and to all the hundreds of people who volunteered to take part in the studies.

The Finnish Defence Forces must also be thanked. One of the original articles present data that was collected at the National Defence College in collaboration with Dr. Vesa Nissinen, and I thank Professor Juhani Sinivuo and Kai Nyman from the Finnish Defence Forces Educational Development Center for providing this opportunity.

The work was primarily carried out at the Department of Psychology at the University of Helsinki. I want to thank the department for providing working facilities. For company and friendship in various social/academic activities related to my studies, thank you Markku Kilpeläinen, Kia Aarnio, Viljami Salmela, Annika Svedholm, Mia Silfver, Ann Ojala, Miika Kabata, Oskari Salmi, and Gari Walkowitz.

My mother and father have always supported me wonderfully – kiitos / tack! Finally, I owe my deepest gratitude to my wife, Sointu Leikas, who besides being a co-author in one of the original papers, has offered love, encouragement, and numerous ideas and feedback at each step of my studies. Thank you Sointu. To paraphrase Neil Young (1975), “In your eyes I can discover / another reason why / I want to live / and make the best / of what I see”.



## 6 REFERENCES

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, *49*, 1621-1630.
- Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology*, *68*, 804-825.
- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Allport, G. W. (1961). *Pattern and growth in personality*. New York: Holt, Rinehart & Winston, Inc.
- Allport, G. W. & Odbert, H. S. (1970) [1936]. Trait names: A psycho-lexical study. *Psychological monographs*, *47*. Princeton, N.J.: Psychological Review Company.
- Asendorpf, J. B., & Van Aken, M. A. (1991). Correlates of the temporal consistency of personality patterns in childhood. *Journal of Personality*, *59*, 689-703.
- Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Chicago: Rand McNally.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.
- Block, J. (1995). A contrarian view of the Five-Factor approach to personality description. *Psychological Bulletin*, *117*, 187-215.
- Block, J., & Block, J. H. (2006). Venturing a 30-year longitudinal study. *American Psychologist*, *61*, 315-327.
- Boie, K., Lee, K., Ashton, M. C., Pascal, S., & Nicol, A. A. M. (2001). The structure of the French personality lexicon. *European Journal of Personality*, *15*, 277-295.
- Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality*, *6*, 199-214.
- Butcher, J. N., & Rouse, S. V. (1996). Personality: Individual differences and clinical assessment. *Annual Review of Psychology*, *47*, 87-111.
- Cattell, R. B. (1943). The description of personality: Foundations of trait measurement. *Psychological Review*, *50*, 559-594.

- Cattell, R. B. (1981). *Sixteen Personality Factor Questionnaire (16PF)*. Helsinki: Psykologien Kustannus OY.
- Cattell, R. B., Eber, H., & Tatsuoka, M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, Illinois: Institute for Personality and Ability Testing, Inc.
- Chaplin, W. F. (1991). The next generation of moderator research in personality psychology. *Journal of Personality, 59*, 143-178.
- Clausen, J. S. (1991). Adolescent competence and the shaping of the life course. *American Journal of Sociology, 96*, 805-842.
- Cohen, J. (1969).  $r_c$ : A profile similarity coefficient invariant over variable reflection. *Psychological Bulletin, 71*, 281-284.
- Comrey, A. L., & Backer, T. E. (1975). Detection of faking on the Comrey Personality Scales. *Multivariate Behavioural Research, 10*, 311-319.
- Conger, A., & Jackson, D. N. (1972). Suppressor variables, prediction, and the interpretation of psychological relationships. *Educational and Psychological Measurement, 32*, 579-599.
- Costa, P. T. Jr., & McCrae, R. R. (1988). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology, 55*, 258-265.
- Costa, P. T. Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.
- Davison, M. L. (1985). Multidimensional scaling versus components analysis of test intercorrelations. *Psychological Bulletin, 97*, 94-105.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behaviour. *Psychological Inquiry, 11*, 227-268.
- DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology, 91*, 1138-1151.
- Digman, J. M. (1997). Higher order factors of the Big Five. *Journal of Personality and Social Psychology, 73*, 1246-1256.
- Dollinger, S. J., & Leong, F. T. (1993). Volunteer bias and the five-factor model. *Journal of Psychology: Interdisciplinary and Applied, 127*, 29-36.

- Donnellan, M. B., Conger, R. D., & Burzette, R. G. (2007). Personality development from late adolescence to young adulthood: differential stability, normative maturity, and evidence for the maturity-stability hypothesis. *Journal of Personality, 75*, 237-263.
- Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996). The validity of non-cognitive measures decays when applicants fake. In J. B. Keyes & L. N. Dosier (Eds.), *Proceedings of the Academy of Management* (pp. 127-131). Madison, WI: Omnipress.
- Dunlop, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*, 170-177.
- 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*. New York: Dryden.
- Epstein, S. (1979). The stability of behaviour: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37*, 1097-1126.
- Feather, N. T. (2002). Values and value dilemmas in relation to judgments concerning outcomes of an industrial conflict. *Personality and Social Psychology Bulletin, 28*, 446-459.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140.
- Finkel, E. J., Burnette, J. L., & Scissors, L. E. (2007). Vengefully ever after: Destiny beliefs, state attachment anxiety, and forgiveness. *Journal of Personality and Social Psychology, 92*, 871-886.
- Goldberg, L. R. (1972). Some recent trends in personality assessment. *Journal of Personality Assessment, 36*, 547-560.
- Goldberg, L. R. (1982). From Ace to Zombie: Some explorations in the language of personality. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessment Vol. 1* (pp. 203-234). Hillsdale, NJ: Erlbaum.
- Goldberg, L. R. (1990). An alternative "Description of Personality": The big-five factor structure. *Journal of Personality and Social Psychology, 59*, 1216-1229.

- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26-42.
- Griffin, B., Hesketh, B., & Grayson, D. (2004). Applicants faking good: Evidence of item bias in the NEO PI-R. *Personality and Individual Differences, 36*, 1545-1558.
- Hathaway, S. R., & McKinley, J. C. (1946). *The Minnesota Multiphasic Personality Inventory*. Minneapolis: University of Minnesota Press.
- Hogan, R. (1983). A socioanalytic theory of personality. In M. M. Page (Ed.), *Nebraska symposium on motivation* (pp. 336–355). Lincoln: University of Nebraska Press.
- Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin, 55*, 243-252.
- John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66-100). New York: Guilford.
- Kluckhohn, C. (1951). Values and value-orientations in the theory of action: An exploration in definition and classification. In T. Parsons & E. Shils (Eds.), *Toward a general theory of action*. Cambridge, MA: Harvard University Press.
- Konstabel, K., Aavik, T., & Allik, J. (2006). Social desirability and consensual validity of personality traits. *European Journal of Personality, 20*, 549-566.
- Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review, 111*, 94-110.
- Lalwani, A. K., Shavitt, S., Johnson, T. (2006). What is the relation between cultural orientation and socially desirable responding? *Journal of Personality and Social Psychology, 90*, 165-178.
- Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*, 34–47.
- Lindeman, M., & Verkasalo, M. (1995a). Sosiaalisesti suotuisa vastaaminen. [Socially desirable responding]. In K. Räikkönen & J.-E. Nurmi (Eds.), *Persoonallisuus, terveys ja hyvinvointi* [Personality, health, and well-being] (pp. 148-162). Helsinki: Suomen psykologinen seura.
- Lindeman, M., & Verkasalo, M. (1995b). Personality, Situation, and positive-negative asymmetry in socially desirable responding. *European Journal of Personality, 9*, 125-134.

- Mäkinen, S. (1991). Isänsä menettäneet pojat varusmiehinä [Fatherless young men as conscripts]. *Annales Medicinae Militari Fenniae, 1a*.
- Malle, B. F., Knobe, J. M., & Nelson, S. E. (2007). Actor-observer asymmetries in explanations of behaviour: New answers to an old question. *Journal of Personality and Social Psychology, 93*, 491-514.
- McAdams, D. P., & Pals, J. L. (2006). A new Big Five : Fundamental principles for an integrative science of personality. *American Psychologist, 61*, 204-217.
- McCrae, R. R., & Costa, P. T. Jr. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882-888.
- Murray, S. L., Holmes, J. G., Dolderman, D., & Griffin, D. W. (2000). What the motivated mind sees: Comparing friends' perspectives to married partners' views of each other. *Journal of Experimental Social Psychology, 36*, 600-620.
- Mustonen, S. (1992). *Survo, an integrated environment for statistical computing and related areas*. Helsinki: Survo Systems Ltd.
- Nyman, K. (2007). *Varusmiesten johtajavalintojen luotettavuus* [The reliability of conscript leader selection.] (Julkaisusarja, 1). Helsinki: Maanpuolustuskorkeakoulu, Käyttäytymistieteiden laitos.
- 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. (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). San Diego, California: Academic Press
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah, New Jersey: Lawrence Erlbaum.
- Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21*, 100-108.

- Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, *84*, 890-904.
- Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality. Special Issue: Defence Mechanisms in Contemporary Personality Research*, *66*, 1025-1060.
- Paulhus, D. L., Robins, R. W., Trzesniewski, K. H., & Tracy, J. L. (2004). Two replicable suppressor situations in personality research. *Multivariate Behavioural Research*, *39*, 3030-328.
- Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, *78*, 582-593.
- Pronin, E., Berger, J., & Molouki, S. (2007). Alone in a crowd of sheep: Asymmetric perceptions of conformity and their roots in an introspection illusion. *Journal of Personality and Social Psychology*, *92*, 585-595.
- Rauste-von Wright, M., & von Wright, J. (1984). On person perception and the perceiver's self-concept. *Jyväskylä Studies in Education, Psychology and Social Research. Special issue: Human action and personality: Essays in honour of Martti Takala*, *54*, 107-119.
- Roberts, B. W., Caspi, A., & Moffitt, T. E. (2001). The kids are alright: Growth and stability in personality development from adolescence to adulthood. *Journal of Personality & Social Psychology*, *81*, 670-683.
- Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality*, *69*, 617-640.
- Rohan, M. J. (2000). A rose by any name? The values construct. *Personality and Social Psychology Review*, *4*, 255-277.
- Rokeach, M. (1973). *The nature of human values*. New York: Free Press.
- Rosenthal, R., & Rosnow, R. L. (1975). *The volunteer subject*. Oxford, England: John Wiley & Sons.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*, 634-644.

- Saucier, G. (2003). Factor structure of English-language personality type-nouns. *Journal of Personality and Social Psychology*, 85, 695–708.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. Zanna, (Ed.), *Advances in experimental social psychology*, 25, 1-65. New York: Academic Press.
- Schwartz, S. H. (2005a). Basic human values: Their content and structure across countries. In A. Tamayo & J. B. Porto (Eds.), *Valores e Comportamento nas Organizações* [Values and Behaviour in Organizations] pp. 21-55. Petrópolis, Brazil: Vozes.
- Schwartz, S. H. (2005b). Robustness and fruitfulness of a theory of universals in individual human values. In A. Tamayo & J. Porto (Eds.), *Valores e trabalho* [Values and work] (pp. 56-95). Brasilia: Editora Universidade de Brasilia.
- Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology*, 53, 550-562.
- Schwartz, S. H., Melech G., Lehmann, A., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology*, 32, 519-542.
- Schwartz, S. H., & Sagiv, L. (1995). Identifying culture specifics in the content and structure of values. *Journal of Cross-Cultural Psychology*, 26, 92-116.
- Schwartz, S. H., Verkasalo, M., Antonovsky, A., & Sagiv, L. (1997). Value priorities and social desirability: Much substance, some style. *British Journal of Social Psychology*, 36, 3-18.
- Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, 78, 966-974.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, 80, 607-620.
- Starzyk, K. B., Holden, R. R., & Fabrigar, L. R. (2006). The personal acquaintance measure: A tool for appraising one's acquaintance with any person. *Journal of Personality and Social Psychology*, 90, 833-847.
- Tarkkonen, L., Vehkalahti, K. (2005). Measurement errors in multivariate measurement scales. *Journal of Multivariate Analysis*, 96, 172-189.

- Tellegen, A. & Waller, N. G. (1992) *Exploring personality through test construction: Development of the Multi-dimensional Personality Questionnaire (MPQ)*. Unpublished manuscript. Department of Psychology, University of Minnesota.
- Tupes, E. C. & Christal, R. E. (1992) [1961]. Recurrent personality factors based on trait ratings. *Journal of Personality*, 60, 225-251.
- Verkasalo, M., & Lindeman, M. (1994). Personal ideals and socially desirable responding. *European Journal of Personality*, 8, 385-393.
- Young, N. (1975). Looking for a love. On *Zuma* [vinyl]. London, UK: Reprise.



## APPENDIX A

BIDR Version 6 – Form 40

Using the scale below as a guide, write a number beside each statement to indicate how much you agree with it.

1-----2-----3-----4-----5-----6-----7  
NOT TRUE                      SOMEWHAT TRUE                      VERY TRUE

- \_\_\_ 1. My first impressions of people usually turn out to be right.
- \_\_\_ 2. It would be hard for me to break any of my bad habits. (R)
- \_\_\_ 3. I don't care to know what other people really think of me.
- \_\_\_ 4. I have not always been honest with myself. (R)
- \_\_\_ 5. I always know why I like things.
- \_\_\_ 6. When my emotions are aroused, it biases my thinking. (R)
- \_\_\_ 7. Once I've made up my mind, other people can seldom change my opinion.
- \_\_\_ 8. I am not a safe driver when I exceed the speed limit. (R)
- \_\_\_ 9. I am fully in control of my own fate.
- \_\_\_ 10. It's hard for me to shut off a disturbing thought. (R)
- \_\_\_ 11. I never regret my decisions.
- \_\_\_ 12. I sometimes lose out on things because I can't make up my mind soon enough. (R)
- \_\_\_ 13. The reason I vote is because my vote can make a difference.
- \_\_\_ 14. My parents were not always fair when they punished me. (R)
- \_\_\_ 15. I am a completely rational person.
- \_\_\_ 16. I rarely appreciate criticism. (R)
- \_\_\_ 17. I am very confident of my judgments.
- \_\_\_ 18. I have sometimes doubted my ability as a lover. (R)
- \_\_\_ 19. It's all right with me if some people happen to dislike me.
- \_\_\_ 20. I don't always know the reasons why I do the things I do. (R)
- \_\_\_ 21. I sometimes tell lies if I have to. (R)
- \_\_\_ 22. I never cover up my mistakes.
- \_\_\_ 23. There have been occasions when I have taken advantage of someone. (R)
- \_\_\_ 24. I never swear.

- \_\_\_ 25. I sometimes try to get even rather than forgive and forget. (R)
- \_\_\_ 26. I always obey laws, even if I'm unlikely to get caught.
- \_\_\_ 27. I have said something bad about a friend behind his or her back. (R)
- \_\_\_ 28. When I hear people talking privately, I avoid listening.
- \_\_\_ 29. I have received too much change from a salesperson without telling him or her. (R)
- \_\_\_ 30. I always declare everything at customs.
- \_\_\_ 31. When I was young I sometimes stole things. (R)
- \_\_\_ 32. I have never dropped litter on the street.
- \_\_\_ 33. I sometimes drive faster than the speed limit. (R)
- \_\_\_ 34. I never read sexy books or magazines.
- \_\_\_ 35. I have done things that I don't tell other people about. (R)
- \_\_\_ 36. I never take things that don't belong to me.
- \_\_\_ 37. I have taken sick-leave from work or school even though I wasn't really sick. (R)
- \_\_\_ 38. I have never damaged a library book or store merchandise without reporting it.
- \_\_\_ 39. I have some pretty awful habits. (R)
- \_\_\_ 40. I don't gossip about other people's business.

*Note.* Items 1 to 20 constitute the SDE scale, and items 21 to 40 the CM scale. Items marked with an (R) are reverse scored.