

SONDERFORSCHUNGSBEREICH 504

Rationalitätskonzepte,
Entscheidungsverhalten und
ökonomische Modellierung

No. 06-03

**Measures of Desirability Beliefs and their Validity as
Indicators for Socially Desirable Responding**

Volker Stocké*
and Christian Hunkler**

April 2006

We gratefully acknowledge stimulating discussions with Hartmut Esser. Financial support from the Deutsche Forschungsgemeinschaft, SFB 504, at the University of Mannheim, is gratefully acknowledged.

*Sonderforschungsbereich 504, email: vstocke@rumms.uni-mannheim.de

**Sonderforschungsbereich 504, email:



Universität Mannheim
L 13,15
68131 Mannheim

Abstract

Social desirability (SD)-bias is a serious threat for survey-data quality, and the respondents' desirability beliefs have proven in many studies to predict the incentives for the strength as well as for the direction of this bias. However, the issue of the relative validity of different ways to measure these incentives has hardly received any attention. We introduced three such measures and discussed the respective tradeoffs between their parsimony on the one hand and the implied assumptions which have to be fulfilled on the other. In the empirical part of our paper, we tested with four questionnaire topics whether and how strong these assumptions are violated and thus how much measurement effort is necessary to obtain a sufficiently good indicator for the perceived incentives for SD-bias. These topics were: (a) having prejudice against the elderly, (b) making charitable donations, (c) participating in political elections and (d) being environmentally conscious. We found the most parsimonious one-point measure to be strongly affected by measurement error in the case of all questionnaire topics. Depending on how strong the assumption of monotony was violated, for some topics the medium elaborated simple difference scores were sufficiently valid, but partly only the most elaborated domain-specific difference scores were a valid predictor for the potential strength and direction of incentives for SD-bias.

1. Introduction

Social desirability (SD)-bias is a serious threat for the validity of survey data. *Firstly*, empirical evidence suggests that the response distributions for a variety of traits, as for example the participation in political elections, the consumption of alcohol or racial attitudes, are biased by social desirability (Belli, Traugott & Beckmann, 2001; Embree & Whitehead, 1993; Krysan, 1998). *Secondly*, and more serious for the aim of survey research is that SD-bias can create spurious or suppress real associations (Ganster, Hennessey & Luthans, 1983). This results when groups of respondents differ in the strength or even direction of SD-bias, and the resulting differences in survey reports artificially create or mask real differentiation between these groups. For instance, men have been found to overreport and women to underreport the number of sex partners, which falsely suggests a correlation between gender and promiscuity (Tourangeau & Smith, 1996). In order to statistically control for such distortions, it is a precondition to measure the strength and direction of SD-bias.

Depending on the theoretical framework, different factors are assumed to predict SD-bias. In Rational-Choice Theory, respondents are assumed to employ impression-management strategies in order to gain social approval from others. Psychological approaches, in contrast, predict SD-bias to result from the aim of ‘self-deception’, where the own person is the addressee of response bias (Paulhus, 2003). Regardless of which kind of SD-bias is under consideration, the respondents’ social desirability beliefs are a key determinant for predicting the strength and direction of this bias. These beliefs, in the literature often referred to as ‘trait desirability’, represent the subjects’ expectations about how different answers are evaluated and thus are suitable for creating a favorable impression in others or themselves. Social norms about which characteristics are desirable in society are an important source of these beliefs, and thus, normative differentiation between respondents causes heterogeneity in the strength and direction of SD-bias (Tourangeau & Smith, 1996).

With few exceptions, individual differences in the perceived incentives for SD-bias have been assessed by collecting desirability ratings of that response option which expresses that a respondent has the characteristic under consideration. This *one-point measure* (OPM) implicitly assumes not having the trait to be evaluated neutrally. As we will argue later, it is an important precondition for the validity of the obtained measure that this assumption – hereafter referred to as ‘*neutrality assumption*’ – is empirically justified. We will show that violations of this assumption require obtaining SD-beliefs for the presence *and* absence of the respective trait. Using *simple difference scores* (SDS) between both ratings is in the case of continuous traits, as for instance the consumption of alcohol, only a *necessary* but *not sufficient*

precondition for an appropriate measure of possible SD-bias. This is the fact because the OPM and SDS assume the desirability to increase or decrease monotonically between the extreme endpoints of the trait-intensity continuum. Violations of this assumption – hereafter called ‘*monotony assumption*’ – always lead to an underestimation of the actual incentives for SD-bias. When this second assumption is violated, for example when moderate drinking is perceived to be more desirable than both abstinence and alcoholism, the valid measurement of incentives for SD-bias additionally requires the desirability rating for a medium trait intensity.

The aim of this paper is *firstly* to introduce the possible measures for respondents’ perceived incentives for SD-bias, to make their underlying assumptions explicit and, in particular, to show the consequences when these assumptions are violated. The empirical part analyzes *secondly* for four continuous traits (a) whether the two assumptions hold and (b) to what degree violations lead to biased predictions about the strength and direction of SD-bias. The analyzed traits are: being prejudiced against the elderly, making charitable donations, voting in political elections and being environmentally conscious.

2. Role of Desirability Beliefs in Predicting Social Desirability Bias

In the conceptual model of answering survey questions proposed by Tourangeau and Rasinski (1988), subjects are assumed to select an answer in the last of four stages, after the comprehension of the question, the retrieval of relevant information from memory and the judgment of the response options. In the fourth selection stage, respondents may edit their answers with respect to social desirability, either due to their desire for *impression management* or because of being susceptible to *self-deception* (Paulhus, 2003). In both cases, the respondents’ SD-beliefs determine how strong and in the direction of which response option incentives for SD-bias are perceived. These incentives, however, only lead to SD-bias when the respondents’ true characteristics differ from those which are regarded as socially desirable. Thus, the response bias predicted on the basis of SD-beliefs is an upper limit of what has to be expected in survey contexts.

The rational-choice theory (RCT) of survey-response behavior is the probably most elaborated approach for explaining *impression management*-based SD-bias (Esser, 1991; Stocké, 2004; Tourangeau, Rips & Rasinski, 2000: 281). Here, survey answers are the result of an instrumentally rational selection between response options. Respondents strive to receive positive and avoid negative evaluation reactions from others in order to maximize their feeling of social approval. Subjects utilize their SD-beliefs in order to form expectations about

how present others will react when one of the available response options is selected (Esser, 1991). In the case of an unknown audience, for instance a survey interviewer, desirability beliefs cannot be based on individualized knowledge about the actually employed evaluation criteria. Then, subjects rely on social norms as a best guess of what evaluation a certain response will provoke (Stocké, 2004). Respondents are expected to deviate increasingly more from their ‘true’ answer when this answer is believed to be less desirable than an alternative response. Whether the less desirable option is negatively evaluated or only less positively than the more desirable alternative, is assumed not to matter for the resulting incentives.

In the case of SD-bias driven by ‘*self-deception*’, others being able to perceive and sanction the response behavior are not assumed to matter for whether a difference between the true score and the most desirable answer leads to SD-bias (Paulhus, 2003). Here, the respondents’ aim is to preserve a positive self-image and thus they themselves are the addressees for socially desirable responding. The SD-beliefs are, however, also the basis for realizing this aim.

3. Measures of Desirability Beliefs and Their Underlying Assumptions

SD-beliefs have been proposed to be either operationalized as *personally or extrinsically expected trait desirability*. In the first case, a sample of respondents evaluate how desirable they personally regard a particular characteristic, and aggregated scores of these judgments are assumed to represent the incentives for SD-bias (Meleddu & Guicciardi, 1998). This is a valid indicator when (a) the underlying personal preferences represent the normative climate in the surveyed population and when (b) the sample of informants is representative for those respondents for which the aim is to predict SD-bias. At least the first precondition has been found not to be given empirically (Crott & Baltes, 1973). In contrast, the concept of extrinsically expected trait desirability directly assesses how the respondents perceive the trait under consideration to be evaluated in society. This measurement concept has been applied in the majority of empirical studies (e.g. Nicotera, 1996).

Incentives for SD-bias based on the extrinsically expected trait desirability can be operationalized in three differently elaborate ways, where, except for the most elaborated one, certain assumptions are made. The standard method – the *one-point measure (OPM)* – was introduced by Edwards (1957) and is the most parsimonious way to measure incentives for SD-bias (e.g. Phillips & Clancy, 1972). Respondents are asked to rate the social desirability of a particular trait (e.g. having voted in a political election), or in the case of continuous characteristics, that of a high trait intensity (e.g. having strong prejudices against the elderly). Usu-

ally bipolar rating scales with a neutral midpoint are used in order to collect the judgments. Negative deviations from the neutral scale midpoint are interpreted as incentives to deny and positive deviations as incentives to ascribe the trait to the own person. The OPM is a valid indicator for the strength and direction of incentives for SD-bias as long the absence of the respective trait is perceived as neutrally evaluated. This *neutrality assumption* and the consequences of its violation are illustrated in figure 1 for the SD-beliefs of four prototypical respondents. In the case of respondent A, the neutrality assumption holds: Whereas the trait ‘often donating blood’ is perceived with a desirability value of +2 to be positively evaluated, the trait ‘never donating blood’ receives a neutral evaluation of zero. Respondent B, in contrast, violates the neutrality assumption: Using an OPM gives a value of +2 and one would predict the same incentives for SD-bias as for respondent A. However, since never donating blood is evaluated mildly negative with -1, this respondent actually perceives stronger incentives for overreporting blood donation than A. Whereas the OPM underestimates the possible SD-bias by one third in the case of respondent B, it results in an overestimation when used for respondent C. Here, a frequent blood donation receives an evaluation of +2, and thus one would assume the same motivation for overreporting donation as in the cases of respondents A and B. Since, however, not donating blood has received a desirability score of +2 as well, actually no incentives for SD-bias are present. Respondent D, as all others, evaluates frequent blood donations with a score of +2. However, since not donating blood is judged to be even more positive, this respondent has actually to be expected to underreport blood donations. These examples show that violations of the neutrality assumption can cause under- and overestimations of SD-bias, as well as a prediction of the wrong direction of this bias.

-- Figure 1 here --

To account for possible violations of the neutrality assumption, *simple difference scores* (SDS) between the judged desirability of the extreme endpoints of the trait-intensity continuum can be used. Here, the neutrality assumption is given up and the incentives for SD-bias are operationalized as how different a high and low trait intensity is evaluated. As in the case of the OPM, the sign of SDS indicates the direction, and the absolute value the strength of the predicted incentives for SD-bias. SDS have rarely been used in applied research (as exception cf. Stocké, 2004, 2005a, 2005b). However, SDS as well as OPM imply a further assumption: They are only valid as long as the perceived desirability increases or decreases monotonically over the trait-intensity continuum. Figure 2 illustrates the consequences of violations of this *monotony assumption*. The prototypical respondent A assumes that donating blood more often is always judged to be more desirable in society than doing this less often. In this exam-

ple, the SDS is +4 scale points and thus predicts incentives for overreporting blood donations. Respondent B, in fact, perceives incentives of the same strength, however here, for reporting a medium frequency of this behavior: The medium frequency is regarded four scale points more desirable than no blood donations and four points more desirable than very frequent donations. However, SDS (like the OPM) predicts the absence of any incentives for SD-bias, since the extreme ends of the frequency continuum are perceived to be equally (un)desirable. In the case of U-shaped or reversed U-shaped desirability profiles, the OPM as well as SDS always lead to an underestimation of SD-bias and only *domain-specific difference scores* (DSDS) are appropriate. This most elaborate measure utilizes the average desirability differences between a low and a medium as well as between a medium and a high trait intensity as an indicator for the total incentives for SD-bias.

-- Figure 2 here --

In sum, the two more parsimonious operationalizations for incentives for SD-bias - the OPM and the SDS - lead to the same predictions as the DSDS and will thus have perfectly convergent validity when the neutrality *and* monotony assumptions hold. When the neutrality assumption is violated, the OPM leads to either an over- or underestimation or a prediction of the wrong direction of incentives for SD-bias, whereas violations of monotony lead to all kind of errors in the case of OPM, while SDS always underestimate the possible SD-bias. It is thus an important question whether and for which questionnaire topics the different violations occur and to what extent they bias the incentive measures.

4. Previous Research

4.1 Predictive Power of SD-Beliefs for Social Desirability Bias

Respondents' desirability beliefs have in many studies been found to be substantially associated with their response behavior. Accordingly, individual differences in how desirable subjects perceived traits were positively correlated with the endorsement of the respective response options for e.g. cheating in academic contexts (Fernandes & Randall, 1992), having the personality trait of argumentativeness (Nicotera, 1996) or symptoms of mental illness (Huang, Liao & Chang, 1998). Another study found the association between endorsement rates and desirability ratings for 176 personality items to be invariant for an American and Turkish college sample (Smith, Smith & Seymour, 1993). Furthermore, SD-ratings of items expressing positive affectivity and responses hereon were stronger related when the respondents had a stronger need for social approval (Chen et al., 1997; cf. Phillips & Clancy, 1972 for conflicting evidence). In all of these studies, the OPM has been utilized.

These correlations are not necessarily indicative for SD-bias and for the explanatory power of SD-beliefs: (a) Asking in the same interview about SD-beliefs and the responses on the same topic may provoke SD-bias which would otherwise not be present (Chen et al., 1997), and (b) the associations may simply result because the respondents truly have socially desirable characteristics (Johnson, 2004). Huang and colleagues (1998) addressed the first objection by recording SD-ratings about 288 items for testing mental health, and the subjects answered these items in a later session. Despite the time between both answers, less symptoms of mental illness were reported when they were judged earlier to be less desirable. The second objection was tested by collecting self-reports and SD-beliefs about the number of police arrests (Wyner, 1980). The subjects' SD-beliefs explained how strong and in which direction the answers differed from the true number of arrests, which were validated with police records later on. Furthermore, SDS for 10 items measuring racial attitudes were found to predict how strong and in which direction responses on the same items differed between experimentally induced privacy differences (Stocké, 2004). Under interviewer- rather than self-administration, subjects answered significantly more in agreement with their SD-beliefs.

4.2 Previous Research about SD-Bias in the Analyzed Topics

We analyzed the SD-beliefs about four traits, frequently measured in survey research: (a) the strength of prejudice toward the elderly, (b) the frequency of charitable donations, (c) the regularity of electoral participation and (d) the degree of environmental consciousness. The following is known about the presence and determinants of SD-bias for these topics.

- *Prejudices toward the elderly*: Several survey studies have assessed the respondents' prejudice toward elderly people with self-report measures.¹ However, no research has been done about whether, under which conditions and for which respondent groups SD-bias affects survey measures about this trait. Yet, several studies found evidence for SD-bias in reports about negative attitudes toward other groups, as for instance African Americans or foreigners. Here, subjects reported less prejudice in interviewer- rather than self-administered interviews (Krysan, 1998) and when subjects believed racial prejudice to be increasingly more socially undesirable (Phillips & Clancy, 1972). In particular, it has been found that the SD-beliefs about positive and negative racial attitudes varied considerably between groups of respondents, established incentives for SD-bias in the direction of both kinds of attitude answers and the respondents' answers followed these incentives (Stocké, 2004).

¹ In the Eurobarometer 37.1, respondents reported their agreement with items expressing stereotypes about the elderly, such as: 'Elderly people are not willing to listen to younger people's views' (ICPSR dataset 9957).

- *Charitable donations*: Questions about whether and to what extent respondents actually do or would donate money to certain kinds of charitable organizations are also frequently asked in surveys.² While some authors suspected that respondents to overreport the extent of their charitable contributions (Hall, 2001), there is not much research about whether this really is the case. It has been shown for donations of time, money and blood that reporting such donations was strongly associated with the respondents' beliefs that significant others expect them to do so (Lee, Piliavin & Call, 1999). More direct evidence for SD-bias has been provided in a study where respondents reported whether they contributed money to an environmental group either in a private mail or a public telephone survey in the last two years. In the mail survey, only 15.5 percent reported such a donation, whereas this rate was with 22.7 percent significantly higher in the telephone survey (Ethier et al., 2000). The authors argue that the insufficient response privacy leads to an impression management-based SD-bias in the direction of overreporting donations.

- *Participation in political elections*: Survey respondents are frequently asked about whether they participated in the most recent elections or in ones held longer ago.³ Research has consistently shown a considerable part of survey participants to report to have voted although they did not. Belli and colleagues (2001) found with data from the American National Election Studies (ANES) for elections between 1964 and 1990 that on average 10.2 percent of all respondents claim to have participated, but this was not confirmed by the validation records. However, only 0.7 percent failed to report a true participation. A few studies tested the hypothesis that SD-bias is the reason for vote overreporting. It has been compared which percentage of respondents claimed to have voted in three federal elections in Germany when their answers were collected either self- or interviewer-administered (Stocké, 2005b). Subjects were significantly more likely to report to have voted when the response privacy was low in interviewer-administered interviews. Furthermore, it was found that under interviewer administration the survey-based measure for electoral turnout differed significantly from the official figures, whereas this was not the case in self-administered interviews.

Empirical evidence suggests that certain groups of respondents are more likely to overreport voting than others. Accordingly, it was found that subjects who overstate their participation in political elections are more educated (Belli et al., 2001), less wealthy (Traugott & Katoch, 1979) and more likely African Americans (Abramson & Claggett, 1991). There is fur-

² In the Eurobarometer 28, respondents answered the question: '... would you be prepared to give money or to give more money than you do now to support some activity to help the third world?' (ICPSR dataset 9082).

³ An example is the question asked in all ANES Studies: 'Do you remember whether you voted in [year] when [Democratic Presidential candidate] ran against [Republican Presidential candidate]?' (ICPSR dataset 8475).

thermore evidence that respondents with higher political involvement (Stocké & Stark, 2006) and a stronger sense of civic duty (Karp & Brockington, 2005) tend more to vote overreporting. An empirically not tested explanation for these differences is that certain respondent groups assume electoral participation to be more socially desirable than others.

- *Environmental consciousness*: Data about the respondents' environmental consciousness has been collected in many surveys.⁴ Here, expert ratings of the social desirability of environmental-friendly attitudes and behavioral intentions were highly correlated with the endorsement of items measuring these traits (Schahn, 2002). Another study showed that the responses on a 45-item scale for measuring environmental consciousness were substantially positively correlated with the subjects' need for social approval (Allen & Ferrand, 1999).

5. Empirical Study

5.1 Data and Measurement

The respondents in our study were a multi-stage, local random-probability sample of residents from a medium-sized city in Germany (about 250,000 inhabitants). Households were listed with a random-walk procedure and individual respondents selected with the 'last-birthday' method. Eligible respondents were at least 18 years old and German citizens. The SD-beliefs about the four analyzed characteristics were collected in two studies. In order to avoid effects due to preceding questions about the desirability of the other topics, respondents in both studies were randomly assigned to one of two conditions where they only answered questions about one of the topics. Study 1 (N=150) covered the topics 'participation in political elections' (N=74) and 'environmental consciousness' (N=76), whereas in study 2 (N=144) the topics 'charitable donations' (N=74) and 'prejudices toward the elderly' (N=70) were included. The response rates were 31.3 percent (study 1) and 24.1 percent (study 2). The interviews in both studies were conducted computer-assisted in the respondents' homes. The questions about the respondents' SD-beliefs were asked in the first third of the interviews, preceded by questions about the respondents' sociodemographic characteristics.

In our study, we utilized the concept of extrinsically expected desirability and thus the respondents' perceptions about how desirable the respective characteristics were judged by others. Respondents were asked how a person would feel when he or she would disclose the particular trait in public.⁵ The responses were recorded using a nine-point response scale,

⁴ The Eurobarometer 58.0 utilized for instance five Likert-type items to measure environmental concern, as for example: 'I don't care about environmental issues' (ICPSR dataset 3661).

⁵ The question wording was as follows: 'We want to find out what is better not said in public because this would provoke negative reactions. In contrast, other statements may arouse positive reactions because they are

ranging from +4 ('saying this in public would be very pleasant') and -4 ('saying this would be very embarrassing'). Aside from the endpoints, the scale midpoint at zero was labeled with: 'this statement would be evaluated neutrally'. For each of the analyzed characteristics, the respondents answered three questions. These questions were *firstly* about how a person would feel when stating to have (a) strong, (b) some or (c) no prejudice against elderly people. *Secondly*, subjects reported what they believed somebody would feel when saying that he/she donates (a) frequently, (b) occasionally or (c) never money for charitable organizations. *Thirdly*, it was recorded what the respondents think how a person would feel when revealing in public to participate (a) always, (b) sometimes and (c) never in political elections. *Fourthly*, subjects reported which feelings they anticipate when somebody revealed in public to have (a) a strong, (b) a certain amount and (c) no environmental consciousness.

On the basis of these desirability ratings, we constructed the different measures for the respondents' perceived incentives for SD-bias. *Firstly*, the OPM is simply the raw rating of a high trait intensity (TD_{HI}) and thus ranges from -4 (very undesirable) to +4 (very desirable). This operationalization is consistent with that used in all other studies in this field of research. *Secondly*, in order to construct SDS, the raw ratings of a low and a high trait intensity were first transformed into a range from 0 (undesirable) to +8 (desirable). Then the transformed rating of a low trait intensity (TD_{LO}) has been subtracted from that of a high intensity (TD_{HI}): $SDS = TD_{HI} - TD_{LO}$. The resulting values ranged between -8 (low trait intensity more desirable) to +8 (high trait intensity more desirable). *Thirdly*, the DSDS are constructed alike the SDS, but here the difference between the SD ratings of a low (TD_{LO}) and a medium (TD_{ME}) trait intensity on the one hand and that between a medium (TD_{ME}) and a high (TD_{HI}) trait intensity on the other were computed: $TD_{ME-LO} = TD_{ME} - TD_{LO}$ and $TD_{HI-ME} = TD_{HI} - TD_{ME}$. Both domain-specific difference scores ranged from -8 (lower trait intensity is more desirable) to +8 (higher trait intensity is more desirable).

In case of all three operationalizations, the strength of the possible SD-bias is defined by the absolute value of the respective measures. In case of the DSDS, this is the sum of the absolute value from the two domains for desirability profiles that do not violate the monotony assumption. In contrast, when this was the case, e.g. in case of u-shaped or reverse u-shaped profiles, this method would overestimate the possible strength of the SD-bias, since here the respondents' 'true' value can only be on one side of the continuum. In these cases, the aver-

in agreement with rules about what can be said openly. Could you please tell me for the following characteristics whether it would be embarrassing or pleasant for a person to reveal them in public? We don't want to know how you judge these characteristics personally, but what can be expressed in public and what not'.

age absolute value of the two domains has been utilized. Thus, in the case of the OPM, the possible strength of SD-bias ranges between 0 and 4, whereas for SDS, this range lies between 0 and 8. The value range of the latter measure is double compared with the former due to the abandonment of the neutrality assumption: In the case of an extreme violation of this assumption, for instance when a high trait intensity is evaluated as very positive (+4) and a low trait intensity as extremely negative (-4), the OPM detects only half of the true incentives for SD-bias. For the DSDS, the strength of SD-bias can vary between 0 and 8 as well.

5.2 Incentives for SD-Bias on the Aggregate Level

In figure 3, it is presented for the aggregate level how the respondents perceived the SD of different intensities of the analyzed trait dimensions. *Firstly*, we found that the judged desirability increased with the trait intensity for charitable donations, electoral participation and for environmental consciousness, whereas more prejudice against the elderly was believed to be less positively evaluated. *Secondly*, the increase or decrease of the trait desirability was by no means found to be linear across the different levels of trait intensity for all characteristics. In the case of charitable donations and environmental consciousness, the trait desirability differed strongly between an absence and a middle strength of these traits, whereas on average, a medium and a strong trait intensity was practically regarded as equally desirable. Similarly, in case of prejudice against the elderly, it did not make a big difference whether a person reports to have strong or medium prejudice, while having no prejudice at all is assumed to be substantially more desirable than the other two trait intensities. In contrast, for the trait dimension of electoral participation, the respondents perceived a nearly linear increase in the SD when somebody reports to have voted more often in political elections. A *third* result was that on the aggregate level the neutrality assumption did not hold for any of the analyzed topics: For prejudice against the elderly, the absence of this trait is perceived to be evaluated statistically significant above the neutral midpoint of the SD-scale, whereas the anticipated evaluation of a low intensity of the other trait dimensions deviated significantly into the negative direction (prejudice: $t=9.0$, $df=69$; donation: $t=-3.2$, $df=73$; participation: $t=-4.9$, $df=73$; environment: $t=-13.1$, $df=75$; all $p \leq .01$). On the aggregate level, the monotony assumption seems to be valid, since, except for a slight and statistically not significant decrease in the SD-ratings of donations between a medium and high trait intensity, the desirability profiles increase or decrease monotonically across the intensity dimensions ($t=1.2$, $df=73$, $p > .10$).

-- Figure 3 here --

When the respondents in the sample perceive incentives for SD-bias into different directions, these differences at least partly cancel each other out on the aggregate level. Thus, we utilized the average absolute values of the different measures in order to compare their predictions about the strength of incentives for SD-bias in the case of the different trait dimensions (cf. table 1). *Firstly*, we found for all analyzed traits the simple difference scores, compared with the one-point measure, to predict significantly stronger incentives for socially desirable responding (prejudice: $t=7.1$, $df=69$; donation: $t=2.4$, $df=73$; participation: $t=6.6$, $df=73$; environment: $t=13.6$, $df=75$; all $p \leq .01$). As we show below, this is the consequence of violations of the neutrality assumption. Comparing how strong the two measures disagree revealed that this was strongest for the trait ‘environmental consciousness’ (2.4 scale points) and weakest for ‘charitable donations’ (0.6 scale points), the other two topics taking an intermediate position (1.8 and 1.5 scale points). This suggests strongest violations of the neutrality assumption in the case of ‘environmental consciousness’ and weakest for ‘charitable donations’. *Secondly*, comparing the SDS and the DSDS showed that the latter measure predicts in two out of four topics significantly stronger incentives for SD-bias (prejudice: $t=2.2$, $df=69$; donation: $t=2.8$, $df=73$; both: $p \leq .05$; participation: $t=0.3$, $df=73$; environment: $t=0.6$, $df=7$, both: $p > .10$). This result suggests, and this will be tested in the following section, that the monotony assumption has been violated more in the case of the first two topics.

-- Table 1 here --

5.3 Prevalence and Effects of Violations of the Neutrality and Monotony Assumptions

Table 2 presents for the analyzed questionnaire topics for which proportion of the total sample (a) the neutrality assumption, (b) the monotony assumption or (c) both assumptions at the same time have been found to be violated. The results have shown *firstly* that in the case of all trait dimensions only a minority of between 28.4 (charitable donations) and 2.6 (environmental consciousness) percent of the respondents have not violated at least one of the two assumptions. *Secondly*, violations of neutrality are by far the quantitatively more extensive problem across all topics, compared with violations of monotony. The latter occurred in between 14.2 percent (‘prejudice against the elderly’) and 33.8 percent (‘charitable donations’) of the cases, and was most often accompanied by violations of neutrality. The proportion of the total sample where neutrality is violated varies between 66.2 percent (‘charitable donations’) and amounts to 94.8 percent (‘environmental consciousness’). *Thirdly*, our data confirmed our assumption that differences in the violation of the neutrality assumption (78.6, 66.2, 79.7, 94.8 percent for the four traits) explain how strong the OPM underestimates the

potential SD-bias compared with the SDS (1.8, 0.6, 1.5, 2.4 scale points for the four traits): The correlation between the two quantities was $r=.96$ and statistically significant ($p \leq .05$). The same was, however, not true for the percentage of violations of the monotony assumption (14.2, 33.8, 17.6, 31.6 percent for the four traits) and the differences in the estimated strength of incentives for SD-bias between the OPM and DSDS (1.9, 1.0, 1.5, 2.5 scale points for the four traits): The association here was $r=-.08$ and did not prove to be significant ($p > .10$).

-- Table 2 here --

In table 3, the consequences of violations of neutrality are shown in detail. *Firstly*, when such a violation occurred, this leads the OPM much more frequently to under- than to overestimate the incentives for SD-bias: On average across the traits, 16.5 percent lead to an overestimation, but in 80.8 percent an underestimation occurred, whereas only in 2.4 percent a wrong direction of incentives was predicted. However, the consequences, when neutrality did not hold, varied considerably across the trait dimensions. Furthermore, the average strength of underestimation (between 2.3 and 3.0 scale points) is for each trait at least as strong as that of an overestimation (between 1.3 and 2.4 scale points). Thus, as already found in the previous paragraph, violations of neutrality cause an underestimation of potential SD-bias.

-- Table 3 here --

Respondents may assume a high, a low, a medium or both extreme trait intensities to be most socially desirable, and thus perceive incentives for SD-bias in different directions. Table 4 presents the direction of incentives for SD-bias which has actually been derived from DSDS. *Firstly*, between 50.0 and 67.6 percent of all respondents assumed that more frequent donations, as well as participating in political elections more often and having a stronger environmental consciousness always lead to more positive evaluations, and 70 percent assumed to have less prejudice against the elderly to have this consequence. Only a very small minority of up to 5.7 percent of the respondents assumed a linear increase of desirability differences in the opposite direction. *Secondly*, in the case of charitable donations and environmental consciousness, we found substantial proportions of respondents to have inverse u-shaped desirability profiles: 27.0 percent assumed sometimes donating and having a medium degree of environmental consciousness to be most socially desirable. In the case of the other two topics, this kind of incentives for SD-bias was perceived by few respondents only. This is true for u-shaped desirability profiles as well: Only between 2.6 and 9.5 percent of all respondents assumed a medium trait intensity to be most undesirable. *Thirdly*, except for the trait 'environmental consciousness', in case of all topics about 10 percent of the subjects per-

ceived a weak, medium and strong trait intensity to be identically evaluated. We can thus conclude that the violation of monotony results for the OPM and the SDS in the case of ‘charitable donations’ and ‘environmental consciousness’ in predictions of wrong direction of the SD-bias.

-- Table 4 here --

When the monotony assumption is violated, OPM and SDS always predict the wrong direction of SD-bias. In table 5a, it is shown which consequences this violation has for the strength of SD-bias predicted by the OPM, when compared with the correct measure obtained on the basis of DSDS. According to our results, the presence of non-linear SD-beliefs leads the OPM in the majority of the cases to overestimate the true strength of SD-bias: On average across all items, this happened in 52.6 percent of the cases, whereas an underestimation was found for 26.7 percent and a correct prediction for 20.7 percent of the respondents.

-- Table 5a here --

The consequences of violations of the monotony assumption were different for the SDS. In case of this measure, incentives for SD-bias were underestimated in the majority of the cases: This was found on average across the trait dimension for 60 percent of the cases, whereas the measurement error was for 33.9 percent of the affected respondents in the direction of an overestimation of incentives, and for 6.1 percent the strength of SD-bias was correctly predicted. In summary, the net effect of violations of monotony on the predicted strength of SD-bias was found to differ between the two less elaborated incentive measures.

-- Table 5b here --

5.4 Convergent Validity in Predicting the Strength of Incentives for SD-Bias

In the final section of this paper, we analyzed under which conditions the more parsimonious OPM and SDS measures predicted more or less the same strength of incentives for SD-bias as the most valid DSDS. Thus, we tested with a series of OLS regressions, the strength of SD-bias predicted from the DSDS being dependent, for which of the questionnaire topics and for the violation of which assumption the more parsimonious measures were more or less consistent with this most elaborated measure. We constructed a pooled dataset consisting of the three incentive measures, obtained for the four questionnaire topics. Thus, the dataset consisted of N=294 cases, where dummy variables indicated from which subsample, and thus from which trait dimension a particular observation has been obtained.

In the *first* step of the analysis, we estimated a base model where only control variables for the respondents’ sociodemographic characteristics and topic dummies were included (cf. table 6a, model 1). The results have shown that the total strength of perceived incentives for

SD-bias did not differ according to the subjects' sex, age or education. The positive regression parameter for the trait dummies indicated that subjects perceived, compared with the trait of charitable donations, significantly stronger incentives for SD-bias when they answered questions about the other traits. However, these differences were only statistically significant for 'environmental consciousness' and 'prejudice against the elderly'. *Secondly*, we tested to what degree the OPM, across all questionnaire topics, predicted the same strength of incentives as the DSDS (cf. table 6a, model 2). This effect was found to be statistically significant and increased the explained variance from 7.2 percent in the base model to 51.8 percent. Thus, the most parsimonious measure captured 44.6 percent of the real variation of incentives for SD-bias. *Thirdly*, we tested whether this convergent validity differed between the questionnaire topics. This was accomplished by adding multiplicative terms between each topic dummy and the OPM into the model. According to the estimated interaction parameters, the OPM agreed significantly stronger with the DSDS in the case of the traits 'electoral participation' and 'prejudice against the elderly', compared with the reference category 'charitable donations'. The convergent validity of the OPM was in the case of 'environmental consciousness', however, not significantly higher than for the reference topic. This result may be due to stronger violations of the neutrality assumption in the case of the traits with less validity. This was tested *fourthly* by controlling for an interaction term between the OPM and how strong the neutrality assumption was violated (cf. table 6a, model 4).⁶ Although the actual strength of incentives predicted from the DSDS was significantly higher, the stronger the neutrality assumption was violated, the latter factor did neither explain the convergent validity of the one-point measure, nor the differences in the convergent validity between the topics: The interaction parameter was not statistically significant, and did only reduce, but not absorb the effect of the topic dummies. However, taking the strength of violations of neutrality into account increased the explained variance substantially from 52.3 to 72.0 percent. If it is not mainly the fact how strong the neutrality assumption has been violated which explains the differences in construct validity between the trait dimension, this may be due to how strong the monotony assumption did not hold. This was tested by additionally including an interaction term between the one-point measure and the magnitude to which the monotony assumption has been violated into the regression equations (cf. table 6a, model 5).⁷ The significant negative interaction parameter indicated that the predictions from

⁶ The strength of this violation is the absolute value of how strongly the judged desirability of a low trait intensity deviates from a neutral evaluation at zero.

⁷ How strongly the monotony assumption has been violated is the absolute value of the difference between the SDS and the DSDS.

the OPM are less consistent with those from the DSDS when the monotony assumption was stronger violated. However, controlling for these differences did still not completely explain the validity differences of the OPM between the questionnaire topics: This measure had higher validity for the trait ‘environmental consciousness’, compared with that of ‘charitable donations’.

-- Table 6a here --

In a next step, we tested whether the SDS were, compared with the OPM, the better indicator for the strength of incentives from SD-bias (cf. table 6b, model 6). The results indicated *firstly* that this measure is much stronger associated with the DSDS values: Including SDS increased the explained variance, taking the base model 1 as a starting point, from 7.2 to 85.1 percent, and thus the SDS captured 77.9 percent of the real strength of incentives for SD-bias. Thus, compared to the 44.6 percent in the case of the OPM, this is a substantial improvement in convergent validity. *Secondly*, the SDS successfully captured the between-item differences in the strength of SD-bias, which have been found in model 1: After controlling for the SDS scores, the topic dummies did not prove to be statistically significant anymore.

In the last step of the analysis, we tested for differences in the strength of the association between the SDS and DSDS between the trait dimensions. We expected a lower convergent validity in the case of questionnaire topics where stronger violations of monotony have been observed. This has been tested by computing interaction terms between the topic dummies and the SDS-measure, which were then included into the regression equation (cf. table 6b, model 7). The results have shown that SDS were a significantly better indicator for the actual incentives for SD-bias in case of the topics ‘electoral participation’ and ‘prejudice against the elderly’, compared to the other topics ‘environmental consciousness’ and ‘charitable donations’. This can be attributed to the fact that the latter trait dimensions have been found in table 4 to be substantially more prone to violations of monotony, and thus the SDS to be much more susceptible to measurement error.

6. Summary and Conclusion

In this paper, we compared the validity of three differently elaborated and thus differently parsimonious measures for the strength and direction of incentives for SD-bias respondents perceive when answering survey questions. We have shown *firstly* that apart from domain-specific difference scores (DSDS) as the most elaborated measure, the validity of the other two depends on whether specific assumptions are fulfilled in each case: Whereas simple dif-

ference scores (SDS) only require the trait desirability to increase or decrease monotonically across the trait-intensity continuum (monotony assumption), the one-point measure (OPM), as the least elaborated operationalization, requires monotony and additionally that the respondents neutrally evaluate the absence of the respective trait as well (neutrality assumption). As we have shown in the theoretical part of our study, violations of both assumptions can be expected to cause error in the predicted strength and direction of SD-bias. It has however not been tested whether such violations occur, whether their prevalence differs between questionnaire topics and whether they have indeed negative effects on the validity of the affected measures. In our empirical part, we thus utilized four trait dimensions, which are frequently measured in survey research, in order to analyze these open questions.

Our results proved *firstly* that the neutrality assumption has been violated by the majority of the respondents, in the case of all analyzed trait dimensions: Having no prejudice against the elderly, making no charitable donations, not voting in political elections and having a weak environmental consciousness were rarely evaluated neutrally. Although this violation seldom biased the direction of predicted SD-bias, it leads in 80 percent of the affected cases the OPM to substantially underestimate the real strength of SD-bias. This measurement error, however, varied between the different questionnaire topics: The strongest error has been found for the trait ‘environmental consciousness’, and the measure for ‘charitable donations’ was least affected. *Secondly*, our analysis revealed that certain groups of respondents did not simply assume that more charitable donations and a stronger environmental consciousness are more socially desirable. Instead, in both cases nearly 30 percent of the sample assumed a medium trait intensity to be most desirable: These subjects perceive incentives for selecting an answer at the middle of the trait continuum. In these cases where the monotony assumption has been violated, both the OPM and SDS predicted the wrong direction of incentives for SD-bias. Furthermore, we found that the OPM suffered in these cases mainly from an overestimation of SD-bias, whereas in the majority of cases, the SDS underestimated the total incentives for SD-bias. *Thirdly*, leaving the quantitatively negligible problem of predictions of a wrong direction of SD-bias aside, we found violations of neutrality to lead to a low convergent validity in case of the OPM: Only about 44.6 percent of the, according to the DSDS actually present incentives for socially desirable response behavior were captured by this measure. This leads to the conclusion that at least the measurement effort associated with SDS, which does not require the neutrality assumption to hold, has to be undertaken in order to obtain a valid measurement of potential SD-bias. In agreement herewith, this measure proved to be much more valid in predicting the strength of incentives for SD-bias: Across all topics,

altogether 77.9 percent of the incentives for SD-bias have been captured by this indicator. In case of the trait dimensions ‘prejudice against the elderly’ and ‘electoral participation’, this indicator can be regarded to be sufficiently elaborated to provide a good prediction for which incentives respondents perceive for SD-bias. However, for the questionnaire topics ‘charitable donations’ and ‘environmental consciousness’, this measure had substantially weaker convergent validity: Here, SDS predict for those respondents with inverse u-shaped desirability profiles the wrong direction of SD-bias. Furthermore, the strength of SD-bias has been substantially underestimated in these cases. Thus, for the affected questionnaire topics, only the most elaborated DSDS are a valid measure for the strength and direction of SD-bias in survey data.

The validity of our results is conditional on the following factors. *Firstly*, even those measures we identified to provide valid measures for incentives for SD-bias must be regarded to provide an indicator for the upper limit of real SD-bias, which has to be expected in survey data. Accordingly, impression management-based SD-bias has been found to be much weaker or even absent when the respondents’ privacy and anonymity are successfully ensured (e.g. Tourangeau et al., 2002). Thus, under these conditions, for instance in self-administered surveys, respondents’ answers can be expected to be less strongly biased by their SD-beliefs than this measure suggests. Furthermore, each of the incentive measures analyzed in this paper presuppose that the respondents’ true characteristics are diametrically opposed to those they perceive as the most desirable. The more, however, the true score and the most desirable response option converge, the less the incentives for SD-bias will actually bias survey data. Accordingly, our incentive measures will be more diagnostic for response bias when the characteristics regarded to be least desirable are increasingly more prevalent in the survey sample.

Secondly, our study provides the first in-depth analysis of which answers respondents perceive as the most desirable in the case of four often used questionnaire topics, and how much measurement effort is necessary to obtain a valid indicator for which incentives respondents perceive for SD-bias. This was done with a small, regionally restricted sample of respondents. Since the representativeness of the data was therefore restricted, our results cannot be simply generalized without further evidence, and therefore must be at present regarded to be tentative. However, within these restrictions, our results clearly suggest a tradeoff between parsimony and validity when differently elaborated measures for incentives for SD-bias are utilized. This issue deserves more attention in future research.

References

- Abramson, P.R. and Claggett, W. (1991). Racial Differences in Self-Reported and Validated Turnout in the 1988 Presidential Election. *The Journal of Politics*, 53, 186-197.
- Allen, J.B. and Ferrand, J.L. (1999). Environmental Locus of Control, Sympathy, and Pro-environmental Behavior. A Test of Geller's Actively Caring Hypothesis. *Environment and Behavior*, 31, 338-353.
- Belli, R.F., Traugott, M.W., and Beckmann, M.N. (2001). What Leads to Voting Overreports? Contrasts of Overreporters to Validated Voters and Admitted Nonvoters in the American National Election Studies. *Journal of Official Statistics*, 17, 479-498.
- Chen, P.Y., Dai, T., Spector, P.E., and Jex, S.M. (1997). Relation Between Negative Affectivity and Positive Affectivity: Effects of Judged Desirability of Scale Items and Respondents' Social Desirability. *Journal of Personality Assessment*, 69, 183-198.
- Crott, H.W. and Baltes, P.B. (1973). Desirability of Personality Dimensions: Auto- and Heteroperceptions by American and German College Students. *The Journal of Social Psychology*, 91, 15-27.
- Edwards, A.L. (1957). *The Social Desirability Variable in Personality Assessment and Research*. New York: The Dryden Press.
- Embree, B.G. and Whitehead, P.C. (1993). Validity and Reliability of Self-Reported Drinking Behavior: Dealing with the Problem of Response Bias. *Journal of Studies on Alcohol*, 54, 334-344.
- Esser, H. (1991). Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als „rational choice“. In R. Wittenberg (Hrsg.), *Person - Situation - Institution - Kultur. Günther Büschges zum 65. Geburtstag* (pp. 59-78). Berlin: Duncker & Humblot.
- Ethier, R.G., Poe, G.L., Schulze, W.D., and Clark, J. (2000). A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green-Pricing Electricity Programs. *Land Economics*, 76, 54-67.
- Fernandes, M.F. and Randall, D.M. (1992). The Nature of Social Desirability Response Effects in Ethics Research. *Business Ethics Quarterly*, 2, 183-205.
- Ganster, D.C., Hennessey, H.W., and Luthans, F. (1983). Social Desirability Response Effects: Three Alternative Models. *Academy of Management Journal*, 26, 321-331.
- Hall, M.H. (2001). Measurement Issues in Surveys of Giving and Volunteering and Strategies Applied in the Design of Canada's National Survey of Giving, Volunteering and Participating. *Nonprofit and Voluntary Sector Quarterly*, 30, 515-526.
- Huang, C., Liao, H., and Chang, S.-H. (1998). Social Desirability and the Clinical Self-

- Report Inventory: Methodological Reconsideration. *Journal of Clinical Psychology*, 54, 517-528.
- Johnson, J.A. (2004). The Impact of Item Characteristics on Item and Scale Validity. *Multivariate Behavioral Research*, 39, 273-302.
- Karp, J.A. and Brockington, D. (2005). Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries. *Journal of Politics*, 67, 825-840.
- Krysan, M. (1998). Privacy and the Expression of White Racial Attitudes: A Comparison Across Three Contexts. *Public Opinion Quarterly*, 62, 506-544.
- Lee, L., Piliavin, J.A., and Call, V.R.A. (1999). Giving Time, Money, and Blood: Similarities and Differences. *Social Psychology Quarterly*, 62, 276-290.
- Meleddu, M. and Guicciardi, M. (1998). Self-Knowledge and Social Desirability of Personality Traits. *European Journal of Personality*, 12, 151-168.
- Nicotera, A.M. (1996). An Assessment of the Argumentativeness Scale For Social Desirability Bias. *Communication Reports*, 9, 23-36.
- Paulhus, D.L. (2003). Self-Presentation Measurement. In R. Fernandez-Ballesteros (ed.), *Encyclopedia of Psychological Assessment* Thousand Oaks, CA: Sage.
- Phillips, D.L. and Clancy, K.J. (1972). Some Effects of "Social Desirability" in Survey Studies. *American Journal of Sociology*, 77, 921-940.
- Schahn, J. (2002). Aufwand und soziale Erwünschtheit als Prädiktoren von Itemmittelnwerten. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 23, 45-54.
- Smith, S.T., Smith, K.D., and Seymour, K.J. (1993). Social Desirability of Personality Items as a Predictor of Endorsement: A Cross-Cultural Analysis. *Journal of Social Psychology*, 133, 43-53.
- Stocké, V. (2004). Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie*, 33, 303-320.
- Stocké, V. (2005a). Cognitive and Motivational Determinants for the Strength and Direction of Social Desirability Bias in Racial Attitude Surveys. In C. van Dijkum, J. Blasius, H. Kleijer, and B. van Hilten (eds.), *Recent Developments and Applications in Social Research Methodology* Amsterdam: Proceedings of the Sixth International Conference on Logic and Methodology, August 17-20, 2004 Amsterdam, The Netherlands (2nd edition).
- Stocké, V. (2005b). *Response Privacy and Elapsed Time Since Election Day as Determinants for Vote Overreporting*. Working Paper No. 05-42: Sonderforschungsbereich 504, Univer-

sity of Mannheim.

- Stocké, V. and Stark, T. (2006). *Political Involvement and Memory Failure as Interdependent Determinants of Vote Overreporting*. Working Paper No. 06-01: Sonderforschungsbereich 504, University of Mannheim.
- Tourangeau, R. and Rasinski, K.A. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin*, 103, 299-314.
- Tourangeau, R., Rips, L.J., and Rasinski, K.A. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Tourangeau, R. and Smith, T.W. (1996). Asking Sensitive Questions. The Impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly*, 60, 275-304.
- Tourangeau, R., Steiger, D.M., and Wilson, D. (2002). Self-Administered Questions by Telephone: Evaluating Interactive Voice Response. *Public Opinion Quarterly*, 66, 265-278.
- Traugott, M.W. and Katosh, J.P. (1979). Response Validity in Surveys of Voting Behavior. *Public Opinion Quarterly*, 43, 359-377.
- Wyner, G.A. (1980). Response Errors in Self-Reported Number of Arrests. *Sociological Methods & Research*, 9, 161-177.

Tables and figures

Figure 1: Illustration of the violation of the neutrality assumption and its consequences

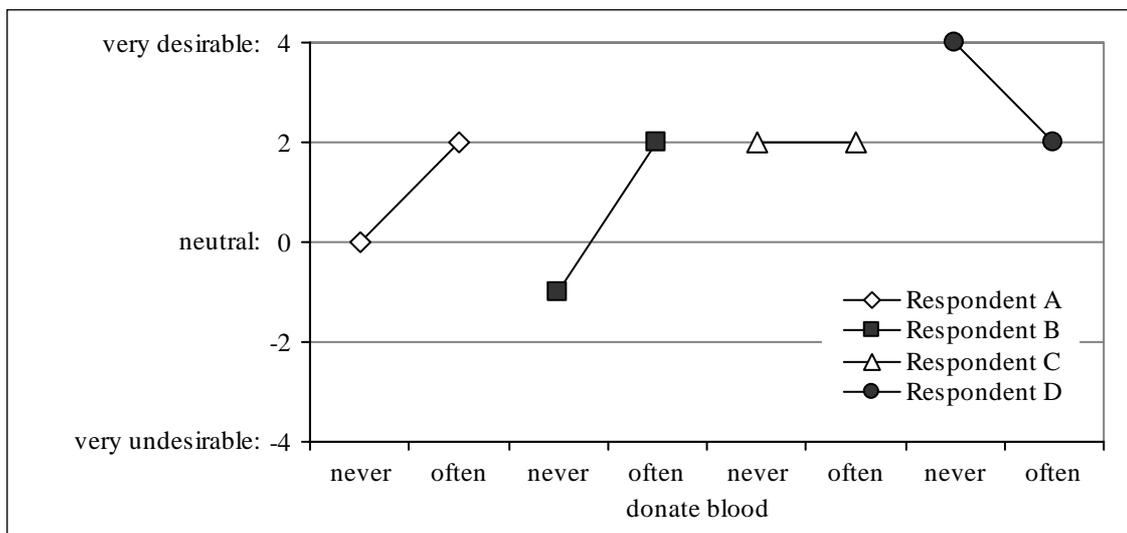


Figure 2: Illustration of the violation of the monotony assumption and its consequences

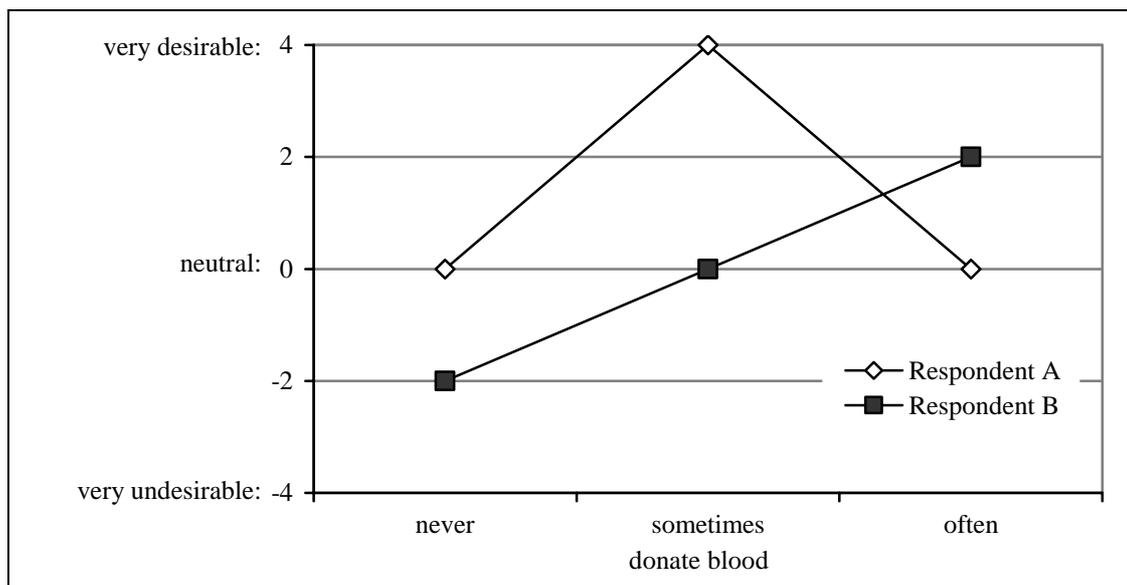
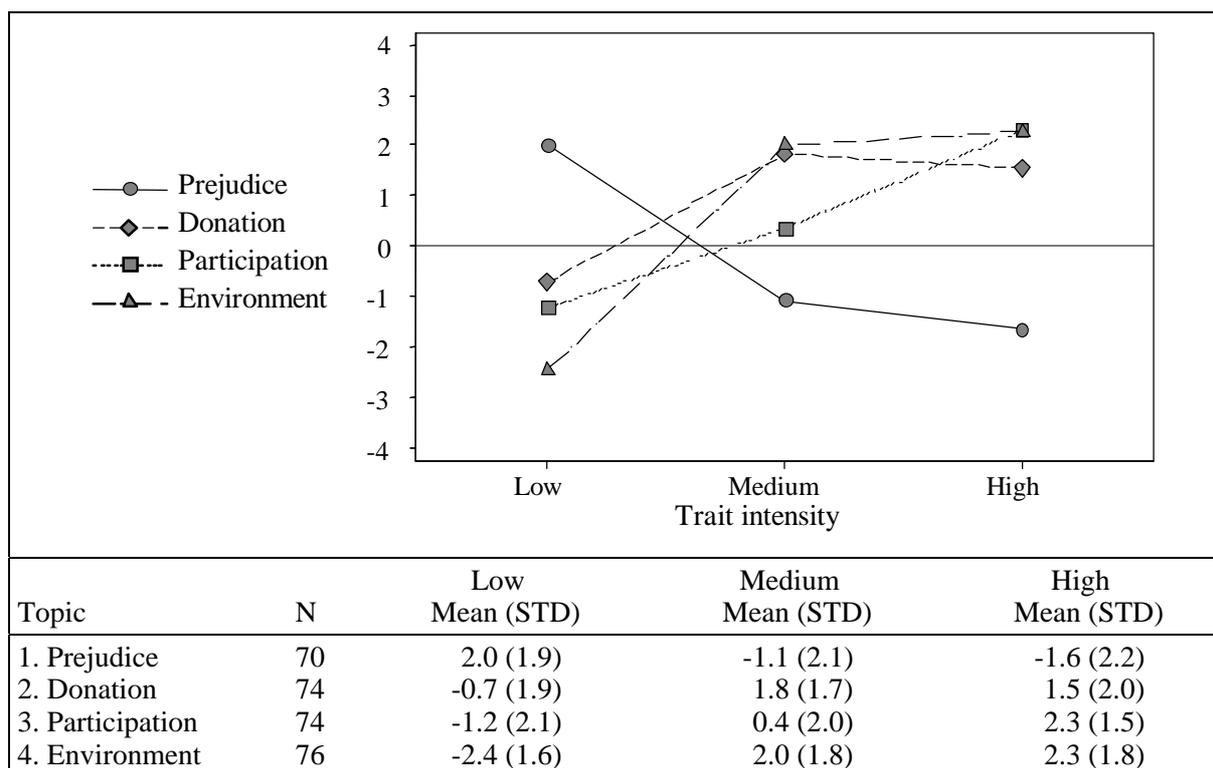


Figure 3: Perceived social desirability of the analyzed trait dimensions on the aggregate level



Response scale ranges between -4 (negatively evaluated) and +4 (positively evaluated).

Table 1: Predicted strength of SD-bias using the different desirability measures

Topic	One-point measure	Difference scores	Domain-specific difference scores		
	$ TD_{HI} $ Mean (STD)	$ TD_{DIFF} $ Mean (STD)	$ TD_{HI-ME} $ Mean (STD)	$ TD_{ME-LO} $ Mean (STD)	DSDS Mean (STD)
1. Prejudice	2.3 (1.5)	4.1 (2.9)	1.0 (1.2)	3.5 (2.6)	4.2 (2.7)
2. Donation	2.2 (1.1)	2.8 (2.1)	1.3 (1.6)	3.0 (2.3)	3.2 (1.9)
3. Participation	2.4 (1.3)	3.9 (2.6)	2.3 (2.2)	2.2 (2.1)	3.9 (2.5)
4. Environment	2.6 (1.3)	5.0 (2.3)	1.7 (1.9)	4.6 (2.3)	5.1 (2.1)
All	2.4 (1.3)	3.9 (2.5)	1.6 (1.7)	3.3 (2.3)	4.1 (2.3)

Table 2: Prevalence of violations of the neutrality and monotony assumptions

Topic	Monotony violated only	Neutrality violated only	Monotony and neutrality violated	No violation	Total
	% (N)	% (N)	% (N)	% (N)	% (N)
1. Prejudice	7.1 (5)	71.5 (50)	7.1 (5)	14.3 (10)	100.0 (70)
2. Donation	5.4 (4)	37.8 (28)	28.4 (21)	28.4 (21)	100.0 (74)
3. Participation	0.0 (0)	62.1 (46)	17.6 (13)	20.3 (15)	100.0 (74)
4. Environment	2.6 (2)	65.8 (50)	29.0 (22)	2.6 (2)	100.0 (76)
All	3.8 (11)	59.3 (174)	20.5 (61)	16.4 (48)	100.0 (294)

Table 3: Consequences of violations of the neutrality assumption for the predicted strength and direction of incentives for SD-bias

<i>Topic</i>	Overestimation			Underestimation			Wrong direction			Total	
	%	(N)	Mean	%	(N)	Mean	%	(N)	Mean	%	(N)
1. Prejudice	12.7	(7)	2.4	85.5	(47)	3.0	1.8	(1)	4.0	100.0	(55)
2. Donation	30.6	(15)	2.3	67.4	(33)	2.3	2.0	(1)	3.0	100.0	(49)
3. Participation	18.6	(11)	1.8	81.4	(48)	2.7	0.0	(0)	--	100.0	(59)
4. Environment	4.2	(3)	1.3	88.9	(64)	2.9	6.9	(5)	2.6	100.0	(72)
All	16.5	(36)	2.0	80.8	(192)	2.7	2.7	(7)	2.4	100.0	(235)

Note: 'Overestimation' and 'underestimation' indicates the one-point measure to predict stronger or weaker incentives for SD-bias, compared with simple difference scores. 'Wrong direction' indicates that the one-point measure predicts incentives in the opposite direction. The means represents the average strength of the deviations between the two measures.

Table 4: Type and prevalence of violations of monotony

	Prejudice	Donation	Participation	Environment
	% (N)	% (N)	% (N)	% (N)
Incentives for high trait intensity	5.7 (4)	50.0 (37)	67.6 (50)	65.8 (50)
Incentives for low trait intensity	70.0 (49)	5.4 (4)	4.0 (3)	1.3 (1)
Incentives u-shaped	8.6 (6)	6.8 (5)	9.5 (7)	2.6 (2)
Incentives inverse u-shaped	5.7 (4)	27.0 (20)	8.1 (6)	29.0 (22)
No incentives	10.0 (7)	10.8 (8)	10.8 (8)	1.3 (1)
Total	100.0 (70)	100.0 (74)	100.0 (74)	100.0 (76)

Table 5a: Consequences of violations of monotony for the one-point measure

<i>Topic</i>	Underestimation		Overestimation		Correct prediction		Total				
	%	(N)	Mean	%	(N)	Mean	%	(N)	Mean		
1. Prejudice	20.0	(2)	2.5	30.0	(3)	2.0	50.0	(5)	0.0	100.0 (10)	1.1
2. Donation	40.0	(10)	2.3	44.0	(11)	2.2	16.0	(4)	0.0	100.0 (25)	1.9
3. Participation	38.5	(5)	2.0	61.5	(8)	2.5	0.0	(0)	--	100.0 (13)	2.3
4. Environment	8.3	(2)	1.8	75.0	(18)	2.7	16.7	(4)	0.0	100.0 (24)	2.2
All (average)	26.7	(19)	2.2	52.6	(39)	2.4	20.7	(9)	0.0	100.0 (72)	1.9

Note: Overestimation and underestimation indicate that the one-point measure predicts stronger or weaker incentives for SD-bias, compared with the domain-specific difference scores. The means indicates the average strength of the deviations between the two measures.

Table 5b: Consequences of violations of monotony for simple difference scores

Topic	Underestimation		Overestimation		Right prediction			Total		
	%	(N) Mean	%	(N) Mean	%	(N) Mean	%	(N) Mean		
1. Prejudice	80.0	(8) 1.4	20.0	(2) 0.5	0.0	(0) --	100.0	(10) 1.2		
2. Donation	64.0	(16) 2.3	16.0	(4) 1.2	20.0	(5) 0.0	100.0	(25) 1.7		
3. Participation	46.2	(6) 1.7	53.9	(7) 1.1	0.0	(0) --	100.0	(13) 1.4		
4. Environment	50.0	(12) 1.8	45.8	(11) 1.4	4.2	(1) 0.0	100.0	(24) 1.5		
All (average)	60.0	1.8	33.9	1.1	6.1	0.0	100.0	1.5		

Note: Overestimation and underestimation indicate that the simple difference scores predict stronger or weaker incentives for SD-bias, compared with the domain-specific difference scores. The means indicates the average strength of the deviations between the two measures.

Table 6a: Capability of the one-point measure for predicting the strength of incentives from SD-bias (OLS-regression results)

	Model 1 B(t)	Model 2 B(t)	Model 3 B(t)	Model 4 B(t)	Model 5 B(t)
Demographic Controls					
- Sex (female)	0.41(1.5)	0.44(2.2)*	0.42(2.1)*	0.33(2.2)*	0.30(2.0)*
- Age (years)	0.00(0.0)	-0.00(-0.2)	-0.00(-0.5)	-0.01(-1.5)	-0.01(-1.2)
- Education (years)	0.04(0.5)	0.03(0.4)	0.02(0.3)	0.02(0.5)	0.02(0.4)
Topic Controls					
- Donation (Ref.Cat.)					
- Participation	0.66(1.7)	0.47(1.7)	-0.72(-1.2)	-0.72(-1.6)	-0.50(-1.1)
- Environment	1.91(5.1)**	1.43(5.2)**	0.63(1.0)	-0.02(-0.0)	-0.50(-1.0)
- Prejudice	1.00(2.6)*	0.90(3.2)**	-0.06(-0.1)	-0.15(-0.3)	-0.13(-0.3)
Explanatory Variables					
- One-point measure (OPM)	--	1.21(16.3)**	0.85(5.0)**	0.62(4.4)**	0.71(5.1)**
- Participation * OPM	--	--	0.53(2.3)*	0.35(2.0)*	0.30(1.8)
- Environment * OPM	--	--	0.36(1.6)	0.25(1.4)	0.40(2.2)*
- Prejudice * OPM	--	--	0.43(2.0)*	0.17(1.0)	0.17(1.0)
- Violation of <u>Neutrality</u>	--	--	--	0.80(7.3)**	0.86(8.0)**
- Neutrality * OPM	--	--	--	0.02(0.6)	0.01(0.2)
- Violations of <u>Monotony</u>	--	--	--	--	0.95(4.0)**
- Monotony * OPM	--	--	--	--	-0.31(-4.6)**
Intercept	2.52(2.1)*	0.08(0.1)	1.05(1.1)	0.43(0.6)	0.11(0.2)
Adjusted R-Squared	0.072	0.518	0.523	0.720	0.738
F-Overall	4.78**	45.89**	33.06**	63.62**	60.03**

Significance: * $p \leq .05$, ** $p \leq .01$; N=294.

Table 6b: Capability of simple difference scores for predicting the strength of incentives from SD-bias (OLS-regression results)

	Model 6 B (t)	Model 7 B (t)
Demographic Controls		
- Sex (female)	-0.07 (-0.6)	-0.08 (-0.8)
- Age (years)	-0.00 (-0.5)	-0.00 (-0.8)
- Education (years)	0.04 (1.0)	0.02 (0.5)
Topic Controls		
- Donation (Ref.Cat.)		
- Participation	-0.25 (-1.6)	-0.96 (-3.7)**
- Environment	-0.04 (-0.2)	-0.13 (-0.4)
- Prejudice	-0.10 (-0.6)	-0.88 (-3.4)**
Explanatory Variables		
- Simple difference scores (SDS)	0.85 (38.7)**	0.69 (13.7)**
- Participation * SDC	--	0.23 (3.5)**
- Environment * SDC	--	0.09 (1.3)
- Prejudice * SDC	--	0.24 (3.9)**
Intercept	0.58 (1.2)	1.27 (2.5)*
Adjusted R-Squared	0.851	0.859
F-Overall	239.50**	179.48**

Significance: * $p \leq .05$, ** $p \leq .01$; N=294.

Nr.	Author	Title
06-03	Volker Stocké Christian Hunkler	Measures of Desirability Beliefs and their Validity as Indicators for Socially Desirable Responding
06-02	Anders Anderson	Is Online Trading Gambling with Peanuts?
06-01	Volker Stocké Tobias Stark	Political Involvement and Memory Failure as Interdependent Determinants of Vote Overreporting
05-43	Volker Stocké Tobias Stark	Stichprobenverzerrung durch Item-Nonresponse in der international vergleichenden Politikwissenschaft
05-42	Volker Stocké	Response Privacy and Elapsed Time Since
05-41	Josef Hofbauer Jörg Oechssler Frank Riedel	Brown-von Neumann-Nash Dynamics: The Continuous Strategy Case
05-40	Markus Glaser Thomas Langer Jens Reynders Martin Weber	Framing Effects in Stock Market Forecasts: The Difference Between Asking for Prices and Asking for Returns
05-39	Tri Vi Dang	Alternating Offer Bargaining with Endogenous Information: Timing and Surplus Division
05-38	Tri Vi Dang	On Bargaining with Endogenous Information
05-37	Patric Andersson	Overconfident but yet well-calibrated and underconfident: A research note on judgmental miscalibration and flawed self-assessment*
05-36	Peter Dürsch Albert Kolb Jörg Oechssler Burkhard Schipper	Rage Against the Machines:
05-35	Siegfried K. Berninghaus Hans Haller Alexander Outkin	Neural Networks and Contagion

SONDERFORSCHUNGSBereich 504 WORKING PAPER SERIES

Nr.	Author	Title
05-34	Jacques Durieu Hans Haller Philippe Solal	Interaction on Hypergraphs
05-33	Markus Glaser Martin Weber	Which Past Returns Affect Trading Volume?
05-32	Zacharias Sautner Martin Weber	Corporate Governance and the Design of Stock Option Programs
05-31	Zacharias Sautner Martin Weber	Subjective Stock Option Values and Exercise Decisions: Determinants and Consistency
05-30	Patric Andersson Richard Tour	How to Sample Behavior and Emotions of Traders:
05-29	Carsten Schmidt Ro'i Zultan	The Uncontrolled Social Utility Hypothesis Revisited
05-28	Peter Albrecht Joachim Coche Raimond Maurer Ralph Rogalla	Optimal Investment Policies for Hybrid Pension Plans - Analyzing the Perspective of Sponsors and Members
05-27	Oliver Kirchkamp Rosemarie Nagel	Learning and cooperation in network experiments
05-26	Zacharias Sautner Martin Weber	Stock Options and Employee Behavior
05-25	Markus Glaser Thomas Langer Martin Weber	Overconfidence of Professionals and Lay Men: Individual Differences Within and Between Tasks?
05-24	Volker Stocké	Determinanten und Konsequenzen von Nonresponse in egozentrierten Netzwerkstudien
05-23	Lothar Essig	Household Saving in Germany:
05-22	Lothar Essig	Precautionary saving and old-age provisions: Do subjective saving motives measures work?
05-21	Lothar Essig	Imputing total expenditures from a non-exhaustive