

## The Interdependence of Determinants for the Strength and Direction of Social Desirability Bias in Racial Attitude Surveys

Volker Stocké<sup>1</sup>

Empirical evidence suggests that respondents' approval motive, their desirability beliefs, and response privacy determine their susceptibility to social desirability (SD) bias. Previous research has analyzed these factors separately and has not taken their interdependence into account. This article examines the prediction made by rational-choice theory that a strong approval motive, clear differences in the perceived desirability of response options, and a lack of privacy are all necessary but not sufficient conditions for SD-bias. Consistent with these predictions, the empirical results of our first study have shown that a three-way interaction between these factors predicts the respondents' answers about their attitudes toward foreigners. In a second, unrelated study, we tested the critical question whether desirability beliefs predict the strength and direction of privacy effects also when the subjects' desirability concerns are not activated due to asking the same respondents about how socially desirable they regard different response options. We confirmed the results from our first study.

*Key words:* Mode of administration; need for social approval; racial attitudes; rational-choice theory; response bias; trait desirability.

### 1. Introduction

Social desirability bias (SD-bias) originates from shared social norms and the resulting systematic error causes univariate response distributions to deviate from true sample characteristics. Furthermore, associations between the subjects' sociodemographic characteristics and the attributes under consideration are likely to be biased when either the strength or direction of SD-bias differs according to these characteristics. In order to prevent such an impairment of data quality or at least to allow the introduction of appropriate statistical controls, precise knowledge about the determinants of SD-bias is an indispensable precondition. Research has made considerable progress in this direction and suggests that three factors are important for explaining SD-bias. The *first* is the situationally stable but individually differently strong need for social approval (NSA) (Crowne and Marlowe 1960). The *second* is the subjects' feeling of response privacy and thus all features of the response situation preventing others from observing the answers

<sup>1</sup> University of Mannheim, Project Program "Rationality Concepts, Decision Behavior and Economic Modeling," D-68131 Mannheim, Germany. Email: vstocke@rumms.uni-mannheim.de

**Acknowledgments:** We gratefully acknowledge stimulating discussions with Hartmut Esser, Herbert Bless, and Stephan Ganter. The author would furthermore like to thank three anonymous reviewers of this journal for valuable comments. This research was supported by a grant of the German Science Foundation (DFG) to the Sonderforschungsbereich 504 at the University of Mannheim.

(Tourangeau, Rips, and Rasinski 2000, p. 275). The *third* determinant for the strength *and* direction of SD-bias are the respondents' beliefs about the desirability of those traits which they ascribe to their own person when answering a survey question (Edwards 1953). Many studies analyzed these three aforementioned factors separately, but none tested the hypothesis that each may be a moderator variable for how strongly the other determinants predict SD-bias (for exceptions cf. Chen et al. 1997; Phillips and Clancy 1972). Accordingly, a strong approval motive, lack of response privacy, and clearly defined social desirability beliefs (SD-beliefs) are all necessary but insufficient preconditions for SD-bias (Esser 1991). Another open question is whether SD-beliefs, which are in all studies found to be associated with response behavior, are a *causal determinant* of SD-bias. Many subjects can be assumed to have desirable traits, and thus a mere correlation between SD-beliefs and survey answers is neither clear evidence for the presence of SD-bias nor for the predictive power of these beliefs (Johnson 2004). Such evidence is, however, provided when the theoretically expected three-way interaction between SD-beliefs and the other two determinants of SD-bias is empirically found. The main aim of the present study is to test empirically whether this interaction effect explains survey responses about attitudes toward foreigners.

## 2. The Rational-choice Explanation of Socially Desirable Response Behavior

In rational-choice theory (RCT), answering a survey question is assumed to be a goal-directed, instrumentally rational selection between response options (Esser 1991; Stocké 2004a; Tourangeau, Rips, and Rasinski 2000, p. 281). The respondents' aim is to achieve positive and avoid negative evaluation reactions from others in order to maximize their feeling of social approval. Thus, the *first* and *motivational* core of this explanation is the respondents' NSA and their concomitant disposition to apply impression-management strategies.<sup>2</sup> The *cognitive* determinants of SD-bias are the subjects' beliefs about whether their answers are likely to provoke evaluative reactions of others and which kind of reactions these will be. Thus, *secondly*, subjects only expect their answers to be instrumental for reaching social approval when potential addressees are both present *and* able to observe as well as to sanction their answers. This implies that the respondents' feeling of privacy and the expectation that socially desirable answers will have consequences for their approval motive depend completely on the objective ability of others to perceive their answers.<sup>3</sup> The *third* necessary precondition for SD-bias is that respondents perceive sufficiently clear *differences* in the desirability of available response options, and selecting a particular one is therefore instrumentally superior for realizing their approval motive (Esser 1991). The SD-beliefs are a predictor for how others will evaluate different answers. In the case of an unknown interviewer, these beliefs cannot be based on individualized knowledge about

<sup>2</sup>The NSA concept thus refers to the dependency on positive reactions from others and captures an extrinsic motive for impression-management strategies. SD-bias may also result from the intrinsic desire for "self-deception," where the aim is to preserve a positive self-image (cf. for a discussion of both concepts: Paulhus 2003). In contrast to RCT, this theoretical perspective only predicts a two-way interaction between the desirability beliefs and the subjects' need for social approval.

<sup>3</sup>The anonymity of answers, that is the probability that they cannot be identified after the interview, may be relevant for an instrumentally rational respondent as well. This is, however, only likely in case of questions about, for instance, illegal drug use, where legal sanctions are possible.

evaluative criteria. Then, subjects are expected to rely on social norms as a proxy for anticipating the most probable evaluation (Stocké 2004a).

From the perspective of RCT, it is a basic precondition for SD-bias as a meaningful concept that “true scores” of the respondents are assumed to exist, and that these characteristics deviate from those perceived as the most socially desirable. For factual survey questions, the concept of a “true score” is straightforward, and what constitutes a deviation is clearly defined. In the case of attitudes, it is much debated whether something like a unique and situationally invariant attitude exists. In “classical” attitude theory, evaluations are assumed to be invariantly associated with an attitude object, and these associations are stored in memory (Eagly and Chaiken 1993: 1). In contrast, in the attitudes-as-constructions approach, evaluations are the consequence of situational activation sequences and thus always context-dependent (Schwarz and Bohner 2001). Other approaches assume that subjects hold multiple attitudes at the same time, and different context factors determine which one dominates (response) behavior (Wilson, Lindsey, and Schooler 2000). From this perspective, the failure of white voting intentions for an African American candidate, when reported to black interviewers, to predict voting behavior, does not result from SD-bias, but from the fact that survey responses and voting behavior are determined by the respondents’ multiple attitudes toward the candidates (Wilson and Hodges 1992). However, much evidence suggests that particularly less crystallized attitudes are less capable of guiding behavior (Bassili and Bors 1997; Fletcher 2000; Huckfeldt and Sprague 2000). Furthermore, respondents have been shown to overreport their participation in political elections with a higher probability when the survey interview was conducted later after the election and therefore the true behavior was less available in memory (Belli et al. 1999; Stocké and Stark 2007). The empirical support for stronger attitudes being less subject to response effects is found to be mixed: Whereas in some studies, response effects proved to be weaker in the case of stronger attitudes (Lavine et al. 1998; Stocké 2004b), this effect was not found in other analyses (Bassili and Krosnick 2000; Krosnick and Schuman 1988). In the present article, we cannot contribute empirical evidence to the ongoing debate about whether stable “true” attitudes exist. For the sake of theoretical consistency, we nevertheless share the assumption from RCT that attitude expressions represent “true scores” when they are cognitively strong and unaffected by social demands in the response situation (Schuman and Presser 1981).

In summary, the RCT predicts that all three preconditions for SD-bias have to be fulfilled simultaneously: A strong approval motive makes evaluative reactions from others relevant for the respondents. They regard such reactions as possible because of insufficient privacy, and their SD-beliefs suggest that choosing one or another option makes a difference. If only one of these conditions is not given, nothing will affect the prevalence of SD-bias, and subjects are assumed to report their “true scores.” Due to the multiplicative combination of all parameters, the total incentives from social desirability (SEU(SD)) become zero if only one of the parameters in the following equation is zero:

$$\text{SEU(SD)} = U_{\text{SD}} \cdot w_{\text{P}} \cdot \Delta w_{\text{TD}}$$

The *first* parameter,  $U_{\text{SD}}$ , represents the strength of the NSA and is one for subjects with a strong NSA and zero when this motive is absent. The *second* parameter,  $w_{\text{P}}$ , stands for the perceived likelihood that answers can affect the satisfaction of the approval motive. This

parameter is zero in the case of complete privacy and one when others are able to perceive the answers. The *third* parameter,  $\Delta w_{TD}$ , represents the SD-beliefs, or more precisely, the desirability differential of the response options. This parameter varies between minus and plus one, depending on which option is regarded as more desirable. The sign of this difference score predicts the *direction*, and the absolute value predicts the *strength* of incentives for SD-bias. Accordingly, RCT predicts, on the level of statistical analysis, that a three-way interaction between the assumed determinants of SD-bias will explain response behavior. In contrast, other theoretical approaches, like the one assuming that SD-bias is driven by “other-deception,” either only predict a two-way interaction, or are not defined clearly enough to predict a specific kind of interplay between the determinants of SD-bias.

### 3. Previous Empirical Research

In an extensive body of research, the respondents’ approval motive has been measured with different social desirability (SD) scales, as for example the Marlowe-Crowne SD-scale (MC-SD) or the Balanced Inventory of Desirable Responding (BIDR).<sup>4</sup> It has been tested whether subjects with a stronger NSA are more likely to endorse desirable and deny undesirable response options (Crowne and Marlowe 1960). The empirical evidence for this hypothesis is mixed. Respondents’ answers about their agreeableness, a desirable dimension of the big-five personality characteristics, and about their noncognitive abilities were found to correlate positively with their BIDR-values (Graziano and Tobin 2002; Schmitt et al. 2003), whereas the answers of subjects from ten countries about their attitudes toward corruption were negatively associated (Bernardi et al. 2003). It has been furthermore shown that the degree to which women underreported weight and overreported height was significantly predictable with their MC-SD-scores (Larson 2000). The evidence from other studies is negative. Neither the subjects’ reports on the social interest index, an instrument for measuring how important friendship, love, and work are regarded, were associated with BIDR-SD-scores (Leak 2004), nor were the reports of sex workers in the Philippines about their condom use found to correlate with MC-SD-values (Morisky, Ang, and Sneed 2002). In another study, subjects answered the items for their big-five personality dimensions, and proxy reports about these traits were obtained from friends (Pauls and Stemmler 2003). The BIDR-NSA-values of the target sample were correlated as expected with their self-reports, but with the proxy-reports as well. It was concluded that the SD-scale does not measure SD-bias, but other personality characteristics.

In the field of racial attitudes, results about the role of subjects’ approval motive are mixed as well. In a study by Mielke (1995), respondents were asked about their racial attitudes using Pettigrew and Meertens’s “Subtle and Blatant Prejudice” scale. Answers on both instruments were correlated negatively with SD-scale scores. In another study by Duck and Hunsberger (1999), answers on the “Manitoba Prejudice” scale, however, were not associated with the respondents’ scores on the MC-SD-scale.

<sup>4</sup> All scales measure the approval motive in a similar way. Respondents report whether they have characteristics which are socially desirable but unlikely to be true for anybody, or negatively evaluated but true for practically everybody. The more positive traits are endorsed and the more negative ones are denied, the higher SD-scores subjects receive.

In split-ballot experiments with different modes of administration, respondents were found to give more desirable answers under insufficient privacy. For instance, subjects in private audio-computer-assisted self-administered (ACASI) interviews have been found to report more episodes of major depression than in paper-and-pencil-interviews administered by interviewers (Epstein, Barker, and Kroutil 2001). In the UMass Tobacco study, adolescent respondents reported their smoking behavior, either with conventional telephone interviewing (CATI) or with telephone audio-computer-assisted self-interviewing (T-ACASI). While in CATI-interviews, the interviewer and other persons can overhear the answers, T-ACASI uses prerecorded questions and answers are recorded by touch-tone entry. This increased response privacy leads to significantly more reports of having smoked in the past year and month. Privacy effects were stronger for adolescents when the parents strongly disapprove of smoking (Currivan et al. 2004). The same administration modes were compared using samples of the customers of a bank and a fast-food chain (Tourangeau, Steiger, and Wilson 2002). In private T-ACASI interviews, subjects in both samples reported less consumer satisfaction, which was taken as evidence of more honest answers. Another study compared the susceptibility to vote overreporting in three federal elections in Germany when respondents answered the questions either self- or interviewer-administered (Stocké 2007). Subjects were significantly more likely to falsely report having voted in interviewer-administered interviews. Furthermore, the aggregate survey measure for electoral turnout did not differ significantly from the official figures under self-administration, whereas interviewer-administration leads to a significant overestimation of the participation rate.

A study from the U.S. tested whether white respondents' reports about their attitudes toward Afro-Americans differ between interviewer- and self-administered interviews (Krysan 1998). Across all items, racial attitude answers were significantly more negative under guaranteed privacy. On the level of the single items, the strength and partly the direction of privacy effects differed: For only nine out of 19 items, private answers were significantly more negative, and for two items the effect showed a tendency in the opposite direction. This may indicate that the SD-beliefs differ according to specific item contents.

All studies we are aware of confirmed that the respondents' SD-beliefs are substantially associated with their response behavior. Individual differences in how desirable subjects perceived cheating in academic contexts to be (Fernandes and Randall 1992) or having the personality trait of argumentativeness (Nicotera 1996) were positively correlated with the endorsement of the relevant response option in question. These associations do not necessarily provide evidence for the presence of SD-bias and for the explanatory power of SD-beliefs: (a) Asking about SD-beliefs and the responses on the same topic in the same interview may provoke SD-bias which would otherwise not be present (Chen et al. 1997), and (b) the response behavior may be simply consistent with the SD-beliefs because the respondents truly have socially desirable characteristics (Johnson 2004). In order to address the first objection, Huang and colleagues (1998) recorded the SD-ratings of 288 items for testing mental health, and the subjects answered these items in a later session. Despite the elapsed time between the two answers, subjects reported fewer symptoms of mental illness when they judged them earlier to be less desirable. The second objection was tested with data in which self-reports about the number of police arrests were validated with police records. The subjects' SD-beliefs explained successfully how strong and in which direction the answers differed from the true number of arrests (Wyner 1980).

Only two studies analyzed the predicted interdependence of the determinants of SD-bias at least partly. Chen and colleagues (1997) asked subjects to rate the desirability of 45 items measuring positive affectivity, and in a second study a new sample answered these items as well as those of the MC-SD scale. The item-level SD-beliefs predicted the item endorsement, and this correlation increased with the approval motive. In contrast, Phillips and Clancy (1972) did not find an interaction between the subjects' NSA scores and their SD-beliefs. The response behavior on seven sensitive questions, in particular about racial prejudice, was significantly, though independently associated with the SD-beliefs and the MC-SD-scale values.

#### 4. Empirical Study

In the first and main part of our study, the predicted interdependence of the three analyzed determinants of SD-bias is tested with the respondents' answers about their attitudes toward foreigners as a dependent variable. Although we assume the existence of "true" attitudes, information about these "true scores" is not available. Thus, we utilize the theoretically predicted pattern of associations between the attitude answers and the determinants of SD-bias as a criterion for the predictive power of these factors. In a second study, we addressed the open question from the first study whether SD-beliefs really *predict* SD-bias or whether the observed associations are an *artifact* of asking about these beliefs and attitudes in the same interview.

##### 4.1. Sample and Data Collection

The respondents in the main and those in the validation study were a multi-stage, local random probability sample of residents from a metropolitan area in Germany (about 300,000 inhabitants). Households were listed with a random-walk procedure, and individual respondents were selected with the "last-birthday" method. Both the 150 interviews in the main study and the 106 interviews in the validation study were conducted computer-assisted in the respondents' homes by professionally trained and paid interviewers, who were uninformed about the hypothesis of the study. Due to the (quasi-)experimental nature of our research design, the low response rates of 31.3 percent in the main and 39.2 percent in the validation study do not pose a threat to the validity of our results.

At the beginning of the interviews of the main study, subjects reported their SD-beliefs, and 50 unrelated questions later in the questionnaire, they completed the instrument for measuring their approval motive. After 35 other filler items, the respondents answered the ten questions about their attitudes toward foreigners. In the validation study, respondents answered the same attitude questions in the second part of the interview, but no reference was made to their SD-beliefs.

##### 4.2. Measures

###### 4.2.1. Responses about Attitudes toward Foreigners

The dependent variable in our study consists of the answers to ten attitude items used in the German General Social Survey (ALLBUS) in 1997 to measure attitudes toward foreigners

(cf. the item wording in Table 1). Responses were recorded using seven-point Likert-scales with endpoints labeled “completely agree” (1) and “completely disagree” (7). All responses were recoded in such a way that higher values represent more positive attitude answers. Respondents from Study 1 as well as those from Study 2 gave rather positive attitude answers, which were expressed by either agreeing with positively formulated items or disagreeing with negative item content: The average responses in Study 1 varied between 3.3 (Item 1) and 5.4 (Item 3), and in Study 2 between 2.8 (Item 1) and 5.2 (Item 3) (cf. Columns 4 and 5 in Table 1). Despite the same sampling process and population, the answers in Study 1 are significantly more positive for Items 1, 4, 6, and 7 than in Study 2 ( $t$ -values reported in Column 6 in Table 1). The internal consistency (Cronbach’s alpha) for the attitudes-toward-foreigners scale is .89 for Sample 1 and .84 for Sample 2.

#### 4.2.2. SD-beliefs and the Relative SD of Positive and Negative Attitudes ( $\Delta w_{TD}$ )

The SD-beliefs were operationalized as the anticipated emotional reactions when a certain opinion is revealed to an unknown audience (Edwards 1957). Subjects were asked to imagine a situation such as a train journey, where, like in a survey interview, a conversation develops between two strangers. They reported how embarrassing it would be for one of the persons to disclose certain opinions about foreigners (cf. the question wording in the Appendix). The 20 opinions presented in this scenario were those expressed when either disagreeing or agreeing with the ten attitude items used in our study. The *SD-beliefs* were recorded with a bipolar response scale from -4 (statement would be very embarrassing) to +4 (statement would be very pleasant). After the responses were recoded ranging between 0 and 8 for each respondent and attitude item, the judged desirability of negative attitude statements was subtracted from the one of positive statements. The resulting *relative SD-values* range from -8 (negative attitudes more desirable) to +8 (positive attitudes more desirable). A value of zero indicates both attitudes to be equally evaluated. On average across all items and respondents, a relative SD-value of +0.3 indicates that positive attitudes were judged to be slightly more desirable than negative attitudes (cf. Table 1, Columns 1–3). The internal consistency of the relative SD for the 10 attitude items, as indicated by Chronbach’s alpha, is .89.

On the item level, we found for Study 1 a strong and significant correlation of  $r = .82$  ( $p \leq .05$ ) between the attitude answers and the relative SD-values: When positive attitude answers were regarded as relatively more desirable, these answers were more prevalent. Does this association hold as well when the responses are not collected together with the SD-beliefs? A significant correlation of  $r = .66$  ( $p \leq .05$ ) between the SD-beliefs from Study 1 and the responses from Study 2 indicates that this is the case.

When the sign of the relative SD-beliefs differs between the respondents, SD-scores aggregated on the item level underestimate the potential SD-bias, since incentives into different directions cancel each other out. This is clearly what happens for all items of our study: On average across all items, 43.2 percent assume positive attitudes, 36.8 percent negative attitudes and 19.9 percent neither to be more desirable (cf. Table 2). The *strength* of perceived incentives is +3.2 scale points for positive and -3.0 points for negative SD-beliefs, and the two are thus very similar. Therefore, the potential SD-bias in opposite directions is only visible when the absolute values of subjects’ SD-beliefs are utilized:

Table 1. Average SD-beliefs and response behavior for the racial attitude items

	Desirability of positive attitudes <sup>a)</sup>	Desirability of negative attitudes <sup>a)</sup>	Relative desirability <sup>b)</sup>	Response behavior <sup>c)</sup> (Study 1)	Response behavior <sup>c)</sup> (Study 2)	Differences in response behavior <sup>d)</sup> <i>t</i> -value
	Mean (Std.)	Mean (Std.)	Mean (Std.)	Mean (Std.)	Mean (Std.)	
Item 1. Foreigners in Germany should adapt their lifestyle more to the one of Germans.	+0.6 (2.1)	+1.3 (1.9)	-0.7 (2.6)	3.3 (1.8)	2.8 (1.6)	2.3*
Item 2. In the case of increasing unemployment, foreigners should be sent back to their home countries.	+0.3 (2.2)	-0.1 (2.5)	+0.4 (3.4)	5.2 (2.0)	4.9 (1.9)	1.2
Item 3. Foreigners in Germany should marry within their own ethnic community.	+0.7 (2.0)	-0.1 (2.4)	+0.8 (3.3)	5.4 (2.0)	5.2 (2.1)	0.7
Item 4. One should forbid any political activities by foreigners in Germany.	+0.3 (2.1)	-0.4 (2.5)	+0.6 (3.3)	5.1 (2.1)	4.6 (2.1)	2.1*
Item 5. Since there are so many foreigners in Germany, one feels like a stranger in one's own country.	+0.6 (2.2)	+0.2 (2.2)	+0.4 (3.2)	5.0 (2.0)	4.5 (2.1)	1.9
Item 6. Foreigners in Germany are a burden for the social security system.	+0.2 (2.1)	+0.2 (2.4)	+0.1 (3.3)	4.6 (2.0)	4.0 (2.0)	2.2*
Item 7. Foreigners in Germany are enrichment for our culture.	+1.1 (2.2)	-0.5 (2.4)	+1.6 (3.7)	5.2 (1.8)	4.6 (1.6)	2.7*
Item 8. Foreigners in Germany commit more criminal offenses than Germans.	+0.2 (2.2)	+0.6 (2.2)	-0.4 (3.7)	4.3 (1.9)	4.1 (1.9)	1.1
Item 9. The presence of foreigners in Germany causes problems in the housing market.	+0.7 (1.8)	+0.3 (1.9)	+0.4 (2.8)	5.0 (1.7)	4.8 (1.8)	1.1
Item 10. Foreigners in Germany should be entitled to the same social welfare and other social security benefits as Germans.	+0.0 (2.3)	+0.1 (2.4)	-0.1 (3.7)	4.7 (2.0)	4.6 (1.9)	0.3
Total	+0.5 (1.3)	+0.2 (1.6)	+0.3 (3.4)	4.8 (2.0)	4.4 (2.0)	

All statistics are based on  $N = 150$ , except the response behavior in Study 2, where the sample size is 106. <sup>a)</sup> Scale ranges from -4 (undesirable) to +4 (desirable). <sup>b)</sup> Scale ranges from -8 (negative attitude more desirable) to +8 (positive attitude more desirable). <sup>c)</sup> Scale ranges from 1 (negative answer) to 7 (positive answer). <sup>d)</sup> Significance:  $*p \leq 0.05$ .

Table 2. Heterogeneity of relative SD-beliefs about positive and negative racial attitudes

	Positive attitudes more desirable <sup>a)</sup>		No differences	Negative attitudes more desirable <sup>a)</sup>		Total incentives <sup>b)</sup>
	%	Mean (Std.)		%	Mean (Std.)	
Item 1	28.0	2.3 (1.4)	24.0	48.0	-2.7 (1.9)	2.0 (1.9)
Item 2	45.3	3.2 (2.3)	21.3	33.3	-3.1 (2.1)	2.5 (2.3)
Item 3	47.3	3.5 (2.3)	20.7	32.0	-2.8 (1.6)	2.6 (2.2)
Item 4	50.0	3.2 (2.1)	19.3	30.7	-3.1 (2.0)	2.5 (2.2)
Item 5	46.7	3.1 (2.0)	14.0	39.3	-2.7 (1.8)	2.5 (2.1)
Item 6	42.7	3.0 (2.0)	17.3	40.0	-3.1 (2.0)	2.5 (2.2)
Item 7	58.7	4.1 (2.2)	18.0	23.3	-3.3 (2.2)	3.1 (2.5)
Item 8	32.7	3.6 (2.1)	18.7	48.7	-3.4 (2.1)	2.8 (2.3)
Item 9	41.3	2.8 (2.0)	27.3	31.3	-2.6 (1.7)	2.0 (2.0)
Item 10	39.3	3.4 (2.1)	18.7	42.0	-3.5 (2.1)	2.8 (2.3)
Total	43.2	3.2 (2.0)	19.9	36.9	-3.0 (2.0)	2.5 (2.2)
N	150					

<sup>a)</sup>Values range from larger than zero to +8 (positive attitudes more desirable) and from smaller than zero to -8 (negative attitudes more desirable). <sup>b)</sup>Total incentives are the average of respondents' absolute values of their relative SD-scores. Subjects with no desirability differences are included in the measure. Variable ranges from 0 (no incentives) to 8 (strong incentives).

The average of this measure, ranging between 0 and 8, varies between 2.0 (Item 1) and 3.1 (Item 7).

#### 4.2.3. Need for Social Approval ( $U_{SD}$ )

Individual differences in the approval motive were measured using a 10-item short form of the MC-SD-scale (cf. question wording in the Appendix). To allow for the subjects' approval motives to be captured as fully as possible by their MC-SD-scores, responses were collected interviewer-administered and thus under high incentives to employ impression-management strategies. Forced-choice responses on all items were recoded in such a way that answers indicating a low approval motive were coded as zero and those indicating a strong motive as one. We computed an additive index, which has a value range from 0 (no approval motive) to 10 (strong approval motive) and a mean of 5.7 (std. = 2.2). The Kuder-Richardson reliability coefficient is .61, which indicates a sufficient internal consistency for dichotomous items.

#### 4.2.4. Privacy of Response Situation ( $w_p$ )

Whether the respondents' attitude answers were private or discernable by others was operationalized in Study 1 as well as in Study 2 by using either self- or interviewer-administered modes of data collection. First, the SD-beliefs and the MC-SD scale were always interviewer-administered (CAPI). Before the questions about the attitudes toward foreigners were asked, the interview software randomly assigned respondents to an interviewer- or self-administrated mode (CASI). In the former case, the interviewer continued to conduct the interviews, whereas under the latter condition, the respondents continued to complete the questionnaire alone. While the respondents read the questions

from the computer screen and typed in the answers, the interviewer remained present, but maintained enough distance not to be able to observe the answers.

## 5. Results

### 5.1. Testing the Hypothesis with Individual-level Data

According to RCT, (a) the sign and absolute value of the subjects' relative SD-beliefs are expected to predict their attitude responses, (b) this correlation represents SD-bias when it is substantially larger under interviewer- compared with self-administration, and (c) the two-way interaction assumed under (b) should be significantly more pronounced when the approval motive is stronger. In statistical terms, this implies that a three-way interaction between the determinants of SD-bias should predict the response behavior. We thus computed a multiplicative term between the three variables, as well as between all their two-way combinations. Before doing so, the continuous SD-beliefs and NSA-scores were *z*-standardized in order to avoid high multicollinearity between the multiplicative terms and the main effects as well as to bring the parameters into a comparable metric (Cronbach 1987). Furthermore, the data was rearranged in such a way that the observations represent the subjects' answers on all ten attitude questions and the same respondents' relative SDs for the respective items. Variables defined on the respondent level, such as response privacy and approval motive, have identical values for each observation belonging to the same respondent. Note that the number of observations in this analysis ( $N = 1,477$ ) is less than 1,500 because a couple of respondents left some of the attitude questions unanswered. Since the observations in this dataset are not independent, and thus standard errors tend to be underestimated, the *t*-statistics in all the following analyses are corrected using Huber-White estimators for robust standard errors with the respondents as clusters (STATA Corporation 1999: 165 ff.).

The results from regression Model 1, with only the subjects' socioeconomic characteristics included, show that males and respondents with only compulsory education reported significantly more negative attitudes than females and subjects with either a university or a vocational college degree (cf. Table 3).<sup>5</sup> Similar group differences have been reported in studies about racial attitudes (Hudson and Hines-Hudson 1999). Model 2 tests the independent effects of the analyzed determinants for SD-bias and thus an incompletely specified model of SD-bias, as it was utilized in previous studies. As a *first* result and consistently with empirical evidence from studies with other questionnaire topics reported in Section 3, the relative SD-beliefs proved to be a strong predictor for the attitude answers: These answers were the more positive, the more the respondents perceived such responses to be more desirable than negative answers. *Secondly*, the approval motive is significantly associated with the response behavior. The negative regression parameter, however, suggests that a stronger approval motive and thus SD-bias leads to more *negative* attitude answers. According to the *third* result, response privacy has

<sup>5</sup> Nine dummy variables for the ten attitude items were included in all regression models in Table 3 in order to control for response differences between the attitude questions. The parameters are not reported because of space limitations.

Table 3. Test of predicted determinants for the racial attitude answers and their interaction (OLS-regression)

	Model 1 B (t-value)	Model 2 B (t-value)	Model 3 B (t-value)
<i>Control variables</i>			
AGE (years)	-.01 (1.2)	-.01 (0.8)	-.01 (0.9)
INCOME (in 1,000 Marks)	.09 (1.9)	.10 (2.2)*	.09 (2.0)*
MALE <sup>a)</sup>	-.42 (2.0)*	-.33 (1.6)	-.32 (1.6)
EDUCATION <sup>b)</sup>			
- Secondary school certificate	.19 (0.7)	.14 (0.6)	.13 (0.5)
- High school certificate	.51 (1.0)	.33 (0.7)	.33 (0.8)
- Vocational college degree	1.15 (3.0)*	.91 (3.1)*	.85 (2.6)*
- University degree	1.06 (3.4)*	.77 (2.8)*	.75 (2.7)*
STATUS <sup>c)</sup>			
- White-collar worker	.46 (1.1)	.45 (1.0)	.49 (1.1)
- Self-employed	.49 (1.0)	.57 (1.1)	.57 (1.1)
- Never been in workforce	.91 (1.7)	.71 (1.2)	.71 (1.3)
<i>Rational-Choice Model</i>			
Relative Desirability	-	.36 (5.2)*	.21 (2.7)*
Need for Social Approval	-	-.25 (2.4)*	-.17 (1.2)
Interviewer Administered Mode <sup>d)</sup>	-	.09 (0.4)	.11 (0.5)
Desirability • Mode	-	-	.25 (2.3)*
Need • Mode	-	-	-.11 (1.6)
Need • Desirability	-	-	-.11 (0.7)
Need • Desirability • Mode	-	-	.30 (3.1)*
Constant	2.63 (4.7)*	2.60 (4.8)*	2.62 (4.9)*
Corrected R <sup>2</sup>	.19	.23	.24
Observations	1,477	1,477	1,477

Significance: \* $p \leq 0.05$ . Omitted categories: <sup>a)</sup>female; <sup>b)</sup>compulsory education; <sup>c)</sup>blue-collar worker; <sup>d)</sup>self-administered.

a weak and nonsignificant effect on response behavior, which suggests a complete *absence* of SD-bias. In Model 3, the three-way interaction term between all determinants for SD-bias and all lower-level interaction effects are entered into the regression equation. As theoretically expected, the three-way interaction parameter is a statistically significant predictor for the respondents' answers about their attitudes toward foreigners. Furthermore, we found a significant positive conditional main effect of desirability beliefs and a positive conditional two-way interaction between these beliefs and the mode of administration. Taking regression Model 1 as a starting point, Model 3 with all theoretically predicted explanation factors included leads to a significantly improved explanation of the attitude answers ( $F(6.0, 1, 247.9) = 13.6; p \leq .01$ ).<sup>6</sup>

In order to allow for an interpretation of the three-way interaction, we computed predicted attitude scores for relevant combinations of the three determinants of SD-bias

<sup>6</sup>The appropriate degrees of freedom for the  $F$ -test were obtained by weighting the inflated degrees of freedom based on the number of observations with a Greenhouse-Geisser's Epsilon of .86. The difference of this value from unity reflects to what degree the variance observed in the data is attributable to within-respondents variation and thus how strongly the assumption of independence of observations is violated (Stevens 1996, 459ff.).

Table 4. Responses about racial attitudes according to the mode of administration, the relative SD-beliefs and subjects' approval motives (predicted values from regression Model 3)

	<i>Weak approval motive</i>			<i>Strong approval motive</i>		
	Mode of administration			Mode of administration		
	Self-administered	Interviewer-administered	Mode difference	Self-administered	Interviewer-administered	Mode difference
<i>Relative desirability of racial attitudes</i>						
Positive attitude more desirable	3.2	3.3	+0.1	2.6	3.2	+0.6
Positive/negative attitudes equally desirable	2.9	3.1	+0.2	2.5	2.5	0.0
Negative attitude more desirable	2.5	2.8	+0.3	2.4	1.9	-0.5

Racial attitude scores vary between 1 (negative attitude) and 7 (positive attitude). Predicted values represent subjects with a standard deviation above (strong approval motive) and below (weak approval motive) the sample mean of MC-SD scores. For the relative desirability values of one standard deviation above (positive attitudes more desirable) and below (negative attitudes more desirable), the population mean is inserted into the equation. The category "positive/negative attitudes equally desirable" represents the sample mean of this dimension. Control variables are fixed on the sample mean for continuous variables and on the reference category for categorical variables.

(cf. Table 4). The results show *firstly* that answers from respondents with a weak approval motive, as indicated by MC-SD-scores of one standard deviation under the sample mean, are influenced only to a very limited extent by the privacy conditions. Nevertheless, observed differences are not in agreement with the direction of the SD-beliefs: Independently of which attitude answer is regarded as more desirable, responses are between 0.1 and 0.3 scale points more positive under interviewer administration. *Secondly*, subjects with MC-SD-values of one standard deviation above the sample mean are much more responsive to differences in response privacy and their relative SD-beliefs. Subjects who assume positive attitudes to be more desirable gave in interviewer- as compared with self-administered interviews 0.6 scale points more positive answers, whereas respondents with opposite SD-beliefs demonstrate an equivalent privacy effect of 0.5 scale points in the negative direction. As theoretically expected, the administration mode is practically irrelevant for subjects who do not perceive any SD-differences between positive and negative attitudes toward foreigners. *Thirdly*, the conditional effects of desirability beliefs and of the two-way interaction between these beliefs and the administration mode are both visible in Table 4 as well. In the case of a weak approval motive and self-administration (both variables have low values), the effect of desirability beliefs is 0.2 scale points stronger than under interviewer-administration. Also, in the case of a strong approval motive, the desirability beliefs have an effect that is 1.1 scale points stronger under interviewer- as compared with self-administration, whereas this difference is only 0.2 scale points into the opposite direction for a weak approval motive. Thus, the two conditional effects indicate a weak effect in an unexpected and a strong effect in the theoretically expected direction.

Our results from Study 1 support the predictions from RCT about the interdependence of all three analyzed determinants of SD-bias. It is a particularly remarkable result that a substantial percentage of subjects believe negative rather than positive attitudes toward foreigners to be more socially desirable *and* that insufficient response privacy leads them to bias their responses in this direction. As it has been criticized in the case of other studies, it is not clear whether SD-beliefs predict the observed privacy effects on the response behavior if the desirability questions are not asked in the same interview. One might suspect that being involved in thinking about the social desirability of attitude responses before reporting on one's own attitudes might cause SD-bias which would not have been present in a "normal" survey interview (Chen et al. 1997). This issue of the external validity of our results is addressed in the following, second part of the study.

In the validation study, a new sample of respondents answered the same attitude questions under varying degrees of response privacy, but were not asked about their SD-beliefs on this topic. Instead, we matched aggregated SD-differences in these beliefs between items and sociodemographic groups, which we found in the first study, with the attitude reports of equivalent groups on the same items in Study 2. Due to the restricted variance in the case of aggregated data, the aim was not a replication of the three-way interaction we found with individual-level data. Instead, we "only" tested the two-way interaction between the SD-beliefs and the response privacy. This, rather than the alternative interaction with the approval motive, was chosen for validation, since the experimental nature of privacy differences excludes the possibility that characteristics confounded with the approval motive would bias the results.

## 5.2. Differences in SD-beliefs between Items and Social Groups

In a first step, we regressed the relative SD-beliefs for all ten attitude items observed in the first study on a set of item dummies and indicators for the respondents' demographic characteristics (cf. Model 4, Table 5). Again, the dependent variable consists of multiple observations from each respondent, and thus the standard errors of the regression parameters were corrected as described in the previous section. According to our *first* result, positive attitude answers were perceived as significantly more desirable in the case of Item 1 (reference category) than in that of Item 7, whereas the remaining items are located on a continuum between these two extremes. The *second* result is that the SD-beliefs do not differ according to the respondents' age, income, or sex. *Thirdly*, blue-collar workers as well as subjects who had never been in the workforce perceived significantly stronger incentives for positive attitude answers than did self-employed and white-collar workers. *Furthermore*, respondents with compulsory education and a vocational college degree held more negative SD-beliefs than did those with a university

Table 5. Effects of items and the respondents' demographic characteristics on their relative SD-beliefs (OLS-regression)

	Model 4 B (t-value)		Model 5 B (t-value)
Item differences <sup>a)</sup>		Item differences <sup>a)</sup>	
- Item 2	1.06 (3.7)*	- Item 2	1.06 (3.7)*
- Item 3	1.42 (4.4)*	- Item 3	1.42 (4.5)*
- Item 4	1.30 (4.3)*	- Item 4	1.30 (4.3)*
- Item 5	1.07 (3.7)*	- Item 5	1.07 (3.7)*
- Item 6	0.72 (2.7)*	- Item 6	0.72 (2.7)*
- Item 7	2.25 (7.0)*	- Item 7	2.25 (7.0)*
- Item 8	0.23 (0.7)	- Item 8	0.23 (0.7)
- Item 9	1.02 (4.1)*	- Item 9	1.02 (4.1)*
- Item 10	0.56 (1.7)	- Item 10	0.56 (1.7)
Group differences		Group differences	
Age (years)	0.00 (0.3)	Age (years)	–
Income (thousand Marks)	–0.12 (0.9)	Income (thousand Marks)	–
Male <sup>b)</sup>	–0.30 (0.7)	Male <sup>b)</sup>	–
Education <sup>c)</sup>		Education <sup>c)</sup>	
- Compulsory education	–1.49 (2.0)*	} Compulsory education & } Vocational college degree	–0.84 (2.3)*
- Vocational college degree	–1.93 (1.8)		
- Secondary school certificate	–1.03 (1.4)		–
- High school certificate	–0.81 (1.0)		–
Social status <sup>d)</sup>		Social status <sup>d)</sup>	
- Blue-collar worker	1.62 (2.3)*	} Blue-collar worker & Never } been in workforce	1.24 (3.2)*
- Never been in workforce	2.00 (2.7)*		
- White-collar worker	0.79 (1.4)		–
Corrected R <sup>2</sup>	0.08		0.06
Observations	1,500		1,500

Significance: \* $p \leq 0.05$ . Omitted categories: <sup>a)</sup> Item 1; <sup>b)</sup> female; <sup>c)</sup> university degree; <sup>d)</sup> self-employed; <sup>e)</sup> secondary school certificate and high school certificate and university degree; <sup>f)</sup> white-collar worker and self-employed.

degree. For the final regression equation, we dropped the characteristics which have proven to be irrelevant, and grouped together those status and educational groups which did not differ significantly. The resulting dummy variables were significant predictors for the SD-beliefs (cf. Model 5, Table 5). Note that the reduced parameter size and explained variance of the final model results from the somewhat mixed reference categories of the remaining dimensions in the analysis.

### 5.3. Results from the Validation Study

In the final step of our study, we *first* computed, separately for each of the ten attitude items, average SD-beliefs for the possible combinations of those demographic characteristics found to be significant predictors for these beliefs. As a result, we obtained 2 (occupational status groups) times 2 (educational status groups) times 10 (attitude items), and thus altogether 40 mean desirability scores. These scores, based on the data from Study 1, varied between  $-1.3$  and  $+4.7$ , with an average value of  $.52$ . In a *second* step, we computed average racial attitude answers for exactly the same combinations of demographic characteristics and attitude items. This was done with data from Study 2 separately for answers obtained interviewer- and self-administered. As a result, we obtained 2 times 40 average attitude scores, which varied under self-administration between 1.9 and 5.5, with an average of 4.2, and for interviewer-administration between 2.5 and 5.8, with a mean value of 4.4. We created a dataset where these 80 attitude scores defined the observations, and a newly created variable indicating whether an observation stems from self- or interviewer-administered interviews (CASI coded 0, CAPI coded 1). A third variable contained the 40 desirability scores, which have identical values for the two privacy conditions.

The SD-beliefs are expected to explain the response behavior significantly more strongly under interviewer- than under self-administration. We thus computed a multiplicative term between both variables and tested its ability to explain the attitude answers. *Firstly*, consistent with the results found with individual-level data in Study 1, the SD-beliefs proved to be a significant predictor of response behavior, but the response privacy alone did not (cf. Table 6, Model 6). *Secondly*, and also consistent with the results from Study 1, the interaction term between the two factors had a significant effect on the attitude responses ( $t = 14.3$ ;  $p \leq .05$ ) (cf. Table 6, Model 7).

In Table 7, the interaction effect from regression Model 7 is presented using predicted attitude responses for subjects who believe positive, negative or neither attitudes to be more desirable and who were either tested self- or interviewer-administered. When positive attitudes are believed to be more socially desirable, this leads to 0.4 scale points more *positive* attitude responses under interviewer- than under self-administration. In the case of SD-beliefs in the reverse direction, the same privacy effect is 0.5 scale points in the direction of more *negative* attitude responses under interviewer-administration. When all attitudes were assumed to be equally desirable, response privacy had hardly any effect on response behavior. Accordingly, the SD-beliefs predict how strongly and in which direction the interviewers' ability to perceive and sanction the answers influences the attitude responses.

Table 6. Aggregated response behavior from Study 2 as a function of administration mode and aggregated relative desirability ratings from Study 1 (OLS-regression)

	Model 6 B (t-value)	Model 7 B (t-value)
Relative Desirability (group means)	.26 (4.9) *	.09 (2.2)
Interviewer-Administered Mode <sup>a)</sup>	.13 (0.4)	-.04 (0.2)
Desirability Mode	-	.33 (14.3) *
Constant	4.10 (15.7) *	4.19 (20.1) *
Corrected $R^2$	0.11	0.16
Observations	80	80

Significance: \*  $p \leq 0.05$ . Omitted category: <sup>a)</sup> self-administered mode

## 6. Summary and Discussion

The aim of this study was to test the hypothesis from Rational-Choice Theory (RCT) that the respondents' approval motive, their beliefs about the relative desirability of response options, and the response privacy are mutually interdependent determinants for the strength and direction of SD-bias. In the first part of our study, this was done with individual-level response data about attitudes toward foreigners. As a *first* result, we found considerable heterogeneity about whether respondents believe positive or negative attitude answers to be evaluated more desirably in society, and how strong these differences are. Thus, an appropriate model of SD-bias needs to take such differences explicitly and in the form of a theoretical parameter into account. *Secondly*, the respondents' attitude answers were found to be strongly associated with their SD-beliefs, and this association was significantly stronger when response privacy was low and subjects had a strong approval motive. This theoretically predicted three-way interaction provides evidence for the appropriateness of RCT as an explanation for impression management-driven SD-bias: The respondents' inner beliefs *and* motives *as well as* the opportunities for self-presentation provided in the response situation are necessary, but on their own not sufficient determinants for SD-bias. This bias has to be expected "only" under the special constellation when respondents perceive clear desirability differences, when they

Table 7. Aggregated racial attitude answers as a function of the administration mode and subjects' relative SD-beliefs (predicted values from regression Model 7)

	Mode of administration		
	Self-administered	Interviewer-administered	Mode difference
<i>Relative desirability of racial attitudes</i>			
Positive attitude more desirable	4.3	4.7	+0.4
Positive and negative attitudes equally desirable	4.2	4.2	0.0
Negative attitude more desirable	4.1	3.6	-0.5

Aggregated racial attitude scores can vary between 1 (negative attitude) and 7 (positive attitude). Predicted values represent subjects with relative desirability scores of one standard deviation above (positive attitudes more desirable) and below (negative attitudes more desirable) the observed desirability distribution. A value of zero has been inserted for the condition 'positive and negative attitudes equally desirable'.

simultaneously regard social approval as an important value, and when others are at the same time able to provide such approval.

A possible objection against our model of SD-bias could be that the predicted interaction between all determinants of SD-bias adds only a small, even though significant proportion of explained variance to a theoretically more parsimonious explanation with main effects only. *First*, however, only the completely specified model represents a notion about which subjective meaning the respondents attach to the process of answering survey questions. This meaning is an instrumentally rational decision between response options. *Second*, an incompletely specified model, as it was utilized in most previous research, has proven in our study to lead to wrong conclusions about the presence, nature and determinants of SD-bias. Thus, the approval motive alone was found to be negatively correlated with the attitude responses, which would have suggested that SD-bias is in general in the direction of *negative* attitude answers. However, our complete model of SD-bias proved that a stronger approval motive intensified response bias in the direction of whatever answer subjects believed to be more desirable. *Third*, the isolated effect of response privacy was found to be insignificant, which would lead to the conclusion of a complete *absence* of SD-bias. However, this result was due to privacy effects in the direction of positive *and* negative attitude answers, which canceled each other out on the aggregate level of the complete sample. This would not have been detected without simultaneously taking the heterogeneity of the respondents' desirability beliefs into account. Thus, beyond the issue of statistical explanation, only the more complex model of SD-bias predicted from RCT leads to a correct and precise picture of the determinants for the strength and direction of SD-bias.

In the first part of our study, we asked the subjects in the same interview about their own attitudes toward foreigners and about how socially desirable they regard these attitudes in society. Answering the desirability questions may have made the underlying social norms more salient than would have otherwise been the case. Thus, the observed associations may be an artifact of our within-subjects research design, and the evidence supporting RCT not externally valid (Chen et al. 1997). In order to exclude this possibility, we undertook a partial replication of the results in a second study. We utilized the aggregated SD-beliefs for different attitude items and demographic groups from Study 1 and tested with a quasi-between-subjects design whether these beliefs predict the strength and direction of privacy effects on the attitude answers of a new sample. This has been found to be the case. Thus, consistently with the results from the first part of our study and with the predictions from RCT, the respondents' answers were significantly more in line with their SD-beliefs when the interviewer was able to perceive and potentially sanction the response behavior.

Our results have consequences for research about the determinants of SD-bias as well as for how to reduce the emergence of this bias during data collection and for which correction strategies can be regarded as appropriate during data analysis (Ellingson, Sackett, and Hough 1999). *First*, results about the isolated effect of the response privacy on the aggregated response distributions have to be interpreted with caution. Weak or absent privacy effects, as they were observed in previous studies (cf. Aquilino 1994), do not provide conclusive evidence for the absence of SD-bias. This issue can only be decided when the strength as well as the direction of the respondents' SD-beliefs and, in particular, group differences in this respect are taken into account at the same time. *Second*, according to the theoretical assumptions of RCT and confirmed in our study, testing the effect of the subjects' need for

social approval on privately collected response behavior does not provide evidence for the presence and direction of impression management-driven SD-bias. Without an “audience,” even a strong approval motive does not bias the response behavior. This explains negative results about the explanatory power of this factor with data from self-completed questionnaires (cf. for instance Leak 2004). *Third*, our results have shown that respondents under ensured privacy answer substantially less in conformity with what they believe to be desirable in society. In the case of topics where the desirability of possible responses differs considerably, survey researchers should always provide a private response situation. *Fourth*, it is common practice in survey research to statistically control for measures of the respondents’ need for social approval in order to eliminate response bias, without taking the response privacy and a possible heterogeneity of the respondents’ SD-beliefs into account (Ellingson, Sackett, and Hough 1999). However, our results have shown that, under most value combinations of these two factors, no SD-bias is to be expected. When there are shared variance components between the SD-scale scores and the target construct, controlling for these SD-scores has the undesirable effect of discarding these variance components.

The predictions obtained from RCT were found to be *correct* in the present study. There is, however, reason to doubt the *completeness* of the theory for explaining all aspects of response behavior. *First*, this approach explains the decision between response options in the fourth stage of the total process of answering survey questions (Tourangeau and Rasinski 1988). Yet, this decision is based on the results from the three stages before: comprehension of the question, retrieval of relevant information from memory, and forming of a judgment. RCT is not able to explain the results from these processes. *Second*, RCT has little to say about how the respondents arrive at their subjective SD-beliefs and how they obtain differently strong approval motives. Here, a theoretical alliance with approaches from social and cognitive psychology is necessary in order to reach a more complete explanation. *Third*, RCT assumes that subjects are always perfectly able and motivated to select the optimal response option. In the case of survey response behavior, when selecting an answer rarely has serious consequences, satisficing rather than maximizing may be the more realistic approach (Krosnick 1991).

Some questions have remained unanswered in our study, which deserve attention in subsequent research. *First*, the respondents’ subjective feelings of privacy were assumed to be completely determined by the objective features of the response situation. Some results, however, have shown that the mere presence of other persons in the interview situation, objectively unable to perceive the answers, can affect the response behavior (Aquilino, Wright, and Supple 2000). This may indicate in certain cases a systematic difference between subjective feelings and objective conditions of privacy. The manipulation of privacy in our study maximized the objective, but not necessarily the respondents’ subjective feeling of privacy. Thus, we cannot exclude that having maximized the difference in subjective and objective privacy at the same time would have led to the detection of an even stronger SD-bias. This issue is worth being explored in future research. *Second*, we operationalized the respondents’ SD-beliefs as the desirability difference between extremely positive and negative attitude answers. Compared with other studies, where only the desirability of one of the extreme endpoints of the desirability continuum is utilized, this is an unusually differentiated measure. However, for other questionnaire topics, a substantial proportion of respondents were found to have inversely U-shaped

desirability profiles across the attribute continuum (Stocké and Hunkler 2007). If this applies in the case of attitudes toward foreigners, our indicator for SD-beliefs would underestimate the incentives from SD. This issue should be resolved empirically. *Third*, our results are based on relatively small local probability samples. Although our main conclusions rest upon the structure of associations centered around experimentally induced privacy conditions, the reported distributions of explanatory variables cannot simply be generalized to the general population. The study should thus be replicated, using representative nationwide samples.

## Appendix

### *Instrument for Measuring the SD-beliefs about Positive and Negative Racial Attitudes*

“Please imagine a person on a train journey, having a discussion with an unknown fellow passenger about foreigners living in Germany. This person has certain opinions about this topic. In the following, I would like to ask you about whether you believe it would be embarrassing or pleasant for this person to state the following opinions.”

SD-beliefs about a strong AGREEMENT and *disagreement* with item (example for Item 1):

“Would it be embarrassing or pleasant for the person mentioned above to express in public the clear opinion that foreigners in Germany should ADAPT/*not be forced to adapt* their lifestyle more to the one of Germans?”

### *Short Version of the MC-SD Scale*

“I would like to ask you, whether the following statements are an accurate description of your person.”

1. “Before voting, I thoroughly investigate the qualifications of all the candidates.”
2. “I sometimes feel resentful when I don’t get my way.”
3. “No matter who I’m talking to, I’m always a good listener.”
4. “I can remember ‘playing sick’ to get out of something.”
5. “There have been occasions when I took advantage of someone.”
6. “I’m always willing to admit when I make a mistake.”
7. “I always try to practice what I preach.”
8. “I am always courteous, even to people who are disagreeable.”
9. “I am sometimes irritated by people who ask me favors.”
10. “I have never deliberately said something that hurt someone’s feelings.”

## 7. References

- Aquilino, W.S. (1994). Interview Mode Effects in Surveys of Drug and Alcohol Use. *Public Opinion Quarterly*, 58, 210–240.
- Aquilino, W.S., Wright, D.L., and Supple, A.J. (2000). Response Effects Due to Bystander Presence in CASI and Paper-and-Pencil Surveys of Drug Use and Alcohol Use. *Substance Use and Misuse*, 35, 845–867.

- Bassili, J.N. and Bors, D.A. (1997). Using Response-Latency to Increase Lead Time in Election Forecasting. *Canadian Journal of Behavioural Science*, 29, 231–238.
- Bassili, J.N. and Krosnick, J.A. (2000). Do Strength-Related Attitude Properties Determine Susceptibility to Response Effects? New Evidence from Response Latency, Attitude Extremity, and Aggregate Indices. *Political Psychology*, 21, 107–132.
- Belli, R.F., Traugott, M.W., Young, M., and McGonagle, K.A. (1999). Reducing Vote Overreporting in Surveys. Social Desirability, Memory Failure, and Source Monitoring. *Public Opinion Quarterly*, 63, 90–108.
- Bernardi, R.A., Delorey, E.L., LaCross, C.C., and Waite, R.A. (2003). Evidence of Social Desirability Response Bias in Ethics Research: An International Study. *Journal of Applied Business Research*, 19, 41–51.
- 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.
- Cronbach, L.J. (1987). Statistical Tests for Moderator Variables: Flaws in Analyses Recently Proposed. *Psychological Bulletin*, 102, 414–417.
- Crowne, D.P. and Marlowe, D. (1960). A New Scale of Social Desirability Independent of Psychopathology. *Journal of Consulting Psychology*, 24, 349–354.
- Currihan, D.B., Nyman, A.L., Turner, C.F., and Biener, L. (2004). Does Telephone Audio Computer-Assisted Self-Interviewing Improve the Accuracy of Prevalence Estimates of Youth Smoking? Evidence from the UMass Tobacco Study. *Public Opinion Quarterly*, 68, 542–564.
- Duck, R.J. and Hunsberger, B. (1999). Religious Orientation and Prejudice: The Role of Religious Proscription. Right-Wing Authoritarianism. *The International Journal for the Psychology of Religion*, 9, 157–179.
- Eagly, A.H. and Chaiken, S. (1993). *The Psychology of Attitudes*. Fort Worth: Harcourt Brace Jovanovich.
- 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: The Dryden Press.
- Ellingson, J.E., Sackett, P.R., and Hough, L.M. (1999). Social Desirability Corrections in Personality Measurement: Issues of Applicant Comparison and Construct Validity. *Journal of Applied Psychology*, 84, 155–166.
- Epstein, J.F., Barker, P.R., and Kroutil, L.A. (2001). Mode Effects in Self-Reported Mental Health Data. *Public Opinion Quarterly*, 65, 529–549.
- Esser, H. (1991). Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als "Rational Choice". In *Person – Situation – Institution – Kultur*. Günter Büschges zum 65. Geburtstag, Hrsg. R. Wittenberg, Berlin: Duncker and Humblot. [In German]
- Fernandes, M.F. and Randall, D.M. (1992). The Nature of Social Desirability Response Effects in Ethics Research. *Business Ethics Quarterly*, 2, 183–205.
- Fletcher, J.F. (2000). Two Timing: Politics and Response Latencies in a Bilingual Survey. *Political Psychology*, 21, 27–55.
- Graziano, W.G., and Tobin, R.M. (2002). Agreeableness: Dimension of Personality or Social Desirability Artifact? *Journal of Personality*, 70, 695–727.

- 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.
- Huckfeldt, R. and Sprague, J. (2000). Political Consequences of Inconsistency: The Accessibility and Stability of Abortion Attitudes. *Political Psychology*, 21, 57–79.
- Hudson, B.J. and Hines-Hudson, B.M. (1999). A Study of the Contemporary Racial Attitudes of Whites and African Americans. *The Western Journal of Black Studies*, 23, 22–34.
- Johnson, J.A. (2004). The Impact of Item Characteristics on Item and Scale Validity. *Multivariate Behavioral Research*, 39, 273–302.
- Krosnick, J.A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5, 213–236.
- Krosnick, J.A. and Schuman, H. (1988). Attitude Intensity, Importance, and Certainty and Susceptibility to Response Effects. *Journal of Personality and Social Psychology*, 54, 940–952.
- Krysan, M. (1998). Privacy and the Expression of White Racial Attitudes: A Comparison across Three Contexts. *Public Opinion Quarterly*, 62, 506–544.
- Larson, M.R. (2000). Social Desirability and Self-Reported Weight and Height. *International Journal of Obesity*, 24, 663–665.
- Lavine, H., Huff, J.W., Wagner, S.H., and Sweeney, D. (1998). The Moderating Influence of Attitude Strength on Susceptibility to Context Effects in Attitude Surveys. *Journal of Personality and Social Psychology*, 75, 359–373.
- Leak, G.K. (2004). Clarification of the Link between Socially Desirable Responding and the Social Interest Index. *Journal of Individual Psychology*, 60, 94–99.
- Mielke, R. (1995). Der Bildungsgrad und die Einstellung gegenüber Ausländern. Universität Bielefeld: Bielefelder Arbeiten zur Sozialpsychologie 173. [In German]
- Morisky, D.E., Ang, A., and Sneed, C.D. (2002). Validating the Effects of Social Desirability on Self-Reported Condom Use Behavior among Commercial Sex Workers. *AIDS Education and Prevention*, 14, 351–360.
- 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 *Encyclopedia of Psychological Assessment*, R. Fernandez-Ballesteros (Ed.). Thousand Oaks, CA: Sage.
- Pauls, C.A. and Stemmler, G. (2003). Substance and Bias in Social Desirability Responding. *Personality and Individual Differences*, 35, 263–275.
- Phillips, D.L. and Clancy, K.J. (1972). Some Effects of “Social Desirability” in Survey Studies. *American Journal of Sociology*, 77, 921–940.
- Schmitt, N., Oswald, F.L., Kim, B.H., Gillespie, M.A., Ramsay, L.J., and Yoo, T.-Y. (2003). Impact of Elaboration on Socially Desirable Responding and the Validity of Biodata Measures. *Journal of Applied Psychology*, 88, 979–988.
- Schuman, H. and Presser, S. (1981). *Questions and Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context*. Thousand Oaks: Sage.
- Schwarz, N. and Bohner, G. (2001). The Construction of Attitudes. In *Blackwell Handbook of Social Psychology*, Vol. 1: Intra-individual Processes, A. Tesser and N. Schwarz, (Eds). Oxford: Blackwell, 436–457.
- STATA Corporation (1999). *STATA. Reference Manual. Release 6. (Vol. 3)*. College Station: Stata Press.

- Stevens, J. (1996). *Applied Multivariate Statistics for the Social Sciences*, Third Edition. Mahwah, N.J.: Erlbaum.
- Stocké, V. (2004a). 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. [In German]
- Stocké, V. (2004b). Measuring Information Accessibility and Predicting Response-Effects: The Validity of Response-Certainties and Response-Latencies. *Metodološki zvezki*, 1, 33–55.
- Stocké, V. (2007). Response Privacy and Elapsed Time Since Election Day as Determinants for Vote Overreporting. *International Journal of Public Opinion Research*, 19, 237–246.
- Stocké, V. and Hunkler, C. (2007). Measures of Desirability Beliefs and Their Validity as Indicators for Socially Desirable Responding. *Field Methods*, 19, 313–336.
- Stocké, V. and Stark, T. (2007). Political Involvement and Memory Failure as Interdependent Determinants of Vote Overreporting. *Applied Cognitive Psychology*, 21, 239–257.
- 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. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Tourangeau, R., Steiger, D.M., and Wilson, D. (2002). Self-Administered Questions by Telephone: Evaluating Interactive Voice Response. *Public Opinion Quarterly*, 66, 265–278.
- Wilson, T.D. and Hodges, S.D. (1992). Attitudes as Temporary Constructions. In *The Construction of Social Judgments*, L.L. Martin, and A. Tesser (Eds.). Hillsdale, N.J.: Erlbaum.
- Wilson, T.D., Lindsey, S., and Schooler, T.Y. (2000). A Modal of Dual Attitudes. *Psychological Review*, 107, 101–126.
- Wyner, G.A. (1980). Response Errors in Self-Reported Number of Arrests. *Sociological Methods and Research*, 9, 161–177.

Received May 2004

Revised March 2007