SONDERFORSCHUNGSBEREICH 504

Rationalitätskonzepte,
Entscheidungsverhalten und
ökonomische Modellierung

No. 04-14

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

Volker Stocké*

April 2004

Financial support from the Deutsche Forschungsgemeinschaft, SFB 504, at the University of Mannheim, is gratefully acknowledged.

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

Universität Mannheim
L 13,15
68131 Mannheim

Abstract[1]

Empirical evidence suggests that the approval motive of respondents, their desirability beliefs and the privacy of the response situation all affect respondents' susceptibility to social desirability bias. Previous research has analyzed the explanatory power of these factors separately and has not taken their possible interdependence as determinants for social desirability bias into account. This article examines the prediction taken from 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 social desirability bias. According to the empirical results from our first study, a three-way interaction between the factors analyzed predicts respondents' racial attitude reports. However, since attitude answers and desirability beliefs were collected in the same interview, the observed associations may be an artifact due to the subjects' sensitization towards social desirability concerns. This possibility is tested in a second study, where only racial attitude answers were collected under conditions of varying response privacy. Aggregated response differences between the utilized attitude items and respondents' social group affiliation were matched with equivalent differences in the desirability beliefs found in the first study. The results from the main study were replicated with this independent sample of respondents.

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

.

[1] **Acknowledgments:** Stimulating discussions with Hartmut Esser, Herbert Bless and Stephan Ganter are gratefully acknowledged. Furthermore, I would like to thank the participants of the 'International Conference on Rational Choice and Social Institutions' in Fukuoka, October 3-5, 2003 for helpful comments on an earlier version of this paper. This research was supported by a grant of the German Science Foundation (DFG) to the Sonderforschungsbereich 504 at the University Mannheim.

1 INTRODUCTION

Social desirability bias (SD-bias) originates from shared social norms and the resulting systematic measurement error causes univariate response distributions to deviate from the true sample characteristics. Furthermore, differences in the strength and direction of SD-bias according to subjects' socio-economic characteristics are likely to suppress or artificially create associations between these characteristics and the attribute under consideration. The precondition for implementing effective measures to prevent SD-bias during data collection, or at least to allow the survey researcher to introduce appropriate statistical controls when analyzing the data, is the detailed knowledge about the determinants for the strength and direction of this bias. Empirical research has made considerable progress in this direction and suggests that three factors are important for respondents' susceptibility to social desirability effects. *Firstly*, several studies have tested the hypothesis that subjects with a stronger need for social approval are more prone to SD-bias. In this line of research respondents' answers deviate from their true scores because of their situationally stable but individually variable motivation to gain approval and avoid disapproval from others (Crowne and Marlowe 1960). The *second* determinant that has received considerable attention in empirical research is response privacy and all the features of the interview situation which enable or prevent others from observing the respondents' answers (for an overview c.f. Tourangeau et al. 2000: 275). A wide range of survey topics have been tested with a view to whether a lack of response privacy can be associated with either more desirable or fewer undesirable answers. A *third* important factor, which is expected to explain the strength *and* direction of SD-bias, are the respondents' beliefs about the desirability of those traits they ascribe to their own person when selecting a particular response option (Edwards 1957). Several researchers have analyzed whether the individual differences in these desirability ratings can predict the probability that a response option is endorsed.

Many studies have analyzed the above-mentioned factors separately, but none have tested the possibility that each may be a moderator variable for the explanatory power of the other determinants of SD-bias (for exceptions c.f.: Chen et al. 1997; Phillips and Clancy 1972). Such an interdependence is predicted by the rational-choice theory (RCT) of response behavior (Esser 1991; Stocké 2001; Tourangeau et al. 2000: 281). In this approach socially desirable responses are explained as the result of the subjects' aim of gaining approval from others *and* their beliefs about whether a certain answer will realize this aim. The theory predicts SD-bias only when the motivational *and* cognitive preconditions are simultaneously

fulfilled. Accordingly, a strong approval motive, a lack of response privacy and clearly defined desirability beliefs are each assumed to be a necessary though on their own insufficient precondition for socially desirable response behavior. When only one of these conditions is not fulfilled, no SD-bias is expected and the other determinants are assumed to be irrelevant. This hypothesis about the complete interdependence of all three determinants for SD-bias has not yet been tested empirically.

Apart from confirming results, several empirical studies failed to find evidence for the strength of the approval motive in explaining the presence of SD-bias. One possible reason for these inconsistent results is the failure to take the role of response privacy into account. In those studies where negative evidence about the explanatory power of the approval motive came to the fore, the responses under consideration were recorded with self-administered questionnaires and therefore under conditions of complete privacy (for instance c.f.: Brunet et al. 1996; Duck and Hunsberger 1999; Vella-Broderick and White 1997). In other studies where the approval motive proved to be irrelevant for the response behavior, no information about the administration mode and the concomitant response privacy has been provided (Ellingson et al. 2001; Laicardi et al. 2001). Whether the predictive power of the approval motive does indeed depend on the response privacy has not yet been explicitly tested.

The respondents' desirability beliefs have been proven in all the studies we are aware of to predict the probability that the respective response option will be endorsed and can therefore be assumed to be a valid predictor for the presence of SD-bias. The observed correlation does not necessarily however indicate the presence of response bias, and may simply be due to the respondents' honest answers about their true characteristics consistent with their desirability beliefs. Desirability beliefs are from the perspective of RCT a determinant of SD-bias only insofar as their correlation with the response behavior differs according to the subjects' approval motive and the privacy of the response situation. Whereas no empirical evidence about the moderating role of privacy differences is available, the results about whether the approval motive is in this respect relevant are inconsistent (Chen et al. 1997; Phillips and Clancy 1972).

This article will examine, based on data from two independent studies, the hypothesis derived from RCT that the respondents' approval motive and their desirability beliefs as well as response privacy must be regarded as interdependent determinants of SD-bias. In an initial study, this is achieved using individual level data on the subjects' answers about their racial attitudes together with detailed information about respondents' desirability judgments about all the attitude items in the questionnaire. Asking the same respondents about their racial

attitudes and desirability beliefs may create SD-Bias as an artifact of the subjects' sensitization regarding social desirability concerns. Because of this reason, the external validity of our results is tested in a second study, where an independent sample answered only the attitude questions under conditions of varying privacy. Here, the predicted pattern of associations has been tested using differences in the desirability beliefs between the attitude items and socioeconomic groups observed in the first study and equivalent differences in the attitude responses from the second study.

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 2001; Tourangeau et al. 2000: 281). The *motivational* core of this explanation is the strength of the respondents' need for social approval and the concomitant differences in their tendency to apply impression management strategies.[2] The *cognitive* determinants of response behavior are subjects' beliefs about whether their answers will provoke evaluative reactions in others and about what kind of reactions can be expected when a particular answer is selected. The total evaluation of each response option is based on the combined result of these motivational and cognitive factors.

The respondents' aim when they select a particular answer in a self-description questionnaire is to achieve positive and avoid negative evaluation reactions in order to maximize their feeling of approval from others. However subjects only expect their answers to be relevant in this respect when others are both present *and* able to observe the answers. In the version of RCT applied in this paper, respondents' feeling of privacy, and the concomitant expectation that the nature of answers may matter for their approval motive, depends completely on the objective ability of others to perceive the answers.[3] A few and not yet fully replicated results

[2] The 'need for social approval' concept stems from personality psychology and is meaningful in two different ways. Firstly, it is used to explain a tendency for 'self deception', where subjects unconsciously bias their reports in the direction of social desirability in order to preserve a positive self-image. Secondly, as a determinant for how important subjects view deceiving others with the aim of creating a positive impression in the eyes of others (c.f. for a discussion of both concepts: Paulhus and Reid 1991). In RCT the concept is used in the second meaning.

[3] The anonymity of answers, defined as the probability that responses are identified after the interview and will cause sanctions at a later point in time, can be relevant in this respect as well. However, we assume that such

about the effects due to bystander-presence seem to indicate that the presence of others may reduce the respondents' subjective feeling of privacy despite their inability to perceive the response content. Thus, adolescents' self-administered and therefore objectively private reports about their alcohol and marijuana consumption have been found to be reduced when parents were present during the answering of the questions (Aquilino et al. 2000). These effects were not replicated in another study (Couper et al. 2003).

A second necessary precondition for SD-bias is that respondents perceive sufficiently clear *differences* in the desirability of available response options and can therefore expect distinct consequences for their approval motive when one or another option is chosen (Esser 1991). According to the rational actor model, the subjects' relevant desirability beliefs are based on their perception of the evaluative criteria of others present. When, however, as with the case of an unknown interviewer, some prior knowledge of the interlocutor is not available, respondents have to rely on more indirect evidence in order to anticipate the most likely evaluation. Subjects may use interviewers' reactions to previous answers or their appearance, which they perceive to be associated with particular evaluation criteria. When none of these more specific cues are available, subjects utilize their knowledge about social norms as a basis for their desirability beliefs (Stocké 2001). Independent from the informational basis of their desirability beliefs, the incentives for socially desirable response behavior are expected to increase, when the subjects assume the available response options to be evaluated more differently.

In summary, RCT predicts that all three preconditions for SD-bias have to be present: a strong approval motive makes evaluation reactions for the respondents relevant, subjects regard such reactions as possible because of insufficient privacy and desirability beliefs are such that choosing one or another option makes a difference. If only one of these conditions is not given, the effect of all other factors on the probability of social desirable responses is assumed to vanish and subjects are expected to answer according to their subjectively true scores. Since this argument assumes a non-compensatory interaction between the single determinants, the strength and direction of total incentives from social desirability (SEU(SD)) can be represented with the multiplicative index shown in the following equation:

$$SEU(SD) = U_{SD} \bullet w_P \bullet \Delta w_{TD}$$

---

expectations are primarily relevant for questions, for instance about illegal drug use, where legal sanctions are possible. In the case of topics, as for example racial attitudes, where 'only' informal, social sanctions have to be expected, the possibility of immediate evaluative reactions of others and therefore the response privacy is more important.

The *first* parameter, U_{SD}, represents the strength of the approval motive. This parameter has a value of one for subjects with a strong approval motive and zero when this motive is absent. The *second* parameter, $w_{P,}$ stands for perceptions about how likely the answers in the particular response situation can affect the satisfaction of the approval motive. This parameter is zero under the condition of complete privacy and one when others are able to perceive the answers. The *third* parameter, Δw_{TD}, represents the subjects' desirability beliefs, or more precisely, the desirability differential between the response options. This parameter varies between minus and plus one, dependent on which option is regarded as being more desirable. Whereas the sign of this differential score predicts the direction of SD-bias, the absolute value represents the strength of perceived incentives.

3 PREVIOUS EMPIRICAL RESEARCH

3.1 The Role Of Need For Social Approval For SD-Bias

In an extensive body of research, the respondents' approval motive has been measured with different social desirability (SD)-scales and further testing was intended to reveal whether these scores can predict answers about a range of sensitive topics.[4] The hypothesis is that subjects with a stronger approval motive are more likely to endorse desirable and deny undesirable response options (Crowne 1979: 153ff.). However, the empirical evidence for the general validity of this hypothesis was found to be very inconsistent.

Hult and colleagues (1999) found in their samples from the U.S., Japan and Sweden that subjects with a strong approval motive, as measured by their scores on the Marlow-Crowne social desirability-scale (MC-SD-scale), reported less consumer ethnocentrism, which was assumed to be undesirable in each of the societies. In another study, military officers were tested with a short form of the MC-SD-scale and answered a self-esteem questionnaire as well as questions about how rational or intuitive their decisions are in general (Thunholm 2001). The MC-SD-scale scores were found to correlate positively both with the self-esteem reported in the questions and the questions on the rationality of decision making. Fisher and Katz (2000) tested the association between respondents' answers about 16 different, though

[4] The approval motive is measured with all SD-scales in a similar way. Respondents are asked whether they have certain characteristics, which are chosen in such a way as to be either socially desirable but unlikely to be true for anybody, or negatively evaluated but true for practically everybody. The more positive traits the respondents claim for themselves and the more negative attributes they deny, the higher SD-scores they receive.

presumably positively, evaluated social values and respondents' scores on the Reynold's SD-scale. Significant correlations were found for 14 out of 16 values when using a sample of customers from two telephone companies. In a meta-analysis of 10 studies about SD-bias in the field of self-reported religiousness, the authors reported significantly more religious responses for all but one when subjects had higher SD-scores (Trimble 1997).

In other studies the expected correlations were not found. Accordingly, subjects' responses about their mental, physical and spiritual wellbeing were found to be completely independent of their MC-SD-scale scores (Vella-Broderick and White 1997). Similarly, Laicardi and colleagues (2001) did not find a significant relationship between reports on a 'life satisfaction' scale and the 'lie scale' of the Eysenck Personality Questionnaire. Furthermore, answers on the 'modernity' scale, a measure for progressivism and optimism, were not related to subjects' MC-SD-scale scores and those on the Edwards SD-scale (Leak 1981).

In the field of racial attitudes, results about the role of subjects' approval motive are mixed as well. In a study from Mielke (1995) respondents were asked about their racial attitudes using Pettigrew and Meertens' 'Subtle and Blatant Prejudice' scale. Answers on both instruments were found to correlate significantly with the scores of the SD-scale from Mummendey and Eifler (1993). In another study, from Duck and Hunsberger (1999), answers on the 'Manitoba Prejudice' scale, which measures prejudice against immigrants, were found not to correlate with the respondents' scores on the MC-SD-scale.

3.2 Privacy Differences And SD-Bias

The effects of response privacy on respondents' susceptibility to SD-bias has been tested in many split ballot experiments where the mode of administration was varied. Most, but not all, studies found positive evidence for the hypothesis that more desirable and fewer undesirable responses are to be expected under conditions of 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 et al. 2001). In another study responses about the number of sex-partners were compared when questions were asked using computer-assisted self-administration (CASI), interviewer-administration (CAPI) or ACASI (Tourangeau and Smith 1996). The characteristic over-reporting of men and the under-reporting of women were substantially reduced under the private CASI and ACSAI conditions. Aquilino (1994)

compared subjects' responses about their illegal drug use and alcohol consumption between self-administrated questionnaires (SAQ), personal interviews undertaken by an interviewer and telephone interviews. For most types of drugs, the answers under the SAQ condition showed a higher prevalence of consumption. However, the differences were only significant for marijuana, cocaine and crack consumption, and only for the sub-sample of black respondents. Another study compared the same administration modes but did not find differences in the reports about marijuana use, shop lifting and church attendance (Reuband and Blasius 1996). Furthermore, respondents were more likely to admit that they had smoked in the previous 30 days in self- rather than in interviewer-administered interviews, but the differences were only significant for 12-17 year olds. In another study questions about nine topics, for example about whether subjects felt lonely, unhappy or proud about their own success, were put to interviewees either in person or with a drop-off/pick-up procedure (Mangione et al. 1982). For six out of nine questions significant mode differences were found, but only four of these differences were in the expected direction.

Two studies from the U.S. have tested whether white respondents' reports about their attitudes towards Afro-Americans differs between interviewer- and self-administered interviews (Krysan et al. 1994; Krysan 1998). Across all items used these studies, racial attitude answers were found on average to be significantly more negative under the condition of guaranteed privacy. On the level of the single items a considerable heterogeneity with respect to the strength and, partly, the direction of privacy effects have been observed: for nine of the nineteen attitude questions were private answers found to be significantly more negative and for two items even a slight tendency towards more positive attitude reports emerged. This may be regarded as evidence for differences in the desirability beliefs according to the specific item contents.

3.3 The Effects Of Desirability Beliefs

All the studies we were aware of have confirmed the hypothesis that respondents' desirability beliefs are substantially related to their response behavior. Gove and Geerken (1977) have shown in an early study from this field of research that differences between subjects as to how desirable they regard a high self-esteem or a positive affect predicts the probability that the respective response options will be endorsed in a personality questionnaire. Another study has analyzed the desirability beliefs of 20 items from the Infante and Rancer's Argumentativeness

Scale (Nicotera 1996). In these results, the respondents' social desirability ratings and their answers about their own argumentativeness were found to be positively correlated. Furthermore, the sex differences in the self-rating of argumentativeness were substantially reduced when differences in the desirability beliefs between men and woman were statistically controlled. In a study from Huang and colleagues (1998), the possible problem of an artificial activation of social desirability concerns when subjects were asked about their characteristics and the judged desirability of these characteristics in the same interview, has been attenuated. Here, the social desirability ratings of 288 items for testing subjects' mental health were recorded in a first interview. In a second session, the same subjects responded to these items. Despite the time distance between both answers, subjects reported fewer symptoms of mental illness when they judged this trait to be less desirable. Another study analyzed the association between beliefs about the desirability of police arrests and the self-reporting about this topic (Wyner 1980). Since the sample was a known group, for which the true number of arrests was obtained from police records, it was not only possible to analyze the correlation between responses and desirability beliefs, but to test the predictive power of these beliefs for the strength and direction of response error. Results have thus shown that desirability belief predicts whether and how strongly subjects have under- and over-reported the number of their police arrests in the past.

In a study by Phillips and Clancy (1972), subjects rated the social desirability of being unprejudiced against ethnic groups other than their own. Subjects reported as well on how bothered they were by meeting persons from other ethnic groups. Respondents who regarded an unprejudiced attitude as being highly desirable reported more positive racial attitudes than those with medium and low desirability ratings. In a German study, respondents rated the social desirability of 37 racial attitude items (Reinecke 1991: 152). On average, subjects judged those responses to be more desirable which expressed more positive attitudes. The desirability judgments, however, differed significantly according to the respondents' age as well as their socioeconomic status and varied between the different attitude questions. Whether desirability belief affects respondents' attitude answers was not tested in this study.

We are only aware of two studies where the interdependence of the three determinants of SD-bias, predicted from RCT, were at least partly analyzed. In a study by Chen and colleagues (1997), respondents initially rated the 45 items of the PANAS positive/negative affectivity scale with respect to social desirability. In the second part of the study an independent sample were asked whether the personality traits described by the PANAS items applied to themselves. Furthermore, subjects' approval motive was measured using the MC-SD-scale.

As a result, the on the level of the items, aggregated desirability beliefs were found to be substantially related to the probability of endorsement for each item. Furthermore, this correlation was significantly stronger for subjects with a strong rather than a weak approval motive. In contrast a study from Phillips and Clancy (1972) has not found an interaction effect between subjects' approval motive and their desirability beliefs. The response behavior to seven sensitive topics, such as for instance the respondents' happiness, the number of friends they have and how unprejudiced they regard themselves to be with respect to ethnic outgroups, were significantly, though independently, associated with desirability beliefs and subjects' approval motive.

4 EMPIRICAL STUDY

In the main part of our study the aim was to test the hypothesis about the interdependence of the three analyzed determinants for SD-bias, with racial attitude answers collected under conditions of varying privacy and individual level data about subjects' approval motive and their desirability beliefs. In a second validation study the external validity of the results from the main study is tested with response data from an independent sample of respondents.

4.1 Sample And Data Collection

The respondents in the main as well as 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 respondents selected using the 'last-birthday' method. Both the 150 interviews in the main study and the 106 interviews in the validation study took place in the respondents' homes. The response rate was 31.3 percent in the main study and 39.2 percent in the validation study.

In the main study the interviews were computer assisted. At the very beginning subjects were asked about their desirability beliefs and then later completed the instrument for measuring their approval motive. Between both blocks of questions respondents answered about 50 questions unrelated to the topic of racial attitudes. In the second part of the questionnaire respondents answered the 10 questions about their racial attitudes. Between these attitude questions and the SD-scale, another 35 filler questions about unrelated topics were asked. The

mode of data collection in the validation study was computer-assisted as well and the same racial attitude questions were asked in the second part of these interviews. In contrast to the main study, respondents did not answer questions about their subjective desirability beliefs in the field of racial attitudes.

4.2 Measures

The dependent variable in our study consists of the respondents' answers to 10 attitude items used in the German General Social Survey (ALLBUS) to measure attitudes towards foreigners (cf. the item wording in table 1).[5] Responses were recorded using seven-point Likert scales with endpoints labeled with ‚completely agree' and ‚completely disagree'. For the following analysis the scoring of all items was recoded in such a way that low values represent negative and high values positive racial attitude answers. The three theoretical constructs were operationalized as follows:

- *Need for social approval (U_{SD}):* Individual differences in the strength of subjects' approval motive were measured using a 10 item, short form of the MC-SD scale (cf. Stocké 2003).[6] An additive index of respondents' answers on this instrument was used as an indicator for the strength of their desire for social approval. Before computing this index, all forced-choice responses were recoded in such a way that answers indicating a low approval motive had a value of zero and answers indicating a strong approval motive were coded as one. The values of the total index ranges between zero (no approval motive) and 10 (strong approval motive). In order to allow the subjects' approval motive to be picked up as fully as possible via their responses on the MC-SD scale, the questions were administered by the interviewer for all respondents and thus under strong pressure to employ impression management strategies.

- *Privacy of response situation (w_P):* Whether respondents' answers about their racial attitudes were private or discernable by others was operationalized by using either a self-administered or an interviewer-administered mode of data collection. In the first part of the interview the respondents' desirability beliefs and the strength of their approval motive was recorded for all respondents interviewer-administered. About halfway through the interviews,

[5] All attitude items were used in the German General Social Survey in 1997 and were adopted from there.

[6] This short form of the scale consists out of items 1, 6, 13, 14, 15, 16, 17, 21, 30 and 33 listed in table 1 in the paper from Crowne and Marlow (1960).

after the administration of the MC-SD scale and before the racial attitude questions, respondents were randomly assigned to one of the two interview modes. In the interviewer-administered mode, the interviewer simply continued to read out the questions and recorded the answers into the laptop computer. In the self-administration mode, the interviewers handed over the laptop computer to the respondents and asked them to complete the interview alone. While the respondents read the questions from the computer screen and typed in the answers, the interviewer remained present in the room and answered, when necessary, questions put by of the respondents. The interviewers were however instructed to maintain enough distance to the respondents so as not to be able to observe their response behavior.

- *Desirability beliefs and the relative desirability of positive and negative racial attitudes (Δw_{TD}):* The desirability beliefs were operationalized as those 'extrinsic' evaluations the respondents anticipated as a typical reaction from others when certain traits were revealed to an unknown counterpart (Edwards 1957).[7] Respondents were asked to imagine a situation such as a train journey, where a conversation develops between two strangers. They were then asked about how embarrassing it would be for one of these persons to disclose different opinions about foreigners (cf. the exact wording of the questions in the appendix). The altogether 20 opinions which were presented to the respondents in this scenario were identical with those either completely disagreeing or completely agreeing with the 10 racial attitude items used in our study. The desirability beliefs were recorded with a bipolar response scale from -4 (this statement would be very embarrassing) to +4 (this statement would be very pleasant). After the responses were recoded into a range of values between 0 and 8 for each respondent and attitude item, the judged desirability of a negative racial attitude statement was subtracted from that of a positive attitude. The relative desirability values obtained for each item and respondent ranges from -8 (negative racial attitudes more desirable) to +8 (positive attitudes more desirable). A value of zero indicates that positive and negative attitudes were regarded as being evaluated equally.

[7] An alternative operationalization would be the subjects' aggregated personal desirability ratings, which may be assumed to represent the normative climate in the population surveyed (Meleddu and Guicciardi 1998). This indicator has however been found to differ from the anticipated, extrinsic desirability judgments (Crott and Roßrucker 1974).

5 RESULTS

5.1 Description Of Variables

Respondents' average beliefs about the desirability of positive racial attitude answers were found to range between +0.01 (item 10) and +1.11 (item 7) on a response scale from +4 (desirable) to –4 (undesirable). The judged desirability of those answers expressing negative attitudes varied between –0.52 (item 7) and +1.26 (item 1) (c.f. table 1, columns 1-2). On average, across all attitude questions, the respondents' assume positive rather than negative attitude answers to be more favorably evaluated in society. However, the desirability differences on the aggregate level between +0.47 (positive attitudes) and +0.16 (negative attitudes) are rather small. This is true for the relative desirability of positive and negative answers as well. Here, the average of the differential scores ranges between –0.66 for item 1 and +1.59 for items 7. For three items we found net incentives for negative attitude responses (item 1, 8 and 10), whereas in the case of the remaining 7 questions positive attitude answers were expected to be on average more positively evaluated (c.f. column 3 in table 1). For the whole attitude scale a mean relative desirability score of +0.30 shows that agreeing with a positive and disagreeing with a negative racial attitude item is regarded as slightly more desirable than answering the other way around.

The respondents from study 1 as well as those from study 2 have answered the racial attitude items in a rather positive way (c.f. columns 4 and 5 in table 1). On the seven-point response scale, with higher values indicating a more positive attitude, we found in study 1 average response values between 3.3 (item 1) and 5.4 (item 3). In study two the average answers vary between 2.8 (item) 1 and 5.2 (item 3). Despite the same sampling process and population but maybe because of the difference in the response rates, the attitude answers in study 1 were for all questions more positive than in study 2. This difference is statistically significant for items 1, 4, 6 and 7 (t-values reported in column 6 in table 1).

Table 1. Average Desirability Beliefs And Response Behavior For The Racial Attitude Items

	Desirability of positive attitudes [a)] Mean (Std)	Desirability of negative attitudes [a)] Mean (Std)	Relative Desirability [b)] Mean (Std)	Response Behavior [c)] (Study 1) Mean (Std)	Response Behavior [c)] (Study 2) Mean (Std)	Differences in Response Behavior [d)] t-value
Item 1. „Foreigners in Germany should adapt their lifestyle more to that of Germans“.	+0.56 (2.1)	+1.26 (1.9)	-0.66 (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“.	+0.32 (2.2)	-0.08 (2.5)	+0.40 (3.4)	5.2 (2.0)	4.9 (1.9)	1.2
Item 3. „Foreigners in Germany should marry within their own ethnic community“.	+0.67 (2.0)	-0.09 (2.4)	+0.76 (3.3)	5.4 (2.0)	5.2 (2.1)	0.7
Item 4. „One should forbid any political activities by foreigners in Germany“.	+0.27 (2.1)	-0.36 (2.5)	+0.63 (3.3)	5.1 (2.1)	4.6 (2.1)	2.1*
Item 5. „Because there are so many foreigners in Germany, one feels like a stranger in one’s own country“.	+0.63 (2.2)	+0.24 (2.2)	+0.41 (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“.	+0.19 (2.1)	+0.16 (2.4)	+0.05 (3.3)	4.6 (2.0)	4.0 (2.0)	2.2 *
Item 7 [R]. „Foreigners in Germany are an enrichment for our culture “.	+1.11 (2.2)	-0.52 (2.4)	+1.59 (3.7)	5.2 (1.8)	4.6 (1.6)	2.7 *
Item 8. „Foreigners in Germany commit more criminal offences than Germans“.	+0.18 (2.2)	+0.61 (2.2)	-0.43 (3.7)	4.3 (1.9)	4.1 (1.9)	1.1
Item 9. „The presence of foreigners in Germany causes problems on the housing market.“	+0.70 (1.8)	+0.33 (1.9)	+0.36 (2.8)	5.0 (1.7)	4.8 (1.8)	1.1
Item 10 [R]. „Foreigners in Germany should be entitled to the same social welfare and other social security benefits as the Germans.“	+0.01 (2.3)	+0.10 (2.4)	-0.11 (3.7)	4.7 (2.0)	4.6 (1.9)	0.3
Total	+0.47 (1.3)	+0.16 (1.6)	+0.30 (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; [a)] Response scale ranges from –4 (attitude undesirable) to +4 (attitude desirable); [b)] Scale ranges between –8 (negative attitude more desirable) and +8 (positive attitude more desirable). [c)] Response scale ranges from 1 (negative attitude answer) to 7 (positive attitude answer). For items marked with ‘R’ the original coding of the responses was recoded in a way that high values expresses a positive racial attitude. [d)] Significance: * $p \leq 0.05$.

On the aggregate level of the items we found for study 1 a strong and statistically significant correlation of r=.82 ($p < .05$) between the response behavior and the relative desirability beliefs. This association between the aggregated answers from study 2 and the desirability beliefs which were reported in study 1 is r= .66 ($p < .05$). Since the two pieces of data stem from different samples we regard this as a remarkably high value and as initial evidence for the validity of the conclusions based on the data from study 1.

Table 2. Heterogeneity Of Relative Desirability Beliefs About Positive And Negative Racial Attitudes

	Positive attitudes more desirable [a)]		No differences	Negative attitudes more desirable [a)]		Total incentives [c)]
	%	Mean (Std)	%	%	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)
Observations				150		

[a)] Values range from greater than zero to +8 (positive attitudes more desirable) and from smaller than zero and -8 (negative attitudes more desirable). [b)]Total incentives are the average of respondents' absolute relative desirability scores. Subjects with no desirability differences are included in this measure. The variable ranges from 0 (no incentives) to 8 (strong incentives).

The relative desirability judgments aggregated on the item level do not reflect the true incentives for socially desirable response behavior when the direction of desirability beliefs differs on the level of the individual respondents: under these conditions a part of the respective incentives cancel each other out. In order to test this possibility, the relative desirability beliefs in our sample are analyzed in table 2, differentiated according the sign of these differential scores on the level of individual respondents. The result clearly indicates that respondents do not agree about whether positive or negative racial attitude answers are more desirable in society: on average across all items 43.2 percent assume positive attitudes to be more desirable, 36.8 percent assume this for negative attitudes and 19.9 do not perceive any desirability differences. Across the entire attitude scale, subjects perceive incentives for a positive racial attitude answer of 3.2 points on the relative desirability scale and for a negative

attitude answer of –3.0 points. Subsamples of respondents therefore perceive social desirability incentives for positive as well as negative racial attitude answers. Since these groups are only 17 percentage points different in size, a substantial part of the total incentives was not visible on the aggregate level of the total sample in table 1. The mean absolute values of individual respondents' relative desirability scores is therefore an appropriate indicator for the strength of *potential* SD-bias. This measure, with a possible value range between 0 and 8, is found to vary between 2.0 for item 1 and 3.1 for item 7.

The strength of the empirically observed approval motive in our sample, as measured with the short form of the MC-SD scale, varies between 0 (no approval motive) and 10 (strong approval motive). The average need for social approval is 5.7 (std. = 2.2).

5.2 Testing The Hypothesis With Individual Level Data

According to the hypothesis derived from RCT, the absolute value and sign of the subjects' relative desirability beliefs are expected to predict the response behavior significantly stronger under interviewer- rather than self-administration. The predictive power of this interaction effect for the attitude answers is furthermore expected to increase with the strength of subjects' approval motive. The following OLS-regression analysis, with the answers on all items as a dependent variable, tests the explanatory power of this three-way interaction.[8]

The results from regression model 1, where only the subjects' socioeconomic characteristics were included, shows that the racial attitude answer differs significantly according to the respondents' sex and education (c.f. table 3).[9] Accordingly, males reported significantly more negative attitudes than females, and respondents with either a university or a vocational college degree answered the questions more positively than subjects with only compulsory secondary education. Similar group differences have been reported in the literature about racial attitudes (Hudson and Hines-Hudson 1999; Sniderman et al. 1991). In model 2 all three

[8] Since respondents' answers on all attitude items are included simultaneously into the analysis, the observations are not independent and the standard error of the parameter tends to be underestimated. This is corrected by computing t-statistics on the basis of Huber-White Sandwich estimators for robust standard errors with the respondents as a cluster variable (Huber 1964; STATA Corporation 1999: 165 ff.).

[9] In addition to the variables shown in table 3, nine dummy variables for the ten attitude items were included in all regression models in order to control for response differences between the attitude questions (parameter not shown). The problem of high multicollinearity because of the inclusion of the multiplicative parameter in model 3 is avoided by z-standardizing the need for social approval values and relative desirability scores before interaction terms were computed (Cronbach 1987). This procedure affects the lower level parameter but leaves those of the highest interaction level unaffected (Aiken and West 1991: 28ff.).

predicted determinants for SD-bias are entered as main effects only into the regression equation.

Table 3. Test Of Predicted Determinants For The Racial Attitude Answers And Their Interaction (OLS Regression Parameter With Robust Standard Errors)

	Model 1 B (t-value)	Model 2 B (t-value)	Model 3 B (t-value)
Control variables			
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 [a)]	-.42 (2.0) *	-.33 (1.6)	-.32 (1.6)
EDUCATION [b)]			
- 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 [c)]			
- 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)
Rational-Choice Model			
RELATIVE DESIRABILITY	--	.36 (5.2)*	.21 (2.7) *
NEED FOR SOCIAL APPROVAL	--	-.25 (2.4)*	-.17 (1.2)
INTERVIEWER ADMINISTERED MODE [d)]	--	.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^2	.19	.23	.24
Observations	1477	1477	1477

Significance: * $p \leq 0.05$; omitted categories: [a)] female; [b)] compulsory education; [c)] blue collar worker; [d)] self-administered

As a *first* result and consistent with the empirical evidence from other studies reported in section 3.3 above, the subjects' relative desirability beliefs proved to be a strong predictor for their racial attitude reports. *Secondly*, the strength of the approval motive was found to be a statistically significant predictor for the response behavior. However, the negative sign of the regression parameter seems to indicate that a strong approval motive is associated with more negative racial attitude reports and therefore that SD-bias pushes respondents in the direction of more negative attitude answers. The results of our analysis will show that such a

conclusion about the general direction of SD-bias would be misleading. This applies to the *third* result of regression model 2 as well. Accordingly, the very weak and non-significant effect of response privacy suggests the absence of any SD-bias. In regression model 3 then, the three-way interaction effect between all determinants for social desirability bias as well as the lower-level interaction terms are entered additionally into the regression equation. As predicted from RCT, the resulting interaction parameter proved to be a statistically significant predictor for the respondents' racial attitude reports (cf. table 3). Furthermore, taking model 1 as a starting point, the fully specified model 3 was found to explain the response data significantly better ($F(6.0, 1247.9)=13.6$; $p \leq 0.01$).[10]

In order to allow an interpretation of the observed three-way interaction, we computed predicted values for relevant combinations of the three determinants of SD-bias included into the analysis (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 only influenced to a very limited extent by the privacy of the response situation. In particular, the response differences do not follow the direction of their desirability beliefs: independent of whether positive or negative racial attitudes or none of both are assumed to be more desirable, response behavior was between .16 and .28 scale points more positive under interviewer-administrated conditions. *Secondly*, the results for subjects with MC-SD values of one standard deviation over the sample mean, and therefore a strong approval motive, are found to differ substantially. In this group the racial attitude answers are much more responsive to differences in the privacy of the response situation and the direction of their relative desirability beliefs. Attitude answers of subjects who assume positive racial attitudes to be more favorably evaluated by society were found to be .55 scale points more positive in interviewer- rather than self-administrated interviews. Subjects who believe that negative attitudes are more desirable demonstrate an equivalent effect in the opposite direction. In this group, responses were found to be .55 scale points more negative when the interviewer was able to observe the answers. The mode of administration proved in contrast to be practically irrelevant for subjects who do not assume any desirability differences for positive and negative racial attitudes.

[10] The appropriate degrees of freedom for the F-test were obtained by weighting the degrees of freedom based on the number of observations with a Greenhouse-Geissers Epsilon of 0.86. This, compared with Huynh-Feldt's Epsilon conservative weighting parameter, takes into account to what degree the data deviates from the assumption of independence of the observations (Stevens 1996: 459ff.).

Table 4. Responses About Racial Attitudes According To The Mode Of Administration, The Relative Desirability Beliefs And Subjects' Approval Motive (Predicted Values From Regression Model 3)

	Weak need for social approval Mode of administration			*Strong need for social approval* Mode of administration		
	Self administered	Interviewer administered	Mode difference	Self administered	Interviewer administered	Mode difference
Relative desirability of racial attitudes						
Positive attitude more desirable	3.18	3.34	+ 0.16	2.62	3.17	+ 0.55
Positive/negative attitudes equal desirable	2.85	3.08	+ 0.23	2.52	2.51	- 0.01
Negative attitude more desirable	2.53	2.81	+ 0.28	2.41	1.86	- 0.55

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 equal 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.

Our results are thus far completely in agreement with the theoretically predicted interdependence of all three analyzed determinants of social desirability bias. It is particularly remarkable that a substantial percentage of subjects believe that negative racial attitudes are more favorably evaluated by society *and* that their answers have been found to be influenced in this direction. Apart from the general methodological issue, this poses the question of whether these respondents would have followed their beliefs as well if the desirability questions had not been asked before the attitude questions in the same interview. This issue will be addressed in the following section.

5.3 Differences In Desirability Beliefs Between Items And Social Groups

Precisely which socioeconomic groups differ according to their desirability beliefs will be analyzed in a first step before we match aggregate measures for respondents' desirability beliefs from study 1 with the response behavior observed in study 2. In regression model 4, shown in table 5, the respondents' relative desirability beliefs about all 10 attitude items are

regressed on the full set of their social characteristics.[11] Differences in the beliefs between the items are held constant with dummy-variables, which were included into the regression equation.

Table 5. Differences In Relative Desirability Beliefs Between Items And Respondents' Social Characteristics (OLS Regression Parameter With Robust Standard Errors)

	Model 4 B (t-Value)		Model 5 B (t-Value)
Item differences [a)]		Item differences [a)]	
- 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 [b)]	-0.30 (0.7)	- Male [b)]	--
Education [c)]		Education [e)]	
- 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 [d)]		Social status [f)]	
- 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^2	0.08		0.06
Observations	1500		1500

Significance: * $p \leq 0.05$; omitted categories: [a)] item 1, [b)] female, [c)] university degree, [d)] self-employed, [e)] secondary school certificate & high school certificate & university degree, [f)] white collar worker & self-employed.

The regression parameter for the item dummies shows *firstly* that the desirability ratings differ significantly between item 1 (reference category) and item 7. The relative desirability beliefs for the remaining items are located on a continuum between these two extremes. The *second* result is that respondents' beliefs about the strength and direction of desirability differences

[11] As in regression models 1-3 the observations are partly not independent and therefore the same correction has

between positive and negative racial attitudes do not differ according to their age, income or sex: none of these variables proved to be a significant predictor for these beliefs. *Thirdly*, we found a differentiation of desirability judgments according to the respondents' affiliation with occupational status groups and their educational degrees. With respect to the first dimension, blue collar workers and subjects who had never been in the workforce were found to perceive significantly stronger incentives for positive attitude answers, compared with self-employed persons. White-collar workers do not differ significantly from the self-employed and the regression parameter indicates that both groups held similar beliefs. On the second dimension, respondents with only compulsory education, and even more those with a vocational college degree, are found to have more negative desirability beliefs compared with subjects who have received a university degree. Because of the small number of persons with a vocational college degree in our sample, the regression parameter for this group is only marginally significant.

In regression model 5 those respondents' characteristics that had proven to be irrelevant in the first analysis were dropped from the regression equation and both the status as well as the educational groups were combined according to the similarity of their desirability beliefs. Significant contrasts were found for the combined group of blue-collar workers and subjects without previous employment and the other occupational status groups. Furthermore, subjects with compulsory education and those with a vocational college degree proved to perceive significantly less incentives for positive attitude responses when compared with the group of respondents with other educational degrees.

5.4 Results From The Validation Study

The data for the following concluding part of our empirical analysis consists, *firstly*, of the mean desirability beliefs of those two occupational status groups and the two groups with different educational degrees identified in the previous section as differing in this dimension. We computed mean desirability scores for the resulting four social groups with respect to each of the 10 racial attitude items. The total of 40 averages have been found to vary between –1.25 and +4.71, with an average value of .52. In a *second* step, we computed the averages of the racial attitude responses observed in study 2 for exactly the same combinations of social

---

been applied as described in footnote 8.

groups and items. This was done separately for the interviewer- and self-administered answers. We were, therefore, able to obtain two times 40 attitude averages which vary for the CASI condition between 1.89 and 5.45, with an average of 4.24. For the CAPI condition the range was between 2.46 and 5.78, with an mean value of 4.37.

In the following analysis the mean attitude responses from study 2 have been regressed onto the mean desirability beliefs reported by the separate sample from study 1.[12] It is expected that the aggregated desirability judgements predict the response behavior significantly stronger under the condition of the interviewer- rather than the self-administrated mode of data-collection. Results from regression model 6, with main effects only, show that the desirability beliefs are a strong and significant predictor for the response behavior, but the different privacy conditions do not explain the aggregated attitude answers (c.f. table 6). This result replicates what we have found with individual level data in section 5.2. Whether the effect of aggregated desirability beliefs on the response behavior differs when the attitude responses have been recorded using either interviewer- or self-administered modes is tested in regression model 7. This hypothesis is confirmed by the significant interaction between the two determinants of SD-bias.

Table 6. Aggregated Response Behavior From Study 2 As A Function Of Administration Mode And Aggregated Relative Desirability Ratings From Study 1 (OLS Regression Parameter With Robust Standard Errors)

	Model 6 B (t-value)	Model 7 B (t-value)
RELATIVE DESIRABILITY (group means)	.26 (4.9) *	.09 (2.2)
INTERVIEWER ADMINISTERED MODE [a)]	.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: a) self-administered mode

In table 7 we have computed the predicted response behavior for aggregated desirability scores of one standard deviation above and below the neutral beliefs at zero. The results indicate that interviewer-administration, compared with self-administration, causes the aggregated racial attitude answers to be .35 points more positive on the response scale when the desirability beliefs favors positive racial attitudes. The same effect, with a reverse sign, is

[12] The mean desirability beliefs in this analysis are not based on independent data for the different items and the privacy conditions. Therefore t-statistics are based on robust standard errors with the 4 social groups as a

found for beliefs which assume negative attitude statements to be more desirable: here, the effect size of .44 scale points is even slightly stronger. For groups of respondents who do not perceive any desirability differences, hardly any differences in the response behavior can be found. We can therefore conclude that our results about the moderating role of response privacy for the effect of desirability beliefs on the response behavior are completely replicated under those conditions where the beliefs and the attitude answers are collected in different interviews.

Table 7. Aggregated Racial Attitude Answers As A Function Of The Administration Mode And Subjects' Relative Desirability Beliefs (Predicted Values From Regression Model 7)

	Mode of administration		
	Self administered	Interviewer administered	Mode difference
Relative desirability of racial attitudes			
Positive attitude more desirable	4.30	4.65	+ 0.35
Positive and negative attitudes equal desirable	4.19	4.15	- 0.04
Negative attitude more desirable	4.08	3.64	- 0.44

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 equal desirable'.

6 SUMMARY AND GENERAL DISCUSSION

The aim of this study was to test the hypothesis that the respondents' approval motive, their beliefs about the relative desirability of response options and the privacy of the response situation are mutually interdependent determinants for the strength and direction of social desirability bias. This general hypothesis and the resulting concrete predictions about the nature of this interdependence have been derived from rational-choice theory of survey response behavior. In the first part of our empirical analysis we tested these predictions with individual level data about the racial attitude answers from a random probability sample of respondents. As a *first* result of our descriptive analysis we found the beliefs about whether positive or negative attitude answers are more desirable to differ considerably in our sample: 43 percent of respondents subjectively perceived positive but 37 percent negative racial

cluster variable (cf. footnote 8).

attitude answers to be more socially desirable, with 20 percent assuming there were not any differences in this respect. *Secondly*, respondents' racial attitude answers were found to be in agreement with their desirability beliefs. But, as expected, both additionally analyzed determinants of SD-bias were found to be relevant with regard to how strongly respondents' desirability beliefs affect the response behavior. Thus subjects' racial attitude reports were significantly more consistent with whatever attitude they perceived to be more favorably evaluated by society when responses were recorded using an interviewer- rather than self-administrated mode. Whereas responses were found to be relatively similar under the condition of self-administration, the interviewers' ability to perceive and evaluate the responses caused the reports, dependent on the sign of desirability differences, to be become either more positive or more negative. Respondents without a clear view as to whether positive or negative racial attitudes are more socially desirable were found to be completely unaffected by differences in response privacy. These results indicate that subjectively perceived desirability is not only a correlate of respondents' true racial attitudes, but represents their *cognition* about the strength and direction of incentives stemming from social desirability.

Thirdly, the explanatory power of this interaction between desirability beliefs and response privacy were found to increase with the subjects' level of *motivation* regarding positive impression management: under conditions of low privacy, a stronger approval motive was found to substantially increase the respondents' orientation toward their own desirability beliefs. In summary, the analysis proved that clear desirability beliefs in a certain direction, a sufficiently strong approval motive and a lack of response privacy are each necessary but not sufficient conditions for SD-bias. This indicates that SD-bias affects survey response behavior under very specific conditions that have to be taken into account when researchers introduce statistical controls for this sort of systematic measurement error.

In a second study, we have addressed the important issue of whether the results found in the main study are externally valid. Although the data from our first study allowed a detailed analysis of individual-level determinants for SD-bias, this made it necessary to ask the respondents in the same interview about their desirability judgments of different racial attitudes and their own attitudes. This leaves the possibility open that respondents' answers only correlated in the manner observed because thinking about their desirability beliefs activated impression management concerns. This was then tested with data from a second and unrelated sample where subjects only answered the attitude questions from study 1 under different privacy conditions. Aggregated differences in the desirability beliefs between the

items and the subjects' affiliation to socioeconomic groups were computed with the data from study 1 and matched with equivalent differences in the response behavior from study 2. The observed pattern of associations in study 1, between the attitude responses on the one hand and desirability beliefs and response privacy on the other, were completely replicated with this aggregated data. It was found that the answers from subjects who believe negative rather than positive racial attitudes to be more desirable are biased in that particular direction.

The results from our study thus have implications for the appropriate design of studies about social desirability bias. We have shown that the failure to take the interdependence of the analyzed determinants for SD-bias into account may cause researchers to draw the wrong conclusions about the presence and even direction of SD-bias. Although our fully specified regression model revealed substantial response effects under the predicted conditions, differences in the response privacy alone have not proved to be a predictor for the racial attitude answers. Without including the heterogeneity of desirability beliefs into the analysis, this result would have suggested the conclusion that there was a complete absence of SD-bias. However, the reason for these contradictory findings is that the attitude answers of similar sized subsamples were simply biased in different directions and these canceled each other out on the aggregate level. Since in studies about the determinants of SD-bias the desirability beliefs are rarely recorded empirically and included in the analysis, some of the weak or absent privacy effects observed in some empirical studies may be the result of desirability beliefs in the opposite direction (Aquilino 1994; Mangione et al. 1982; Reuband and Blasius 1996). In our analysis the respondents' approval motive have been found to correlate *negatively* with the attitude answers, which would suggest, without taking the other determinants of SD-bias into account, that a stronger impression management motive causes racial attitude answers to be more negative. According to the full explanation model this would have been an invalid conclusion: more motivation for impression management increases a subjects' conformity with whatever subjective belief they held about social norms in society.

In the present study detailed hypotheses from rational-choice theory about the determinants of SD-bias have been empirically confirmed. This approach is fruitful since it allows one to make clear predictions about the interrelatedness of subjects' cognitions and their motivation. We do not however regard this theory as being sufficiently complete for a comprehensive explanation of response behavior. *Firstly,* the decision to select a specific answer in the fourth stage of the total process of selecting an answer is based on the subjects' comprehension of the question, the retrieval of relevant information from memory and the judgment of the

response options (Tourangeau and Rasinski 1988). Several cognitive processes, for instance the priming of information through the preceding question, have been found to affect response behavior (cf. for example: Todorov 2000). *Secondly*, RCT has little to say about how respondents arrive at their subjective beliefs, a factor which necessarily constitutes an important part of the explanation. In our view, this approach has to be extended using findings from cognitive psychology about the determinants of subjects' beliefs and subjects' motivations in the survey context. This knowledge can be used to predict the values of those parameters which form the input into the rational-choice approach. Such a theoretical 'alliance' would contribute to a deeper understanding of response behavior. *Thirdly*, it may be necessary, in particular when the theory is applied in the field of response behavior where subjects are typically insufficiently motivated regarding elaborated information processing, to attenuate the RCT's strong rationality assumption. Thus, future studies should test whether the predictive power for the determinants of SD-bias increases when elements from satisficing-theory are included (Krosnick 1991).

Some questions have remained unanswered in our study which deserves attention in subsequent research. In the present study it has been assumed that the respondents' subjective feelings of privacy are completely determined by the objective features of the response situation. This assumption has not been tested empirically in ours nor in any other study we are aware of. It would therefore be worthwhile to analyze how strongly objective conditions and subjective feelings are in fact associated, and whether a direct measure of privacy feelings would improve the explanatory power of the model for explaining SD-bias presented in this paper. Furthermore, we have operationalized respondents' desirability beliefs using the difference in the perceived desirability of extremely positive and negative racial attitude answers. Compared with the usually utilized measures of desirability beliefs, which are restricted to the evaluation of one end of the attitude continuum, this is a relatively elaborated indicator. However, our operationalization assumes that these beliefs increase or decrease monotone across the attitude continuum. For other questionnaire topics this assumption has been found to be invalidated, since a substantial proportion of respondents perceived inversely u-shaped desirability profiles across the analyzed attribute continuum (Stocké and Hunkler 2004). Since under these conditions the indicator used in our study would underestimate incentives stemming from social desirability, the presence of non-monotone desirability beliefs should be tested empirically. The results from our main as well as those from the validation study are based on relatively small, local probability samples. Although our 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. We would therefore suggest a replication of these results by using a representative nationwide sample.

REFERENCES

Aiken, L.S. and West, S.G. (1991). *Multiple Regression: Testing and Interpreting Interactions.* Newbury Park: Sage.

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 & Misuse, 35,* 845-867.

Brunet, A., Boucher, C., and Boyer, R. (1996). Social Desirability in the Assessment of Trauma. *Psychological Reports, 79,* 511-514.

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.

Couper, M.P., Singer, E., and Tourangeau, R. (2003). Understanding the Effects of Audio-CASI on Self-Reports of Sensitive Behavior. *Public Opinion Quarterly, 67,* 385-395.

Cronbach, L.J. (1987). Statistical Tests for Moderator Variables: Flaws in Analyses Recently Proposed. *Psychological Bulletin, 102,* 414-417.

Crott, H.W. and Roßrucker, K. (1974). Erwünschtheit von Eigenschaften in Abhängigkeit von Alter, Geschlecht und Schichtzugehörigkeit. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 6,* 241-261.

Crowne, D.P. (1979). *The Experimental Study of Personality.* New York: John Wiley.

Crowne, D.P. and Marlowe, D. (1960). A New Scale of Social Desirability Independent of Psychopathology. *Journal of Consulting Psychology, 24,* 349-354.

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.

Edwards, A.L. (1953). The Relationship Between the Judged Desirability of a Trait and the Probability That the Trait Will Be Endorsed. *The Journal of Applied Psychology, 37,* 90-

93.

Ellingson, J.E., Smith, D.B., and Sackett, P.R. (2001). Investigating the Influence of Social Desirability on Personality Factor Structure. *Journal of Applied Psychology, 86,* 122-133.

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: R. Wittenberg (ed.), *Person - Situation - Institution - Kultur. Günter Büschges zum 65. Geburtstag*. Berlin: Duncker & Humblot.

Fisher, R.J. and Katz, J.E. (2000). Social-Desirability Bias and the Validity of Self-Reported Values. *Psychology & Marketing, 17,* 105-120.

Gove, W.R. and Geerken, M.R. (1977). Response Bias in Surveys of Mental Health: An Empirical Investigation. *American Journal of Sociology, 82,* 1289-1317.

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.

Huber, P.J. (1964). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics, 35,* 73-101.

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.

Hult, G.T.M., Keillor, B.D., and Lafferty, B.A. (1999). A Cross-National Assessment of Social Desirability Bias and Consumer Ethnocentrism. *Journal of Global Marketing, 12,* 29-43.

Krosnick, J.A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology, 5,* 213-236.

Krysan, M. (1998). Privacy and the Expression of White Racial Attitudes. *Public Opinion Quarterly, 62,* 506-544.

Krysan, M., Schuman, H., Scott, L.J., and Beatty, P. (1994). Response Rates and Response Content in Mail Versus Face-to-Face Surveys. *Public Opinion Quarterly, 58,* 381-399.

Laicardi, C., Baldassarri, F., and Artistico, D. (2001). Unidimensionality of Life Satisfaction and its Relation to Social Desirability: A Confirmatory Study of a Short Form of the Life Satisfaction Scale. *Psychological Report, 88,* 253-261.

Leak, G.K. (1981). Susceptibility of the Modernity Scale to Socially Desireable Response Sets. *Psychological Reports, 48,* 689-690.

Mangione, T.W., Hingson, R., and Barrett, J. (1982). Collecting Sensitive Data. A Comparison of Three Survey Strategies. *Sociological Methods & Research, 10,* 337-346.

Meleddu, M. and Guicciardi, M. (1998). Self-Knowledge and Social Desirability of Personality Traits. *European Journal of Personality, 12*, 151-168.

Mielke, R. (1995). *Der Bildungsgrad und die Einstellung gegenüber Ausländern.* Universität Bielefeld: Bielefelder Arbeiten zur Sozialpsychologie 173.

Mummendey, H.D. and Eifler, S. (1993). *Eine neue Skala zur Messung sozialer Erwünschtheit.* Universität Bielefeld: Bielefelder Arbeiten zur Sozialpsychologie 167.

Nicotera, A.M. (1996). An Assessment of the Argumentativeness Scale For Social Desirability Bias. *Communication Reports, 9,* 23-36.

Paulhus, D.L. and Reid, D.B. (1991). Enhancement and Denial in Socially Desirable Responding. *Journal of Personality and Social Psychology, 60,* 307-317.

Phillips, D.L. and Clancy, K.J. (1972). Some Effects of "Social Desirability" in Survey Studies. *American Journal of Sociology, 77,* 921-940.

Reinecke, J. (1991). *Interviewer- und Befragtenverhalten. Theoretische Ansätze und methodische Konzepte.* Opladen: Westdeutscher Verlag.

Reuband, K.-H. and Blasius, J. (1996). Face-to-Face-, telephonische und postalische Befragungen. Ausschöpfungsquoten und Antwortmuster in einer Großstadt-Studie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 48,* 296-318.

Sniderman, P.M., Piazza, T., Tetlock, P.E., and Kendrick, A. (1991). The New Racism. *American Journal of Political Science, 35,* 423-447.

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. (2001). *Socially Desirable Response Behavior as Rational Choice: The Case of Attitudes Towards Foreigners.* Arbeitspapier No. 01-22: Sonderforschungsbereich 504, Universität Mannheim.

Stocké, V. (2003). Informationsverfügbarkeit und Response-Effects: Die Prognose von Einflüssen unterschiedlich kategorisierter Antwortskalen durch Antwortsicherheiten und Antwortlatenzen. *ZA-Information, 52,* 6-36.

Stocké, V. and Hunkler, C. (in press). Die angemessene Erfassung der Stärke und Richtung von Anreizen durch soziale Erwünschtheit. *ZA-Information.*

Thunholm, P. (2001). Social Desirability in Personality Testing of Military Officers. *Military Psychology, 13,* 223-234.

Todorov, A. (2000). The Accessibility and Applicability of Knowledge: Predicting Context

Effects in National Surveys. *Public Opinion Quarterly, 64,* 429-451.

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. 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.

Trimble, D.E. (1997). The Religious Orientation Scale: Review and Meta- Analysis of Social Desirability Effects. *Educational and Psychological Measurement, 57,* 970-986.

Vella-Broderick, D.A. and White, V. (1997). Response Set of Social Desirability in Relation to the Mental, Physical and Spiritual Well-being Scale. *Psychological Reports, 81,* 127-130.

Wyner, G.A. (1980). Response Errors in Self-Reported Number of Arrests. *Sociological Methods & Research, 9,* 161-177.

APPENDIX

Instrument for measuring the desirability beliefs about positive and negative racial attitudes

‘In all societies certain opinions are likely to cause negative reactions in others when expressed in public. Other opinions are in agreement with the unwritten rules about what is allowed to be stated in public and are therefore typically positive evaluated. 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. Please don’t tell me your own opinion about this topic or the viewpoint of the majority in Germany, but what you believe can be said in public and what not. Please use this response scale, ranging between minus four, ‘to express this opinion in public would be very embarrassing’, and plus four, ‘to express this opinion in public would be very pleasant’, to answer the following questions. A value of zero means that this would be neutral for the person.’

Desirability beliefs about a strong agreement with racial attitude item 1

(for all other items asked in a equivalent way)

‘Would it be embarrassing or pleasant for the person mentioned above, to express in public the convinced opinion that foreigners in Germany should adapt their lifestyle more to that of the Germans?’

Desirability beliefs about a strong disagreement with racial attitude item 1

(for all other items asked in a equivalent way)

‘Would it be embarrassing or pleasant for that person to express in public the convinced opinion that foreigners in Germany should not be forced to adapt their lifestyles more to that of Germans?’

Nr.	Author	Title
04-18	Alexander Zimper	Dominance-Solvable Lattice Games
04-17	Volker Stocké Birgit Becker	DETERMINANTEN UND KONSEQUENZEN DER UMFRAGEEINSTELLUNG. Bewertungsdimensionen unterschiedlicher Umfragesponsoren und die Antwortbereitschaft der Befragten
04-16	Volker Stocké Christian Hunkler	Die angemessene Erfassung der Stärke und Richtung von Anreizen durch soziale Erwünschtheit
04-15	Elena Carletti Vittoria Cerasi Sonja Daltung	Multiple-bank lending: diversification and free-riding in monitoring
04-14	Volker Stocké	The Interdependence of Determinants for the Strength and Direction of Social Desirability Bias in Racial Attitude Surveys
04-13	Mei Wang Paul Fischbeck	Evaluating Lotteries, Risks, and Risk-mitigation Programs Ño A Comparison of China and the United States
04-12	Alexander Ludwig Torsten Sløk	The relationship between stock prices, house prices and consumption in OECD countries
04-11	Jens Wüstemann	Disclosure Regimes and Corporate Governance
04-10	Peter Albrecht Timo Klett	Referenzpunktbezogene risikoadjustierte Performancemaße: Theoretische Grundlagen
04-09	Alexander Klos	The Investment Horizon and Dynamic Asset Allocation - Some Experimental Evidence
04-08	Peter Albrecht Cemil Kantar Yanying Xiao	Mean Reversion-Effekte auf dem deutschen Aktienmarkt: Statistische Analysen der Entwicklung des DAX-KGV
04-07	Geschäftsstelle	Jahresbericht 2003
04-06	Oliver Kirchkamp	Why are Stabilisations delayed - an experiment with an application to all pay auctions