



Estimating the prevalence of xenophobia and anti-Semitism in Germany: A comparison of randomized response and direct questioning

Ivar Krumpal*

University of Leipzig, Institute of Sociology, Beethovenstrasse 15, 04107 Leipzig, Germany

ARTICLE INFO

Article history:

Received 13 July 2011

Revised 24 May 2012

Accepted 31 May 2012

Available online 15 June 2012

Keywords:

Xenophobia

Anti-Semitism

Sensitive questions

Social desirability bias

Randomized response technique

Survey design

ABSTRACT

An experimental CATI-survey ($N = 2041$), asking sensitive questions about xenophobia and anti-Semitism in Germany, was conducted to compare the randomized response technique (RRT) and the direct questioning technique. Unlike the vast majority of RRT surveys measuring the prevalence of socially undesirable behaviors, only few studies have explored the effectiveness of the RRT with respect to the disclosure of socially undesirable opinions. Results suggest that the RRT is an effective method eliciting more socially undesirable opinions and yielding more valid prevalence estimates of xenophobia and anti-Semitism than direct questioning ('more-is-better' assumption). Furthermore, the results indicate that with increasing topic sensitivity, the benefits of using the RRT also increase. Finally, adapted logistic regression analyses show that several covariates such as education and generalized trust are related to the likelihood of being prejudiced towards foreigners and Jews.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Many surveys include so-called 'sensitive questions', i.e. questions that ask about socially undesirable answers indicating the violation of a social norm. Due to self-presentation concerns, many respondents misreport on sensitive behaviors such as drug use (McAuliffe et al., 1991; Näher and Krumpal, 2011), abortion (Lara et al., 2004, 2006), shoplifting (Coutts and Jann, 2011), plagiarism (Coutts et al., 2011; Jann et al., 2012), noncompliance with social benefit rules (van den Hout et al., 2010) and voting (Holbrook and Krosnick, 2010), as well as sensitive attitudes such as racism (Krysan, 1998), sexism (Roese and Jamieson, 1993), and xenophobia (Stocké, 2004, 2007). Facing an interviewer, respondents tend to underreport socially undesirable behaviors or attitudes and overreport socially desirable ones. As a consequence of misreporting, prevalence estimates are often distorted by social desirability bias.

To combat deliberate misreporting in sensitive surveys, the randomized response technique (RRT; Warner, 1965) was developed to protect the respondent's privacy in the interview situation and obtain more valid self-reports to sensitive questions. All RRT schemes rely on the principle that a random mechanism is used in order to generate a probabilistic relationship between the sensitive question posed and the response given. The respondent uses a randomizer (e.g. coins or dice) to determine whether he will answer to the sensitive question or a surrogate. Since the interviewer is unaware of the outcome of the random experiment, a given answer does not reveal anything definite about the respondent's true status. Given the assumption that respondents understand the RRT scheme and comply with the procedure, more accurate self-reports to sensitive questions are expected compared to direct questioning.

In the following, Warner's original scheme is described to illustrate the rationale of RRT schemes in general: The respondent is confronted with two statements, the socially undesirable one (e.g., 'I think that Jews, more than other people, work

* Fax: +49 341 97 35669.

E-mail address: krumpal@sozio.uni-leipzig.de

with tricks and deception') and its negation (e.g., 'I do not think that Jews, more than other people, work with tricks and deception'). The interviewer asks 'Do you somewhat agree with the following statement?' Using a randomizer, the respondent determines which of the two statements he will answer. For example, the respondent may be given a box of seven colored marbles, four red and three blue marbles, and told to take one marble out of the box and to respond to the first statement if a red marble is selected, but to respond to the second statement if a blue marble is selected. Without revealing the outcome of the random experiment to the interviewer, the respondent answers with either a 'yes' or a 'no' according to his attitude towards Jews. Since the meaning of the answer given is obscured, Warner (1965) assumed that the respondent's cooperation and the validity of self-reports in sensitive surveys would increase.

In Warner's design, the prevalence of the unsocial opinion can be estimated on the basis of elementary probability theory: The expected value ϕ of observing a 'yes' answer can be modeled as $\phi = p\pi + (1 - p)(1 - \pi)$, where π is the unknown proportion of anti-Semites in the population, and $p(p \neq 0.5)$ is the probability that the statement 'I think that Jews, more than other people, work with tricks and deception' is selected. Since the observed sample proportion of 'yes' answers is an estimate of ϕ , and the selection probability p is given by design, the population prevalence of the unsocial opinion π can be estimated. Such probabilistic link between the observed answer and the respondent's true status is also at the heart of alternative RRT schemes (overviews of proportion and variance estimators for different RRT schemes can be found in Fox and Tracy, 1986).

The gains of protection from RRT come at higher data collection costs. The procedure introduces additional random error to the data and increases standard errors of the parameter estimates. Thus, parameter estimates generated by RRT schemes have lower statistical efficiency compared to estimates generated by direct questioning (Fox and Tracy, 1986). Larger sample sizes are necessary to increase statistical power. In addition, RRT questions are more complex compared to more conventional interview methods. They impose a higher cognitive burden on the question-and-answer process (Stem and Steinhilber, 1984).

A recent meta-analysis of RRT research summarizes the results of six validation studies and 32 experimental studies without validation data comparing the RRT with other interview methods such as direct questioning (Lensvelt-Mulders et al., 2005a). In the first type of studies, individual 'true scores' were available so that researchers could assess the proportions of correct answers and compare these proportions across experimental conditions. In the second type of studies, prevalence estimates yielded via different data collection procedures were compared with respect to an increase of socially undesirable answers ('more-is-better' assumption). Overall, the results indicate that self-reports on sensitive issues were more accurate and more socially undesirable answers were elicited respectively when RRT was employed. However, other studies have found no superiority of the RRT and standard direct questioning sometimes elicited more socially undesirable answers than did the RRT (Holbrook and Krosnick, 2010). Unlike the vast majority of RRT surveys measuring the prevalence of socially undesirable behavior like abortion or delinquency (for overviews see Fox and Tracy, 1986; Holbrook and Krosnick, 2010), only few studies have explored the effectiveness of the RRT with respect to the disclosure of unsocial attitudes such as racism (Ostapczuk et al., 2009).

This article shows results from an experimental CATI-survey yielding prevalence estimates of xenophobia and anti-Semitism in Germany and comparing the RRT with direct questioning. In the next section, a short and selective review of the research on xenophobia and anti-Semitism will be presented. In the subsequent section, the study design and the implementation of the RRT will be described. The analysis section presents prevalence estimates and logistic regression models for RRT data to analyze the dependence of agreement with a xenophobic or an anti-Semitic statement respectively on a set of covariates. Finally, the results will be discussed.

2. Asking sensitive questions: xenophobia and anti-Semitism

Since the end of the Second World War, an extensive research interest in the origins of anti-Semitism, xenophobia, and right-wing extremism has emerged from the traditional research line on authoritarianism in Germany (Adorno et al., 1950; Bergmann and Erb, 1986, 1991; Beyer and Krumpal, 2010; Decker and Brähler, 2005, 2006; Zick et al., 2008). Against the background of the recent electoral successes of rightist extremist parties on the federal state level (e.g. in the state elections in 2004, 2009 in Saxony and in 2006 in Mecklenburg–Vorpommern, the National Democratic Party of Germany, NPD received 9.2%, 5.6% and 7.3% of the total votes and thus achieved representation in the 'Landtag', the German legislative assembly at the state level), public concern has risen and growing attention has been directed to the investigation of the causes of these parties' recent popularity. Besides the investigation of determinants of xenophobic and anti-Semitic attitudes, social scientists aim for providing accurate prevalence estimates in order to monitor the development of these attitudes in the general population.

While the term 'anti-Semitism' subsumes negative stereotypes and prejudice against Judaism and Jewish people, the term 'xenophobia' refers to the various ways of depreciation and rejection of immigrants and foreign nationals. Although anti-Semitism and xenophobia are strongly interrelated, they differ in terms of motives, prevalence and public acceptance (see Bergmann, 1997 for an overview): Prejudices toward Jews in Germany today are often being analyzed within the context of a historic perspective. The Holocaust and the subsequent decline of National Socialism resulted in reparations and feelings of guilt which in turn shaped attitudes toward Jewish people. Stereotypes about Jews and Judaism often assume an exorbitant financial strength and too much influence of obscure and strong forces using unethical means in order to manipulate media, economy and the political system. In an empirical study on right-wing extremism in Germany after re-unification

the following two items were used to measure this dimension (Decker and Brähler, 2006): 'Even today the influence of the Jews is too high' and 'The Jews, more than other people, work with tricks and deception in order to achieve what they want' respectively. On five-point answer scales, respondents were asked to state how strongly they would agree/disagree with these statements.

Another source of prejudice against Jews as a group involves attempts to allocate the blame and belittle the responsibility for crimes against Jews during the Second World War. Items reflecting this complex issue were included in the German General Social Survey (ALLBUS) 2006. The ALLBUS is a biennial national data collection program that has been collecting data on social structure in Germany since 1980. The main topic of the ALLBUS 2006 was attitudes to different ethnic groups in Germany. The survey included item-batteries on acceptance of immigration, of ethnic minorities, and of Jews in Germany. For example, one item was worded as follows: 'The persecution of the Jews is to a certain extent their own fault due to their behavior'. Another anti-Semitic stereotype involves the suspicion that Jews intentionally profit from German history: 'Many Jewish people today try to take personal advantage of what happened during the Nazi era and make Germans pay for it.' On seven-point answer scales, respondents rated how strongly they would agree/disagree with these statements.

By contrast, competition on the labor market, entitlement to welfare services and feelings of social distance have been the major issues shaping sentiments towards migrant workers and foreigners in post-war Germany (Bergmann, 1997). Prejudices against lower-class foreigners are primarily based on social competition and conflict and differ from the traditional stereotypes imputed to middle-class, socially integrated Jews. The negative attitudes toward foreign nationals tend to involve accusations such as they would take away the jobs from the Germans. Furthermore, fears of foreign influences threatening the German culture pose another source of xenophobic hostility, in particular toward Turks and Arabs. Finally, foreigners are often suspected of committing social fraud and to intentionally exploit the welfare system. The prevalence and distribution of such prejudices in the German population have been measured via the degree of agreement to items reflecting clichés and negative stereotypes such as (Decker and Brähler, 2006): 'If jobs are in short supply, immigrants should be sent home', or 'Foreigners only come to Germany to take advantage of our welfare state'.

Preliminary findings indicate that a considerable proportion of Germans still share resentments toward immigrants and Jews, whereas xenophobia is more prevalent compared to the spread of anti-Jewish attitudes (Decker and Brähler, 2006). In addition, different degrees of public sanctioning of anti-Semitic and xenophobic opinions can be observed (Bergmann, 1997): A strong social taboo against anti-Semitism has emerged and Holocaust denial has been prosecuted since the German post-war era of 'reeducation'. The collective responsibility for the persecution of Jews has reinforced the public norm of 'anti-anti-Semitism' resulting in heightened public awareness and strict public censure of anti-Semitism in politics, media, and the public opinion. On the contrary, such a strong public taboo has not emerged with respect to xenophobia. Resentment of immigrants and non-German groups are more common compared to resentments of Jews and Judaism. Nevertheless, clear social norms disapproving xenophobic opinions and xenophobia in general are shared by the majority of the Germans in the 21st century. Consequently, regarding the degree of sensitivity, questions asking the respondents to disclose their opinions toward Jews are considered more sensitive compared to questions asking about attitudes toward foreigners.

Since social norms against xenophobic and anti-Semitic prejudices can be observed in the general public, survey answers indicating such prejudices are deemed socially undesirable. The disclosure of such unsocial opinions can result in consequences such as disapproval of the interviewer or social sanctions from third parties beyond the survey setting. Due to self-presentation concerns and the fear of social ostracism, a subset of respondents sharing xenophobic or anti-Semitic attitudes but at the same time being aware of the public norms against these attitudes is expected to conceal what they really think and to distort their self-reports toward the public norm. As a consequence of underreporting xenophobic and anti-Semitic attitudes on the individual level, survey estimates on the aggregate level will be distorted by social desirability bias underestimating the prevalence of such unsocial attitudes. Different empirical studies have investigated socially desirable responding in surveys measuring unsocial attitudes such as anti-Semitism, right-wing extremism and racism (Beyer and Krumpal, 2010; Decker and Brähler, 2006; Ostapczuk et al., 2009).

3. The study

This study compares the RRT with the direct questioning method in order to analyze the extent of social desirability bias in prevalence estimates of xenophobia and anti-Semitism in Germany. Because of the expected systematic underreporting of xenophobic and anti-Semitic opinions, it is assumed that, in comparing prevalence estimates of the two questioning techniques, the higher estimates will be the more valid ones ('more-is-better' assumption, see Lensvelt-Mulders et al., 2005a; Tourangeau and Yan, 2007; Krumpal, 2011).¹ The first hypothesis can be stated as follows: *Since anonymity of the question-and-answer process is greater in the RRT condition, more valid self-reports on the socially sensitive topics are expected when data*

¹ Note that demonstrating validity of measures of subjective phenomena is a challenging task (Roese and Jamieson, 1993). Unlike answers to factual questions (e.g. social fraud, abortion) that can, in principle, be verified (e.g. in the context of a record check study), there is no possibility to validate answers to non-factual, attitudinal questions with clear and objective 'true scores'. Thus, the validity of answers to attitudinal questions can only be assessed indirectly via empirical consistency criteria such as a consistent and replicable effect of the RRT across different questions and indication that this effect occurs in the theoretically expected direction, i.e. more socially undesirable opinions under conditions of anonymity. Although the 'more-is-better' assumption is highly plausible in the context of this study – xenophobic and anti-Semitic prejudices are socially sanctioned and therefore very likely to be underreported in an interviewer-administered survey – it is still just an assumption.

collection is conducted via the RRT. Furthermore, former research indicates that the degree of topic sensitivity moderates the effectiveness of the RRT (Lensvelt-Mulders et al., 2005a). More specifically, with increasing topic sensitivity the benefits of making the interview conditions more anonymous should also increase. This *second hypothesis* will be examined by comparing the ratios of the prevalence estimates under RRT to the prevalence estimates under direct questioning: *With increasing topic sensitivity the ratios of the prevalence estimates under RRT to the prevalence estimates under direct questioning are expected to increase.*

An additional aim of this study is the analysis of explanatory variables that are associated with xenophobia and anti-Semitism: First, former research shows that opinions about foreigners and Jews in Germany differ for different sociodemographic characteristics such as education, age, and employment status (Decker and Brähler, 2006). Second, psychological and situational variables (e.g. trust, perceived questions' sensitivity, and the presence of bystanders during the interview) have been found important predictors of the respondent's willingness to disclose sensitive characteristics in a survey interview (Tourangeau and Yan, 2007). In the data analysis section, more details will be given about the covariates and the specific hypotheses relating these covariates to the prejudiced attitudes toward foreigners and Jews.

3.1. Research design

The survey's target population included all German citizens of the general population who resided in private households with a landline phone connection and who, on the day of the interview, were at least 18 years old. Subjects were selected using a national, two-stage random sampling procedure: In the first sampling stage, private households were randomly selected via drawing landline phone numbers from the 'ADM telephone sample system', a sampling frame for CATI surveys containing both phone numbers listed in directories and phone numbers not included in any such directories.² The sampling frame is based on the 'Gabler-Häder' design, a random digit dialing (RDD) procedure which has been adapted to the specific German structure of landline phone numbers (Gabler and Häder, 1997, 2002). In the second sampling stage, target persons within households were randomly selected via a random selection card (Kish, 1965).³

The data collection and field work was conducted by a social-science survey institute (USUMA, Berlin). Overall, 2041 subjects completed the CATI interviews. The response rate was 36.4% (AAPOR, RR3). The survey was introduced as being a scientific study concerning the topic of 'opinions on different segments of the population'. 38 interviewers carried out the interviews from January 27 to March 20, 2009. The interviews were conducted in German. At the beginning of the interview, the respondents were randomly assigned to one of the two experimental conditions at a ratio of 2 (RRT) to 1 (direct questioning). The RRT was oversampled because it has lower statistical power than direct questioning.

3.2. The two experimental conditions

The RRT is more difficult to communicate to the respondents than direct questioning, especially in a CATI survey. Therefore, a suitable instrument based on pretests was developed by: (a) holding discussions with a group of university students in the context of a master-level research seminar, (b) consulting experts in RRT research, and (c) conducting qualitative interviews ($N = 20$). Based on these pretests, the RRT instructions were simplified and ambiguous formulations were clarified. Furthermore, a small pilot-study ($N = 60$) was conducted in order to identify and to correct errors in conjunction with the sampling procedure, the wording and sequence of questions, the code structure, and the routing instructions of the questionnaire.⁴ In order to become familiar with the RRT, interviewers received comprehensive training. For the purposes of quality control, members of the research team were present during the data collection phase, closely monitored the field work (via headphones and screen) and evaluated the interviewers' performance. The results of the supervision were satisfactory.

In the 'direct questioning' – condition ($N = 678$), a confidentiality assurance was read out loud to the respondent. The specific introduction of the sensitive items is visualized in Fig. 1.

² The ADM ('Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.') is the representation of private market and social research agencies in Germany.

³ Due to budget constraints, only households with landline telephone numbers were sampled. This sampling scheme excludes certain subpopulations, in particular no-phone and mobile-only populations. In Germany, the estimated rate of the no-phone population is 1% (Busse and Fuchs, 2012) and the estimated rate of the mobile-only population (which is the main source of non-coverage of landline telephone samples) is 8% (Kunz and Fuchs, 2011). Furthermore, estimates of the relative coverage bias in traditional landline telephone surveys in Germany indicate that the potential bias caused by the mobile-only population is relatively stronger than bias introduced by the no-phone population which is estimated to be virtually zero (these estimates were based on two sociodemographic variables, namely 'age' and 'type of community'; see Busse and Fuchs, 2012). With regard to the present study, the non-coverage of the mobile-only population could potentially influence the absolute levels of the prevalence estimates of the sensitive attitudes. However, non-coverage is not expected to influence the comparisons of the two data collection techniques resp. the assessment of differences between the experimental groups. To account for these potential sources of survey error, all analyses reported in the Tables 2–5 were recalculated using weights. Besides design weighting (accounting for the unequal selection probabilities resulting from different household sizes) post-stratification weights were used to adjust the sample distributions according to the German population distributions for gender, age and region (see Gabler et al., 1994; Lensvelt-Mulders et al., 2006; the population distributions can be obtained from the German statistical offices). Overall, differences between the weighted and unweighted analyses were negligible and the conclusions remain the same. In the following, only the unweighted analyses are reported.

⁴ Except for the differing experimental part, the two questionnaire versions had the same structure: (a) general introduction; (b) sociodemographic characteristics; (c) specific introduction of the experimental part (RRT instructions versus direct questioning); (d) sensitive items reflecting xenophobic and anti-Semitic stereotypes; (e) follow-up questions measuring psychological and situational variables (such as perceived questions' sensitivity, trust, and presence of third persons during the interview); (f) the interviewer's observations of the respondents' behavior which were documented in a separate questionnaire after the regular interviews were finished.

“Now we would like to know your personal opinion on different segments of the population. One of these segments is foreigners living in Germany. We are aware of the fact that many people are very hesitant about giving their personal opinion on topics like this because they are very private. With this in mind, the University of Leipzig would like to assure you once again that all answers given will be kept confidential and will not be passed on to anyone else. I will now read aloud some statements to you which you might have heard at some point before. Please tell me each time, whether or not you would somewhat agree with the statement. We will now start with the first statement. Would you somewhat agree with the following statement? “There are too many foreigners in Germany.” (...).”

Fig. 1. Direct questioning.

After reading the introduction, three items reflecting xenophobic views and two items reflecting anti-Semitic resentments were read out loud to the respondent. The sensitive items were dichotomous with the possible answer categories ‘Yes’ and ‘No’. Before reading an item, the interviewer each time asked: “Would you somewhat agree with the following statement?”. The five sensitive items were worded as follows:

- *Xenophobia 1*: “There are too many foreigners in Germany.”
- *Xenophobia 2*: “The German Federal Republic is dangerously swamped by foreign influences because of all the foreigners living in the country.”
- *Xenophobia 3*: “Foreigners only come to Germany to take advantage of our welfare state.”
- *Anti-Semitism 1*: “Even today the influence of the Jews is too high.”
- *Anti-Semitism 2*: “The persecution of the Jews is to a certain extent their own fault due to their behavior.”

In the RRT condition ($N = 1363$), the ‘forced response design’ was implemented (Boruch, 1971). Compared to alternative variants of the RRT, the implementation of the ‘forced response design’ is less complex and thus is particularly suitable for CATI applications. Unlike Warner’s method (Warner, 1965), the respondent is required to answer to only one statement (instead of two). Moreover, the ‘forced response design’ yields the most efficient statistical estimation compared to alternative RRT schemes (Lensvelt-Mulders et al., 2005b).

Three coins were used as a randomizer. This has the advantage that the randomizer is under the control of the respondent making it impossible for the interviewer to discover the result of the random experiment and to decode the meaning of a given answer. The interviewers were instructed to check the respondents’ understanding by means of an example. If the respondent made mistakes, the rules had to be repeated and the practice example had to be carried through again. After making sure that the respondent correctly applied the rules, the interviewer emphasized the importance to comply with the rules. Furthermore, if a respondent was distrustful or had objections, the interviewers were trained to identify the doubts and to show that they were unfounded. If after several attempts the respondent was unable or not willing to get the coins, the interviewer was instructed to continue with direct questioning. Respondents in the RRT-condition which rejected the RRT for direct questioning got a separate code ‘RRT-direct’. The introduction of the ‘coin-flip-method’ is visualized in Fig. 2.

3.3. Methods of data analysis

The probability distribution of having to answer to a sensitive item and having to give a surrogate answer respectively is known by design: The probability of being directed to give an automatic ‘yes’-answer (three tails) is $p_1 = 0.5^3 = 0.125$, the probability of being directed to give an automatic ‘no’-answer (three heads) is $p_2 = 0.5^3 = 0.125$, and the probability of being requested to give a truthful answer (two heads and one tail; one head and two tails) is $p_3 = 1 - p_1 - p_2 = 1 - 0.5^3 - 0.5^3 = 0.750$. Given the design probabilities, the expected value ϕ of observing a ‘yes’-answer can be written as $\phi = p_1 + p_3\pi$, where π is the unknown population proportion of subjects agreeing with a sensitive item. Since the observed sample proportion of ‘yes’-answers is an estimate of ϕ , the population prevalence of the unsocial opinion π can be estimated via (Boruch, 1971; Fox and Tracy, 1986):

$$\hat{\pi} = \frac{\hat{\phi} - p_1}{p_3} \quad (1)$$

where $\hat{\phi}$ is the observed sample proportion of ‘yes’-answers. The sampling variance of the estimator in the forced response design $\text{Var}(\hat{\pi})$ is given by:

$$\text{Var}(\hat{\pi}) = \frac{\hat{\phi}(1 - \hat{\phi})}{np_3^2} \quad (2)$$

Based on the above formulas, 90%-confidence intervals (z values: -1.65 ; $+1.65$) and 95%-confidence intervals (z values: -1.96 ; $+1.96$) respectively were generated for the estimates of the proportions of subjects who agreed with the socially undesirable statements. Further analyses were as follows: First, a descriptive analysis and a comparison of the sociodemographic

“Now we would like to know your personal opinion on different segments of the population. One of these segments is foreigners living in Germany. We are aware of the fact that many people are very hesitant about giving their personal opinion on topics like this because they are very private. With this in mind, the University of Leipzig has developed a novel question technique that guarantees your privacy and makes the interview more comfortable.

When answering the following questions, you can keep your personal opinion secret by flipping a coin. This might sound a bit unusual, however, I would like to ask you to help us and try out this new method together with us. Could you please get three coins as well as a piece of paper and a pen? (...)

Please flip the three coins each time before I ask you a question. However, please do not tell me the results! Depending on the result of the coin flip, please answer as follows. I am happy to give you some time to write down the rules, if you would like:

- *If you flip tails 3 times, please always answer “yes”.*
- *If you flip heads 3 times, please always answer “no”.*
- *If you flip a combination of heads and tails, for example tails 2 times and heads 1 time, please always tell your true personal opinion.*

As you can see, coincidence will decide whether you answer the question truthfully or whether you give a predetermined answer. This way your privacy will always be protected. I will not know the result of your coin flip and therefore I will never know why your answer is “yes” or “no”. Did you understand the coin-flip method? (...)

Sometimes you will answer “yes” or “no” due to the result of your coin flip, even though this is not your real personal opinion. Please do not worry about that. You are doing the right thing if you follow the rules of the coin-flip method and always answer according to the result of the coin-flip.

I will now read out aloud some statements to you which you might have heard at some point before. Please tell me each time, according to your coin flip, whether or not you would somewhat agree with the statement. We will now start with the first statement. (...) Please flip your three coins now without telling me the result. According to your coin flip, would you somewhat agree with the following statement? “There are too many foreigners in Germany” (...).”

Fig. 2. Randomized response technique (RRT).

distributions between the two experimental subsamples were conducted. Second, multiple logistic regressions for RRT data were estimated to study the relationship between explanatory variables and a dichotomous response variable generated via the RRT. Both groups, RRT and direct questioning, were analyzed within the same logistic regression model using the STATA module ‘rrlogit’ (Jann, 2005). This procedure is equivalent to the maximum likelihood method described in Maddala, 1983, pp. 54–56), Scheers and Dayton (1988) and van der Heijden et al. (2000, p. 529).⁵ The binary logistic regression model is defined as:

$$\pi(\mathbf{x}_i) = \frac{\exp(b_0 + \sum_j x_{ij} b_j)}{1 + \exp(b_0 + \sum_j x_{ij} b_j)} \quad (3)$$

where \mathbf{x}_i denotes a vector of covariates of respondent i , where $\pi(\mathbf{x}_i)$ denotes the conditional probability of answering ‘yes’ to the socially undesirable statement given \mathbf{x}_i , where b_j denotes the regression coefficient for covariate j , and where x_{ij} denotes the j th element of \mathbf{x}_i . In the case of the RRT design, the adjusted log likelihood function is defined as:

$$\log L = \sum_i n_{i1} \log[p_1 + p_3 \pi(\mathbf{x}_i)] + \sum_i n_{i0} \log[p_2 + p_3 (1 - \pi(\mathbf{x}_i))] \quad (4)$$

$$\text{with } \pi(\mathbf{x}_i) = \frac{\exp(b_0 + \sum_j x_{ij} b_j)}{1 + \exp(b_0 + \sum_j x_{ij} b_j)} \quad (5)$$

where $1 - \pi(\mathbf{x}_i)$ denotes the conditional probability of answering ‘no’ to the socially undesirable statement given \mathbf{x}_i , and where p_1 and p_2 denote the design probabilities of giving automatic ‘yes’-answers and ‘no’-answers respectively. Furthermore, n_{i1} denotes the number of respondents answering ‘yes’ to the sensitive item and n_{i0} denotes the number of respondents which answer ‘no’. The term $p_1 + p_3 \pi(\mathbf{x}_i)$ refers to the probability of observing a ‘yes’ answer given the characteristics \mathbf{x}_i and the term $p_2 + p_3 (1 - \pi(\mathbf{x}_i))$ refers to the corresponding conditional probability of observing a ‘no’ answer. Note that in the case of direct questioning, the design probabilities p_1 and p_2 both equal zero reducing the RRT logistic regression model to the regular logistic regression model with unadjusted log likelihood function. In the RRT design specified

⁵ Note that, besides maximum likelihood analyses which work sufficiently well to achieve the goals defined in this study, it is also possible to use a Bayesian framework to analyze data generated via the RRT (see van den Hout et al. (2010) for an innovative approach).

above, the probabilities p_1 and p_2 both equal 0.125. Data from the RRT and from direct questioning can be jointly analyzed within the same logistic regression model through observation-specific values for p_1 and p_2 . The log likelihood has to be maximized over the parameters b_j using an iterative process. For greater detail, refer to the sources cited above.

Based on the adjusted logistic regression model, odds ratios $\exp(b_j)$ quantifying the effect of the covariates were calculated along with the corresponding standard errors and p -values.⁶ The RRT odds ratios can be interpreted just like the odds-ratios from a regular logistic regression model. Furthermore, log-likelihood based indicators of model fit are reported: McFadden's Pseudo R^2 comparing a model with all parameters to a model with just the intercept is presented along with the likelihood-ratio χ^2 value that, too, is based on the difference between the log-likelihoods for the full and intercept-only models (Long and Freese, 2006, pp. 99–102, 109).

4. Results

4.1. Sociodemographic characteristics

The sociodemographic distribution of the sample by experimental split and methodology used is visualized in Table 1. To test for the presence of initial differences between the two experimental groups, sociodemographic characteristics of respondents which were assigned to the direct questioning-group ($N = 678$) are compared with characteristics of those which were assigned to the RRT version ($N = 1363$). As expected, random assignment yields no initial differences on the variables sex, age, education, marital status, employment status, and region. Pearson χ^2 -tests of independence on each variable show no significant difference between the two experimental groups.

During the introduction of the RRT, a subset of respondents assigned to use the RRT did not comply with the interviewer's request to get the coins and refused to use the RRT ('RRT-direct'). The reasons for rejecting the RRT for direct questioning can be summarized as follows: Some respondents insisted to state their personal opinion directly. They said that they had nothing to conceal. Some elderly respondents were unable to follow the RRT procedure due to cognitive problems. Other respondents doubted the benefit of the RRT. They were unwilling to get the coins and to follow the procedure respectively (e.g. too much effort).

χ^2 Tests of independence show that respondents in the 'RRT-direct' group ($N = 226$) significantly differ from respondents assigned to direct questioning ($N = 678$) as well as respondents in the RRT-group completing the RRT version ($N = 1137$) on the variables age, education, marital status, and employment status ($p < 0.01$). Respondents who switched from the RRT to direct questioning were more likely to be older, to have lower education, to be separated or widowed, and to be retired. Because the 'RRT-direct' group could undermine the randomization and bias the comparison of the two methods,⁷ multiple logistic regression analyses were conducted to account for these (potentially confounding) covariates.

This covariate control takes also account of the potentially biasing effect of break offs: In the direct questioning split, only few respondents broke off the interview while being introduced to the sensitive part of the survey ($N = 18$). In contrast, a relatively higher share of respondents broke off the interview in the RRT split ($N = 138$). The majority of break offs in the RRT split occurred during the introduction of the RRT ($N = 114$), the rest occurred in the sensitive questions' part ($N = 18$) resp. in subsequent parts of the interview ($N = 6$). Further analyses (not displayed) show that subjects who broke off the interview in the RRT split resemble respondents in the 'RRT-direct' group in regards to several sociodemographic characteristics, i.e. subjects who broke off the interview were more likely to be older, to have lower education, to be separated or widowed, and to be retired. The increased break off rate in the RRT split indicates that despite its carefully designed CATI implementation, the RRT might be too cumbersome for certain subpopulations, apparently those with only low tolerance for the additional cognitive effort required by the technique.

4.2. Prevalence estimates of xenophobia and anti-Semitism

The proportions of respondents agreeing with the sensitive items are displayed in Table 2. The responses of subjects who rejected the RRT ('RRT-direct') are reported separately. Overall, the prevalence of xenophobic attitudes is relatively high ranging between 27.0% and 50.7%. By contrast, lower prevalence estimates of anti-Semitism can be observed ranging between 12.1% and 22.5%. As expected, the estimates of the proportions of socially undesirable 'yes' – answers are higher for respondents who accepted to use the RRT compared to those who were originally selected for direct questioning.

More specifically, three out of five differences (xenophobia 2: 34.5%–27.0% = 7.5%; anti-Semitism 1: 17.4%–12.1% = 5.3%; anti-Semitism 2: 22.5%–15.7% = 6.8%) indicate a significant increase of socially undesirable answers in the RRT using group, the 90%-confidence intervals do not overlap between the groups. With regard to the other two differences (xenophobia 1:

⁶ Statistical significance (p -values) was determined on the basis of the Wald test, which is a z -test that equals to the estimate of the regression parameter b_j divided by its standard error.

⁷ Table 1 also displays separate tests of the difference between the direct questioning group and the RRT (interview via RRT) group. The results show that respondents assigned to direct questioning ($N = 678$) significantly differ from respondents in the RRT-group completing the RRT version ($N = 1137$) on the variables age ($p < 0.01$) and employment status ($p < 0.1$). Thus, there is some evidence that simply omitting the 'RRT-direct' group would undermine the randomization.

Table 1

Sociodemographic characteristics of respondents (relative frequencies) by experimental group and questioning technique used.

Variable	Direct questioning	RRT (total)	RRT (interview via RRT)	RRT-direct (interview via direct questioning)	χ^2
Sex					1.2 ^a 0.6 ^b 3.0 ^c
Male	50.4	47.8	48.6	43.8	
Female	49.6	52.2	51.4	56.2	
Total (N)	100.0 (678)	100.0 (1363)	100.0 (1137)	100.0 (226)	
Age group (years)					5.7 16.4 ^{***} 76.9 ^{***}
<29	15.0	16.8	19.3	4.4	
29–38	14.9	13.1	13.3	12.4	
39–48	19.8	22.7	23.8	16.8	
49–58	17.4	16.3	16.7	14.6	
59–68	14.9	15.5	14.1	22.1	
>68	18.0	15.6	12.8	29.7	
Total (N)	100.0 (678)	100.0 (1363)	100.0 (1137)	100.0 (226)	
Education (years)					1.8 4.4 45.4 ^{***}
<9	3.9	4.4	3.8	7.5	
9	21.0	20.8	18.2	33.8	
10	32.5	32.5	32.8	30.7	
11–12	5.2	5.4	5.6	4.4	
13	10.9	12.3	13.6	5.8	
>13	26.5	24.6	26.0	17.8	
Total (N)	100.0 (676)	100.0 (1362)	100.0 (1137)	100.0 (225)	
Marital status					1.3 3.7 39.4 ^{***}
Married	48.8	46.7	46.7	46.4	
Separated	2.8	3.2	2.8	4.9	
Single	29.3	30.3	32.8	17.7	
Divorced	8.0	8.9	8.5	11.1	
Widowed	11.1	10.9	9.2	19.9	
Total (N)	100.0 (678)	100.0 (1361)	100.0 (1135)	100.0 (226)	
Employment status					1.0 9.0 [*] 66.4 ^{***}
Employed	52.9	53.9	56.8	39.4	
Student	4.9	5.6	6.6	0.4	
Retiree	30.1	28.3	24.2	49.1	
Unemployed	3.8	3.8	4.0	2.7	
Other	8.3	8.4	8.4	8.4	
Total (N)	100.0 (677)	100.0 (1362)	100.0 (1136)	100.0 (226)	
City size (population)					6.6 6.1 10.5
<20,000	12.3	13.9	13.1	17.7	
20–50,000	12.2	11.9	12.0	11.5	
50–100,000	13.1	9.6	9.5	10.2	
100–500,000	32.3	33.0	33.2	31.9	
>500,000	30.1	31.6	32.2	28.7	
Total (N)	100.0 (678)	100.0 (1363)	100.0 (1137)	100.0 (226)	
Region					1.5 2.1 2.6
Western Germany	79.9	82.2	82.7	79.7	
Eastern Germany	20.1	17.8	17.3	20.3	
Total (N)	100.0 (678)	100.0 (1363)	100.0 (1137)	100.0 (226)	

* $p < 0.10$.** $p < 0.05$.*** $p < 0.01$.^a Test of independence: direct questioning versus RRT (total).^b Test of independence: direct questioning versus RRT (interview via RRT).^c Test of independence: direct questioning versus RRT (interview via RRT) versus RRT-direct (interview via direct questioning).

Table 2
Prevalence estimates of xenophobia and anti-Semitism.

Topic	Group	Prevalence estimate (%)	Standard error (%)	90%-Confidence interval	95%-Confidence interval	N
Xenophobia 1	Direct	38.8	1.9	35.7–41.9	35.1–42.5	662
	RRT	41.3	2.0	38.1–44.6	37.5–45.2	1129
	RRT-direct	50.7	3.4	45.1–56.3	44.0–57.4	215
Xenophobia 2	Direct	27.0	1.7	24.2–29.9	23.7–30.4	662
	RRT	34.5	1.9	31.3–37.7	30.7–38.3	1131
	RRT-direct	35.9	3.2	30.6–41.3	29.5–42.3	220
Xenophobia 3	Direct	31.6	1.8	28.6–34.6	28.0–35.2	655
	RRT	30.1	1.9	26.9–33.2	26.3–33.8	1130
	RRT-direct	43.0	3.4	37.4–48.6	36.3–49.7	214
Anti-Semitism 1	Direct	12.1	1.3	10.0–14.2	9.6–14.7	636
	RRT	17.4	1.7	14.5–20.2	14.0–20.8	1124
	RRT-direct	21.0	2.9	16.2–25.8	15.3–26.7	200
Anti-Semitism 2	Direct	15.7	1.4	13.3–18.1	12.9–18.5	638
	RRT	22.5	1.8	19.5–25.5	18.9–26.0	1124
	RRT-direct	21.6	2.9	16.8–26.4	15.8–27.4	199

Note: Xenophobia 1: "There are too many foreigners in Germany." Xenophobia 2: "The German Federal Republic is dangerously swamped by foreign influences because of all the foreigners living in our country." Xenophobia 3: "Foreigners only come to Germany to take advantage of our welfare state." Anti-Semitism 1: "Even today the influence of the Jews is too high." Anti-Semitism 2: "The persecution of the Jews is to a certain extent their own fault due to their behavior." All items are dichotomous with possible answer categories 1 = 'yes' or 0 = 'no'.

Direct: direct questioning; RRT: interview via RRT; RRT-direct: switch from RRT to direct questioning.

41.3%–38.8% = 2.5%; xenophobia 3: 30.1%–31.6% = –1.5%), the 90%-confidence intervals show a considerable overlap between the groups thus indicating a non-significant result.

As the data in Table 2 indicate, the proportions of respondents agreeing with the unsocial statements are highest in the RRT-direct group: Compared to respondents originally assigned to direct questioning, all five differences ranging between 5.9% (anti-Semitism 2) and 11.9% (xenophobia 1) indicate higher prevalence of xenophobic and anti-Semitic attitudes in the RRT-direct group. Compared to respondents which were interviewed via the RRT, four out of five differences ranging between –0.9% (anti-Semitism 2) and 12.9% (xenophobia 3) indicate higher point estimates in the RRT-direct group. These results suggest that the RRT-direct group systematically differs from the other two groups. One possible explanation of these results could be that genuine xenophobes and anti-Semites with a strong internalization of these attitudes have no subjective benefit from using the RRT and from hiding their opinions of which they are firmly convinced. Rather they may feel satisfied while expressing publicly their private and personal feelings about foreigners and Jews. It seems apparent that the RRT-direct group with its clear overrepresentation of xenophobes and anti-Semites respectively results in an underestimation of the effect of the RRT in reducing social desirability bias. If an equivalent subgroup dropped out in the direct questioning split, the differences between the RRT group and the direct group would increase. Despite these distortions, the empirical results demonstrate that the effectiveness of the RRT is still large enough and yields more valid estimates of socially sensitive attitudes compared to direct questioning.

Regarding the degree of sensitivity, different measures were used to assess the five sensitive items (see Table 2): First, the degree of sensitivity was assessed post hoc via item nonresponse. It is assumed that the item nonresponse rate increases with item sensitivity. Higher rates of missing data can be observed for anti-Semitism (range: 5.9–6.2% in the direct questioning version) indicating a lower sensitivity for xenophobia (range: 2.4–3.4% in the direct questioning version).⁸

Second, respondents were asked to judge the feelings of hypothetical persons publicly disclosing their unsocial opinions: "In your opinion, how unpleasant would it be for most people to agree in public to the statements about foreigners [about Jews], which were read out to you earlier? [Answer categories: not unpleasant at all; somewhat unpleasant; very unpleasant]" With regard to xenophobia, 11% of the respondents reported that for most people it would be 'not unpleasant at all', 54% stated that it would be 'somewhat unpleasant', and 35% judged the feeling of most people to be 'very unpleasant' (N = 1999). Concerning anti-Semitism, 12% thought that for most people it would be 'not unpleasant at all', 36% supposed it would be 'somewhat unpleasant', and 52% believed that answering the questions would make most people feel 'very unpleasant' (N = 1978). As with item nonresponse, this measure also indicates higher sensitivity for anti-Semitism.⁹

⁸ A similar pattern of item nonresponse can be observed in the two RRT groups: Higher rates of missing data can be observed for anti-Semitism (range: 1.1–1.1% in the RRT and 11.5–12.0% in the RRT-direct group) indicating a lower sensitivity for xenophobia (range: 0.5–0.7% in the RRT and 2.7–5.3% in the RRT-direct group). It is noticeable that the RRT-direct group has the highest level of item nonresponse. Besides potentially different sensitivity perceptions, different shares of 'non-attitudes', i.e. 'no opinion' respondents, could be responsible for the different levels of item nonresponse observed across the two RRT groups. The RRT-direct group with its clear overrepresentation of respondents with lower cognitive skills (as measured by the formal education level and age) seems to have a higher share of 'non-attitudes' (which were also coded as item nonresponse in the present study) than the RRT group; see also Krosnick et al. (2002) for a discussion of the association between inclination to 'no opinion' and lower levels of cognitive skills.

⁹ In addition, this sensitivity measure was compared across the three subgroups (direct questioning vs. RRT vs. RRT-direct). χ^2 tests of independence confirm the null hypothesis showing no evidence for systematic differences in the distributions of the sensitivity perceptions between the three subgroups ($\chi^2 = 6.91$, $p = 0.14$ for xenophobia; $\chi^2 = 5.79$, $p = 0.22$ for anti-Semitism).

Third, respondents were asked to judge the reactions of a hypothetical crowd observing a person openly expressing xenophobic or anti-Semitic attitudes: “Imagine the following situation: On a city market square filled with people somebody would say out loud in public the statements about foreigners [about Jews], which were read out to you earlier. How negatively would the surrounding people judge that person? [Answer categories: not negatively at all; somewhat negatively; very negatively]” With respect to xenophobia, 10% of the respondents believed that the judgements would be ‘not negatively at all’, 49% guessed that they would be ‘somewhat negatively’, and 41% thought they would be ‘very negatively’ ($N = 1980$). Regarding anti-Semitism, 10% imagined the judgements to be ‘not negatively at all’, 34% believed they would be ‘somewhat negatively’, and 56% expected the judgements to be ‘very negatively’ ($N = 1979$).¹⁰ These results are consistent with prior research (Bergmann, 1997). Questions asking about anti-Semitic attitudes are widely seen as very sensitive in Germany. Comparisons of the ratios of the prevalence estimates in the RRT group to the prevalence estimates in the direct group provide evidence of the hypothesis that with increasing topic sensitivity, the effectiveness of the RRT also increases (see Table 2): Compared to xenophobia which turned out to be moderately sensitive (xenophobia 1: $41.3/38.8 = 1.06$; xenophobia 2: $34.5/27.0 = 1.28$; xenophobia 3: $30.1/31.6 = 0.95$), the observed ratios for the highly sensitive topic anti-Semitism are considerably greater (anti-Semitism 1: $17.4/12.1 = 1.44$; anti-Semitism 2: $22.5/15.7 = 1.43$).

4.3. Associations with explanatory variables

In this section, results of the analyses of the relationships between explanatory variables and prejudices toward foreigners and Jews are reported. To start with, a brief overview of the variables which were included in the multiple logistic regression analysis and the hypothesized relationships between these variables will be provided (for a comprehensive treatise on anti-Semitism and xenophobia in Germany after unification, see Kurthen et al., 1997): The response variables are the five sensitive items *xenophobia 1–3* and *anti-Semitism 1–2* which are described in Table 2. A socially undesirable answer was coded 1 = yes, a socially desirable answer was coded 0 = no. Separate models were estimated for each of the five sensitive items. The experimental block encompasses three dummy variables which indicate the experimental group and the question technique used: *Direct*: 1 = direct questioning and 0 = otherwise; *RRT*: 1 = interview via RRT and 0 = otherwise; *RRT-direct*: 1 = assigned to RRT but interviewed via direct questioning and 0 = otherwise.

The first block of covariates includes different sociodemographic characteristics: *gender* (0 = male, 1 = female), *age* (in years), *education* (measures increasing levels of education: 1 = low, 6 = high), *region* (0 = Western Germany; 1 = Eastern Germany), and *city size* (increasing: 1 = small, 5 = big).¹¹ Former studies (Bergmann and Erb, 2000, pp. 409, 419; Decker and Brähler, 2006, pp. 48–50) showed that men have a slightly increased propensity for being prejudiced toward foreigners and Jews respectively. However, the relationship between gender and prejudices turned out to be statistically insignificant. These studies also showed that the older generation in Germany (60 years and more) has a significantly higher probability to have anti-Semitic resentments. Nevertheless, “as the number of those old enough to have been actively involved in the Nazi period continues to decrease, the attitudes of subsequent generations are also becoming more and more similar, so that the dominant influence of age is disappearing” (Bergmann, 1997, p. 23). A robust result that has been observed across various empirical studies is a negative relationship between the subjects’ level of education and his or her inclination to be prejudiced. Subjects with a high educational level are clearly less prejudiced compared to less educated subjects (Bergmann, 1997; Beyer and Krumpal, 2010; Kurthen, 1997; Mielke and Mummendey, 1995; Ostapczuk et al., 2009; Wagner and Zick, 1995): Institutions of formal education in Germany actively promote a culture of Holocaust remembrance. Additional education provides a person with more information about the history of anti-Semitism. More knowledge of the historical facts is expected to increase empathy with the fate of persecuted Jewish people. Furthermore, education can be assumed to improve people’s cognitive capabilities to recognize the dangers associated with prejudices against minority groups and to reject them. Various factors have been found to moderate the complex relationship between education and xenophobia (Ostapczuk et al., 2009, p. 928): “A highly educated person who is—on average—more intelligent, more open to experience, of higher self-esteem, less economically deprived, sensing a smaller incongruity [sic] between his or her own culture and the one of the outgroup, and who has also had more (positive) contact with members of the ethnic

¹⁰ Again, this sensitivity measure was compared across the three subgroups (direct questioning vs. RRT vs. RRT-direct). χ^2 tests of independence confirm the null hypothesis showing no evidence for systematic differences in the distributions of the sensitivity perceptions between the three subgroups regarding xenophobia ($\chi^2 = 4.83$, $p = 0.31$). However, regarding anti-Semitism, systematic differences in the sensitivity perceptions can be observed ($\chi^2 = 20.45$, $p = 0.00$): 11% of the respondents in the direct questioning group (8% in the RRT group; 16% in the RRT-direct group) believed that the judgements about Jews would be ‘not negatively at all’, 34% in the direct questioning group (33% in the RRT group; 40% in the RRT-direct group) believed they would be ‘somewhat negatively’, and 55% in the direct questioning group (59% in the RRT group; 44% in the RRT-direct group) thought they would be ‘very negatively’. This comparison provides some evidence that respondents in the RRT-direct group view the questions asking about anti-Semitic attitudes less sensitive than respondents in the other two subgroups. One possible explanation of these results could be that genuine anti-Semites which are clearly overrepresented in the RRT-direct group (see Table 2) may feel an intrinsic benefit from openly communicating their prejudices of which they are convinced and which they regard as normal. These different perceptions of normality may in turn influence the perceptions of sensitivity as well as the subjective benefit from using the RRT.

¹¹ The empirical distribution of the sociodemographic variables is displayed in Table 1. In the multiple logistic regression models, the variable *age* is not a group variable but is introduced in its original form (years: mean = 48.7, SD = 17.6). The categories of the variable *education* are as follows: 1 = left school without completing secondary general school (<9 years), 2 = completed secondary general school (in Germany: ‘Haupt-/Volksschulabschluss’, 9 years), 3 = completed intermediate school (‘Realschulabschluss/Mittlere Reife’, 10 years), 4 = obtained the qualification for admittance for studies at German universities of applied sciences (‘Fachhochschulreife’, 11–12 years), 5 = obtained the subject-linked or general higher education entrance qualification (‘fachgebundene Hochschulreife/allgemeine Hochschulreife/Abitur’, 13 years), 6 = obtained a degree from a university of applied sciences or a university degree (‘Fachhochschul-/Universitätsabschluss’, >13 years). The variable *city size*, measuring whether the respondent lives in a rural versus an urban area, was generated as follows: 1 = (<20,000), 2 = (20–50,000), 3 = (50–100,000), 4 = (100–500,000), 5 = (>500,000).

Table 3

Logistic regression models showing odds ratios (standard errors in parentheses) for experimental group and question technique used predicting xenophobia and anti-Semitism (baseline models).

	Xenophobia 1	Xenophobia 2	Xenophobia 3	Anti-Semitism 1	Anti-Semitism 2
RRT (<i>Ref. Direct</i>)	1.110 (0.126)	1.421*** (0.174)	0.930 (0.115)	1.527** (0.262)	1.560*** (0.235)
RRT (<i>Ref. RRT-direct</i>)	0.685** (0.109)	0.940 (0.155)	0.570*** (0.094)	0.791 (0.167)	1.052 (0.212)
RRT-direct (<i>Ref. Direct</i>)	1.620*** (0.256)	1.512*** (0.250)	1.632*** (0.264)	1.930*** (0.409)	1.483* (0.302)
N	2006	2013	1999	1960	1961
Log likelihood	−1364.2	−1283.1	−1286.8	−976.1	−1061.3
Model χ^2	9.4***	10.7***	12.1***	11.4***	9.5***
McFadden R^2	0.003	0.004	0.005	0.005	0.004

Note: Response variables are dichotomous (1 = yes, 0 = no); xenophobia 1 (“There are too many foreigners in Germany.”); xenophobia 2 (“The German Federal Republic is dangerously swamped by foreign influences because of all the foreigners living in our country.”); xenophobia 3 (“Foreigners only come to Germany to take advantage of our welfare state.”); anti-Semitism 1 (“Even today the influence of the Jews is too high.”); anti-Semitism 2: (“The persecution of the Jews is to a certain extent their own fault due to their behavior.”).

The dummy variables indicating the experimental group and the question technique used are: Direct (1 = direct questioning, 0 = otherwise); RRT (1 = interview via RRT, 0 = otherwise); RRT-direct (1 = switch from RRT to direct questioning, 0 = otherwise). Refer to footnote 12 for further details.

Statistical significance:

* $p < 0.1$.

** $p < 0.05$.

*** $p < 0.01$ (two-sided test).

minority, is not unlikely to really be, and not just pretending to be, more xenophile.” With regard to the variable city size which measures the degree of urbanity, Bergmann and Erb (2000, p. 420) observed a mixed pattern of association with anti-Semitism yielding no clear predictions. Finally, empirical evidence indicates differences between attitudes toward foreigners and those toward Jews across regions. In East Germany the prevalence of antiforeigner attitudes is higher than in West Germany, but in West Germany relatively higher prevalence rates of anti-Semitism can be observed (Bergmann, 1997; Bergmann and Erb, 2000, pp. 420–423; Decker and Brähler, 2006, p. 43). The higher prevalence rates of xenophobia in Eastern Germany can be explained mainly by social fears and economic competition on the labor market. In the newly formed German states (formerly GDR), job offers are noticeably scarcer and the rate of unemployment is considerably higher compared to the old West German states. In contrast, the reasons for the empirically observable lower prevalence of anti-Semitic resentments in Eastern Germany can be found in the exposition of the population to two different political systems before German re-unification (Kurthen, 1997, p. 47): “Until 1989 the official ideology of the socialist rulers proclaimed East Germany an inheritor of German antifascism which represented a clear break with the past of National Socialism (...). The government provided the population with a consistent interpretation of the past, but prevented an unrestricted debate about German’s guilt and responsibility (...). Stringent antifascist education seems to have had rather a suppressing effect on anti-Jewish resentments in East Germany (...).”

The second block of covariates is based on different follow-up questions measuring psychological and situational variables which are assumed to influence response behavior in sensitive surveys: Based on the indicator from the World Value Survey, the following question was used to measure *generalized trust* (0 = one cannot be too careful, 1 = most people can be trusted): ‘In general, do you think that most people can be trusted, or cannot you be too careful in dealing with people?’ The concept of generalized trust is known to be an important ingredient of social capital which focuses on those democratic values that foster citizens to reciprocate, to tolerate otherness, and to cooperate with each other (Putnam, 1993). Generalized attitudes of trust are characterized by a potential readiness to cooperate with others not only within specific personal contexts in which the exchange partner is already known, but to extend trust beyond these contexts and also being prepared to cooperate with people who are strangers. Previous research has shown that trusting individuals are more inclined to engage in mutually beneficial exchanges with others. In addition, they have a higher social activity and tolerance, they are engaged in civic endeavors more frequently, and they are more likely to support minority rights (Stolle, 2002, pp. 397–399). Consequently, trustful people who are also expected to be more tolerant and open-minded toward minorities can be hypothesized to be less inclined to xenophobia and anti-Semitism than distrustful individuals. On the other hand, distrustful xenophobes and anti-Semites respectively could be more inclined to conceal their true opinions in the interview situation and to give socially desirable answers due to increased fears of negative consequences should the sensitive information become public. This issue was considered in a question measuring the degree of *trust in data protection* (measures decreasing levels of trust: 1 = very much, 2 = much, 3 = not so much, 4 = not much at all): ‘How much do you trust in your data being protected in this interview?’ Furthermore, respondents were asked to rate their own feelings of unpleasantness in conjunction with answering to the sensitive items (*unpleasantness*: 1 = not unpleasant at all, 2 = somewhat unpleasant, 3 = very unpleasant): ‘How unpleasant were the questions read out to you earlier to you personally?’ It is assumed that the questions are more sensitive for those for whom the potential answer is ‘yes’ (Fowler, 1995, p. 29), i.e. for potential xenophobes and anti-Semites respectively. In contrast, questions asking about attitudes toward foreigners and Jews are expected to be less sensitive for unprejudiced respondents, since they do not violate a public norm and thus have no reason to fear disapproval and social ostracism. Finally, empirical research shows that the presence of third persons (e.g. parents, spouse, or children) in the interview situation, besides the interviewer and the respondent, may affect the accuracy of survey responses (Aquilino et al., 2000; Reuband, 1987, 1992). Thus, information about the *presence of third persons* (0 = yes, 1 = no) was gathered via the following question: ‘Were there other persons but you in the room during the interview?’

Table 4

Logistic regression models showing odds ratios (standard errors in parentheses) for experimental group, question technique used and covariates predicting xenophobia and anti-Semitism (full models).

	Xenophobia 1	Xenophobia 2	Xenophobia 3	Anti-Semitism 1	Anti-Semitism 2
RRT (<i>Ref. Direct</i>)	1.144 (0.151)	1.494*** (0.204)	1.017 (0.141)	1.737*** (0.315)	1.714*** (0.273)
RRT (<i>Ref. RRT-direct</i>)	0.928 (0.180)	1.474** (0.288)	0.830 (0.163)	1.286 (0.305)	1.519* (0.341)
RRT-direct (<i>Ref. Direct</i>)	1.232 (0.232)	1.014 (0.194)	1.224 (0.233)	1.350 (0.319)	1.129 (0.254)
Gender	1.489*** (0.190)	1.399*** (0.185)	0.923 (0.124)	0.633*** (0.113)	0.722** (0.112)
Age	0.997 (0.007)	1.000 (0.007)	0.999 (0.007)	1.017* (0.009)	1.014* (0.008)
Education	0.709*** (0.030)	0.759*** (0.034)	0.790*** (0.035)	0.778*** (0.046)	0.868*** (0.044)
Region	1.594*** (0.248)	1.212 (0.189)	3.029*** (0.468)	0.692* (0.150)	0.854 (0.164)
City size	0.895** (0.040)	0.973 (0.044)	0.928 (0.043)	0.968 (0.057)	0.962 (0.051)
Generalized trust	0.436*** (0.060)	0.399*** (0.061)	0.439*** (0.066)	0.642** (0.128)	0.476*** (0.085)
Trust in data protection	1.148* (0.094)	1.169* (0.099)	1.210** (0.103)	1.316** (0.149)	1.179* (0.116)
Unpleasantness	1.812*** (0.250)	1.828*** (0.241)	1.311** (0.175)	1.619*** (0.258)	1.210 (0.187)
Presence of third persons	1.126 (0.171)	0.875 (0.132)	0.982 (0.150)	1.011 (0.201)	1.092 (0.200)
Willingness	0.812 (0.136)	0.901 (0.149)	1.069 (0.172)	0.962 (0.193)	0.708 (0.151)
Indignation	0.442*** (0.094)	0.658** (0.135)	0.667** (0.134)	0.945 (0.180)	0.679* (0.145)
N	1937	1944	1929	1895	1897
Log likelihood	−1174.5	−1136.6	−1130.9	−892.2	−978.7
Model χ^2	295.3***	223.8***	239.9***	120.3***	103.2***
McFadden R^2	0.112	0.090	0.096	0.063	0.050

Note: Response variables are dichotomous (1 = yes, 0 = no): xenophobia 1 (“There are too many foreigners in Germany.”); xenophobia 2 (“The German Federal Republic is dangerously swamped by foreign influences because of all the foreigners living in our country.”); xenophobia 3 (“Foreigners only come to Germany to take advantage of our welfare state.”); anti-Semitism 1 (“Even today the influence of the Jews is too high.”); anti-Semitism 2: (“The persecution of the Jews is to a certain extent their own fault due to their behavior.”).

The dummy variables indicating the experimental group and the question technique used are: Direct (1 = direct questioning, 0 = otherwise); RRT (1 = interview via RRT, 0 = otherwise); RRT-direct (1 = switch from RRT to direct questioning, 0 = otherwise). Refer to footnote 12 for further details.

The covariates are: gender (0 = male, 1 = female); age (in years); education (1 = low, 6 = high), region (0 = Western Germany; 1 = Eastern Germany); city size (increasing: 1 = small, 5 = big); generalized trust (0 = one cannot be too careful, 1 = most people can be trusted); trust in data protection (1 = very much, 4 = not much at all); respondent's feelings of unpleasantness (1 = not unpleasant at all, 3 = very unpleasant); presence of third persons (0 = yes, 1 = no). The interviewers' observations are: respondent's willingness to answer the questions (1 = very willing, 3 = hardly willing); respondent's indignation at the questions (1 = not indignant, 3 = highly indignant). The models additionally control for the variables marital status and employment status (results for these two variables are not displayed).

Statistical significance:

* $p < 0.1$.

** $p < 0.05$.

*** $p < 0.01$ (two-sided test).

The third block includes further control variables which are based on the interviewer's observations of the respondent's behavior: The respondent's *willingness* (1 = very willing, 2 = somewhat willing, 3 = hardly willing) to cooperate was measured via the following indicator: ‘How willing was the interviewee to answer the questions?’ Finally, the interviewer rated the degree of respondent's *indignation* (1 = not indignant, 2 = somewhat indignant, 3 = highly indignant) at the questions. Table 3 shows the baseline-models including the experimental group and the question technique used only.¹²

Table 4 shows the full models including the three blocks of covariates described above. The full models additionally control for the variables *marital status* and *employment status* (results regarding these two variables are not displayed).

Empirical results show that respondents interviewed via the RRT reveal more xenophobic and anti-Semitic opinions respectively compared to respondents which were assigned to direct questioning: In the full models, three out of five odds ratios (OR) indicate that the odds of reporting a socially undesirable attitude among those who were interviewed via the RRT were significantly higher compared to the odds among those who were assigned to direct questioning: OR = 1.494, $p < 0.01$ (xenophobia 2); OR = 1.737, $p < 0.01$ (anti-Semitism 1); OR = 1.714, $p < 0.01$ (anti-Semitism 2). In addition, controlling for covariates results in a slight increase of the odds ratios between the two groups ‘RRT’ and ‘direct’ (OR in the baseline model in Table 3 versus OR in the full model in Table 4): 1.110 versus 1.144 (xenophobia 1); 1.421 versus 1.494 (xenophobia 2); 0.930 versus 1.017 (xenophobia 3); 1.527 versus 1.737 (anti-Semitism 1); 1.560 versus 1.714 (anti-Semitism 2). Furthermore, controlling covariates which were found to be significantly related to the rejection of the RRT clearly diminishes the odds ratios between the two groups ‘RRT-direct’ and ‘direct’ which are no longer statistically significant in the full model (OR in the baseline model in Table 3 versus OR in the full model in Table 4): 1.620 versus 1.232 (xenophobia 1); 1.512 versus 1.014 (xenophobia 2); 1.632 versus 1.224 (xenophobia 3); 1.930 versus 1.350 (anti-Semitism 1); 1.483 versus 1.129 (anti-Semitism 2). Finally, covariate control results in significantly higher odds of agreeing with a sensitive statement for those who were interviewed via the RRT relative to those who rejected the RRT for direct questioning. In the full models (Table 4), two out of five comparisons indicate significantly higher odds for the RRT-group compared to odds of the RRT-direct group:

¹² Regarding the three dummy variables for the interview mode (direct, RRT, RRT-direct), two separate logistic regression models were estimated. In the first model, the baseline category was direct questioning (*Ref. Direct*). In the second model which is identical to the first model, the baseline category was RRT-direct (*Ref. RRT-direct*). In Tables 3 and 4, results for the baseline category ‘direct questioning’ on the basis of the first model are displayed (*Ref. Direct*). In order to obtain a significance test for the third contrast RRT vs. RRT-direct, results for the baseline category ‘RRT-direct’ on the basis of the second model are displayed in an additional row (*Ref. RRT-direct*). The odds-ratios of the covariates (see Table 4) were estimated on the basis of the first model.

Table 5

Logistic regression models showing odds ratios (standard errors in parentheses) for experimental group treating the RRT-followers and RRT-deniers as one joint group (baseline models and full models).

	Xenophobia 1	Xenophobia 2	Xenophobia 3	Anti-Semitism 1	Anti-Semitism 2
<i>Baseline models</i>					
RRT (total) (Ref. Direct)	1.223* (0.129)	1.445*** (0.164)	1.088 (0.121)	1.640*** (0.255)	1.539*** (0.217)
N	2006	2013	1999	1960	1961
<i>Full models</i>					
RRT (total) (Ref. Direct)	1.166 (0.142)	1.346** (0.171)	1.069 (0.136)	1.606*** (0.270)	1.531*** (0.230)
N	1937	1944	1929	1895	1897

Note: Response variables are dichotomous (1 = yes, 0 = no): xenophobia 1 ("There are too many foreigners in Germany."); xenophobia 2 ("The German Federal Republic is dangerously swamped by foreign influences because of all the foreigners living in our country."); xenophobia 3 ("Foreigners only come to Germany to take advantage of our welfare state."); anti-Semitism 1 ("Even today the influence of the Jews is too high."); anti-Semitism 2: ("The persecution of the Jews is to a certain extent their own fault due to their behavior.").

The dummy variable indicating the experimental group treats the RRT-followers and RRT-deniers as one joint group: RRT (total): 1 = interview via RRT or RRT-direct, 0 = direct questioning.

All other aspects of the model specification are identical with the ones specified in the Tables 3 and 4. Results for the covariates and indicators of model fit are not displayed.

Statistical significance:

* $p < 0.1$.

** $p < 0.05$.

*** $p < 0.01$ (two-sided test).

xenophobia 2: OR = 1.474, $p < 0.05$ (baseline comparison in Table 3: OR = 0.940, not significant p -value); anti-Semitism 2: OR = 1.519, $p < 0.1$ (baseline comparison in Table 3: OR = 1.052, not significant p -value). Overall, five out of 10 odds ratios in the full model comparing the 'RRT'-group with either the 'direct'-group or the 'RRT-direct' group indicate a significant increase of socially undesirable answers in the RRT-group. The other five odds ratios are statistically insignificant regarding conventional levels. The results of the multiple logistic regression analyses demonstrate the effectiveness of the RRT in eliciting more socially undesirable opinions compared to direct questioning. Table 5 displays additional results treating the RRT-followers and the RRT-deniers as one joint group.¹³

With regard to the relationships between the sociodemographic characteristics and the odds of having prejudiced opinions toward foreigners and Jews respectively the following results can be observed (see Table 4): Female respondents have significantly higher odds of reporting xenophobic attitudes than male respondents (xenophobia 1: OR = 1.489, $p < 0.01$; xenophobia 2: OR = 1.399, $p < 0.05$). In contrast, women have significantly lower odds of reporting anti-Semitic attitudes compared to men (anti-Semitism 1: OR = 0.633, $p < 0.01$; anti-Semitism 2: OR = 0.722, $p < 0.05$). The odds of being prejudiced towards Jews significantly increase with age (anti-Semitism 1: OR = 1.017, $p < 0.1$; anti-Semitism 2: OR = 1.014, $p < 0.1$).¹⁴ This result is in accordance with the findings of Decker and Brähler (2006, pp. 48–50) and Bergmann and Erb (2000, pp. 409, 419), indicating that members of the older generation in Germany have a higher chance of being prejudiced towards Jews. Furthermore, the robust association between education and prejudices which was observed in former empirical studies can be replicated (Bergmann, 1997; Beyer and Krumpal, 2010; Kurthen, 1997; Mielke and Mummendey, 1995; Ostapczuk et al., 2009; Wagner and Zick, 1995). With increasing levels of formal education, the odds of being prejudiced towards foreigners and Jews significantly decrease (range of OR: 0.709–0.868, $p < 0.01$). Regarding the variable region, the empirical findings confirm the results of former empirical studies (Bergmann, 1997; Bergmann and Erb, 2000, pp. 420–423; Decker and Brähler, 2006, p. 43) reporting higher prevalence rates of antiforeigner attitudes in East Germany, but higher prevalence rates of anti-Semitism in West Germany. East Germans have significantly higher odds of being prejudiced towards foreigners than West Germans (xenophobia 1: OR = 1.594, $p < 0.01$; xenophobia 3: OR = 3.029, $p < 0.01$). On the other hand, East Germans have significantly lower odds of being anti-Semitic compared to West Germans (anti-Semitism 1: OR = 0.692, $p < 0.1$). Finally, there is some evidence that respondents from urban areas are less xenophobe compared to respondents from rural areas. With increasing city size, the odds of being prejudiced toward foreigners significantly decrease (xenophobia 1: OR = 0.895, $p < 0.05$).

¹³ An alternative data analysis strategy is to treat the RRT-followers and the RRT-deniers as one joint group and to compare this group ('RRT total') with the direct group, as this was the way the sampling design was set up initially. These additional analyses are displayed in Table 5. The conclusions are essentially the same as with the models treating the RRT-followers and the RRT-deniers as separate groups: In the baseline models, four out of five odds ratios (OR) indicate that the odds of reporting a socially undesirable attitude among those who were initially assigned to the RRT were significantly higher compared to the odds among those who were assigned to direct questioning: OR = 1.223, $p < 0.1$ (xenophobia 1); OR = 1.445, $p < 0.01$ (xenophobia 2); OR = 1.640, $p < 0.01$ (anti-Semitism 1); OR = 1.539, $p < 0.01$ (anti-Semitism 2). In the full models, three out of five odds ratios (OR) indicate that the odds of reporting a socially undesirable attitude increase significantly among those who were initially assigned to the RRT: OR = 1.346, $p < 0.05$ (xenophobia 2); OR = 1.606, $p < 0.01$ (anti-Semitism 1); OR = 1.531, $p < 0.01$ (anti-Semitism 2). In the main analyses, it was decided to treat the RRT-followers and the RRT-deniers as separate groups. From a sociological perspective, it is more interesting to compare these two groups to get better insights in the data generation process, i.e. in the reasons why some respondents do not follow the design (see also Weissman et al., 1986).

¹⁴ The results (not displayed) indicate no specific relationship between marital status and the likelihood of being prejudiced. With regard to employment status, students have significantly lower odds of being xenophobic, relative to employed respondents (xenophobia 1: OR = 0.431, $p < 0.05$; xenophobia 2: OR = 0.511, $p < 0.1$). Furthermore, respondents who are retired show significantly higher odds of reporting antiforeigner attitudes compared to the odds of employed respondents (xenophobia 3: OR = 1.487, $p < 0.1$). The remaining groups display a heterogeneous pattern of associations regarding the likelihood of being prejudiced yielding inconclusive results.

The analyses of the relationships between psychological variables and the odds of agreeing with the sensitive statements yielded the following results: As expected, respondents with a high degree of generalized trust have significantly lower odds of being prejudiced compared to respondents with a low degree of generalized trust (range of OR: 0.399–0.642, $p < 0.05$). This result confirms the hypothesis that generalized trust which is an important dimension of social capital fosters tolerance and at the same time suppresses prejudices against societal minorities (Stolle, 2002, pp. 397–399). Surprisingly, with decreasing levels of respondents' trust in the data protection of this study the odds of being prejudiced are significantly increasing (range of OR: 1.148–1.316, $p < 0.1$). These results indicate that xenophobes and anti-Semites respectively appear to be very distrustful people. Furthermore, with the respondents' growing intensity of unpleasantness when answering to the sensitive items the odds of being prejudiced significantly increase (xenophobia 1: OR = 1.812, $p < 0.01$; xenophobia 2: OR = 1.828, $p < 0.01$; xenophobia 3: OR = 1.311, $p < 0.05$; anti-Semitism 1: OR = 1.619, $p < 0.01$). These observations are in line with the considerations of Fowler (1995, p. 29) who states that questions are more sensitive for those who have something to hide and who have to fear repercussions as a consequence of revealing self-discrediting information. Finally, the presence of third parties in the interview situation did not significantly affect the respondents' answer behavior.

With regard to the variables measuring the interviewers' impressions of the question-and-answer process the following associations can be observed: The respondent's willingness to answer to the questions is not significantly related to the odds of having prejudiced opinions toward foreigners and Jews respectively. However, with increasing respondent's indignation at the questions the odds of agreeing with the sensitive statements significantly decrease (xenophobia 1: OR = 0.442, $p < 0.01$; xenophobia 2: OR = 0.658, $p < 0.05$; xenophobia 3: OR = 0.667, $p < 0.05$; anti-Semitism 2: OR = 0.679, $p < 0.1$). The analyses show that several sociodemographic characteristics and psychological variables respectively are significantly related to the likelihood of being prejudiced.¹⁵

4.4. A closer look at the RRT-followers and the RRT-deniers

This section discusses further analyses and results regarding the differences between the RRT-followers and the RRT-deniers: It can be shown that the rejection of the RRT not only is related to sociodemographic characteristics (see Table 1) but also to psychological variables of respondents and survey related factors. Table 6 displays the conditional distributions of some psychological variables (generalized trust, trust in data protection, unpleasantness) and survey related factors (presence of third person, willingness, indignation) of the RRT-followers and the RRT-deniers.

A comparison of the distributions between the two groups (p -values on the basis of χ^2 tests of independence in brackets) indicates that RRT-deniers are less trustful ($p < 0.01$ for both generalized trust and trust in data protection), less cooperative ($p < 0.01$ for willingness), and, somewhat surprising, more indignant at the sensitive questions ($p < 0.01$ for both xenophobia and anti-Semitism). The first two results are in line with the study of Landsheer et al. (1999) providing empirical evidence that respondents with lower cognitive skills develop less trust in the RRT compared to respondents who have higher cognitive skills and who turn out to be more cooperative in using the RRT.¹⁶ The subgroup of RRT-deniers, with a clear overrepresentation of older and less educated respondents, seems to be more inclined to misunderstand the principle of the RRT and, as a consequence, to develop less trust and cooperation. In contrast, the third result, indicating higher indignation at the sensitive questions in the subgroup of RRT-deniers is somewhat counterintuitive. Because of the relatively high share of prejudiced respondents among the RRT-deniers, one would expect less sensitization and hence a lower level of indignation in this group. These findings indicate, however, that the composition of the RRT-direct group is quite heterogeneous, also encompassing respondents who react strongly defensive and emotional if confronted with xenophobic resp. anti-Semitic statements.

5. Discussion

This study investigated the viability of the RRT with respect to the measurement of socially sensitive attitudes. In this context, a national CATI-survey of the general population was conducted in Germany estimating the prevalence of xenophobia and anti-Semitism. Difficulties in the data collection process were similar to those observed in former RRT telephone studies (Weissman et al., 1986): Of 1363 respondents assigned to the RRT-group, 16.6% were definitively unwilling to use a randomizer in the question-and-answer process rejecting the RRT for direct questioning.¹⁷ Sociodemographic analyses

¹⁵ In addition, interaction effects between the three subgroups (direct, RRT, and RRT-direct) and covariates (education, generalized trust, trust in data protection, unpleasantness) that are strongly related to xenophobia and anti-Semitism were explored. Overall, no robust and consistent interaction patterns could be found across subgroups and across the sensitive items and, for the most part, the results were not statistically significant. These inconclusive findings regarding interaction patterns could be explained by the low efficiency of the RRT estimator. One general problem of sensitive-question techniques that induce anonymity by adding noise to the data (as does the RRT) is that larger samples are required to achieve a high level of statistical power. Here, the statistical power is too low to analyze interaction effects (i.e. differences in effect sizes across interview conditions) successfully by multivariate methods. A bigger sample size would be necessary to analyze interaction effects between the interview conditions and key covariates. For the sake of clarity and conciseness, results regarding interaction patterns are not displayed but can be obtained upon request.

¹⁶ In general, the RRT imposes a higher cognitive burden on the respondents than a standard direct question (Stem and Steinhurst, 1984). Therefore, additional effort (such as special interviewer training and extra pretesting and tailoring the wording of the RRT instructions) is necessary for a credible and comprehensive communication of the RRT and for keeping the group of RRT-deniers as small as possible.

¹⁷ In the study of Weissman et al. (1986), 55 (47.8%) of the 115 respondents which were randomly assigned to use the RRT were actually unwilling to use the technique and preferred to be interviewed via direct questioning.

Table 6

Psychological variables of respondents and survey related factors (relative frequencies) within the RRT split (RRT-followers versus RRT-deniers).

Variable	RRT (interview via RRT)	RRT-direct (interview via direct questioning)	χ^2
Generalized trust			9.5 ^{***a}
Most people can be trusted	37.5	26.6	
One cannot be too careful	62.5	73.4	
Total (N)	100.0 (1124)	100.0 (218)	
Trust in data protection			64.8 ^{***}
Very much	6.5	4.2	
Much	41.7	19.9	
Not so much	41.3	48.6	
Not much at all	10.5	27.3	
Total (N)	100.0 (1127)	100.0 (216)	
Unpleasantness			1.3
Not unpleasant at all	78.8	75.4	
Somewhat unpleasant	19.1	21.9	
Very unpleasant	2.1	2.7	
Total (N)	100.0 (1137)	100.0 (224)	
Presence of third persons			0.03
Yes	22.1	22.7	
No	77.9	77.3	
Total (N)	100.0 (1135)	100.0 (225)	
Willingness			116.6 ^{***}
Very willing	88.9	64.6	
Somewhat willing	10.7	27.9	
Hardly willing	0.4	7.5	
Total (N)	100.0 (1135)	100.0 (226)	
Indignation (xenophobia)			24.4 ^{***}
Not indignant	93.9	84.5	
Somewhat indignant	5.9	14.6	
Highly indignant	0.2	0.9	
Total (N)	100.0 (1135)	100.0 (226)	
Indignation (anti-Semitism)			33.1 ^{***}
Not indignant	91.1	80.1	
Somewhat indignant	7.7	13.7	
Highly indignant	1.2	6.2	
Total (N)	100.0 (1135)	100.0 (226)	

^{*} $p < 0.10$.

^{**} $p < 0.05$.

^{***} $p < 0.01$.

^a Test of independence: RRT (interview via RRT) versus RRT-direct (interview via direct questioning).

showed that respondents in the 'RRT-direct' group were – on average – older and less educated. Furthermore, covariates were identified which were related to both the rejection of the RRT and the likelihood of being prejudiced. Multiple logistic regression analyses controlling for these covariates indicate that respondents interviewed via the RRT were more likely to reveal a prejudiced attitude compared to those interviewed via direct questioning.¹⁸ The results demonstrate that the RRT is a beneficial instrument for the purposes of reducing social desirability problems in sensitive attitudinal data. In addition, the study shows that it is important to gather data on those unwilling to use the RRT and to incorporate this information in the statistical analysis improving our knowledge of the data generation process and allowing for sociologically meaningful analyses of the 'RRT-direct' group. This is, in fact, the first large-scale survey systematically collecting empirical data on those unwilling to use the

¹⁸ One reviewer of this article pointed out that answers might be subject to acquiescence, i.e. that a subset of respondents might answer affirmatively irrespective of the sensitive items' content (see Schaeffer and Presser (2003) for an overview). Acquiescence could potentially impact the absolute levels of affirmative answers. However, there is no obvious reason to believe that the extent of acquiescence (if present) would differ across the two questioning modes and therefore acquiescence is not expected to bias the assessment of differences in reported rates of antiforeigner or anti-Semitic prejudices between the experimental groups. Furthermore, even if there was acquiescence, it would probably be less severe with the RRT than with direct questioning: Cumulative evidence indicates that RRT formats regularly face the problem of self-protective 'no' answers, i.e. a subset of respondents (irrespective of whether they are prejudiced or not) are tempted to deviate from the RRT instructions and provide evasive 'no' answers even if the outcome of the randomizer directs them to say 'yes' (Böckenholt et al., 2009; Ostapczuk et al., 2009; van den Hout et al., 2010; Holbrook and Krosnick, 2010; Coutts and Jann, 2011). Thus, the respondents' propensity to provide self-protective 'no' answers with the RRT would work against possible acquiescence tendencies in this interview mode. In other words, acquiescence cannot explain the higher rates of self-reported socially undesirable prejudices yielded with the RRT and, even more importantly, it cannot explain that with increasing topic sensitivity, the benefits of the RRT also increase (with acquiescence, the effect of the RRT should have remained constant for different levels of topic sensitivity).

RRT and explicitly integrating this 'RRT-direct' group in a multiple logistic regression model accounting for possible self-selection problems when comparing the RRT with the direct questioning method.

In summary, this study shows that measures of xenophobia and anti-Semitism in general population surveys are susceptible to social desirability bias, i.e. systematic underreporting of these frowned-upon attitudes. As expected, susceptibility for underreporting was higher for the more sensitive items measuring anti-Semitic attitudes compared to the less sensitive items measuring antiforeigner prejudices. Social scientists who aim to measure the prevalence of sensitive attitudes in the general population should take this potential source of measurement error into account. Future studies concerned about measurements of sensitive attitudes are encouraged to extend this research and to replicate the present findings in other cultural settings and countries. Moreover, it would be worthwhile to compare the RRT with alternative methods of data collection which also aim to increase respondents' privacy in CATI surveys. A recent experimental study showed that telephone audio-computer-assisted self-interviewing (T-ACASI), a questioning method in sensitive surveys that substitutes a computer for human interviewers, yielded more 'intolerant' and less 'socially liberal' responses compared to interviews conducted by human interviewers (Harmon et al., 2009). More specifically, more support of traditional gender roles and corporal punishment, and less support of same-gender sex and integrated neighborhoods could be observed in the group which was interviewed via the more anonymous T-ACASI. It would be worthwhile to implement and contrast the three techniques (RRT, T-ACASI, and 'conventional' CATI) in an empirical study.

Acknowledgments

The author would like to thank Ben Jann, Karl-Dieter Opp, Stephan Poppe, and Thomas Voss for their helpful comments. Research for this article was supported by the German Research Foundation (DFG) within the project 'Asking Sensitive Questions: Possibilities and Limits of Randomized Response and Other Techniques in Different Survey Modes' [VO 684/11 to Thomas Voss and Karl-Dieter Opp].

References

- Adorno, T.W., Frenkel-Brunswick, E., Levinson, D.J., Sanford, R.N., 1950. *The Authoritarian Personality*. Harper and Brothers, New York.
- Aquilino, W.S., Wright, D.L., 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.
- Bergmann, W., 1997. Antisemitism and Xenophobia in Germany since unification. In: Kurthen, H., Bergmann, W., Erb, R. (Eds.), *Antisemitism and Xenophobia in Germany after Unification*. Oxford University Press, New York, pp. 21–38.
- Bergmann, W., Erb, R., 1986. Kommunikationslatenz, Moral und öffentliche Meinung. Theoretische Überlegungen zum Antisemitismus in der Bundesrepublik Deutschland. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38, 223–246.
- Bergmann, W., Erb, R., 1991. Mir ist das Thema Juden irgendwie unangenehm. Kommunikationslatenz und die Wahrnehmung des Meinungsklimas im Fall des Antisemitismus. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 43, 502–519.
- Bergmann, W., Erb, R., 2000. Antisemitismus in der Bundesrepublik Deutschland 1996. In: Alba, R., Schmidt, P., Wasmer, M. (Eds.), *Deutsche und Ausländer: Freunde, Fremde oder Feinde? Empirische Befunde und theoretische Erklärungen*, Westdeutscher Verlag, Wiesbaden, pp. 401–438.
- Beyer, H., Krumpal, I., 2010. "Aber es gibt keine Antisemiten mehr": Eine experimentelle Studie zur Kommunikationslatenz antisemitischer Einstellungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62, 681–705.
- Böckenholt, U., Barlas, S., van der Heijden, P.G.M., 2009. Do randomized-response designs eliminate response biases? An empirical study of non-compliance behavior. *Journal of Applied Econometrics* 24, 377–392.
- Boruch, R.F., 1971. Assuring confidentiality of responses in social research: a systematic analysis. *The American Psychologist* 26, 413–430.
- Busse, B., Fuchs, M., 2012. The Components of Landline Telephone Survey Coverage Bias. The Relative Importance of Nophone and Mobile-only Populations. *Quality & Quantity* 46, 1209–1225.
- Coutts, E., Jann, B., 2011. Sensitive questions in online surveys: experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). *Sociological Methods and Research* 40, 169–193.
- Coutts, E., Jann, B., Krumpal, I., Näher, A.-F., 2011. Plagiarism in student papers: prevalence estimates using special techniques for sensitive questions. *Journal of Economics and Statistics* 231 (5+6), 749–760.
- Decker, O., Brähler, E., 2005. Rechtsextreme Einstellungen in Deutschland. *Aus Politik und Zeitgeschichte* 42, 8–17.
- Decker, O., Brähler, E., 2006. *Vom Rand zur Mitte. Rechtsextreme Einstellungen und ihre Einflussfaktoren in Deutschland*. Friedrich-Ebert-Stiftung, Berlin.
- Fowler, F.J.J., 1995. *Improving Survey Questions: Design and Evaluation*. Sage, Thousand Oaks, CA.
- Fox, J.A., Tracy, P.E., 1986. *Randomized Response: A Method for Sensitive Surveys*. Sage, Beverly Hills.
- Gabler, S., Häder, S., 1997. Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland. *ZUMA-Nachrichten* 41, 7–18.
- Gabler, S., Häder, S., 2002. Idiosyncrasies in telephone sampling – the case of Germany. *International Journal of Public Opinion Research* 14, 339–345.
- Gabler, S., Hoffmeyer-Zlotnik, J.H.P., Krebs, D., 1994. *Gewichtung in der Umfragepraxis*. VS Verlag für Sozialwissenschaften, Wiesbaden.
- Harmon, T., Turner, C.F., Rogers, S.M., Eggleston, E., Roman, A.M., Villarroel, M.A., Chromy, J.R., Ganapathi, L., Li, S., 2009. Impact of T-ACASI on survey measurements of subjective phenomena. *Public Opinion Quarterly* 73, 255–280.
- Holbrook, A.L., Krosnick, J.A., 2010. Measuring voter turnout by using the randomized response technique: evidence calling into question the method's validity. *Public Opinion Quarterly* 74, 328–343.
- Jann, B., 2005. *RRLOGIT: Stata module to estimate logistic regression for randomized response data*. Boston College Department of Economics, Boston.
- Jann, B., Jerke, J., Krumpal, I., 2012. Asking sensitive questions using the crosswise model: an experimental survey measuring plagiarism. *Public Opinion Quarterly* 76, 32–49.
- Kish, L., 1965. *Survey Sampling*. John Wiley, New York.
- Krosnick, J.A., Holbrook, A.L., Berent, M.K., Carson, R.T., Hanemann, W.M., Kopp, R.J., Mitchell, R.C., Presser, S., Ruud, P.A., Smith, V.K., Moody, W.R., Green, M.C., Conaway, M., 2002. The impact of "no opinion" response options on data quality: non-attitude reduction or an invitation to satiate? *Public Opinion Quarterly* 66, 371–403.
- Krumpal, I., 2011. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*. <http://dx.doi.org/10.1007/s11135-011-9640-9> (Online First).
- Krysan, M., 1998. Privacy and the expression of white racial attitudes – a comparison across three contexts. *Public Opinion Quarterly* 62, 506–544.
- Kunz, T., Fuchs, M., 2011. Pre-call Validation of RDD Cell Phone Numbers. A Field Experiment. Conference Paper AAPOR 2011, Phoenix, Arizona.

- Kurthen, H., 1997. Antisemitism and Xenophobia in United Germany: how the burden of the past affects the present. In: Kurthen, H., Bergmann, W., Erb, R. (Eds.), *Antisemitism and Xenophobia in Germany after Unification*. Oxford University Press, New York, pp. 39–87.
- Kurthen, H., Bergmann, W., Erb, R., 1997. *Antisemitism and Xenophobia in Germany after Unification*. Oxford University Press, New York.
- Landsheer, J.A., van der Heijden, P.G.M., van Gils, G., 1999. Trust and understanding, two psychological aspects of randomized response. *Quality & Quantity* 33, 1–12.
- Lara, D., Strickler, J., Olavarrieta, C.D., Ellertson, C., 2004. Measuring induced abortion in Mexico. A comparison of four methodologies. *Sociological Methods and Research* 32, 529–558.
- Lara, D., García, S.G., Ellertson, C., Camlin, C., Suárez, J., 2006. The measure of induced abortion in Mexico using random response technique. *Sociological Methods and Research* 35, 279–301.
- Lensvelt-Mulders, G.J.L.M., Hox, J.J., van der Heijden, P.G.M., Mass, C.J.M., 2005a. Meta-analysis of randomized response research. Thirty-five years of validation. *Sociological Methods and Research* 33, 319–348.
- Lensvelt-Mulders, G.J.L.M., Hox, J.J., van der Heijden, P.G.M., 2005b. How to improve the efficiency of randomized response designs. *Quality and Quantity* 39, 253–265.
- Lensvelt-Mulders, G.J.L.M., van der Heijden, P.G.M., Laudy, O., 2006. A validation of a computer-assisted randomized response survey to estimate the prevalence of fraud in social security. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169, 305–318.
- Long, J.S., Freese, J., 2006. *Regression Models for Categorical Dependent Variables Using Stata*, second ed. Stata Press, College Station, Texas.
- Maddala, G.S., 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, New York.
- McAuliffe, W.E., Breer, P., Ahmadifard, N.W., Spino, C., 1991. Assessment of drug abuser treatment needs in Rhode Island. *American Journal of Public Health* 81, 365–371.
- Mielke, R., Mummendey, H.D., 1995. Wenn Normen zu sehr wirken – Ausländerfeindlichkeit, Bildungsgrad und soziale Erwünschtheit. *Bielefelder Arbeiten zur Sozialpsychologie* 175, 1–9.
- Näher, A.F., Krumpal, I., 2011. Asking sensitive questions: the impact of forgiving wording and question context on social desirability bias. *Quality & Quantity*. <http://dx.doi.org/10.1007/s11135-011-9469-2> (Online First).
- Ostapczuk, M., Musch, J., Moshagen, M., 2009. A randomized-response investigation of the education effect in attitudes towards foreigners. *European Journal of Social Psychology* 39, 920–931.
- Putnam, R., 1993. *Making Democracy Work*. Princeton University Press, Princeton, NJ.
- Reuband, K.H., 1987. Unerwünschte Dritte beim Interview: Erscheinungsformen und Folgen. *Zeitschrift Für Soziologie* 16, 303–308.
- Reuband, K.H., 1992. On 3rd persons in the interview situation and their impact on responses. *International Journal of Public Opinion Research* 4, 269–274.
- Roose, N.J., Jamieson, D.W., 1993. 20 years of bogus pipeline research – a critical-review and metaanalysis. *Psychological Bulletin* 114, 363–375.
- Schaeffer, N.C., Presser, S., 2003. The science of asking questions. *Annual Review of Sociology* 29, 65–88.
- Scheers, N.J., Dayton, M.C., 1988. Covariate randomized response models. *Journal of the American Statistical Association* 83, 969–974.
- Stem, D.E., Steinhorst, R.K., 1984. Telephone interview and mail questionnaire applications of the randomized response model. *Journal of the American Statistical Association* 79, 555–564.
- Stocké, V., 2004. Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift Für Soziologie* 33, 303–320.
- Stocké, V., 2007. The interdependence of determinants for the strength and direction of social desirability bias in racial attitude surveys. *Journal of Official Statistics* 23, 493–514.
- Stolle, D., 2002. Trusting strangers – the concept of generalized trust in perspective. *Österreichische Zeitschrift für Politikwissenschaft* 31, 397–412.
- Tourangeau, R., Yan, T., 2007. Sensitive questions in surveys. *Psychological Bulletin* 133, 859–883.
- Van den Hout, A., Böckenholt, U., van der Heijden, P.G.M., 2010. Estimating the prevalence of sensitive behavior and cheating with a dual design for direct questioning and randomized response. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 59, 723–736.
- Van der Heijden, P.G.M., van Gils, G., Bouts, J., Hox, J.J., 2000. A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning – eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods and Research* 28, 505–537.
- Wagner, U., Zick, A., 1995. The relationship of formal education to ethnic prejudice. Its reliability, validity and explanation. *European Journal of Social Psychology* 24, 41–56.
- Warner, S.L., 1965. Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* 60, 63–69.
- Weissman, A.N., Steer, R.A., Lipton, D.S., 1986. Estimating illicit drug use through telephone interviews and the randomized response technique. *Drug and Alcohol Dependence* 18, 225–233.
- Zick, A., Wolf, C., Küpper, B., Davidov, E., Schmidt, P., Heitmeyer, W., 2008. The syndrome of group-focused enmity: the interrelation of prejudices tested with multiple cross-sectional and panel data. *Journal of Social Issues* 64, 363–383.