Public Opinion Quarterly, Vol. 74, No. 1, Spring 2010, pp. 37–67

SOCIAL DESIRABILITY BIAS IN VOTER TURNOUT REPORTS
TESTS USING THE ITEM COUNT TECHNIQUE

ALLYSON L. HOLBROOK*
JON A. KROSNICK

Abstract Surveys usually yield rates of voting in elections that are higher than official turnout figures, a phenomenon often attributed to intentional misrepresentation by respondents who did not vote and would be embarrassed to admit that. The experiments reported here tested the social desirability response bias hypothesis directly by implementing a technique that allowed respondents to report secretly whether they voted: the "item count technique." The item count technique significantly reduced turnout reports in a national telephone survey relative to direct self-reports, suggesting that social desirability response bias influenced direct self-reports in that survey. But in eight national surveys of American adults conducted via the Internet, the item count technique did not significantly reduce turnout reports. This mode difference is consistent with other evidence that the Internet survey mode may be less susceptible to social desirability response bias because of self-administration.

Self-reports in surveys have often overestimated voter turnout, and researchers have speculated that this may occur partly because some respondents intentionally misreport that they voted because they wish to portray themselves in admirable ways (Corbett 1991; Lyons and Scheb 1999; Aarts 2002; Brockington and Karp 2002; Andolina et al. 2003; Lutz 2003; Blais et al.

ALLYSON L. HOLBROOK is an associate professor of public administration and psychology at the Survey Research Laboratory of the University of Illinois at Chicago, 412 S. Peoria St. (MC336), Chicago, IL 60607, USA. JON A. KROSNICK is Professor of Communication, Political Science, and Psychology, Stanford University, 434 McClatchy Hall, 450 Serra Mall, Stanford, CA, 94305, USA, and University Fellow at Resources for the Future. Study 3's data were collected through Time-Sharing Experiments for the Social Sciences (TESS), supported by the National Science Foundation (SES-0094964). Studies 1, 2, and 4's data were collected by the Stanford Institute for the Quantitative Study of Society. The authors thank Young Ik Cho and Anthony Greenwald for comments and suggestions and thank Norman Nie, Douglas Rivers, and LinChiat Chang for their roles in making Studies 1, 2, and 4 possible. *Address correspondence to Allyson L. Holbrook; e-mail: allyson@uic.edu; or Jon A. Krosnick; e-mail: krosnick@stanford.edu

doi:10.1093/poq/nfp065 Advance Access publication November 5, 2009

2004).[1] Previous attempts to reduce overreporting of turnout by reducing social desirability pressures have generally been unsuccessful (e.g., Presser 1990; Abelson, Loftus, and Greenwald 1992), but the strategies used to reduce social desirability pressures in past studies have generally not been validated, so it is not clear whether they were effective. Therefore, their failure to reduce turnout overreporting could be attributed to the strategies' failure to reduce social desirability pressures.

Methods that have effectively reduced social desirability response bias in other domains have not yet been implemented in turnout studies. We explored whether one such strategy (the item count technique) reduced turnout reports, providing more direct tests of the hypothesis that social desirability response bias is responsible for overreporting in telephone and internet surveys. We begin below by documenting turnout overreporting and reviewing past studies examining its causes. Then we outline the experiments we conducted, describe our results, and discuss their implications.

The Problem

Evidence that surveys overestimate the proportion of people who voted is of two sorts. At the aggregate level, the proportion of respondents who report they voted has often been larger than the proportion of voters who were officially recorded to have voted (e.g., Clausen 1968; Traugott and Katosh 1979). For example, the 1976 American National Election Study's (ANES) turnout estimate was 72 percent, and the Census Bureau's Voting Supplement to the November 1976 Current Population Survey estimated 59 percent, whereas the government's official turnout rate was only 54 percent (Traugott and Katosh 1979). Similarly, the 1964 ANES estimated turnout to be 78 percent, in contrast to the government's official turnout rate of 63 percent (Clausen 1968).

Other evidence of turnout overreporting comes from comparisons of individuals' self-reports with official records of their voting behavior. When citizens go to the polls to vote, their vote choices are anonymous, but officials keep a record of whether they voted, and these records can be used to determine whether a respondent voted in a particular election. When researchers have compared respondents' self-reports with their official turnout records, the proportion of people who reported that they voted has been consistently larger than the proportion for whom evidence of turnout was found. For example, Traugott and Katosh (1979) reported that 78 percent of ANES respondents reported that they voted in 1976, but official records confirmed that only 61 percent of those individuals had actually voted. If systematic misreporting produces such discrepancies, they may distort the conclusions of research attempting to identify the causes or consequences of turnout.

1. Bernstein, Chadha, and Montjoy (2001) have also suggested that nonvoters sometimes feel guilty about not voting, and this guilt motivates misreporting.

Sources of Measurement Error not Related to Reporting Accuracy

Discrepancies between survey assessments of turnout rates and actual turnout rates may have a number of causes, many of which do not involve intentional misreporting by respondents.

ERRORS IN VALIDATED TURNOUT MEASURES

Although collecting official records of respondent turnout is very expensive, such assessments are not without errors themselves (Traugott 1989; Presser, Traugott, and Traugott 1990). When checking turnout records, it is much more likely that the record of an actual voter will not be found (e.g., because it has been misplaced or the voter was registered at a different address) than that a record incorrectly indicating that a nonvoter voted will be found. As a result, the discrepancy between self-reported and official indicators of turnout may be partly due to error in the latter.

INCORRECT DENOMINATORS

The population of "potential" voters included in the denominator when calculating turnout has sometimes been incorrect. Clausen (1968) found that the denominator used in the government's official report of turnout in the 1964 presidential election included homeless, jailed, and institutionalized individuals—"potential" voters who were not included in the pool of possible respondents for any survey. McDonald and Popkin (2001) showed that observed declines in official estimates of turnout since 1972 were the result of increases in the number of people in these not-interviewed groups (see also McDonald 2003). Because these people are unlikely or unable to vote, failing to include them in the denominator of survey turnout estimates inflates these estimates relative to official rates of turnout among the population over age 18.

SURVEY NONRESPONSE

Failure of some sample members to be interviewed also impacts the accuracy of survey turnout assessments. For example, in the 1964 ANES, 19 percent of eligible potential respondents were not interviewed preelection, and an additional 6 percent were not interviewed postelection. Clausen (1968) found that respondents who were interviewed both pre- and postelection were more likely to have voted than people interviewed only preelection, and controlling for unit nonresponse reduced the discrepancy between official turnout estimates and survey estimates. More recently, Burden (2000) argued that increasing unit

nonresponse in the ANES over time has been responsible for increasing overestimation of turnout rates, although other researchers have disagreed (Martinez 2003; McDonald 2003).

THE EFFECT OF PREELECTION INTERVIEWS

Many surveys that gathered postelection turnout self-reports involved interviewing respondents before the election as well (e.g., the ANES). This raises the possibility that being interviewed preelection might increase people's actual turnout rate, thus inflating the proportion of survey respondents who voted relative to the nation as a whole (Clausen 1968; Kraut and McConahay 1973; Yalch 1976). Several studies have demonstrated that being interviewed before an election increases the probability that a person will vote (see Traugott and Katosh 1979 for a review). And some evidence suggests that simply being asked to predict whether one will vote prior to election day sometimes makes a person more likely to vote (e.g., Greenwald et al. 1987), although other studies have failed to find this effect (e.g., Greenwald et al. 1988; Smith, Gerber, and Orlich 2003; Mann 2005).

THE COMBINED IMPACT OF CONFOUNDING FACTORS

To assess the combined impact of some confounding factors, Clausen (1968) compared self-reported turnout in an ANES survey to reports in a Census Bureau survey done at the same time. The ANES survey involved pre- and postelection interviews with the same respondents, whereas the Census Bureau survey involved only postelection interviews. The unit nonresponse rate was much higher for the ANES (25 percent) than for the Census Bureau survey (4 percent), and the Census Bureau survey included residents of "boarding-houses," whereas the ANES sample did not. Thus, the Census Bureau survey minimized all three sources of error relative to the ANES, and the Census Bureau estimated a turnout rate considerably less than the ANES did. Nonetheless, the Census Bureau survey's turnout figure was higher than the official turnout figure by 5.5 percentage points (Clausen 1968). Traugott and Katosh (1979) found very similar results in parallel analyses of other ANES and Census Bureau surveys.

Sources of Measurement Error Related to Reporting Accuracy

This remaining overestimation may have occurred because some respondents reported that they voted when in fact they did not. There are at least three possible explanations for such overreporting: (1) some respondents might

answer yes/no turnout questions affirmatively due to acquiescence response bias, (2) some respondents may unintentionally misremember that they voted when they did not, or (3) some respondents may intentionally misreport having voted in order to present themselves in a socially desirable way.

ACQUIESCENCE

Abelson, Loftus, and Greenwald (1992) asked some respondents (selected randomly) whether they had voted and asked other respondents whether they had "missed out" on voting. This latter question reverses the impact of acquiescence response bias by associating an affirmative answer with not having voted. These two questions yielded equivalent percentages of people saying that they voted, challenging the notion that acquiescence response bias contributes to overreporting.

MEMORY ERRORS

Unintentional misremembering can occur because of mistakes made when respondents consult their long-term memories to retrieve stored information about whether they voted in a particular election (Belli, Traugott, and Rosenstone 1994). When asked whether he or she voted in a particular election, a person may retrieve memories of elections in which he or she did vote. Then, the individual must decide when each of those retrieved instances occurred: at the time of the election being asked about or at the time of a different election. People who often voted in the past but did not vote in the most recent election may be especially likely to retrieve memories of having voted, and if these people make dating errors, this may lead them to incorrectly say they voted recently (Belli, Traugott, and Rosenstone 1994). In addition, people who considered voting but did not end up doing so may confuse this thought with carrying out the action (Belli, Traugott, and Rosenstone 1994). This is called "source confusion."

This hypothesis is consistent with evidence that, as compared to respondents who usually do not vote, respondents who usually vote are more likely to say that they voted recently when in fact they did not (Abelson, Loftus, and Greenwald 1992). Also consistent is evidence indicating that turnout overreporting may increase as the time between the election and the interview increases (Belli, Traugott, and Beckmann 2001, although see Belli, Moore, and Van Hoewyk 2006).

SOCIAL DESIRABILITY

Even if memory errors account for some of the survey overestimation of turnout rates, it is possible that social desirability response bias accounts for some overreporting as well. Voting is an admired and valued civic behavior (see

Holbrook, Green, and Krosnick 2003), so some people may be reluctant to admit that they did not live up to their civic duty. A great deal of evidence suggests that survey respondents sometimes intentionally present themselves in inaccurate but socially admirable ways (Warner 1965; Sigall and Page 1971; Pavlos 1972; Evans, Hansen, and Mittlemark 1977; Himmelfarb and Lickteig 1982; Paulhus 1984; see Demaio 1984 for a review), and researchers studying voting behavior have speculated that this may be one source of overreporting (e.g., Silver, Anderson, and Abramson 1986).

Many question wording experiments on turnout have attempted to reduce social desirability pressures. However, almost none of these experiments yielded statistically significant evidence of intentional misreporting, and each study is open to alternative explanations for its failure(s). For example, in an experiment by Abelson, Loftus, and Greenwald (1992), half of the respondents (selected randomly) were asked whether they had voted in the most recent Congressional elections, and the other half were first asked whether they had voted in previous elections and were then asked whether they had voted in the most recent elections. Allowing respondents to report previous voting enabled them to communicate that they usually lived up to the norm of civic responsibility, which might decrease any pressure they might feel to distort reports of recent turnout behavior. This manipulation had no significant impact on turnout reports, challenging the social desirability hypothesis. An experiment by Presser (1990) yielded a similar result. However, perhaps answering the first question affirmatively (about voting in previous elections) made respondents feel pressure to appear consistently civic minded, thereby enhancing the felt need to report having turned out in the most recent elections.

In another experiment by Abelson, Loftus, and Greenwald (1992), half of the respondents were told that there were two elections in the fall of 1988 (the primaries and the presidential election) and were asked whether they had voted in the primaries. The other half of the respondents were told that "most people aren't able to get to vote in every election," were asked whether they voted in both elections, and then were asked whether they had voted in the primaries. The latter approach was intended to communicate to respondents that they were not expected to vote all the time and that it would therefore be reasonable to respond negatively to the question about voting in the primaries. This manipulation also did not alter turnout reports significantly. However, pressure to appear consistently civic minded is an alternative explanation for this finding as well.

Presser (1990) tried another approach to reducing social desirability pressures. Half of the respondents were asked whether they had voted, and the other half were first asked the location of the place where they go to vote before being asked whether they voted. Presser (1990) thought that many overreporters would not know the location of their voting place, and compelling them to acknowledge that first would reduce their inclination to claim they voted when in fact they did not. However, this question order manipulation had no significant

effect, perhaps because most overreporters knew the location of their polling place, since overreporters usually vote.[2]

The failures of these manipulations to affect turnout reports may be attributable to two possible explanations: social desirability response bias does not influence reports of turnout, or the manipulations used in these studies did not successfully reduce social desirability pressures. Because the latter cannot be fully ruled out, these manipulations do not allow us to reject social desirability response bias as an explanation for turnout overreporting.

MEMORY ERRORS AND SOCIAL DESIRABILITY

Other researchers have attempted to reduce turnout overreporting using manipulations designed to reduce *both* social desirability and memory errors, but with mixed success (Belli, Traugott, and Rosenstone 1994; Belli et al. 1999). Belli, Traugott, and Rosenstone (1994) randomly assigned postelection survey respondents to be asked either a simple question about whether they voted in the most recent election or a more complex "experimental" question designed to reduce source confusion by (1) explicitly making people aware of the problem of source confusion, (2) encouraging people to think carefully about whether they voted, and (3) offering four response options: "I did not vote in the November 8th election"; "I thought about voting this time, but didn't"; "I usually vote, but didn't this time"; and "I am sure I voted in the November 8th election." Voting records were checked to validate self-reports.[3] The proportion of respondents who said they voted was about the same among people asked the standard turnout question (87.8 percent) as among people asked the experimental turnout question (87.1 percent). Furthermore, the proportion of respondents who accurately reported whether they voted was about the same among people asked the standard turnout question (94.5 percent) and among those asked the experimental question (95.2 percent; Belli, Traugott, and Rosenstone 1994).

However, recent similar experiments have been more successful (Belli et al. 1999; Belli, Moore, and Van Hoewyk 2006). In three studies, some respondents (selected randomly) were asked a simple and direct turnout question, and other respondents were asked an "experimental" question that emphasized people's tendency to have trouble remembering, encouraged careful thought about the specifics of election day, and offered respondents the same four answer choices

2. The characteristics of over-reporters have been identified by comparing respondents who inaccurately claimed to have voted (when official records suggest they did not) to validated voters (respondents who said they voted and official records show that they did) and admitted nonvoters (respondents who said they did not vote and official records verify that; e.g., Belli, Traugott, and Beckmann 2001).

3. Because the respondents were selected from lists of registered voters in a limited geographic area (Ann Arbor and Ypsilanti, Michigan), researchers were able to rule out many of the sources of error typically associated with validation in national surveys (e.g., Presser, Traugott, and Traugott 1990; Traugott 1989).

employed by Belli, Traugott, and Rosenstone (1994). This question wording reduced overreporting, but it is difficult to know why, because the lengthy question introduction and answer choices were designed to reduce *both* memory errors *and* social desirability pressures.

The ICT: A Method for Reducing Social Desirability Pressures

All this evidence suggests that some of the errors in turnout reports may be due to social desirability response bias. However, a number of well-established techniques to reduce social desirability response bias have not been tested in the context of voter turnout. Demonstrating that such a technique reduces turnout reports would provide the strongest evidence about the effects of social desirability on turnout reports. We therefore set out to test whether allowing respondents to report turnout secretly would reduce overreporting, using the item count technique (ICT).

The logic of this approach is as follows. Social desirability response bias is presumed to result from a desire among some respondents to misrepresent themselves in admirable ways. Thus, when asked to report something embarrassing directly and explicitly, these individuals may choose to answer inaccurately to avoid being judged negatively by an interviewer or researcher. If the respondent could report an embarrassing fact anonymously and confidentially, then he or she would have no motivation to lie and would tell the truth. If a method to elicit anonymous self-reports can be implemented, then responses under those conditions can be compared to responses provided explicitly by a comparable group of respondents. If the group reporting anonymously and confidentially acknowledges possessing a socially undesirable attribute more often than the group reporting directly and explicitly, this suggests that social desirability response bias affected direct reports of the attribute. If no difference is observed, that would suggest that no lying occurred in the direct reports.

Past Studies of The ICT has been used for this purpose for decades (Miller 1984; Miller, Harrel, and Cisin 1986; Droitcour et al. 1991; Tsuchiya, Hirai, and Ono 2007) and has sometimes been called the "unmatched count technique" (e.g., Dalton, Wimbush, and Daily 1994; Dalton, Daily, and Wimbush 1997) or the "list technique" (Kuklinski et al. 1996; Sniderman and Grob 1996; Kuklinski, Cobb, and Gilens 1997; Kuklinski and Cobb 1998; Cobb 2001). Half of a sample (selected randomly) are asked to report the number of items on a list that fit a particular criterion. For example, a respondent can be given a list of three behaviors and asked how many of them he or she has performed. The other half of the respondents can be given the same list of three plus one additional behavior and asked the same question. Subtracting the average number of behaviors reported by the first group of respondents from the average number of behaviors reported by the second group estimates the proportion of people given the longer list who said they performed the added behavior. Because

respondents know that the researcher cannot know which of the behaviors they performed, their answers to this question should be undistorted by social desirability response bias.

In a number of past studies, the ICT yielded more reports of undesirable attributes than did direct self-report questions. For example, the ICT has indicated significantly more illegal drug use (Miller 1984; Miller, Harrel, and Cisin 1986), more unethical workplace behavior (Dalton, Wimbush, and Daily 1994), more employee theft (Wimbush and Dalton 1997), more risky sexual behavior (e.g., LaBrie and Earleywine 2000; the difference was significant in 3 of 4 tests), more hate crime victimization (Rayburn, Earleywine, and Davison 2003a, 2003b; the difference was significant for 14 of 18 behaviors), and more shop-lifting (Tsuchiya, Hirai, and Ono 2007) than did direct self-reports. Furthermore, several studies have shown that the ICT yielded similar estimates to direct reports for behaviors with minimal social desirability connotations, such as professional auctioneers' reports of making audio and video recordings of auctions and of charging bidders a buyer's premium (Dalton, Wimbush, and Daily 1994), college students getting drunk (LaBrie and Earleywine 2000), giving blood (Tsuchiya, Hirai, and Ono 2007), and endorsing the belief that greater employee theft occurs during the night shift (Wimbush and Dalton 1997). In contrast, only two studies examined behaviors tinged with social desirability connotations and found no difference in prevalence estimates generated by the ICT and direct self-reports (e.g., Droitcour et al. 1991; Ahart and Sackett 2004). Combining across all past studies, the ICT yielded significantly higher prevalence estimates of undesirable behaviors and attitudes than did direct self-report measures in 63 percent of forty-eight comparisons, much more than would be expected by chance alone.[4] Thus, these studies yielded much evidence suggesting that the ICT may improve the validity of self-reports by reducing social desirability pressure.[5]

Past studies have also found that the predictors of sensitive behaviors measured with the ICT are different than the predictors of those behaviors measured with direct self-reports. For example, Corstange (2006) found that according to direct self-reports of support for voting rights in Lebanon, Shia respondents were more supportive of these rights than were Christian respondents, and respondents of higher social class were more likely to support voting rights. But the ICT method did not reveal either of these relations. And the ICT method suggested that the more respondents shared class identities with Christians, the

4. Cobb (2001) compared estimates from ICT measures to those from direct self-reports but did not report significance tests of the differences observed, so we do not discuss those findings here.
5. Tourangeau and Yan (2007) reported a meta-analysis of some of these studies and found that the ICT did not have a significant impact on reported frequencies across the studies they examined. But they also found significant heterogeneity among studies, suggesting that it was not appropriate to combine them in a meta-analysis. Our review of these studies suggests many reliable effects of the ICT on reports.

less likely they were to support voting rights, but this relation did not appear using direct self-reports.

This Paper

Despite the widespread application of the ICT to measurement of a wide range of sensitive attitudes and behaviors, no studies have used this technique to study reports of turnout. Indeed, most applications of the ICT have focused on stigmatized behaviors rather than those that have positive social desirability connotations (e.g., behaviors that may be overestimated using direct self-reports). Our goal in conducting the research reported here was to explore further the possibility that social desirability might bias reports of turnout in surveys of large samples of American adults. Study 1 involved a telephone survey of a representative national sample of American adults. Studies 2 and 3 involved Internet surveys of representative national samples. Study 4 involved six Internet surveys conducted with nonprobability samples. In each study, we compared turnout estimates yielded by direct self-reports and the ICT. We also explored whether the demographic predictors of turnout vary depending on the measurement strategy used.

We followed a number of guidelines when designing our ICT experiments. First, the behaviors to be used on an ICT list should be such that few respondents have performed all or none of them, because giving one of those responses voids the anonymity to be provided by the ICT. We therefore used recent national surveys (e.g., the General Social Survey) to choose behaviors that were likely to have been performed by some but not all respondents. Second, as Tsuchiya, Hirai, and Ono (2007) recommended, we had separate groups of respondents answer the direct self-report and ICT questions, and we did not provide a demonstration of the ICT before they were asked to implement it.

Methods[6]

STUDY 1

Respondents: A representative national RDD sample of 898 American adults was interviewed by telephone by Schulman, Ronca, and Bucuvalas, Inc. (AAPOR Response Rate 3 was 35.6 percent).

Experimental conditions: Twenty percent of respondents were randomly assigned to be asked the traditional ANES voter turnout question ($N = 176$)[7]: "In talking to people about elections, we often find that a lot of people were not

6. Additional methodological details on our studies are presented in Appendix A.
7. One respondent did not provide a substantive response to this question and was excluded from our turnout analyses.

able to vote because they weren't registered, they were sick, or they just didn't have time. How about you—did you vote in the Presidential election held on November 7, 2000?"[8]

Another 20 percent of the respondents were randomly assigned to the 4-item ICT condition ($N = 186$): "Here is a list of four things that some people have done and some people have not. Please listen to them and then tell me HOW MANY of them you have done. Do not tell me which you have and have not done. Just tell me how many. Here are the four things: Owned a gun; given money to a charitable organization; gone to see a movie in a theater; written a letter to the editor of a newspaper. How many of these things have you done?"

Another 20 percent of the respondents were randomly assigned to a 5-item ICT condition ($N = 168$) and were asked the same question as the 4-item ICT respondents with the addition of "Voted in the Presidential election held on November 7, 2000."[9] The percent of respondents who voted was estimated by subtracting the mean for the 4-item ICT condition from the mean for the 5-item ICT condition.

Demographics: Gender, age, education, and race were measured using questions similar to ones from the U.S. Census Bureau's Current Population Survey (see Appendix A).

STUDY 2

Respondents: Study 2's survey was administered by Knowledge Networks (KN), whose representative national panel of American adults was recruited via RDD telephone interviews. People who did not have Internet access were given equipment to access the Internet using their televisions. Panelists were sent weekly emails inviting them to complete questionnaires (see Knowledge Networks 2006). A total of 1,533 panelists were randomly drawn from the KN panel and invited to complete this survey. A total of 1,175 responded to the invitation and 1,137 of these completed the survey, yielding a final stage completion rate of 74.2 percent. The recruitment rate for this study was 35.1 percent; the profile rate was 58.7 percent; the cumulative response rate was 15.3 percent.

8. The NES turnout question wording has varied over the years; this wording was employed in 1952–1960, 1964–1998, and 2002.

9. One respondent did not provide a substantive answer to this question and was excluded from our turnout analyses. The remaining respondents in this study were assigned to experimental conditions not analyzed in this paper.

Experimental conditions: Ten percent of respondents were randomly assigned to be asked the traditional ANES voter turnout question ($N = 117$).[10] Twenty percent were randomly assigned to the 4-item ICT condition ($N = 234$),[11] and 20 percent were randomly assigned to the 5-item ICT condition ($N = 232$).[12] Procedures and instructions were like Study 1 but adapted for Internet administration (see Appendix A).[13]

Demographics: Gender, age, education, and race were again measured (see Appendix A).

STUDY 3

Respondents: Study 3's survey was also administered by Knowledge Networks. A total of 9,894 panelists were randomly drawn from the KN panel and invited to complete this study. A total of 6,094 responded to the invitation, yielding a final stage completion rate of 61.6 percent. The recruitment rate for this study was 47.6 percent; the profile rate was 61.8 percent; the cumulative response rate was 18.2 percent.

Experimental conditions:[14] One-third of respondents were randomly assigned to be asked the traditional ANES direct self-report turnout question ($N = 2{,}018$).

The ICT was implemented slightly differently than in Studies 1 and 2. This study used 3- and 4-item lists, and some different behaviors were used. One-sixth ($N = 1{,}017$) of respondents were randomly assigned to the 3-item ICT condition, and another one-sixth ($N = 1{,}012$) were randomly assigned to the 4-item ICT condition. The behaviors on the 3-item list were "given money to a charitable organization," "served in the military in any capacity," and "written a letter to the editor of a newspaper." For the 4-item list, "voted in the elections held on November 5, 2002" was added.[15]

10. Two respondents did not provide substantive responses to this question and were excluded from our turnout analyses.
11. Four respondents did not provide substantive responses to this question and were excluded from our turnout analyses.
12. Eight respondents did not provide substantive responses to this question and were excluded from our turnout analyses.
13. The remaining respondents were assigned to experimental conditions not included in our analyses.
14. Thirty-five respondents did not have valid data for either the variable that assigned respondents to a turnout condition or to any of the turnout questions. These were partial interviews and breakoffs where respondents stopped participating before the turnout questions, and these respondents were excluded from our turnout analyses.
15. The remaining respondents were assigned to experimental conditions not included in our analyses.

Demographics: Measures of gender, age, education, and race were gathered during the profile survey that respondents completed when they first joined the Knowledge Networks panel (see Appendix A).

STUDY 4

Study 4 implemented the ICT in six Internet surveys conducted by six different companies that routinely collect data from panels of volunteer respondents: Firm 1 ($N = 1{,}129$), Firm 2 ($N = 1{,}223$), Firm 3 ($N = 2{,}664$), Firm 4 ($N = 1{,}137$), Survey Direct ($N = 1{,}323$), and Firm 5 ($N = 1{,}103$; Firm 5 provided the sample, and another company administered the questionnaire online), and Firm 6 ($N = 1{,}323$). Each company was asked to draw a representative national sample for the survey (see Appendix A for more details). Random assignment to experimental condition was done as in Study 2.[16] Demographics were measured as in Study 2.

Results

SAMPLE COMPOSITION

Shown in table 1 are the unweighted demographic characteristics of respondents in Studies 1, 2, and 4 and of weighted demographic characteristics of respondents in the U.S. Census Bureau's March Current Population Survey (CPS) sample interviewed in the same year, 2004. Study 3's sample is described in column 6, and the CPS Sample interviewed in the same year, 2003, is described in column 5. In general, the samples are similar to their respective populations, although the Study 4 sample deviated sharply from the Census estimates in terms of education.

SUCCESS OF RANDOM ASSIGNMENT

To assess whether random assignment produced comparable groups of respondents, we compared the distributions of gender, race, age, and education across conditions in each study. Only one of fifty such comparisons yielded a statistically significant difference, less than would be expected by chance alone. Therefore, random assignment was effective.

16. Across organizations, a total of 705 respondents did not have valid data for either the variable assigning respondents to a turnout condition or to any of the turnout questions. These were partial interviews and breakoffs where respondents stopped participating before the turnout questions, and these respondents were excluded from our turnout analyses.

Table 1. Demographic Characteristics of Respondents and CPS Demographic Distributions

	Studies conducted in 2004				Studies conducted in 2003	
	CPS	Study 1	Study 2	Study 4	CPS	Study 3
Gender						
Male	48.2%	43.3%	47.9%	44.5%	48.3%	46.9%
Female	51.8	56.7	49.0	49.2	51.7	53.1
Missing	0.0	0.0	3.1	6.3	0.0	0.0
Total	100%	100%	100%	100%	100%	100%
Race						
White	82.2%	78.2%	76.1%	72.6%	82.0%	78.1%
Nonwhite	17.8	20.5	20.1	20.5	18.0	21.9
Missing	0.0	1.3	3.8	6.9	0.0	0.0
Total	100%	100%	100%	100%	100%	100%
Hispanic origin						
Hispanic	12.2%	4.9%	8.9%	6.2%	12.4%	7.4%
Nonhispanic	87.8	94.0	88.0	83.8	87.6	92.6
Missing	0.0	1.1	3.1	10.0	0.0	0.0
Total	100%	100%	100%	100%	100%	100%
Age						
18–24	12.9%	7.8%	7.2%	10.4%	13.0%	7.4%
25–34	18.5	14.8	15.5	17.3	18.3	14.6
35–44	20.7	17.6	20.9	17.5	20.3	22.7
45–54	18.9	22.8	18.7	19.0	19.1	19.7
55–64	12.9	15.9	15.1	15.9	13.2	17.0
65–74	8.5	11.1	11.0	10.9	8.5	11.6
75 and older	7.6	7.2	7.0	2.6	7.6	6.9
Missing	0.0	2.7	4.5	6.3	0.0	0.0
Total	100%	100%	100%	100%	100%	100%
Education						
High school graduate or less	48.1%	38.1%	45.4%	18.1%	47.5%	43.1%
At least some college	51.8	60.8	51.3	73.2	52.5	56.9
Missing	0.0	1.1	3.2	8.7	0.0	0.0
Total	100%	100%	100%	100%	100%	100%
N	149,844[a]	898[b]	1,175[b]	8,579[b]	148,180[a]	6,094[b]

[a]The sample sizes reported for the 2003 and 2004 CPS data were obtained from the unweighted survey data. The percentages reported are weighted using person-level expansion weights provided for each March CPS survey to weight the sample to the size of the estimated total population.

[b]The demographic characteristics of the survey samples include the full sample of respondents (including those assigned to conditions other than the standard ANES question wording and ICT conditions).

DIRECT QUESTION

The proportions of respondents asked the direct self-report question who reported voting were 72.0 percent (Study 1), 66.1 percent (Study 2), and 69.9 percent (Study 4) for the 2000 Election and 59.5 percent (Study 3) for the 2002 election (see row 3 of table 2).

ITEM COUNT TECHNIQUE

According to Study 1's telephone data, the ICT yielded a turnout estimate of 52.4 percent (see row 7 of table 2),[17] significantly lower than the direct self-report question's 72.0 percent (z-statistic = 1.66, $p < .05$).[18] This suggests that social desirability pressures were present when the traditional ANES question was asked in this telephone survey and that these pressures were reduced by using the ICT.

According to the Internet data from Studies 2, 3, and 4, the ICT turnout rates were not significantly different from the direct self-report rates (compare rows 3 and 7 in table 2).[19] The figures for the ICT and direct self-reports were 66.4 percent and 66.1 percent, respectively, for Study 2 (z-statistic = .009, ns), 58.1 percent and 59.5 percent for Study 3 (z-statistic = .31, ns), and 66.8 percent and 69.9 percent for Study 4 (z-statistic = .81, ns). A meta-analysis of these three studies indicated no significant difference between the ICT and direct self-report measurements ($z = .67$, ns). This suggests that social desirability pressures did not distort turnout reports in the Internet surveys.

In a meta-analysis of all four studies, the difference between the turnout rates yielded by the direct self-reports and the ICT was marginally significantly greater in the telephone data than in the Internet data ($z = 1.41$, $p < .10$).

17. In the five-item condition, 3.0 percent of respondents reported they had done zero behaviors, and 7.2 percent reported they had done all five of the behaviors.

18. To compare turnout estimates across conditions, we computed z-tests for proportions: $z = (\text{abs}(p_1\text{-}p_2))/(\text{sqrt}(\text{SE}_1{}^2+\text{SE}_2{}^2))$. The standard error of a proportion obtained from a direct self-report was: $\text{SE}_\text{p} = \text{Square Root}((p^*(1-p))/n)$. The standard error of a proportion obtained with the ICT was the standard error of the difference between the means for the two ICT conditions. This approach takes into account error in the ICT proportion estimate due to random assignment to condition and variance in the prevalence of nonsensitive behaviors included in the ICT lists. As Tsuchiya, Hirai, and Ono (2007) noted, the variance in the ICT estimates is likely to increase as a function of variance in the prevalence of the nontarget items on the list and as a function of the length of the list. Our approach takes into account the variances of the means from the two ICT conditions in comparing the ICT estimate to the direct self-report estimate. For all tests of directional hypotheses, we report one-tailed p's when the difference was in the expected direction. All other reported p's are two-tailed.

19. In the five-item condition in Study 2, 3.6 percent of respondents said they had done zero behaviors, and 8.0 percent said they had done all five of the behaviors. In the four-item condition in Study 3, 12.9 percent of respondents said they had done zero behaviors, and 3.6 percent said they had done all four of the behaviors. In the five-item condition in Study 4, 1.3 percent of respondents said they had done zero behaviors, and 12.3 percent said they had done all five of the behaviors.

Table 2. Estimates of Turnout

	Study 1	Study 2	Study 3	Study 4	All studies	Internet studies only
Traditional NES wording						
Number "YES"	126	76	1,200	599		
Sample size	175	115	2,018	857		
Estimated turnout	**72.0%**	**66.1%**	**59.5%**	**69.9%**		
ICT						
Mean for short list	2.4	2.3	1.2	2.5		
Mean for long list	2.9	2.9	1.7	3.2		
Sample size	353	454	2,029	3,077		
Estimated turnout	**52.4%**	**66.4%**	**58.1%**	**66.8%**		
Standard error of difference	**11.7%**	**9.9%**	**3.9%**	**3.7%**		
95% confidence interval	**29.4–75.3%**	**46.9–85.8%**	**50.4–65.7%**	**59.5–74.1%**		
Comparison of estimates from ICT and traditional NES wording						
Conditions: z-test[a]	1.66*	.009	.31	.81		
Combined z					1.41+	.67

NOTE.—Sample sizes reported in this table indicate the number of valid responses.

[a]Z-tests were used to test proportion differences in turnout estimates between conditions. $^{+}p < 0.10$, $^{*}p < 0.05$.

Thus, it appears that social desirability pressures were less when respondents answered questions via the Internet than when they spoke over the telephone.[20]

DEMOGRAPHIC PREDICTORS OF TURNOUT

To explore whether using the ICT instead of direct self-reports changes the apparent demographic predictors of turnout, we conducted one analysis of the telephone data and a second analysis using the combined Internet data (from Studies 2, 3, and 4). With the direct self-reports, we conducted logistic regressions (because turnout is dichotomous) using education, age, age squared, race, Hispanic origin, and gender as predictors. With the ICT data, we conducted OLS regressions predicting the count provided by respondents with a dummy variable indicating whether the respondent received the short list or the long list, the demographics, and interactions of the list length dummy variable with each demographic. With the Internet data, another predictor was added: a dummy variable indicating whether the data came from Study 3 (involving the 3- and 4-item lists) or from Studies 2 and 4 (involving the 4- and 5-item lists).

Because respondents were randomly assigned to conditions, respondents in each of the two ICT conditions were equally likely to have performed each of the nonsensitive behaviors. Therefore, the interactions in these regressions test whether the demographics predicted the magnitude of the impact of adding turnout to the behavior list. In other words, these interactions test whether the difference between the reported number of behaviors in the two conditions (which indicates the proportion of people who voted) was larger in some groups of respondents than in other groups. The bigger the interaction, the more of the respondents in the specified group voted. For example, a positive interaction

20. By design, the ICT uses two lists of behaviors, one that is longer than the other list. If simply offering a longer list induces some respondents to give larger numeric answers, regardless of the items on the list, this would cause apparent differences between the experimental conditions that do not reflect accurate reporting of attitudes, beliefs, or behaviors. To test this possibility, we conducted an experiment in an Internet survey of a non-representative sample of American adults who volunteered to do surveys. The sample was provided by Firm 5 (see Appendix A for a description of their methodology), and the completion rate was 72 percent (field dates: May 7 to 13, 2008). Seven hundred sixty nine respondents were randomly assigned to be asked the following question: "Here is a list of four things that some people have done and some people have not. Please read them and then report below HOW MANY of them you have done. Do not report which you have and have not done. Just report how many. Here are the four things: owned a gun, given money to a charitable organization, gone to see a movie in a theater, written a letter to the editor of a newspaper. How many of these things have you done?" Seven hundred and forty three respondents were asked the same question with an additional fifth behavior: "Taken a vacation in the country of Tantatoula." Because Tantatoula does not exist, any increase in the numeric answers given to this list as compared to the shorter list can be attributed to list length alone. The average numbers of behaviors reported in response to the two lists were 1.77 and 1.86, respectively, which are not significantly different ($t(1,510) = 1.40$, n.s.), suggesting that list length itself does not cause illusory changes in responses.

between education and the list length dummy variable would indicate that the difference between the ICT conditions in the mean number of behaviors reported (which is the estimate of turnout) was larger among more educated respondents. So the interactions can be viewed as comparable to main effects of the demographics when predicting direct self-reports of turnout.

Consistent with past research (e.g., Campbell 1979; Rosenstone and Hansen 1993; Holbrook et al. 2001), the direct self-reports indicated that more educated respondents were more likely to have voted, as were older respondents (see columns 1 and 2 of table 3). However, the effect of age on direct self-reports was nonlinear: increasing age was strongly associated with increasing turnout early in the life cycle and more weakly later in life. These relations of directly reported turnout with education and age were evident in both the telephone and Internet data, though the coefficients were notably stronger in the telephone data. And in the telephone data, but not in the Internet data, whites were more likely to say that they had voted than did nonwhites (see columns 1 and 2 of table 3).

The interactions in the ICT data manifested similar patterns: more turnout with increasing education, more turnout with increasing age (less strongly late in the life cycle), and more turnout among whites than among nonwhites (see columns 3 and 4 of table 3). However, none of these effects were significant in the telephone data, and nonlinearity in the age effect was not significant in the Internet data, though the linear age effect, education effect, and race effect were significant in those data.

To see whether the significance tests with the telephone data were handicapped by the relatively small sample size, we conducted a simulation repeating the regression specifying a sample size of 5,574, to match the ICT Internet sample. As expected, the effects of education ($b = .09$, SE $= .05$, $p < .10$), age ($b = 1.33$, SE $= .49$, $p < .01$) and race ($b = .13$, SE $= .05$, $p < .05$) became marginally significant or significant, exactly as they were in the Internet ICT data. This suggests that the ICT yielded only one change in the demographic predictors of turnout: nonlinearity in the age relation disappeared.

Comparison to Official Turnout Estimates

Although it may seem reasonable to compare turnout estimates from each condition to the official turnout rate for the election, the telephone and Internet survey samples are not directly comparable to any official figures that can be calculated from publicly available data. For example, the RDD sample in Study 1 could have included noncitizens and convicted felons who were not in jail (who were not eligible to vote in some states). This would misleadingly depress the survey estimates of turnout. On the other hand, the survey samples did not include American citizens who were overseas and were eligible to vote, people who did not speak English, people living in institutional settings (e.g., college

Table 3. Effects of Demographic Variables on Turnout Reports

	Direct self-report		ICT	
Predictor	Telephone	Internet	Telephone	Internet
List condition			−.06	.02
			(.39)	(.09)
Education	**1.05***	**1.03****	.43**	.22**
	(.45)	(.09)	(.15)	(.04)
Age	**15.26****	**6.04****	2.14	2.39**
	(4.13)	(.81)	(1.31)	(.32)
Age^2	**−13.80****	**−2.82***	−2.84	−1.83**
	(5.29)	(1.11)	(1.74)	(.41)
White	**1.80****	.14	.56**	.09$^+$
	(.51)	(.12)	(.17)	(.05)
Hispanic	−.20	.01	−.53	−.14*
	(.95)	(.18)	(.40)	(.08)
Male	.26	.01	.43**	.24**
	(.45)	(.08)	(.15)	(.03)
Education × List condition			.09	**.15****
			(.22)	(.05)
Age × List condition			1.33	**1.74****
			(2.01)	(.45)
Age^2 × List condition			−.41	−.91
			(2.67)	(.60)
White × List condition			.13	**.11**$^+$
			(.27)	(.07)
Hispanic × List condition			−.04	.10
			(.57)	(.11)
Male × List condition			.08	.02
			(.21)	(.05)
Study				−1.38**
				(.03)
R^2			.29	.49
N	170	2,975	338	5,474

NOTE.—Coefficients from logistic regressions are shown in columns 1 and 2; coefficients from OLS regressions are shown in columns 3 and 4; standard errors are shown in parentheses. The variable "Study" was coded 1 for the Study 3 data and 0 for the data from Studies 2 and 4. This variable was included to control for the differences in list length across these studies.

$^+p < .10$ $^*p < .05$ $^{**}p < .01$.

dorms or group homes), and people without a working landline telephone, and these omissions could misleadingly inflate our estimates of turnout. Similar problems with sample composition also make comparing the Internet surveys' turnout estimates to official estimates problematic. Therefore, we should not expect a perfect match between the survey estimates and official numbers.

Nonetheless, it is interesting to note that official estimates suggest that 51.3 percent of Americans voted in the 2000 Presidential election, and the ICT in the telephone survey yielded a weighted turnout rate of 47.1 percent in Study 1, not significantly different from 51.3 percent ($z = .34$, ns).[21] This could be viewed as showing that eliminating social desirability response bias completely eliminated the discrepancy between the telephone survey result and official estimate.

However, considerable discrepancies existed between the Internet samples and the population figures. In contrast to the official turnout figures of 51.3 percent for the 2000 election and 37.4 percent for the 2002 election, the two KN survey ICTs yielded weighted turnout rates of 65.1 percent and 54.7 percent, respectively, both significantly larger than the official estimates ($z = 2.14$, $p < .05$ and $z = 5.30$, $p < .001$, respectively). Weighted estimated turnout from the ICT in the nonprobability Internet samples of Study 4 was 60.4 percent, again significantly larger than the official 51.3 percent figure ($z = 2.28$, $p < .01$).

Discussion

EFFECTIVENESS OF THE ITEM COUNT TECHNIQUE

The ICT reduced turnout estimates (relative to direct self-reports) when interviews were conducted via telephone (in Study 1). This is consistent with the hypothesis that social desirability response bias inflated direct self-reports in this mode. The ICT did not reduce turnout in the self-administered questionnaires completed via the Internet. This is consistent with the argument that social desirability response bias did not inflate direct self-reports in surveys conducted via the Internet (in Studies 2–4).

MODE DIFFERENCES

Our evidence of social desirability bias in telephone interviews but not in Internet questionnaire responses is consistent with the findings of other past mode comparison studies suggesting that social desirability response bias is more common in interviewer-administered surveys (e.g., telephone surveys)

21. Official turnout figures reported here are based on work by McDonald (2003; see http://elections.gmu.edu) and are most similar to the proportion of the voting age population (rather than the voting eligible population) who voted. We excluded prisoners from the denominator when estimating official turnout (assuming, as McDonald did, that all prisoners are felons) because they could not have voted, nor could they have been included in our samples. In order to permit comparisons to official rates, the survey numbers reported in this section were computed after weighting the sample for probability of selection and to match the CPS demographics shown in table 1. Tests of statistical significance compared the ICT's estimate of the proportion of people who voted with the proportion according to official estimates.

than in self-administered surveys (e.g., mail surveys; e.g., Tourangeau and Smith 1996; see Tourangeau, Rips, and Rasinski 2000 for a review). Our evidence contributes to a growing body of literature specifically comparing answers to sensitive questions in telephone and Internet surveys. For example, Chang and Krosnick (in press) showed that reports of socially desirable attitudes and behaviors were more prevalent in a national telephone survey than in a parallel national survey done via the Internet. Likewise, Chang and Krosnick (in press) showed in a laboratory experiment that socially desirable attitudes and behaviors were reported more often when respondents were interviewed orally by intercom than when completing the same questionnaire on a computer. As in most past research, these investigators presumed that the reduced prevalence of reports of socially desirable attributes under self-administration conditions was evidence of reduced social desirability pressures.

In the present research, the ICT provided such documentation. That is, the ICT removed respondents' motivation to intentionally misrepresent themselves in self-reports. The impact of the ICT on reports in the telephone interviews and the absence of such impact in the Internet data tie the mode difference directly to the reduction in social desirability pressures.[22]

Our conclusion that social desirability pressures are minimal in Internet surveys contrasts with findings reported by Tsuchiya, Hirai, and Ono (2007), who found social desirability-driven distortions of reports of shop lifting in an Internet survey. There are a number of possible explanations for the apparent inconsistency between our findings and Tsuchiya, Hirai, and Ono (2007) in this regard. First, shoplifting is socially *undesirable*, whereas voting is socially *desirable*. The mechanisms that result in underreporting of socially undesirable behaviors and those that result in overreporting of socially desirable behaviors may be different. Second, Tsuchiya, Hirai, and Ono's (2007) research was conducted with residents of Japan, whereas our studies were conducted with residents of the United States. Japanese culture is more collectivistic (e.g., more concerned with group membership, belonging, and interdependence) than U.S. culture (e.g., Hofstede 1980), and higher levels of collectivism are associated with more socially desirable responding, particularly for the purposes of impression management (e.g., Lalwani, Shavitt, and Johnson 2006). So although social desirability pressures may not influence answers to Internet surveys of Americans, they may cause distortions in answers from Japanese.

22. An alternative explanation for the apparent mode difference might seem to be different list length and/or differences in the non-sensitive behaviors inquired about in the ICT across studies. Tsuchiya, Hirai, and Ono (2007) suggested that the length of the list of behaviors used and the particular non-sensitive behaviors may influence the apparent impact of the ICT. However, list length and the non-sensitive behaviors were identical in Study 1 (involving telephone interviewing) and Studies 2 and 4 (involving internet data collection), although a shorter list with different behaviors was used in Study 3. Because we observed that the ICT was effective at reducing turnout reports in Study 1 but not in Studies 2, 3, or 4, list length and the prevalence of non-sensitive behaviors cannot explain the mode differences we observed.

PRIOR STUDIES OF SOCIAL DESIRABILITY AND TURNOUT

Many previous efforts to reduce social desirability response bias in turnout reports have done so in telephone surveys (Presser 1990; Abelson, Loftus, and Greenwald 1992). Yet these attempts consistently failed to reduce turnout reports. One might be tempted to infer that this evidence disconfirmed the social desirability hypothesis, and that might seem inconsistent with our documentation of social desirability bias in telephone interviews. But we see no inconsistency here. We suspect that Abelson, Loftus, and Greenwald (1992) and Presser (1990) failed to find evidence of reduced overreporting because their manipulations failed to reduce social desirability pressures, not because such pressures were absent in their surveys.

OVERESTIMATION OF TURNOUT IN THE INTERNET SURVEYS

The ICT administered by telephone yielded a turnout rate essentially identical to the official turnout rate for that election, which is encouraging about the accuracy of those reports. But the ICT technique's turnout estimates in the Internet surveys were considerably higher than the official turnout rates for the relevant elections. Because the ICT eliminated all incentive for respondents to lie intentionally, this overestimation is very unlikely to be attributable to social desirability response bias. Instead, we suspect, the samples of individuals who participated in the Internet surveys may in fact have voted at higher rates than the general public. This may be attributable to these respondents' participation in many preelection surveys about politics, or it may be attributable to the similarity of investing a little effort to express one's preferences in surveys and in voting booths, which may lead Internet survey samples to overrepresent actual voters. Whatever the reason, the ICT measurement of turnout for these samples may be accurate, reflecting this truly higher propensity to participate in elections. Or these reports may have overestimated turnout due to source confusion or other reporting errors.

PREDICTORS OF TURNOUT

Many analysts presume that because techniques such as the ICT do not yield precise measurements of the variable of interest for each respondent, this sort of approach cannot be used to explore the predictors of that variable. But as we demonstrated, it is possible to estimate the parameters of a regression equation using the ICT to identify such predictors. The predictors of turnout measured by direct self-reports and by the ICT were very similar, but nonlinearity in the effect of age was apparent in the direct self-report data and not in the ICT data.

Thus, perhaps the nonlinearity was due to shifts in social desirability motives across the life cycle, not differences in actual turnout.

AN ANALYTIC ADVANCE

Our study represents an advance beyond the existing ICT literature in terms of analytic methodology. In most past publications that compared ICT results to direct self-report results, the statistical analyses reported did not properly model all sources of error in the ICT measurements (e.g., Dalton, Wimbush, and Daily 1994, 1997; Wimbush and Dalton 1997; Cobb 2001; although see Tsuchiya, Hirai, and Ono 2007). Specifically, ICT assessments of behavior prevalence are influenced by error due to (1) random assignment of respondents to one of the two lists (short vs. long), and (2) variance in the prevalence of the nonsensitive behaviors. We properly incorporated this error by using the standard error of the difference between means for the two ICT conditions as the standard error for the ICT turnout estimate when estimating the confidence interval for the ICT turnout estimates. Because many past ICT studies have not done this, their results may cause scholars to underestimate the sample size necessary to obtain reasonably small confidence intervals with the ICT. We hope that future ICT studies will consistently employ proper computational methods.

Conclusion

The evidence reported here suggests that social desirability response bias is partly responsible for distorted turnout reports in telephone surveys, but social desirability response bias may not distort turnout reports in Internet surveys. These findings attest to the value of the item count technique for measuring attitudes and behaviors laced with social desirability implications and attest to the value of Internet surveys for achieving accurate measurement.

Appendix A: Methodological Details

STUDY 1

Procedures: Interviewing was done between June 15, 2004, and September 16, 2004. Of 6,990 initial phone numbers in the sample, reverse lookup procedures identified addresses for 2,518 of these numbers, and prenotification letters were sent to these addresses (36 percent of the sample), alerting recipients that an interviewer would be calling them. Up to twelve attempts were made to each number, and one refusal conversion attempt was made for each number if needed. Partway through the field period, another letter was mailed to 879 households who had not yet been reached to complete the survey and

for which addresses could be obtained. Letters were also sent to another 95 households that had refused to be interviewed and for whom addresses were available. The noncontact and refusal conversion letters offered a $10 incentive for completing the survey.

Measures: Gender was recorded by the interviewer and was coded 0 for women and 1 for men. Age was measured by asking respondents "In what year were you born?" and was coded to range from 0 to 1, with 0 meaning age 18 (the youngest age) and 1 meaning age 104 (the highest age). Education was measured by asking respondents an open-ended question: "What is the highest level of school you have completed or the highest degree you have received?" Interviewers recorded responses in one of the following categories: less than 1st grade; 1st grade; 2nd grade; 3rd grade; 4th grade' 5th grade; 6th grade; 7th grade; 8th grade; 9th grade; 10th grade; 11th grade; 12th grade with no diploma; high school diploma or an equivalent, such as a GED; some college but no degree; associate degree from an occupational/vocational program; associate degree from an academic program; bachelor's degree, such as B.A., B.S., or A.B.; master's degree, such as M.A., M.S., Masters in Engineering, Masters in Education, or Masters in Social Work; professional school degree, such as M.D., D.D.S., or D.V.M.; or doctorate degree, such as Ph.D. or Ed.D. Education was coded 0 for respondents with a high school education or less and 1 for respondents with at least some college education. Race was measured by asking respondents, "Which of the following races do you consider yourself to be: White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, or Other?" A variable called "White" was coded 1 for White respondents, and 0 for all other respondents. Respondents were also asked, "Are you Spanish, Hispanic, or Latino?", and a variable called "Hispanic" was coded 1 for people who answered affirmatively and zero for all others.

STUDY 2

Procedures: The data were collected by Knowledge Networks (KN). KN recruited panel members through random digit dialing (RDD) telephone interviewing. Before the initial telephone calls were made, households for which KN was able to recover a valid postal address were sent letters saying that they had been randomly selected to participate in the survey panel, they would not incur any cost, confidentiality was assured, and a KN staff member would call them within a week. During the telephone interview, respondents were told they had been selected to participate in an important national study. Households without Internet access were offered an Internet appliance and an Internet service connection in exchange for their participation in surveys. Potential panel

members who had access to the Internet were asked to use their own equipment and were given points for participation that could be redeemed for cash.

KN panel members were sent an e-mail inviting them to participate in each survey. Embedded in each e-mail was a hyperlink that took panel members directly to the questionnaire. Respondents could complete the questionnaire whenever they liked, and people could stop before completing it and return to it later.

E-mails inviting respondents to complete our survey were sent on June 18, 2004, and no responses were accepted after July 2, 2004.

Measures: Demographics were measured at the time of the survey. Gender was measured by asking, "Are you male or female?" and was coded 0 for women and 1 for men. Age was measured and coded as in Study 1. Education was measured by asking respondents: "What is the highest level of school you have completed or the highest degree you have received?" Education was coded as in Study 1. Race and Hispanic origin were measured using questions identical to that used in Study 1, and these variables were coded as in Study 1.

STUDY 3

Procedures: This study was also conducted by KN. E-mails inviting three groups of potential respondents to complete our survey were sent on November 15, November 20, and November 26, 2002, respectively, and no responses were accepted after December 5, 2002.

Measures: The demographics were measured when each person joined the KN panel. Gender and age were measured by asking respondents to "Please enter your age on your last birthday and whether you are male or female in the spaces below." Gender and age were coded as in Studies 1 and 2. Education was measured by asking respondents: "What is the highest degree or level of education that you have completed?" Respondents chose from the following list: less than high school; some high school, no diploma; graduated from high school—diploma or equivalent (GED); some college, no degree; associate degree (for example: AA, AS); bachelor's degree; master's degree; professional degree (for example: MD, DDS, LLB, JD); and doctorate degree. Education was coded as in Study 1. Race was measured by asking respondents to "Please check one or more categories below to indicate what race(s) you consider yourself to be: White, Black, African American or Negro, American Indian or Alaska Native, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian, Native Hawaiian, Guamanian or Chamorro, Samoan, Other Pacific Islander, or Some other race." A dummy variable "White" was coded 1 for white respondents and 0 for all other respondents. Hispanic origin was measured by asking respondents "Now we would

like to ask you about your Hispanic ethnicity. Are you of Spanish, Hispanic, or Latino descent?" Respondents who answered affirmatively were coded 1, and all others were coded 0.

STUDY 4

Data for Study 4 were collected via the Internet by six different companies using nonrepresentative panels of people who volunteered to do surveys. Each organization was asked to provide a survey sample that was representative of adults from the fifty U.S. states, and the methodologies used are described below. For organizations that maintained a panel of respondents, we computed the participation rate by dividing the number of panel members who fully or partially completed the questionnaire by the number of panel members who were invited to do so. This is comparable to AAPOR's cooperation rate 2.

Firm 1: Firm 1 collected data from members of a panel of approximately 2.2 million people. Panel members opted into the panel via the Internet and were recruited via many methods through Firm 1's affiliates, including text links in newsletters, banner ads, e-mail invitations, and word of mouth. On average, respondents had been in the panel for six months. They were invited to complete no more than one survey every week. Panel attrition was approximately 30 percent per year. Respondents were given $1 for completing a questionnaire that could be obtained in cash via Paypal or buy.com or could be used to pay for music downloads. Panel members were invited to participate in our survey in proportions matching quotas (reflecting Census estimates of the population) for household income, ethnicity, education, and gender. E-mails inviting 2,123 people to complete our questionnaire were sent on July 26, 2004, and no responses were accepted after August 1, 2004. A total of 1,129 (53.2 percent) of these people completed the questionnaire.

Firm 2: Firm 2 collected data via the Internet from members of a panel of over 1.7 million respondents. Panel members opted into the panel via the Internet after being recruited via embedded text links, editorial inclusion, targeted opt-in e-mail lists, word of mouth, co-registrations, and online promotions. On average, panelists had been in the panel for 18 months. They completed no more than one survey every two weeks and were invited to participate in one or two surveys per week. Panel attrition was approximately 32 percent per year. Respondents were given a chance to enter a weekly drawing for a $10,000 prize as compensation for questionnaire completion. Panel members were invited to complete our questionnaire in proportions matching quotas for gender, age, and region. E-mails inviting people to complete the survey were sent on June 11, 2004, and no responses were accepted after June 14,

2004. A total of 50,000 panel members were invited to participate, and 1,223 (2.4 percent) did so.

Firm 3: Firm 3 conducted surveys via the Internet with members of a panel of approximately 5 million panel members in the United States, who were recruited though Internet sign-ups. Panel members were typically invited to participate in two to three surveys per month. Respondents were given 100 points in a monthly $10,000 sweepstakes as compensation for completing questionnaires. Panel members were invited to participate in our survey in proportions matching quotas for age, gender, and region, with oversampling so that Hispanics and African Americans would participate at rates mirroring their presence in the U.S. population. E-mails inviting respondents to complete the questionnaire were sent on June 11, 2004, and no responses were accepted after June 21, 2004. E-mail invitations were sent to 11,530 veteran panel members and 45,014 new panel members (who had not completed a prior survey). 2,001 of the former group (17.4 percent) and 663 of the latter group (1.5 percent) completed our survey.

Firm 4: Firm 4 recruited respondents via advertisements on ISP Web sites inviting visitors to participate in a survey. Respondents who clicked on an ad were first asked a series of demographic screening questions. Based on this information, potential respondents were assigned to participate in one of a series of surveys via quota sampling in proportions to match the population on age, gender, income, and region. A respondent could participate in only two surveys per month and in only one survey every 90 days on a particular topic. Respondents were offered either a $4.50 credit on their monthly ISP bill or 300 frequent flier miles for completing a survey. A total of 1,137 respondents began doing the survey between June 16, 2004, and July 1, 2004, and 1,013 completed it (89 percent).

Firm 5: Data were collected from a panel of approximately 1.6 million members maintained by Firm 5. Panel members were recruited via the Internet, RDD invitations, referrals, and banner ads. Panel members completed one or two surveys per month, and they received no more than one or two e-mails per week inviting them to complete a survey. Panel attrition was approximately 20–25 percent per year, and people had been panel members for one year on average. Panel members were invited to complete our questionnaire in proportions matching 2001 CPS estimates for gender, age, and income. People were offered a chance to enter a monthly prize pool to win one of 114 prizes worth $10,000 in exchange or completing our questionnaire. E-mails inviting respondents to complete the questionnaire were sent on June 23, 2004, and no responses were accepted after June 30, 2004. Of the 9,921 panel members invited to participate, 1,103 (11.1 percent) did so.

Firm 6: Firm 6 collected data from a panel of 2.5 million potential respondents who had been recruited through more than 400 Web sites. Panel members were typically sent no more than four to six invitations to participate in survey per month. On average, people were on the panel for a total of 18 months. No incentive was offered for completing our survey. The sample of invited panel members was drawn to match the population in terms of age, gender, and geography. E-mails inviting respondents to complete the questionnaire were sent on August 25, 2004, and no responses were accepted after September 1, 2004. Survey Direct invited 14,000 panel members to participate, and 1,323 (9.5 percent) did so.

Measures: Question wordings and the coding of all variables were identical to those used in Study 2.

References

Aarts, Kees. 2002. "Electoral Turnout in West-European Democracies." Paper presented at the Annual Meeting of the American Political Science Association, Boston, MA, USA.

Abelson, Robert P., Elizabeth F. Loftus, and Anthony G. Greenwald. 1992. "Attempts to Improve the Accuracy of Self-Reports of Voting." In *Questions about Questions*, ed. Judith M. Tanur, pp. 138–53. New York: Russell Sage Foundation.

Ahart, Allison M., and Paul R. Sackett. 2004. "A New Method of Examining Relationships between Individual Difference Measures and Sensitive Behavior Criteria: Evaluating the Unmatched Count Technique." *Organizational Research Methods* 7:101–14.

Andolina, Molly, Scott Keeter, Cliff Zukin, and Krista Jenkins. 2003. "A Guide to the Index of Civic and Political Engagement." College Park, MD: The Center for Information and Research on Civic Learning and Engagement.

Belli, Robert F., Sean E. Moore, and John Van Hoewyk. 2006. "An Experimental Comparison of Question Forms Used to Reduce Vote Overreporting." *Electoral Studies* 25:751–9.

Belli, Robert F., Michael W. Traugott, and Matthew N. Beckmann. 2001. "What Leads to Vote Overreports? Contrasts of Overreporters to Validated Voters and Admitted Nonvoters in the American National Election Studies." *Journal of Official Statistics* 17:479–98.

Belli, Robert F., Michael W. Traugott, Margaret Young, and Katherine A. McGonagle. 1999. "Reducing Vote Over-Reporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring." *Public Opinion Quarterly* 63:90–108.

Belli, Robert F., Santa Traugott, and Steven J. Rosenstone. 1994. "Reducing Over-Reporting of Voter Turnout: An Experiment Using a Source Monitoring Framework." NES Technical Reports Number 35 (see available technical reports at http://www.umich.edu/~nes/).

Bernstein, Robert, Anita Chadha, and Robert Montjoy. 2001. "Overreporting Voting: Why It Happens and Why It Matters." *Public Opinion Quarterly* 65:22–44.

Blais, André, Elisabeth Gidengil, Neil Nevitte, and Richard Nadeau. 2004. "Where Does Turnout Decline Come from?" *European Journal of Political Research* 43:221–36.

Brockington, David, and Jeffrey Karp. 2002. "Social Desirability and Response Validity: A Comparative Analysis of Over-Reporting Turnout in Five Countries." Paper presented at the Annual Meeting of the American Political Science Association, Boston, MA, USA.

Burden, Barry C. 2000. "Voter Turnout and the National Election Studies." *Political Analysis* 8:389–98.

Campbell, Bruce A. 1979. *The American Electorate: Attitudes and Action*. New York: Holt Rinehart and Winston.

Chang, LinChiat, and Jon A. Krosnick. "National surveys via RDD telephone interviewing vs. the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly* doi:10.1093/poq/nfp075.

Clausen, Aage. 1968. "Response Validity: Vote Report." *Public Opinion Quarterly* 32:588–606.

Cobb, Michael D. 2001. *Unobtrusively Measuring Racial Attitudes: The Consequences of Social Desirability Effects*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Corbett, Michael. 1991. *American Public Opinion*. White Plains, NY: Longman.

Corstange, Daniel. 2006. "Sensitive Questions, Truthful Answers?: Modeling the List Experimental Multivariately with LISTIT." *Political Analysis* 17:45–63.

Dalton, Dan R., Catherine M. Daily, and James C. Wimbush. 1997. "Collecting 'Sensitive' Data in Business Ethics Research: A Case for the Unmatched Count Technique (UCT)." *Journal of Business Ethics* 16:1049–57.

Dalton, Dan R., James C. Wimbush, and Catherine M. Daily. 1994. "Using the Unmatched Count Technique (UCT) to Estimate Base Rates for Sensitive Behavior." *Personnel Psychology* 47:817–28.

Demaio, Theresa J. 1984. "Social Desirability and Survey Measurement: A Review." In *Surveying Subjective Phenomena*, eds. Charles F. Turner and Elizabeth Martin, vol. 2, pp. 257–82. New York: Russell Sage Foundation.

Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, Teresa L. Parsley, Wendy Visscher, and Trena M. Ezzati. 1991. "The Item Count Technique as a Method of Indirect Questioning: A Review of Its Development and a Case Study Application." In *Measurement Errors in Surveys*, eds. Paul B. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, pp. 185–210. New York: Wiley.

Evans, Richard I., William B. Hansen, and Maurice B. Mittlemark. 1977. "Increasing the Validity of Self-Reports of Smoking Behavior in Children." *Journal of Applied Psychology* 62: 521–3.

Greenwald, Anthony G., Catherine G. Carot, Rebecca Beach, and Barbara Young. 1987. "Increasing Voting Behavior by Asking People if They Expect to Vote." *Journal of Applied Psychology* 72:315–8.

Greenwald, Anthony G., Mark R. Klinger, Mark E. Vande Kamp, and K. L. Kerr. 1988. "The Self-Prophecy Effect: Increasing Voter Turnout by Vanity-Assisted Consciousness Raising." Unpublished manuscript, University of Washington.

Himmelfarb, Samuel, and Carl Lickteig. 1982. "Social Desirability and the Randomized Response Technique." *Journal of Personality and Social Psychology* 43:710–7.

Hofstede, Geert. 1980. *Cultures Consequences: International Differences in Work-Related Values*. Newbury Park, CA: Sage.

Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone vs. Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias." *Public Opinion Quarterly* 67:79–125.

Holbrook, Allyson L., Jon A. Krosnick, Penny S. Visser, Wendi L. Gardner, and John T. Cacioppo. 2001. "Attitudes toward Presidential Candidates and Political Parties: Initial Optimism, Inertial First Impressions, and a Focus on Flaws." *American Journal of Political Science* 45:930–50.

Knowledge Networks. 2006. "Knowledge Networks Remains Only Online Research Firm to Deliver Six Elements of Accuracy." Press release available at: http://www.knowledgenetworks.com/news/releases/2006/100406_knpanel.htm.

Kraut, Robert E., and John B. McConahay. 1973. "How Being Interviewed Affects Voting: An Experiment." *Public Opinion Quarterly* 37:398–406.

Kuklinski, James H., and Michael D. Cobb. 1998. "When White Southerners Converse about Race." In *Perception and Prejudice*, eds. Jon Hurwitz and Mark Peffley. New Haven: Yale University Press.

Kuklinski, James H., Michael D. Cobb, and Martin Gilens. 1997. "Racial Attitudes and the 'New South.'" *Journal of Politics* 59:323–49.

Kuklinski, James H., Paul M. Sniderman, Kathleen Knight, Thomas Piazza, Philip E. Tetlock, Gordon R. Lawrence, and Barbara Mellers. 1996. "Racial Prejudice and Attitudes toward Affirmative Action." *American Journal of Political Science* 41:402–19.

LaBrie, Joseph W., and Mitchell Earleywine. 2000. "Sexual Risk Behaviors and Alcohol: Higher Base Rates Revealed Using the Unmatched-Count Technique." *Journal of Sex Research* 37:321–6.

Lalwani, Ashok K., Sharon Shavitt, and Timothy Johnson. 2006. "What Is the Relation between Cultural Orientation and Socially Desirable Responding?" *Journal of Personality and Social Psychology* 90:165–78.

Lutz, George. 2003. "Participation, Cognitive Involvement, and Democracy: When Do Low Turnout and Low Cognitive Involvement Make a Difference, and Why?" Paper presented at the European Consortium for Political Research Joint Sessions of Workshops, Edinburgh, UK.

Lyons, William, and John M. Scheb. 1999. "Early Voting and the Timing of the Vote: Unanticipated Consequences of Electoral Reform." *State and Local Government Review* 31:147–52.

Mann, Christopher B. 2005. "Unintentional Voter Mobilization: Does Participation in Pre-election Surveys Increase Voter Turnout?" *Annals of the American Academy of Political and Social Science* 601:155–68.

Martinez, Michael D. 2003. "Comment on 'Voter Turnout and the National Election Studies.'" *Political Analysis* 11:187–92.

McDonald, Michael P. 2003. "On the Over-Report Bias of the National Election Study Turnout Rate." *Political Analysis* 11:180–6.

McDonald, Michael P., and Samuel L. Popkin. 2001. "The Myth of the Vanishing Voter." *American Political Science Review* 95:963–74.

Miller, Judith D. 1984. *A New Survey Technique For Studying Deviant Behavior*. Unpublished doctoral dissertation, George Washington University.

Miller, Judith D., Adele V. Harrel, and Ira A. Cisin. 1986. "A New Technique for Surveying Deviant Behavior: Item-Count Estimates of Marijuana, Cocaine, and Heroin." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Petersburg, FL, USA.

Paulhus, Delroy L. 1984. "Two-Component Models of Socially Desirable Responding." *Journal of Personality and Social Psychology* 46:598–609.

Pavlos, Andrew J. 1972. "Racial Attitude and Stereotype Change with Bogus Pipeline Paradigm." Proceedings of the 80th Annual Convention of the American Psychological Association, vol. 7, 292.

Presser, Stanley. 1990. "Can Context Changes Reduce Vote Over-Reporting?" *Public Opinion Quarterly* 54:586–93.

Presser, Stanley, Michael W. Traugott, and Santa Traugott. 1990. "Vote 'Over' Reporting in Surveys: The Records or the Respondents?" Paper presented at the International Conference on Measurement Errors, Tucson, AZ, USA.

Rayburn, Nadine Recker, Mitchell Earleywine, and Gerald C. Davison. 2003a. "An Investigation of Base Rates of Anti-gay Hate Crimes Using the Unmatched-Count Technique." *Journal of Aggression, Maltreatment and Trauma* 6:137.

Rayburn, Nadine Recker, Mitchell Earleywine, and Gerald C. Davison. 2003b. "Base Rates of Hate Crime Victimization Among College Students." *Journal of Interpersonal Violence* 18: 1209–21.

Rosenstone, Steven J., and John Mark Hansen. 1993. *Mobilization, Participation, and Democracy in America*. New York: Macmillan.

Sigall, Harold, and Richard Page. 1971. "Current Stereotypes: A Little Fading, a Little Faking." *International Journal of Public Opinion Research* 7:157–71.

Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson. 1986. "Who Overreports Voting?" *American Political Science Review* 80:613–24.

Smith, Jennifer K., Alan S. Gerber, and Anton Orlich. 2003. "Self-Prophecy Effects and Voter Turnout: An Experimental Replication." *Political Psychology* 24:593–604.

Sniderman, Paul M., and Douglas B. Grob. 1996. "Innovations in Experimental Design in Attitude Surveys." *Annual Review of Sociology* 22:377–99.

Tourangeau, Roger, Lance J. Rips, and Kenneth A. Rasinski. 2000. *The Psychology of Survey Response.* Cambridge: Cambridge University Press.

Tourangeau, Roger, and Tom W. Smith. 1996. "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context." *Public Opinion Quarterly* 60:275–304.

Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133:859–83.

Traugott, Michael W., and John P. Katosh. 1979. "Response Validity in Surveys of Voting Behavior." *Public Opinion Quarterly* 43:359–77.

Traugott, Santa. 1989. "Validating Self-Reported Vote: 1964–1988." Paper presented at the Annual Meeting of the American Statistical Association, Washington, DC, USA.

Tsuchiya, Takahiro, Yoko Hirai, and Shigeru Ono. 2007. "A Study of the Properties of the Item Count Technique." *Public Opinion Quarterly* 71:253–72.

Warner, Stanley L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60:63–9.

Wimbush, Dan C., and Donald R. Dalton. 1997. "Base Rate for Employee Theft: Convergence of Multiple Methods." *Journal of Applied Psychology* 82:756–63.

Yalch, Richard F. 1976. "Pre-election Interview Effects on Voter Turnout." *Public Opinion Quarterly* 40:331–6.