Sensitive Questions in Surveys

Roger Tourangeau
University of Maryland and University of Michigan

Ting Yan
University of Michigan

Psychologists have worried about the distortions introduced into standardized personality measures by social desirability bias. Survey researchers have had similar concerns about the accuracy of survey reports about such topics as illicit drug use, abortion, and sexual behavior. The article reviews the research done by survey methodologists on reporting errors in surveys on sensitive topics, noting parallels and differences from the psychological literature on social desirability. The findings from the survey studies suggest that misreporting about sensitive topics is quite common and that it is largely situational. The extent of misreporting depends on whether the respondent has anything embarrassing to report and on design features of the survey. The survey evidence also indicates that misreporting on sensitive topics is a more or less motivated process in which respondents edit the information they report to avoid embarrassing themselves in the presence of an interviewer or to avoid repercussions from third parties.

Keywords: sensitive questions, mode effects, measurement error, social desirability

Over the last 30 years or so, national surveys have delved into increasingly sensitive topics. To cite one example, since 1971, the federal government has sponsored a series of recurring studies to estimate the prevalence of illicit drug use, originally the National Survey of Drug Abuse, later the National Household Survey of Drug Abuse, and currently the National Survey on Drug Use and Health. Other surveys ask national samples of women whether they have ever had an abortion or ask samples of adults whether they voted in the most recent election. An important question about such surveys is whether respondents answer the questions truthfully. Methodological research on the accuracy of reports in surveys about illicit drug use and other sensitive topics, which we review in this article, suggests that misreporting is a major source of error, more specifically of bias, in the estimates derived from these surveys. To cite just one line of research, Tourangeau and Yan (in press) reviewed studies that compared self-reports about illicit drug use with results from urinalyses and found that some 30%–70% of those who test positive for cocaine or opiates deny having used drugs recently. The urinalyses have very low false positive rates (see, e.g., Wish, Hoffman, & Nemes, 1997), so those deniers who test positive are virtually all misreporting.

Most of the studies on the accuracy of drug reports involve surveys of special populations (such as enrollees in drug treatment programs or arrestees in jail), but similar results have been found for other sensitive topics in samples of the general population. For instance, one study compared survey reports about abortion from respondents to the National Survey of Family Growth (NSFG) with data from abortion clinics (Fu, Darroch, Henshaw, & Kolb, 1998). The NSFG reports were from a national sample of women between the ages of 15 and 44. Both the survey reports from the NSFG and the provider reports permit estimates of the total number of abortions performed in the U.S. during a given year. The results indicated that only about 52% of the abortions are reported in the survey. (Unlike the studies on drug reporting, the study by Fu et al., 1998, compared the survey reports against more accurate data in the aggregate rather than at the individual level.) Another study (Belli, Traugott, & Beckmann, 2001) compared individual survey reports about voting from the American National Election Studies with voting records; the study found that more than 20% of the nonvoters reported in the survey that they had voted. National surveys are often designed to yield standard errors (reflecting the error due to sampling) that are 1% of the survey estimates or less. Clearly, the reporting errors on these topics produce biases that are many times larger than that and may well be the major source of error in the national estimates.

In this article, we summarize the main findings regarding sensitive questions in surveys. In the narrative portions of our review, we rely heavily on earlier attempts to summarize the vast literature on sensitive questions (including several meta-analyses). We also report three new meta-analyses on several topics for which the empirical evidence is somewhat mixed. Sensitive questions is a broad category that encompasses not only questions that trigger social desirability concerns but also those that are seen as intrusive by the respondents or that raise concerns about the possible repercussions of disclosing the information. Possessing cocaine is not just socially undesirable; it is illegal, and people may misreport in a drug survey to avoid legal consequences rather than merely to avoid creating an unfavorable impression. Thus, the first issue we deal with is the concept of sensitive questions and its relation to the
concept of social desirability. Next, we discuss how survey respondents seem to cope with such questions. Apart from misreporting, those who are selected for a survey on a sensitive topic can simply decline to take part in the survey (assuming they know what the topic is), or they can take part but refuse to answer the sensitive questions. We review the evidence on the relation between question sensitivity and both forms of nonresponse as well as the evidence on misreporting about such topics. Several components of the survey design seem to affect how respondents deal with sensitive questions; this is the subject of the third major section of the article. On the basis of this evidence, we discuss the question of whether misreporting in response to sensitive questions is deliberate and whether the process leading to such misreports is nonetheless partly automatic. Because there has been systematic research on the reporting of illicit drug use to support the design of the federal drug surveys, we draw heavily on studies on drug use reporting in this review, supplementing these findings with related findings on other sensitive topics.

What Are Sensitive Questions?

Survey questions about drug use, sexual behaviors, voting, and income are usually considered sensitive; they tend to produce comparatively higher nonresponse rates or larger measurement error in responses than questions on other topics. What is it about these questions that make them sensitive? Unfortunately, the survey methods literature provides no clear answers. Tourangeau, Rips, and Rasinski (2000) argued that there are three distinct meanings of the concept of “sensitivity” in the survey literature.

Intrusiveness and the Threat of Disclosure

The first meaning of the term is that the questions themselves are seen as intrusive. Questions that are sensitive in this sense touch on “taboo” topics, topics that are inappropriate in everyday conversation or out of bounds for the government to ask. They are seen as an invasion of privacy, regardless of what the correct answer for the respondent is. This meaning of sensitivity is largely determined by the content of the question rather than by situational factors such as where the question is asked or to whom it is addressed. Questions asking about income or the respondent’s religion may fall into this category; respondents may feel that such questions are simply none of the researcher’s business. Questions in this category risk offending all respondents, regardless of their status on the variable in question.

The second meaning involves the threat of disclosure, that is, concerns about the possible consequences of giving a truthful answer should the information become known to a third party. A question is sensitive in this second sense if it raises fears about the likelihood or consequences of disclosure of the answers to agencies or individuals not directly involved in the survey. For example, a question about use of marijuana is sensitive to teenagers when their parents might overhear their answers, but it is not so sensitive when they answer the same question in a group setting with their peers. Respondents vary in how much they worry about the confidentiality of their responses, in part based on whether they have anything to hide. In addition, even though surveys routinely offer assurances of confidentiality guaranteeing nondisclosure, survey respondents do not always seem to believe these assurances, so concerns about disclosure may still be an important factor in the misreporting of illegal or socially undesirable behaviors (Singer & Presser, in press; but see also Singer, von Thurn, & Miller, 1995).

Sensitivity and Social Desirability

The last meaning of question sensitivity, closely related to the traditional concept of social desirability, is the extent to which a question elicits answers that are socially unacceptable or socially undesirable (Tourangeau et al., 2000). This conception of sensitivity presupposes that there are clear social norms regarding a given behavior or attitude; answers reporting behaviors or attitudes that conform to the norms are deemed socially desirable, and those that report deviations from the norms are considered socially undesirable. For instance, one general norm is that citizens should carry out their civic obligations, such as voting in presidential elections. As a result, in most settings, admitting to being a nonvoter is a socially undesirable response. A question is sensitive when it asks for a socially undesirable answer, when it asks, in effect, that the respondent admit he or she has violated a social norm. Sensitivity in this sense is largely determined by the respondents’ potential answers to the survey question; a question about voting is not sensitive for a respondent who voted. Social desirability concerns can be seen as a special case of the threat of disclosure, involving a specific type of interpersonal consequence of revealing information in a survey—social disapproval.

The literature on social desirability is voluminous, and it features divergent conceptualizations and operationalizations of the notion of socially desirable responding (DeMaio, 1984). One fundamental difference among the different approaches lies in whether they treat socially desirable responding as a stable personality characteristic or a temporary social strategy (DeMaio, 1984). The view that socially desirable responding is, at least in part, a personality trait underlies psychologists’ early attempts to develop various social desirability scales. Though some of these efforts (e.g., Edwards, 1957; Philips & Clancy, 1970, 1972) recognize the possibility that social desirability is a property of the items rather than (or as well as) of the respondents, many of them treat socially desirable responding as a stable personality characteristic (e.g., Crowne & Marlowe, 1964; Schuessler, Hittle, & Cardascia, 1978). By contrast, survey researchers have tended to view socially desirable responding as a response strategy reflecting the sensitivity of specific items for specific individuals; thus, Sudman and Bradburn (1974) had interviewers rate the social

---

1 In addition, the relevant norms may vary across social classes or subcultures within a society. T. Johnson and van der Vijver (2002) provided a useful discussion of cultural differences in socially desirable responding. When there is such variation in norms, the bias induced by socially desirable responding may distort the observed associations between the behavior in question and the characteristics of the respondents, besides affecting estimates of overall means or proportions. For instance, the norm of voting is probably stronger among those with high levels of education than among those with less education. As a result, highly educated respondents are both more likely to vote and more likely to misreport if they did not vote than are respondents with less education. This differential misreporting by education will yield an overestimate of the strength of the relationship between education and voting.
The desirability of potential answers to specific survey questions. Paulhus's (2002) work encompasses both viewpoints, making a distinction between socially desirable responding as a response style (a bias that is “consistent across time and questionnaires”); Paulhus, 2002, p. 49) and as a response set (a short-lived bias “attributable to some temporary distraction or motivation”); Paulhus, 2002, p. 49).

A general weakness with scales designed to measure socially desirable responding is that they lack “true” scores, making it difficult or impossible to distinguish among (a) respondents who are actually highly compliant with social norms, (b) those who have a sincere but inflated view of themselves, and (c) those who are deliberately trying to make a favorable impression by falsely reporting positive things about themselves. Bradburn, Sudman, and Associates (1979, see chap. 6) argued that the social desirability scores derived from the Marlowe–Crowne (MC) items (Crowne & Marlowe, 1964) largely reflect real differences in behaviors, or the first possibility we distinguish above:

> We consider MC scores to indicate personality traits . . . MC scores [vary] . . . not because respondents are manipulating the image they present in the interview situation, but because persons with high scores have different life experiences and behave differently from persons with lower scores. (p. 103)

As an empirical matter, factor analyses of measures of socially desirable responding generally reveal two underlying factors, dubbed the Alpha and Gamma factors by Wiggins (Wiggins, 1964; see Paulhus, 2002, for a review). Paulhus’ early work (Paulhus, 1984) reflected these findings, dividing social desirability into two components: self-deception (corresponding to Wiggins’ Alpha factor and to the second of the possibilities we distinguish above) and impression management (corresponding to Gamma and the third possibility above). (For related views, see Messick, 1991; Sackheim & Gur, 1978.) The Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1984) provides separate scores for the two components.

Paulhus’s (2002) later work went even further, distinguishing four forms of socially desirable responding. Two of them involve what he calls “egoistic bias” (p. 63), or having an inflated opinion of one’s social and intellectual status. This can take the form of self-deceptive enhancement (that is, sincerely, but erroneously, claiming positive characteristics for oneself) or more strategic agency management (bragging or self-promotion). The other two forms of socially desirable responding are based on “moralistic bias” (p. 63), an exaggerated sense of one’s moral qualities. According to Paulhus, self-deceptive denial is a relatively unconscious tendency to deny one’s faults; communion (or relationship) management is more strategic, involving deliberately minimizing one’s mistakes by making excuses and executing other damage control maneuvers. More generally, Paulhus (2002) argued that we tailor our impression management tactics to our goals and to the situations we find ourselves in. Self-promotion is useful in landing a job; making excuses helps in avoiding conflicts with a spouse.

**Assessing the Sensitivity of Survey Items**

How have survey researchers attempted to measure socially desirability or sensitivity more generally? In an early attempt, Sudman and Bradburn (1974) asked coders to rate the social desirability of the answers to each of a set of survey questions on a 3-point scale (no possibility, some possibility, or a strong possibility of a socially desirable answer). The coders did not receive detailed instructions about how to do this, though they were told to be “conservative and to code ‘strong possibility’ only for questions that have figured prominently in concern over socially desirable answers” (Sudman & Bradburn, 1974, p. 43). Apart from its vagueness, a drawback of this approach is its inability to detect any variability in the respondents’ assessments of the desirability of the question content (see DeMaio, 1984, for other criticisms).

In a later study, Bradburn et al. (1979) used a couple of different approaches to get direct respondent ratings of the sensitivity of different survey questions. They asked respondents to identify questions that they felt were “too personal.” They also asked respondents whether each question would make “most people” very uneasy, somewhat uneasy, or not uneasy at all. They then combined responses to these questions to form an “acute anxiety” scale, used to measure the degree of threat posed by the questions. Questions about having a library card and voting in the past election fell at the low end of this scale, whereas questions about bankruptcy and traffic violations were at the high end. Bradburn et al. subsequently carried out a national survey that asked respondents to judge how uneasy various sections of the questionnaire would make most people and ranked the topics according to the percentage of respondents who reported that most people would be “very uneasy” or “moderately uneasy” about the topic. The order was quite consistent with the researchers’ own judgments, with masturbation rated the most disturbing topic and sports activities the least (Bradburn et al., 1979, Table 5).

In developing social desirability scales, psychological researchers have often contrasted groups of participants given different instructions about how to respond. In these studies, the participants are randomly assigned to one of two groups. Members of one group are given standard instructions to answer according to whether each statement applies to them; those in the other group are instructed to answer in a socially desirable fashion regardless of whether the statement actually applies to them (that is, they are told to “fake good”; Wiggins, 1959). The items that discriminate most sharply between the two groups are considered to be the best candidates for measuring social desirability. A variation on this method asks one group of respondents to “fake bad” (that is, to answer in a socially undesirable manner) and the other to “fake good.” Holbrook, Green, and Krosnick (2003) used this method to identify five items prone to socially desirable responding in the 1982 American National Election Study.

If the consequences of sensitivity were clear enough, it might be possible to measure question sensitivity indirectly. For example, if sensitive questions consistently led to high item nonresponse rates, this might be a way to identify sensitive questions. Income questions and questions about financial assets are usually considered to be sensitive, partly because they typically yield very high rates of missing data—rates as high as 20%–40% have been found across surveys (Juster & Smith, 1997; Moore, Stinson, & Welniak, 1999). Similarly, if sensitive questions trigger a relatively controlled process in which respondents edit their answers, it should take more time to answer sensitive questions than equally demanding nonsensitive ones. Holgraves (2004) found that response times got longer when the introduction to the questions heightened social desirability concerns, a finding consistent with the operation of an
editing process. Still, other factors can produce high rates of missing data (see Beatty & Herrmann, 2002, for a discussion) or long response times (Bassili, 1996), so these are at best indirect indicators of question sensitivity.

Consequences of Asking Sensitive Questions

Sensitive questions are thought to affect three important survey outcomes: (a) overall, or unit, response rates (that is, the percentage of sample members who take part in the survey), (b) item nonresponse rates (the percentage of respondents who agree to participate in the survey but who decline to respond to a particular item), (c) and response accuracy (the percentage of respondents who answer the questions truthfully). Sensitive questions are suspected of causing problems on all three fronts, lowering overall and item response rates and reducing accuracy as well.

Unit Response Rates

Many survey researchers believe that sensitive topics are a serious hurdle for achieving high unit response rates (e.g., Catania, Gibson, Chitwood, & Coates, 1990), although the evidence for this viewpoint is not overwhelming. Survey texts often recommend that questionnaire designers keep sensitive questions to the end of a survey so as to minimize the risk of one specific form of nonresponse—break-offs, or respondents quitting the survey part way through the questionnaire (Sudman & Bradburn, 1982).

Despite these beliefs, most empirical research on the effects of the topic on unit response rates has focused on topic saliency or interest rather than topic sensitivity (Groves & Couper, 1998; Groves, Presser, & Dipko, 2004; Groves, Singer, & Corning, 2000). Meta-analytic results point to topic saliency as a major determinant of response rates (e.g., Heberlein & Baumgartner, 1978); only one study (Cook, Heath, & Thompson, 2000) has attempted to isolate the effect of topic sensitivity. In examining response rates to Web surveys, Cook et al. (2000) showed that topic sensitivity was negatively related to response rates, but, though the effect was in the expected direction, it was not statistically significant.

Respondents may be reluctant to report sensitive information in surveys partly because they are worried that the information may be accessible to third parties. Almost all the work on concerns about confidentiality has examined attitudes toward the U.S. census. For example, a series of studies carried out by Singer and her colleagues prior to Census 2000 suggests that many people have serious misconceptions about how the census data are used (Singer, Van Hoewyk, & Neugebauer, 2003; Singer, Van Hoewyk, & Tourangeau, 2001). Close to half of those surveyed thought that other government agencies had access to names, addresses, and other information gathered in the census. (In fact, the confidentiality of the census data as well as the data collected in other federal surveys is strictly protected by law; data collected in other surveys may not be so well protected.) People with higher levels of concern about the confidentiality of the census data were less likely to return their census forms in the 1990 and 2000 censuses (Couper, Singer, & Kulka, 1998; Singer et al., 2003; Singer, Mathiowetz, & Couper, 1993). Further evidence that confidentiality concerns can affect willingness to respond at all comes from experiments conducted by the U.S. Census Bureau that asked respondents to provide their Social Security Numbers in a mail survey; this lowered the unit response rate and raised the level of missing data among those who did mail back the questionnaire (Dillman, Sinclair, & Clark, 1993; see also Guarino, Hill, & Woltman, 2001). Similarly, an experiment by Junn (2001) showed that when confidentiality issues were made salient by questions about privacy, respondents were less likely to answer the detailed questions on the census long form, resulting in a higher level of missing data.

Thus, concerns about confidentiality do seem to contribute both to unit and item nonresponse. To address these concerns, most federal surveys include assurances about the confidentiality of the data. A meta-analysis by Singer et al. (1995) indicated that these assurances generally boost overall response rates and item response rates to the sensitive questions (see also Berman, McCombs, & Boruch, 1977). Still, when the data being requested are not all that sensitive, elaborate confidentiality assurances can backfire, lowering overall response rates (Singer, Hippler, & Schwarz, 1992).

Item Nonresponse

Even after respondents agree to participate in a survey, they still have the option to decline to answer specific items. Many survey researchers believe that the item nonresponse rate increases with question sensitivity, but we are unaware of any studies that systematically examine this hypothesis. Table 1 displays item nonresponse rates for a few questions taken from the NSFG Cycle 6 Female Questionnaire. The items were administered to a national sample of women who were from 15 to 44 years old. Most of the items in the table are from the interviewer-administered portion of the questionnaire; the rest are from the computer-administered portion. This is a survey that includes questions thought to vary widely in sensitivity, ranging from relatively innocuous sociodemographic items to detailed questions about sexual behavior. It seems apparent from the table that question sensitivity has some positive relation to item nonresponse rates: The lowest rate of missing data is for the least sensitive item (on the highest grade completed), and the highest rate is for total income question. But the absolute differences across items are not very dramatic, and lacking measures of the sensitivity of each item, it is difficult to assess the strength of the overall relationship between item nonresponse and question sensitivity.

Table 1

<table>
<thead>
<tr>
<th>Item</th>
<th>Mode of administration</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total household income</td>
<td>ACASI</td>
<td>8.15</td>
</tr>
<tr>
<td>No. of lifetime male sexual partners</td>
<td>CAPI</td>
<td>3.05</td>
</tr>
<tr>
<td>Received public assistance</td>
<td>ACASI</td>
<td>2.22</td>
</tr>
<tr>
<td>No. of times had sex in past 4 weeks</td>
<td>CAPI</td>
<td>1.37</td>
</tr>
<tr>
<td>Age of first sexual intercourse</td>
<td>CAPI</td>
<td>0.87</td>
</tr>
<tr>
<td>Blood tested for HIV</td>
<td>CAPI</td>
<td>0.65</td>
</tr>
<tr>
<td>Age of first menstrual period</td>
<td>CAPI</td>
<td>0.39</td>
</tr>
<tr>
<td>Highest grade completed</td>
<td>CAPI</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Note. ACASI = audio computer-assisted self-interviewing; CAPI = computer-assisted personal interviewing.
As we noted, the item nonresponse rate is the highest for the total income question. This is quite consistent with prior work (Juster & Smith, 1997; Moore et al., 1999) and will come as no surprise to survey researchers. Questions about income are widely seen as very intrusive; in addition, some respondents may not know the household’s income.

Response Quality

The best documented consequence of asking sensitive questions in surveys is systematic misreporting. Respondents consistently underreport some behaviors (the socially undesirable ones) and consistently overreport others (the desirable ones). This can introduce large biases into survey estimates.

Underreporting of socially undesirable behaviors appears to be quite common in surveys (see Tourangeau et al., 2000, chap. 9, for a review). Respondents seem to underreport the use of illicit drugs (Fendrich & Vaughn, 1994; L. D. Johnson & O’Malley, 1997), the consumption of alcohol (Duffy & Waterton, 1984; Lemmens, Tan, & Knibbe, 1992; Locander, Sudman & Bradburn, 1976), smoking (Bauman & Dent, 1982; Murray, O’Connell, Schmid, & Perry, 1987; Patrick et al., 1994), abortion (E. F. Jones & Forrest, 1992), bankruptcy (Locander et al., 1976), energy consumption (Warinner, McDougall, & Claxton, 1984), certain types of income (Moore et al., 1999), and criminal behavior (Wynner, 1980). They underreport racist attitudes as well (Krysan, 1998; see also Devine, 1989). By contrast, there is somewhat less evidence for overreporting of socially desirable behaviors in surveys. Still, overreporting has been found for reports about voting (Belli et al., 2001; Locander et al., 1976; Parry & Crossley, 1950; Traugott & Katosh, 1979), energy conservation (Fuji, Hennessy, & Mak, 1985), seat belt use (Stulginskas, Verreault, & Pless, 1985), having a library card (Locander et al., 1976; Parry & Crossley, 1950), church attendance (Presser & Stinson, 1998), and exercise (Tourangeau, Smith, & Rasinski, 1997). Many of the studies documenting underreporting or overreporting are based on comparisons of survey reports with outside records (e.g., Belli et al., 2001; E. F. Jones & Forrest, 1992; Locander et al., 1976; Moore et al., 1999; Parry & Crossley, 1950; Traugott & Katosh, 1979; Wynner, 1980) or physical assays (Bauman & Dent, 1982; Murray et al., 1987). From their review of the empirical findings, Tourangeau et al. (2000) concluded that response quality suffers as the topic becomes more sensitive and among those who have something to hide but that it can be improved by adopting certain design strategies. We review these strategies below.

Factors Affecting Reporting on Sensitive Topics

Survey researchers have investigated methods for mitigating the effects of question sensitivity on nonresponse and reporting error for more than 50 years. Their findings clearly indicate that several variables can reduce the effects of question sensitivity, decreasing item nonresponse and improving the accuracy of reporting. The key variables include the mode of administering the questions (especially whether or not an interviewer asks the questions), the data collection setting and whether other people are present as the respondent answers the questions, and the wording of the questions. Most of the studies we review below compare two or more different methods for eliciting sensitive information in a survey. Many of these studies lack validation data and assume that whichever method yields more reports of the sensitive behavior is the more accurate method; survey researchers often refer to this as the “more is better” assumption. Although this assumption is often plausible, it is still just an assumption.

Mode of Administration

Surveys use a variety of methods to collect data from respondents. Traditionally, three methods have dominated: face-to-face interviews (in which interviewers read the questions to the respondents and then record their answers on a paper questionnaire), telephone interviews (which also feature oral administration of the questions by interviewers), and mail surveys (in which respondents complete a paper questionnaire, and interviewers are not typically involved at all). This picture has changed radically over the last few decades as new methods of computer administration have become available and been widely adopted. Many national surveys that used to rely on face-to-face interviews with paper questionnaires have switched to computer-assisted personal interviewing (CAPI); in CAPI surveys, the questionnaire is no longer on paper but is a program on a laptop. Other face-to-face surveys have adopted a technique—audio computer-assisted self-interviewing (ACASI)—in which the respondents interact directly with the laptop. They read the questions on-screen and listen to recordings of the questions (typically, with earphones) and then enter their answers via the computer’s keypad. A similar technique—interactive voice response (IVR)—is used in some telephone surveys. The respondents in an IVR survey are contacted by telephone (generally by a live interviewer) or dial into a toll-free number and then are connected to a system that administers a recording of the questions. They provide their answers by pressing a number on the keypad of the telephone or, increasingly, by saying aloud the number corresponding to their answer. In addition, some surveys now collect data over the Internet.

Although these different methods of data collection differ along a number of dimensions (for a thorough discussion, see chap. 5 in Groves, Fowler, et al., 2004), one key distinction among them is whether an interviewer administers the questions. Interviewer administration characterizes CAPI, traditional face-to-face interviews with paper questionnaires, and computer-assisted telephone interviews. By contrast, with traditional self-administered paper questionnaires (SAQs) and the newer computer-administered modes like ACASI, IVR, and Web surveys, respondents interact directly with the paper or electronic questionnaire. Studies going back nearly 40 years suggest that respondents are more willing to report sensitive information when the questions are self-administered than when they are administered by an interviewer (Hochstim, 1967). In the discussion that follows, we use the terms computerized self-administration and computer administration of the questions interchangeably. When the questionnaire is electronic but an interviewer reads the questions to the respondents and records their answers, we refer to it as computer-assisted interviewing.

2 Of course, this assumption applies only to questions that are subject to underreporting. For questions about behaviors that are socially desirable and therefore overreported (such as voting), the opposite assumption is adopted—the method that elicits fewer reports is the better method.
Table 2 (adapted from Tourangeau & Yan, in press) summarizes the results of several randomized field experiments that compare the effects of self-administered questionnaires with interviewer-administered questionnaires. The table includes information on the method of data collection, the drug being studied, the month and year of data collection, and the ratio of estimated prevalence under self-administration compared to interviewer-administration.

Note. Each study compares a method of self administration with a method of interviewer administration. Dashes indicate that studies did not include questions about the illicit use of drugs during the prior month.

A key finding of our analysis was that computer instruments reduced social desirability distortion when these instruments were used as a substitute for face-to-face interviews, particularly when the interviews were asking respondents to reveal highly sensitive personal behavior, such as whether they used illegal drugs or engaged in risky sexual practices. (Richman et al., 1999, p. 770)

An important feature of the various forms of self-administration is that the interviewer (if one is present at all) remains unaware of the respondent’s answers. Some studies (e.g., Turner,Lessler,& Devore, 1992) have used a hybrid method in which an interviewer reads the questions aloud, but the respondent records the answers on a separate answer sheet; at the end of the interview, the respondent seals the answer sheet in an envelope. Again, the interviewer is never aware of how the respondent answered the questions. This method of self-administration seems just as effective as more conventional SAQs and presumably helps respondents with poor reading skills. ACASI offers similar advantages as more conventional SAQs and presumably helps respondents with poor reading skills. ACASI offers similar advantages.

A general issue with mode comparisons is that the mode of data collection may affect not only reporting but also the level of unit or item nonresponse, making it difficult to determine whether the difference between modes reflects the impact of the method on nonresponse, reporting, or both. Most mode comparisons attempt to deal with this problem either by assigning cases to an experimental group after the respondents have agreed to participate or by controlling for any observed differences in the makeup of the different mode groups in the analysis.

---

Table 2 (adapted from Tourangeau & Yan, in press) summarizes the results of several randomized field experiments that compare different methods for collecting data on illicit drug use. The figures in the table are the ratios of the estimated prevalence of drug use under a self-administered mode to the estimated prevalence under some form of interviewer administration. For example, Corkrey and Parkinson (2002) compared IVR (computerized self-administration by telephone) with computer-assisted interviewer administration over the telephone and found that the estimated rate of marijuana use in the past year was 58% higher when IVR was used to administer the questionnaire. Almost without exception, the seven studies in Table 2 found that a higher proportion of respondents reported illicit drug use when the questions were self-administered than when they were administered by an interviewer. The median increase from self-administration is 30%. These results are quite consistent with the findings from a meta-analysis done by Richman, Kiesler, Weisband, and Drasgow (1999), who examined studies comparing computer-administered and interviewer-administered questionnaires. Their analysis focused on more specialized populations (such as psychiatric patients) than the studies in Table 2 and did not include any of the studies listed there. Still, they found a mean effect size of −.19, indicating greater reporting of psychiatric symptoms and socially undesirable behaviors when the computer administered the questions directly to the respondents:
viewer will learn the respondent’s true status on the variable in question regardless of what he or she reports (Clark & Tifft, 1966; see E. E. Jones & Sigall, 1971, for an early review of studies using the bogus pipeline, and Roese and Jamieson, 1993, for a more recent one). Researchers have used a variety of means to convince the respondents that they can detect false reports, ranging from polygraph-like devices (e.g., Tourangeau, Smith, & Rasinski, 1997) to biological assays that can in fact detect false reports (such as analyses of breath or saliva samples that can detect recent smoking; see Bauman & Dent, 1982, for an example). The bogus pipeline presumably increases the respondent’s motivation to tell the truth—misreporting will only add the embarrassment of being caught out in a lie to the embarrassment of being exposed as an illicit drug user, smoker, and so forth.

Roese and Jamieson’s (1993) meta-analysis focused mainly on sensitive attitudinal reports (e.g., questions about racial prejudice) and found that the bogus pipeline significantly increases respondents’ reports of socially undesirable attitudes (Roese & Jamieson, 1993). Several studies have also examined reports of sensitive behaviors (such as smoking, alcohol consumption, and illicit drug use). For example, Bauman and Dent (1982) found that the bogus pipeline increased accuracy in reports by teens about smoking. They tested breath samples to determine whether the respondents had smoked recently; in their study, the “bogus” pipeline consisted of warning respondents beforehand that breath samples would be used to determine whether they had smoked. The gain in accuracy came solely from smokers, who were more likely to report that they smoked when they knew their breath would be tested than when they did not know it would be tested. Murray et al. (1987) reviewed 11 studies that used the bogus pipeline procedure to improve adolescents’ reports about smoking; 5 of the studies found significant effects for the procedure. Tourangeau, Smith, and Rasinski (1997) examined a range of sensitive topics in a community sample and found significant increases under the bogus pipeline procedure in reports about drinking and illicit drug use. Finally, Wish, Yacoubian, Perez, and Choyka (2000) compared responses from adult arrestees who were asked to provide urine specimens either before or after they answered questions about illicit drug use; there was sharply reduced underreporting of cocaine and marijuana use among those who tested positive for the drugs in the “test first” group (see Yacoubian & Wish, 2001, for a replication). In most of these studies of the bogus (or actual) pipeline, nonresponse is negligible and cannot account for the observed differences between groups. Researchers may be reluctant to use the bogus pipeline procedure when it involves deceiving respondents; we do not know of any national surveys that have attempted to use this method to reduce misreporting.

Forgiving Wording and Other Question Strategies

Surveys often use other tactics in an attempt to improve reporting of sensitive information. Among these are “forgiving” wording of the questions, assurances regarding the confidentiality of the data, and matching of interviewer–respondent demographic characteristics.

Most of the standard texts on writing survey questions recommend “loading” sensitive behavioral questions to encourage respondents to make potentially embarrassing admissions (e.g., Fowler, 1995, pp. 28–45; Sudman & Bradburn, 1982, pp. 71–85). For example, the question might presuppose the behavior in question (“How many cigarettes do you smoke each day?”) or suggest that it is quite common (“Even the calmest parents get angry at their children sometimes. Did your children do anything in the past seven days to make you yourself angry?”). Surprisingly few studies have examined the validity of these recommendations to use “forgiving” wording. Holtgraves, Eck, and Lasky (1997) reported five experiments that varied the wording of questions on sensitive behaviors and found few consistent effects. Their wording manipulations had a much clearer effect on respondents’ willingness to admit they did not know much about various attitude issues (such as global warming or the GATT treaty) than on responses to sensitive behavioral questions. Catania et al. (1996) carried out an experiment that produced some evidence for increased reporting (e.g., of extramarital affairs and sexual problems) with forgiving wording of the sensitive questions than with more neutral wording, but Abelson, Loftus, and Greenwald (1992) found no effect for a forgiving preamble (“. . . we often find a lot of people were not able to vote because they weren’t registered, they were sick, or they just didn’t have the time”) on responses to a question about voting.

There is some evidence that using familiar wording can increase reporting; Bradburn, Sudman, and Associates (1979) found a significant increase in reports about drinking and sexual behaviors from the use of familiar terms in the questions (“having sex”) as compared to the more formal standard terms (“sexual intercourse”). In addition, Tourangeau and Smith (1996) found a context effect for reports about sexual behavior. They asked respondents to agree or disagree with a series of statements that expressed either permissive or restrictive views about sexuality (“It is only natural for people who date to become sexual partners” versus “It is wrong for a married person to have sexual relations with someone other than his or her spouse”). Contrary to their hypothesis, Tourangeau and Smith found that respondents reported fewer sexual partners when the questions followed the attitude items expressing permissive views than when they followed the ones expressing restrictive views; however, the mode of data collection had a larger effect on responses to the sex partner questions with the restrictive than with the permissive items, suggesting that the restrictive context items had sensitized respondents to the difference between self- and interviewer administration. Presser (1990) also reported two studies that manipulated the context of a question about voting in an attempt to reduce overreporting; in both cases, reports about voting were unaffected by the prior items.

Two other question-wording strategies are worth noting. In collecting income, financial, and other numerical information, researchers sometimes use a technique called unfolding brackets to collect partial information from respondents who are unwilling or unable to provide exact amounts. Item nonrespondents (or, in some cases, all respondents) are asked a series of bracketing questions (“Was the amount more or less than $25,000?”, “More or less than $100,000?”) that allows the researchers to place the respondent in a broad income or asset category. Heeringa, Hill, and Howell

---

6 In such cases, of course, the method might better be labeled the true pipeline. We follow the usual practice here and do not distinguish versions of the technique that actually can detect false reports from those that cannot.
(1993) reported that this strategy reduced the amount of missing financial data by 60% or more, but Juster and Smith (1997) reported that more than half of those who refused to provide an exact figure also refused to answer the bracketing questions. Apparently, some people are willing to provide vague financial information but not detailed figures, but others are unwilling to provide either sort of information. Hurd (1999) noted another drawback to this approach. He argued that the bracketing questions are subject to acquiescence bias, leading to anchoring effects (with the amount mentioned in the initial bracketing question affecting the final answer). Finally, Press and Tanur (2004) have proposed and tested a method—the respondent-generated intervals approach—in which respondents generate both an answer to a question and upper and lower bounds on that answer (values that there is “almost no chance” that the correct answer falls outside of). They used Bayesian methods to generate point estimates and credibility intervals that are based on both the answers and the upper and lower bounds; they demonstrated that these point estimates and credibility intervals are often an improvement over conventional procedures for eliciting sensitive information (about such topics as grade point averages and SAT scores).

Other Tactics

If respondents misreport because they are worried that the interviewer might disapprove of them, they might be more truthful with interviewers whom they think will be sympathetic. A study by Catania et al. (1996) provides some evidence in favor of this hypothesis. Their experiment randomly assigned some respondents to a same-sex interviewer and others to an opposite-sex interviewer; a third experimental group got to choose the sex of their interviewer. Catania et al. concluded that sex matching produces more accurate reports, but the findings varied across items, and there were interactions that qualify many of the findings. In contrast to these findings on live interviewers, Couper, Singer, and Tourangeau (2004) found no effects of the sex of the voice used in an IVR study nor any evidence of interactions between that variable and the sex of the respondent.

As we noted earlier in our discussion of question sensitivity and nonresponse, many surveys include assurances to the respondents that their data will be kept confidential. This seems to boost response rates when the questions are sensitive and also seems to reduce misreporting (Singer et al., 1995). Cannell, Miller, and Oksenberg (1981) reviewed several studies that examined various methods for improving survey reporting, including two tactics that are relevant here; instructing respondents to provide exact information and asking them to give a signed pledge to try hard to answer the questions increased the accuracy of their answers. In an era of declining response rates, making added demands on the respondents is a less appealing option than it once was, but at least one national survey—the National Medical Expenditure Survey—used a commitment procedure modeled on the one developed by Cannell et al.

Rasinski, Visser, Zagatsky, and Rickett (2005) investigated an alternative method to increase respondent motivation to answer truthfully. They used a procedure that they thought would implicitly prime the motive to be honest. The participants in their study first completed a task that required them to assess the similarity of the meaning of words. For some of the participants, four of the target words were related to honesty; for the rest, none of the target words were related to honesty. The participants then completed an ostensibly unrelated questionnaire that included sensitive items about drinking and cheating. In line with the hypothesis of Rasinski et al., the participants who got the target words related to honesty were significantly more likely to report various drinking behaviors than were the participants who got target words unrelated to honesty.

One final method for eliciting sensitive information is often mentioned in survey texts: having respondents complete a self-administered questionnaire and placing their completed questionnaire in a sealed ballot box. The one empirical assessment of this method (Krosnick et al., 2002) indicated that the sealed ballot box does not improve reporting.

Summary

Several techniques consistently reduce socially desirable responding: self-administration of the questions, the randomized response technique, collecting the data in private (or at least with no parents present), the bogus pipeline, and priming the motivation to be honest. These methods seem to reflect two underlying principles. They either reduce the respondent’s sense of the presence of another person or affect the respondent’s motivation to tell the truth or both. Both self-administration and the randomized response technique ensure that the interviewer (if one is present at all) is unaware of the respondent’s answers (or of their significance). The presence of third parties, such as parents, who might disapprove of the respondents or punish him or her seems to reduce respondents’ willingness to report sensitive information truthfully; the absence of such third parties promotes truthful reporting. Finally, both the bogus pipeline and the priming procedure used by Rasinski et al. (2005) seem to increase the respondents’ motivation to report sensitive information. With the bogus pipeline, this motivational effect is probably conscious and deliberate; with the priming procedure, it is probably unconscious and automatic. Confidentiality assurances also have a small impact on willingness to report and accuracy of reporting, presumably by alleviating respondent concerns that the data will end up in the wrong hands.

How Reporting Errors Arise

In an influential model of the survey response process, Tourangeau (1984) argued that there are four major components to the process of answering survey questions. (For greatly elaborated versions of this model, see Sudman, Bradburn, & Schwarz, 1996; Tourangeau et al., 2000.) Ideally, respondents understand the survey questions the way the researcher intended, retrieve the necessary information, integrate the information they retrieve using an appropriate estimation or judgment strategy, and report their answer without distorting it. In addition, respondents should have taken in (or “encoded”) the requested information accurately in the first place. Sensitive questions can affect the accuracy of the answers through their effects on any of these components of the response process.

Different theories of socially desirable responding differ in part in which component they point to as the source of the bias. Paulhus’s (1984) notion of self-deception is based on the idea that
some respondents are prone to encode their characteristics as positive, leading to a sincere but inflated opinion of themselves. This locates the source of the bias in the encoding component. Holtgraves (2004) suggested that several other components of the response process may be involved instead; for example, he suggested that some respondents may bypass the retrieval and integration steps altogether, giving whatever answer seems most socially desirable without bothering to consult their actual status on the behavior or trait in question. Another possibility, according to Holtgraves, is that respondents do carry out retrieval, but in a biased way that yields more positive than negative information. If most people have a positive self-image (though not necessarily an inflated one) and if memory search is confirmatory, then this might bias responding in a socially desirable direction (cf. Zuckerman, Knee, Hodgins, & Miyake, 1995). A final possibility is that respondents edit their answers before reporting them. This is the view of Tourangeau et al. (2000, chap. 9), who argued that respondents deliberately alter their answers, largely to avoid embarrassing themselves in front of an interviewer.

Motivated Misreporting

Several lines of evidence converge on the conclusion that the main source of error is deliberate misreporting. First, for many sensitive topics, almost all the reporting errors are in one direction—the socially desirable direction. For example, only about 1.3% of voters reported in the American National Election Studies that they did not vote; by contrast, 21.4% of the nonvoters reported that they voted (Belli et al., 2001; see Table 1). Similarly, few adolescents claim to smoke when they have not, but almost half of those who have smoked deny it (Bauman & Dent, 1982; Murray et al., 1987); the results on reporting of illicit drug use follow the same pattern. If forgetting or misunderstanding of the questions were the main issue with these topics, we would expect to see roughly equal rates of error in both directions. Second, procedures that reduce respondents’ motivation to misreport, such as self-administration or the randomized response techniques, sharply affect reports on sensitive topics but have few or relatively small effects on nonsensitive topics (Tourangeau et al., 2000, chap. 10). These methods reduce the risk that the respondent will be embarrassed or lose face with the interviewer. Similarly, methods that increase motivation to tell the truth, such as the bogus pipeline (Tourangeau, Smith, & Rasinski, 1997) or the priming technique used by Rasinski et al. (2005) have greater impact on responses to sensitive items than to nonsensitive items. If respondents had sincere (but inflated) views about themselves, it is not clear why these methods would affect their answers. Third, the changes in reporting produced by the bogus pipeline are restricted to respondents with something to hide (Bauman & Dent, 1982; Murray et al., 1987). Asking a nonsmoker whether they smoke is not very sensitive, because they have little reason to fear embarrassment or punishment if they tell the truth. Similarly, the gains from self-administration seem larger the more sensitive the question (Tourangeau & Yan, in press).

So, much of the misreporting about sensitive topics appears to result from a motivated process. In general, the results on self-administration and the privacy of the setting of the interview suggest that two distinct motives may govern respondents’ willingness to report sensitive information truthfully. First, respondents may be reluctant to make sensitive disclosures to an interviewer because they are afraid of embarrassing themselves (Tourangeau et al., 2000, chap. 9) or of losing face (Holtgraves, Eck, & Lasky, 1997). This motive is triggered whenever an interviewer is aware of the significance of their answers (as in a telephone or face-to-face interview with direct questions). Second, they may be reluctant to reveal information about themselves when bystanders and other third parties may learn of it because they are afraid of the consequences; these latter concerns generally center on authority figures, such as parents (Aquilino et al., 2000) or commanding officers (Rosenfeld et al., 1996).

There is evidence that respondents may edit their answers for other reasons as well. For example, they sometimes seem to tailor their answers to avoid offending the interviewer, giving more pro-feminist responses to female interviewers than to male interviewers (Kane & Macaulay, 1993) or reporting more favorable attitudes towards civil rights to Black interviewers than to White ones (Schuman & Converse, 1971; Schuman & Hatchett, 1976).

Retrieval Bias Versus Response Editing

Motivated misreporting could occur in at least two different ways (Holtgraves, 2004). It could arise at a relatively late stage of the survey response process, after an initial answer has been formulated; that is, respondents could deliberately alter or edit their answers just before they report them. Misreporting could also occur earlier in the response process, with respondents either conducting biased retrieval or skipping the retrieval step entirely. If retrieval were completely skipped, respondents would simply respond by choosing a socially desirable answer. Or, if they did carry out retrieval, respondents might selectively retrieve information that makes them look good. (Schaeffer, 2000, goes even further, arguing that sensitive questions could trigger automatic processes affecting all the major components of the response process; see her Table 7.10.) Holtgraves’s (2004) main findings—that heightened social desirability concerns produced the longer response times whether or not respondents answered in a socially desirable direction—tend to support the editing account. If respondents omitted the retrieval step when the question was sensitive, they would presumably answer more quickly rather than more slowly; if they engaged in biased retrieval, then response times would depend on the direction of their answers. Instead, Holtgraves’s findings seem to suggest that respondents engaged in an editing process prior to reporting their answers, regardless of whether they ultimately altered their answers in a socially desirable direction (see Experiments 2 and 3; for a related finding, see Paulhus, Graf, & van Selst, 1989).

Attitudes, Traits, and Behaviors

The psychological studies of socially desirable responding tend to focus on misreporting about traits (beginning with Crowne & Marlowe, 1964; Edwards, 1957; and Wiggins, 1964, and continuing with the work by Paulhus, 1984, 2002) and attitudes (see, e.g., the recent outpouring of work on implicit attitude measures, such as the Implicit Attitudes Test [IAT], for assessing racism, sexism, and other socially undesirable attitudes; Greenwald & Banaji, 1995; Greenwald, McGhee, & Schwartz, 1998). By contrast, the survey studies on sensitive questions have focussed on reports about behaviors. It is possible that the sort of editing that leads to misreporting about sensitive behaviors in surveys (like drug use or
sexual behaviors) is less relevant to socially desirable responding on trait or attitudinal measures.

A couple of lines of evidence indicate that the opposite is true—that is, similar processes lead to misreporting for all three types of self-reports. First, at least four experiments have included both standard psychological measures of socially desirability responding (such as impression management scores from the BIDR) and sensitive survey items as outcome variables (Couper et al., 2003, 2004; and Tourangeau et al., 2003, Studies 1 and 2). All four found that changes in mode of data collection and other design variables tend to affect both survey reports and social desirability scale scores in the same way: thus, there seems to be considerable overlap between the processes targeted by the classic psychological measures of socially desirable responding and those responsible for misreporting in surveys. Similarly, the various implicit measures of attitude (see, e.g., Devine, 1989; Dovidio & Fazio, 1992; Greenwald & Banaji, 1995; Greenwald et al., 1998) are thought to reveal more undesirable attitudes than traditional (and explicit) attitude measures because the implicit measures are not susceptible to conscious distortion whereas the explicit measures are. Implicit attitude measures, such as the IAT, assess how quickly respondents can carry out some ostensibly nonattitudinal task, such as identifying or classifying a word; performance on the task is facilitated or inhibited by positive or negative attitudes. (For criticisms of the IAT approach, see Karpinski & Hilton, 2001; Olson & Fazio, 2004.) In addition, respondents report more socially undesirable attitudes measures (such as race prejudice) on explicit measures of these attitudes when the questions are administered under the bogus pipeline than under conventional data collection (Roese & Jamieson, 1993) and when they are self-administered than when they are administered by an interviewer (Krysan, 1998). Misreporting of undesirable attitudes seems to result from the same deliberate distortion or editing of the answers that produces misreporting about behaviors.

**Controlled or Automatic Process?**

The question remains to what extent this editing process is automatic or controlled. We argued earlier that editing is deliberative, suggesting that the process is at least partly under voluntary control, but it could have some of the other characteristics of automatic processes (e.g., happening wholly or partly outside of awareness or producing little interference with other cognitive processes). Holgraves (2004) provided some evidence on this issue. He administered the BIDR, a scale that yields separate scores for impression management and self-deception. Holgraves argued (as did Paulhus, 1984) that impression management is largely a controlled process, whereas self-deception is mostly automatic. He found that respondents high in self-deception responded reliably faster to sensitive items, but not to nonsensitive items, than did those low in self-deception, consistent with the view that high self-deception scores are the outcome of an automatic process. However, he did not find evidence that respondents high in impression management took significantly more time to respond than did those low in impression management, even though they were more likely to respond in a socially desirable manner than their low IM counterparts. This evidence seems to point to the operation of a fast, relatively effortless editing process.

The notion of a well-practiced, partly automatic editing process is also consistent with studies on lying in daily life (DePaulo et al., 2003; DePaulo, Kashy, Kirkenol, Wyer, & Epstein, 1996). Lying is common in everyday life and it seems only slightly more burdensome cognitively than telling the truth—people report that they do not spend much time planning lies and that they regard their everyday lies as of little consequence (DePaulo et al., 1996, 2003).

By contrast, applications of subjective expected utility theory to survey responding (e.g., Nathan, Sirken, Willis, & Esposito, 1990) argue for a more controlled editing process, in which individuals carefully weigh the potential gains and losses from answering truthfully. In the survey context, the possible gains from truthful reporting include approval from the interviewer or the promotion of knowledge about some important issue; potential losses include embarrassment during the interview or negative consequences from the disclosure of the information to third parties (Rasinski, Baldwin, Willis, & Jobe, 1994; see also Schaeffer, 2000, Table 7.11, for a more detailed list of the possible losses from truthful reporting). Rasinski et al. (1994) did a series of experiments using vignettes that described hypothetical survey interviews; the vignettes varied the method of data collection (e.g., self- versus interviewer administration) and social setting (other family members are present or not). Participants rated whether the respondents in the scenarios were likely to tell the truth. The results of these experiments suggest that respondents are sensitive to the risk of disclosure in deciding whether to report accurately, but the studies do not give much indication as to how they arrive at these decisions (Rasinski et al., 1994).

Overall, then, it remains somewhat unclear to what extent the editing of responses to sensitive questions is an automatic or a controlled process.

**Misreporting Versus Item Nonresponse**

When asked a sensitive question (e.g., a question about illicit drug use in the past year), respondents can choose to (a) give a truthful response, (b) misreport by understating the frequency or amount of their illicit drug use, (c) misreport by completely denying any use of illicit drugs, or (d) refuse to answer the question. There has been little work on how respondents select among the latter three courses of action. Beatty and Herrmann (2002) argued that three factors contribute to item nonresponse in surveys—the cognitive state of the respondents (that is, how much they know), their judgment of the adequacy of what they know (relative to the level of exactness or accuracy the question seems to require), and their communicative intent (that is, their willingness to report). Respondents opt not to answer a question when they do not have the answer, when they have a rough idea but believe that the question asks for an exact response, or when they simply do not want to provide the information. Following Schaeffer (2000, p. 118), we speculate that in many cases survey respondents prefer misreporting to not answering at all, because refusing to answer is often itself revealing—why would one refuse to answer a question about, say, cocaine use if one had not ever used cocaine at all?

**Conclusions**

According to models of the survey response process (e.g., Tourangeau, Rips, & Rasinski, 2000), response errors in surveys can
occur because respondents misunderstand the questions, cannot retrieve all the relevant information, use inaccurate estimation and judgment strategies, round their answers, or have difficulty mapping them onto one of the response categories. Answers to sensitive questions are subject to these normal sources of reporting errors, but they have an additional problem as well—respondents simply do not want to tell the truth.

This article summarizes the main findings regarding sensitive questions in surveys. There is some evidence that asking sensitive questions lowers response rates and boosts item nonresponse and reporting errors. There is even stronger evidence that misreporting about sensitive topics is quite common in surveys and that the level of misreporting is responsive to features of the survey design. Much of the misreporting on sensitive topics seems not to involve the usual suspects when it comes to reporting error in surveys, but rather results from a more or less deliberate process in which respondents edit their answers before they report them. Consequently, the errors introduced by editing tend to be in the same direction, biasing the estimates from surveys rather than merely increasing their variance. Respondents are less likely to overreport socially desirable behaviors and to underreport socially undesirable ones when the questions are self-administered, when the randomized response technique or bogus pipeline is used, and when the data are collected in private (or at least away from the respondent’s parents). Respondents in surveys seem to lie for pretty much the same reasons they lie in everyday life—to avoid embarrassment or possible repercussions from disclosing sensitive information—and procedures that take these motives into account are more likely to elicit accurate answers. The methodological findings suggest that socially desirable responding in surveys is largely contextual, depending both on the facts of the respondent’s situation and on features of the data collection situation such as the degree of privacy it offers.

Future survey procedures to encourage honest reporting are likely to involve new forms of computerized self-administration. There is some evidence (though nonsignificant) that computerization increases the reporting of sensitive information in surveys relative to paper questionnaires, though the difference between the two may depend on other variables (such as whether the computer is a stand-alone machine or part of a network). It remains to be seen whether Web administration of the questions will retain the advantages of other forms of computer administration. Moreover, as ever more elaborate interfaces are adopted, computerization may backfire by conveying a sense of social presence. Future research is needed to determine what features of computerized self-administration are likely to encourage or discourage candid reporting.

Even when the questions are self-administered, whether by computer or on paper, many respondents still misreport when they answer sensitive questions. Thus, another topic for future research is the development of additional methods (or new combinations of methods) for increasing truthful reporting. In a period of declining response rates (e.g., de Leeuw & de Heer, 2002; Groves & Couper, 1998), it is likely that respondents will be more reluctant than they used to be to take part in surveys of sensitive topics and that when they do take part, they will be less inclined to reveal embarrassing information about themselves. The need for methods of data collection that elicit accurate information is more urgent than ever.

References

References marked with an asterisk indicate studies included in the meta-analysis on computerized versus paper self-administration. Those marked with a tilde indicate studies included in the meta-analysis on bystander effects. Finally, those marked with a dagger indicate studies included in the meta-analysis on the item count technique.


