

## The Eleventh Morris Hansen Lecture Public Perceptions of Confidentiality

*Eleanor Singer*<sup>1</sup>

### 1. Introduction

Thank you for inviting me to take part in this panel discussion about issues related to confidentiality. It is a pleasure to be here on this very special occasion.

The data I am going to present today are outdated, not because they are so old but because they were collected both before the events of September 11 and before the subsequent adoption of heightened security and surveillance measures by our government. Because of this, we do not know what the specific data points might be if the surveys and experiments I am reporting on were repeated today. Nevertheless, I am willing to bet on the continued validity of the following four general points:

1. People have concerns about the confidentiality of personal information collected by government agencies.
2. These concerns about confidentiality have consequences for their behavior.
3. Respondents' perceptions of the risk of confidentiality breaches are only imperfectly related to the assurances of confidentiality provided by researchers.
4. However, perceptions of risk, and of the ratio of risk to benefit – however erroneous – significantly and substantially predict people's expressed willingness to participate in a survey.

After considering the evidence for those statements, I speculate about some conclusions to be drawn from this research.

### 2. Trends in Concerns about Confidentiality

The mail return rate to the U.S. census decreased substantially between 1980 and 1990. One hypothesis put forward to explain this decline was increased public concern about privacy, documented in a series of surveys by the Harris Organization (Westin 1990), and about confidentiality.<sup>2</sup> For example, although the Outreach Evaluation Study, carried out by the U.S. Census Bureau in 1990, found that the large majority of respondents believed that census data are kept confidential (Fay, Bates, and Moore 1991:18), and that such beliefs had not declined since the last decennial census, it also documented a significant change in the relationship between trust in the U.S. Census Bureau's assurance

<sup>1</sup> Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, U.S.A. E-mail: [esinger@isr.umich.edu](mailto:esinger@isr.umich.edu)

<sup>2</sup> By concern about confidentiality, I refer to a desire to keep information already given to one agent out of the hands of others; by concern about privacy, I refer to a desire to keep information out of the hands of others altogether (Singer, Mathiowetz, and Couper 1993).

of confidentiality and self-reported census return rate. Whereas trust was not predictive of self-reported returns in 1980, it was predictive of such returns in 1990 (*ibid.* and Table 4). As a result, in order to counter declining response rates, the U.S. Census Bureau in the early 1990's embarked on a program of privacy-related research, including the sponsorship of four surveys of attitudes toward privacy, confidentiality, and data sharing between 1995 and 2000. The first two points in my outline are based on data from these surveys.

The four surveys were carried out by telephone among respondents 18 and over. The first survey was developed in consultation with the U.S. Census Bureau as part of the University of Maryland's 1995 Joint Program in Survey Methodology practicum and carried out with funding from the National Science Foundation. The second, which used a questionnaire virtually identical to that in 1995, was carried out by Westat in 1996 under contract to the U.S. Census Bureau. The third, fielded in July through October 1999, just before the start of the public relations campaign and nationwide field recruiting for U.S. Census 2000, and the fourth, fielded from April to July of 2000, after delivery of census forms to U.S. households, were done by the University of Michigan under contract with the U.S. Census Bureau, with data collected by The Gallup Organization. All four surveys used virtually identical methods and achieved very similar response rates. All were random digit dialed surveys with one member of the household aged 18 or over randomly selected after household listing by the interviewer. The response rates were 61 percent, 60 percent, 62 percent, and 61 percent for 1995, 1996, 1999, and 2000, respectively.<sup>3</sup> Achieved sample sizes were 1,443, 1,215, 1,677, and 1,978.

Each interview began with the census short-form questions: sex, age, race, ethnicity, and (in 1995 and 1996) marital status (which was not asked on the 2000 Census). Subsequent questions about confidentiality and data sharing referred back to the content of these five items in order to anchor them in the specific context of the information requested.

Trends in attitudes about confidentiality from 1995 to 2000 are reported elsewhere (Singer 2001), and I will summarize only the highlights here.

1. First, a large fraction of the population is unsure whether or not the U.S. Census Bureau keeps information confidential, or whether or not other agencies can get answers to questions along with the respondent's name and address. This response is given by 77.5 percent to 40.7 percent of the population, depending on the question and the year, with more people professing uncertainty when the question is asked about confidentiality than when it is asked about other agencies getting names and addresses along with the answers to the question. The large proportion of "Not Sure" responses may partly reflect the fact that most people had not thought about these issues before the survey. But it also reflects the fact that the question included an explicit "Not Sure" category among the response options. When, in 1996, such an option was not offered to a random half of the sample, the "Not Sure" rate for the question asking whether other agencies could get the data identified by name and

<sup>3</sup> Response rates were calculated as interviews divided by the total sample less businesses, nonworking numbers, and (in 1995) numbers that were never answered after a minimum of 20 calls or (in 1996, 1999, and 2000) the estimated number of ineligible among the noncontacts.

address dropped from 46.8 to 7.7 percent. However, the ratio of correct to incorrect responses did not change.

2. Second, among those who did claim to know, the most frequent response was that such data *are* shared with other agencies – this response was given by 42 percent to 50.1 percent of the sample, with less inaccuracy toward the end of the decade. Respondents who said the U.S. Census Bureau protects confidentiality outnumber those who said it does not, but the proportion saying “protects” reaches only 25.1 percent in 2000.
3. Third, accuracy in public perceptions of the laws governing U.S. Census Bureau confidentiality practices increased between 1996 and 2000, with some 49 percent agreeing in 2000 that there is a law forbidding data sharing with other agencies, up from 28.3 percent in 1996. The largest increase occurred between 1996 and 1999.
4. Fourth, trust in the U.S. Census Bureau to keep the information confidential did not change between 1996 and 2000, remaining steady at two thirds of those who accurately perceived that the law forbids the sharing of data, who were the only ones asked the question.
5. Fifth, concerns about privacy, as distinct from confidentiality, appear to have declined slightly but significantly between 1995 and 2000, as measured by an index consisting of five general questions about privacy. The proportion regarding the census as an invasion of privacy also declined significantly, from 23.5 percent in 1995 to 20.9 percent in 2000.
6. Although concerns about privacy appear to have declined as of 2000, negative reactions to possible breaches of confidentiality increased during this period, with the proportion of those saying they would be bothered “a lot” if their answers to the census were not kept confidential increasing from 36.6 percent in 1995 to 46.4 percent in 1999 and 49.6 percent in 2000. It would be interesting to have a current data point for this question.
7. Perhaps because of this trend, answers to questions that probe respondents’ willingness to have other agencies share data with the U.S. Census Bureau, and their willingness to provide a Social Security Number (SSN) to facilitate such sharing, show a steady decline of 13–20 percentage points between 1995 and 1999, with a leveling off in 2000. In every year, those strongly opposed to data sharing outnumbered those strongly in favor by almost two to one.

Two kinds of reasons are associated with expressed reluctance to provide one’s SSN to the U.S. Census Bureau. First, there are reasons related to beliefs about the census: People who were less aware of the census, who considered it less important, and who were less favorable toward the idea of data sharing were significantly less willing to provide their SSN. Low levels of education are also associated with these characteristics. Second is a set of beliefs and attitudes concerning privacy, confidentiality, and trust: People who were more concerned about privacy, who had less trust in the Bureau’s maintenance of confidentiality, and who were less trusting of government in general were much less likely to say they would provide their SSN to the U.S. Census Bureau. Women are in general more concerned about privacy than men, and they were also less willing to say they would provide their SSN.

Attitudes about data sharing are very likely to fall into the category of “nonattitudes,” or “pseudo-attitudes” (Bishop et al. 1980). Schuman and Presser (1981:153), however, point out that, when confronted with an issue they have not encountered before, people tend to assimilate it to information and attitudes which they do, in fact, possess. Drawing on their arguments, Singer, Schaeffer, and Raghunathan (1997) showed that opinions about data sharing are related in predictable ways to trust in government, to confidence in the U.S. Census Bureau’s promise of confidentiality, to feelings of political effectiveness, and to a more general inclination to share or withhold personal information. It would obviously be highly instructive to have comparable measures of these attitudes today.

### 3. Attitudes and Behavior

So far, I have discussed trends in people’s *attitudes* about confidentiality and data sharing. However, concerns about breaches of confidentiality also have consequences for *behavior*. In the 1990 Census, people who were concerned about confidentiality and saw the census as an invasion of privacy were significantly less likely to return their census form by mail than those who had fewer privacy and confidentiality concerns (Singer, Mathiowetz, and Couper 1993; Couper, Singer, and Kulka 1998). Although such attitudes explained a relatively small proportion of the variance in census returns (1.3 percent), this proportion represented a significant number of people who had to be followed up in person in order to collect their census information.

Analysis of the mail returns of survey respondents in the 2000 Census yielded similar results. Once again, respondents with greater privacy and confidentiality concerns were less likely to return their census form by mail. The variance in census returns explained by attitudes toward privacy and confidentiality was very similar to that obtained in 1990: namely, 1.19 percent (Singer, Van Hoewyk, and Neugebauer 2003). In 2000, respondents with greater privacy and confidentiality concerns were also significantly less likely to provide an address to Gallup interviewers for the purpose of matching their survey responses to the file of census returns, and they were much less likely to respond to a question about their income.

### 4. (Mis)perceptions of Risk

Breaches of confidentiality, and their possible consequences, pose perhaps the major risk of harm to social science research subjects, and so concerns about confidentiality are understandable. For this reason, two years ago I undertook a small study of risk perception and its effect on survey participation (cf. Singer 2003). The study was carried out by means of an experiment embedded in the Survey of Consumer Attitudes, an RDD survey carried out monthly by the Survey Research Center at the University of Michigan. Approximately three hundred people are interviewed for the first time each month. The response rate in 2001 was about 65%.

The aim of my study was to survey a representative sample of potential respondents to social surveys to find out what they understood by risks and harms, and how these perceptions influenced their willingness to participate in the research.

The study was designed to answer research questions such as the following: Who do people think will see their survey answers? What does “confidentiality” mean to them?

How concerned are they about a breach of confidentiality – how much do they mind if it occurs? What kinds of consequences do they imagine would follow on a breach of confidentiality? Does the utility of the information compensate respondents for possible risks to which they may be exposed? Do perceptions of risks and benefits predict willingness to participate in research? How does a request for a signed consent form affect willingness to participate?

To answer these questions, respondents to the SCA were presented with hypothetical but realistic introductions to two ongoing studies at the University of Michigan: The National Survey of Family Growth (NSFG), and the Health and Retirement Study (HRS). (Note that the data presented here come from *respondents* to the SCA, and they may or may not generalize to nonrespondents.)

The introductions were very similar to the actual descriptions of these studies provided to potential respondents, but we tried to make the statements about risks and benefits in the two introductions as comparable as possible. In addition, the introductions systematically varied the request for a signature on a consent form, with half the introductions mentioning such a request and the other half not mentioning it. Each respondent was asked about both studies; if they were asked for a signed consent form for one study, they were also asked for such a form for the other. Half the respondents were asked first about the HRS and then about the NSFG, and the other half were asked about them in the reverse order. The questions appeared at the end of the regular SCA interview, just before the demographic questions, and were introduced by interviewers as follows:

“Now for something a little different. We are trying to learn how to better describe surveys to respondents . . . Imagine that the interviewer is talking with the respondent in person, in the respondent’s home, and describes the first study as follows . . .”

Both introductions included several sentences assuring respondents of the confidentiality of their answers. For example, the relevant portion of the HRS introduction stated:<sup>4</sup>

“Your answers to our questions are used for research purposes only. Any information you give us will be kept confidential. The researchers who use our data see only statistics. We never give out names and addresses to anyone. You will not be individually identifiable in any reports.”

Following the introduction, respondents were first asked how likely it was that they would be willing to participate in the study described. Then they were asked a number of questions about how likely they thought it was that various categories of others would gain access to their answers; how much they would mind if each of these did manage to see their answers, along with their name and address; how much they thought various groups, and they themselves, would benefit from the study; and whether they thought the risks of the study outweighed its benefits, or whether the benefits outweighed the risks.

We used responses to these questions to construct indicators of perceived risk – that is, the likelihood that others would gain access to their answers. We also constructed indicators of perceived harm – that is, how much they would mind if others did gain such access; and of what we called “perceived threat” – that is, the product of risk and harm.

In spite of the fact that respondents had been given what we believed to be fairly strong assurances of the confidentiality of their answers – assurances that were repeated in

<sup>4</sup> The complete introductions are given in Singer (2003).

various ways in five sentences of the introduction – a substantial number of them believed that one or more of the groups we asked about had some chance of gaining access to their answers, identified by name and address. On a scale where 0 means there is no chance that a particular group will be able to see their answers and 10 means that others are certain to see their answers, the average score for family members seeing their answers was 2.6; for employers, 2.9; for businesses trying to market something, 4.2; and for law enforcement agencies, 4.9. The average risk score for all four groups combined is 3.9 for the HRS and 3.4 for the NSFG. This difference between studies is significant, even with other variables controlled ( $p < 0.05$ ), but the order in which the introductions were read to respondents is not. No other variable had a significant effect on perceptions of risk.

It seems fairly clear that respondents either do not hear, do not believe, or do not remember what interviewers tell them about the confidentiality of their replies, which is why this section is titled “(Mis)perceptions of Risk.” A similar conclusion emerged from a much earlier study of the effect of informed consent on people’s participation in surveys (Singer 1979). In that study, respondents were asked, in a self-administered debriefing questionnaire after the main study, what the interviewer had said about the confidentiality of their answers. Although the question simply required them to select the correct answer from among several choices, only between 20 and 63 percent of the respondents, depending on the experimental condition, chose the correct response. When the interviewer had given them an unconditional assurance of confidentiality, less than half (49 percent) correctly answered that the results would be made public but their answers would remain confidential.

## 5. Perceptions of Risk and Harm as Predictors of Survey Participation

The measures of perceived risk, harm, and threat described above were entered, along with the perceived risk-benefit ratio and several control variables (age, gender, education, study, and order of presentation of the vignettes), into regression equations predicting willingness to participate in the survey described.

These measures proved to be both significant and substantial predictors of expressed willingness to participate. Results are shown in Table 1. Perceived risk had a significant

Table 1. The effect of perceived risk, perceived benefit, and perceived risk-benefit ratio\* on willingness to participate

Variable	Parameter Estimate	(Standard Error)	<i>p</i>
Risk	−0.190	(0.053)	< .001
Benefit	0.338	(0.068)	< .001
Risk-benefit ratio	−3.550	(0.328)	< .001
Age	−0.006	(0.010)	n.s.
Education	0.482	(0.340)	n.s.
Sex	−0.113	(0.299)	n.s.
Study	1.282	(0.235)	< .001
Order	0.188	(0.302)	n.s.
Signature	−0.230	(0.311)	n.s.
Intercept	1.617	(1.428)	n.s.

\*Measured independently of perceived risk and perceived benefit.

negative effect on willingness to participate, and perceived benefits a significant positive effect; the risk-benefit ratio, measured independently, also had a significant negative effect.

We repeated the model in Table 1, substituting perceived threat (as measured by perceived risk multiplied by perceived harm) for perceived risk but retaining the other variables in the equation. The results are identical to those reported above, with threat a highly significant negative predictor of willingness to participate in the survey.

Clearly, potential respondents are sensitive to the variables normally included in informed consent statements – i.e., the perceived risks and benefits associated with participation in research. It is also clear that the perceived risks *differ* from those actually described. In the introduction, the researchers went to great pains to assure respondents of the confidentiality of their replies – an assurance that apparently failed to convince at least some of them.

Another line of research, however, suggests that survey respondents *do* differentiate among assurances of confidentiality provided by researchers. Singer, Hippler, and Schwarz (1992) showed that when the research itself was not especially sensitive (i.e., did not ask for information whose disclosure might harm respondents), elaborate assurances of confidentiality resulted in less expressed willingness to participate than minimal assurances did. In addition, respondents who were offered more elaborate assurances of confidentiality under these circumstances expected more sensitive questions on the survey and were more concerned about the research falling into the wrong hands. Thus, the elaborate assurance of confidentiality appeared to increase the threat associated with the survey, rather than reassuring respondents that the confidentiality of their answers would be protected.

Following up on this research with a metaanalysis of existing studies, Singer, Von Thurn, and Miller (1995) showed that strong assurances of confidentiality *did* improve survey cooperation – either in the form of higher response rates, or in the form of less item nonresponse – when the topic of the survey was sensitive, but that such assurances had no effect on cooperation when the survey dealt with nonsensitive topics (see also Boruch and Cecil 1979; Singer 1978). Thus, perceptions of risk, and therefore willingness to participate in a survey, appear to be related to the interaction of the confidentiality assurance and the sensitivity of the data requested. We plan to carry out further research to clarify these issues.

There is no indication in these data that the request for a signature to document consent had a significant negative effect on expressed willingness to participate in the survey. However, some 13 percent of those who said they were willing to participate in the research also said they would not sign the consent form, which means that in a real survey situation they would be excluded from participation.

## 6. Concluding Observations

The research reported here (as well as earlier research – cf. Singer 1978) indicates that respondents do not hear, understand, or remember everything we tell them in the introduction to a survey, suggesting that neither consent nor refusal may be very well informed.

However, *given their perceptions*, respondents are rational. Their perceptions of risk,

benefit, and the risk-benefit ratio significantly predict their expressed willingness to participate in the survey described to them. And their concerns about confidentiality predict not only expressed willingness to participate but actual participation as well (Singer, Mathiowetz, and Couper 1993; Singer, Van Hoewyk, and Neugebauer 2003), accounting for between 1.2 and 1.3 percent of the explained variance in census returns in both 1990 and 2000. Respondents also appear to discriminate among assurances of confidentiality and to assess the appropriateness of the assurance to the sensitivity of the data requested. When the data are sensitive, respondents are more likely to cooperate if they are given strong, rather than weak, assurances of confidentiality; when the data are not sensitive, strong assurances of confidentiality have no effect or are, if anything, counterproductive.

Thus, the key to improving survey participation would seem to lie in more persuasive descriptions of the absence of risk (where there is none) and of the presence of benefits where these exist. We also need to know what the most effective assurance of confidentiality is, given respondents' perceptions of the risks involved, keeping in mind that more is not necessarily better.

Although I have confidence in the general truth of the four points with which I started this talk, I confess to not knowing at all how general social conditions, and new laws, might affect respondents' levels of concern about confidentiality, or their assessment of the risk of a breach of confidentiality. And on the legislative front, there is, as you know, both good news and bad news. The good news is that legislation protecting the confidentiality of statistical data has, after many years of effort, finally cleared Congress. The bad news is that the Patriot Act may still trump such legislation.

In the final analysis, the health of the statistical system depends on the trust and cooperation of respondents. If we fail to treat them with honesty and respect, we will forfeit both.

## 7. References

- Bishop, G.F., Oldendick, R.W., Tuchfarber, A.J., and Bennett, S.E. (1980). Pseudo-Opinions on Public Affairs. *Public Opinion Quarterly*, 44, 198–209.
- Boruch, R.F. and Cecil, J.S. (1979). *Assuring the Confidentiality of Social Research Data*. Philadelphia: University of Pennsylvania Press.
- Couper, M.P., Singer, E., and Kulka, R. (1998). Participation in the 1990 Decennial Census: Politics, Privacy, Pressures. *American Politics Quarterly*, 26, 59–80.
- Fay, R.L., Bates, N., and Moore, J. (1991). Lower Mail Response in the 1990 Census: A Preliminary Interpretation. *Proceedings of the Annual Research Conference*, Washington, DC: U.S. Census Bureau.
- Schuman, H. and Presser, S. (1981). *Questions and Answers in Attitude Surveys*. New York: Academic Press.
- Singer, E. (1978). Informed Consent: Consequences for Response Rate and Response Quality in Social Surveys. *American Sociological Review*, 43, 144–162.
- Singer, E. (1979). Informed Consent Procedures in Surveys: Some Reasons for Minimal Effects on Response. In M.L. Wax and J. Cassell (eds.) *Federal Regulations: Ethical Issues and Social Research*. Boulder, Colorado: Westview Press.

- Singer, E. (2001). Perceptions of Confidentiality and Attitudes toward Data Sharing by Federal Agencies. In P. Doyle, J. Lane, J.J.M Theeuwes, and L. Zayatz (eds.) *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. Washington, DC: U.S. Census Bureau.
- Singer, E. (2003). Exploring the Meaning of Consent: Participation in Research and Beliefs about Risks and Benefits. *Journal of Official Statistics*, 19, 273–285.
- Singer, E., Hippler, H.-J., and Schwarz, N. (1992). Confidentiality Assurances in Surveys: Reassurance or Threat. *International Journal of Public Opinion Research*, 4, 256–268.
- Singer, E., Mathiowetz, N., and Couper, M.P. (1993). The Role of Privacy and Confidentiality as Factors in Response to the 1990 Census. *Public Opinion Quarterly*, 57, 465–482.
- Singer, E., Von Thurn, D.R., and Miller, E.R. (1995). Confidentiality Assurances and Survey Response: A Review of the Experimental Literature. *Public Opinion Quarterly*, 59, 266–277.
- Singer, E., Van Hoewyk, J., and Neugebauer, R.J. (2003). Attitudes and Behavior: The Impact of Privacy and Confidentiality Concerns on Survey Participation. *Public Opinion Quarterly*, 67, 367–384.
- Singer, E., Schaeffer, N.C., and Raghunathan, T. (1997). Public Attitudes toward Data Sharing by Federal Agencies. *International Journal for Public Opinion Research*, 9, 277–285.
- Westin, A.F. (1990). *The Equifax Report on Consumers in the Information Age*. Atlanta: Equifax.

Received February 2003