Respondents' Understandings of Confidentiality in a Changing Privacy Environment

Eleanor Gerber[1], Ashley Landreth[2]
[1]Statistical Research Division, US Census Bureau, 5K313, Washington, DC, 20233
[2]Population Division, US Census Bureau, HQ-6H174A, Washington, DC, 20233

Abstract

Pledges of confidentiality are widely viewed as important in motivating respondent participation in surveys. Recent developments such as new privacy laws, increased media attention to identity theft, and highly publicized data leakages have changed the environment in which respondents interpret such a pledge. The aim of this paper is to examine respondent beliefs about confidentiality in this new environment. New data is compared to findings from interviews about privacy conducted during Census 2000. During 2006, fifty cognitive interviews were carried out regarding a cover letter intended for use in the 2008 census test. Three different versions of the letter were examined. These interviews assess the confidentiality language, including an informed consent statement about uses of administrative records from other government agencies to augment census results. Now, more respondents regard the pledge as inherently conditional: that is, they do not regard it as an absolute promise that their data will not be revealed. They focus on the policy of the agency to keep data confidential, but are highly aware that leaks occur through mistake or malfeasance. Communication about administrative records use is widely misunderstood as a two-way exchange between agencies. Some respondents believe this is a breach of confidentiality, but others accept it because they assume that all government agencies share data.

KEY WORDS: Confidentiality, Privacy beliefs, Census letters

1. Introduction

The aim of this paper[1] is to discuss recent findings about respondents' views of confidentiality, and to examine continuity and change in these ideas since research that we did in the year 2000.[2]

The focus of this research was a cognitive assessment of confidentiality language in cover letters, designed for use with the decennial mail out-mail back package. The aim of the language tested in these letters was two-fold:
1. To reassure respondents about the confidentiality of their answers, and therefore to motivate them to respond; and
2. To include information which provides respondents with a more complete understanding the uses of census data, including the use of administrative records to augment census data.

Reassurance generally included various ways of explaining about non-identification of respondents and their households, the restrictions on access to the data, and the legal authority and penalties that protect Census Bureau data.

Information about the uses of census data was included in this letter to allow respondents more complete understanding of what happens to their data once it is collected. The information refers to long-standing Census Bureau practices; however, this is the first time that such informational language has been part of the decennial communication to respondents. Two new elements were included in these letters: the use of administrative records data from other government agencies in the analysis of Census results, and the qualification that Census records become public after 72 years, when they are available in the National Archives.

[1] This paper is released to inform parties of research and to encourage discussion. The views expressed are those of the author and not necessarily those of the Census Bureau.

[2] The previous study in 2000 (Gerber, 2002, 2003) was more ethnographic in intent, and did not restrict the discussions to any particular text. Rather, we were interested in respondents' privacy concerns in general.

It is clear that the inclusion of the new information makes communicating the main confidentiality message more complex than it has been in previous census letters. Because there was more information to convey, our strategy in two of the letters was to limit the confidentiality paragraph in the front of the letter to a simple statement with one clear message: 'your data is confidential.' The more complex supports for this idea, and the new informational language, occur on the reverse side of the letter. We were not certain that this strategy would work, and thus, the third version of the letter presents all of the necessary ideas, in very abbreviated form, on the front of the letter.

2. Methods

We carried out 50 cognitive interviews in DC metro area, St. Louis area and Honolulu. Since we were concerned with wide readability of the letters, we tried to concentrate our recruiting on respondents with no more than a high school education, although the range of education was from 9th grade through graduate degrees. All interviews were carried out in English. Our group of respondents included persons of White, Black, Hispanic, Asian and Native Hawaiian descent.

Three versions of the letter were tested. Each respondent was asked to read two letters, in randomized order. Cognitive probing was designed to see if respondents understood the concepts being presented, to assess their reactions to the letters, and to establish their preferences between the two letters they saw.

3. Findings

3.1. The Term "Confidential" is Familiar to Respondents.

Our letters did not use the term "confidentiality" because previous research had indicated that some respondents were unsure of the meaning of the word in this form (Landreth, 2001.) However, the term "confidential" itself is familiar, and is generally readable to these respondents. We did not discover any respondent who could not read the term, or claimed not to understand its meaning.

3.2. Respondents Use Several Strategies to Define "Confidential."

Although respondents sometimes said that "confidential" meant "secret" or "privileged" the most commonly used term in this context was "private." For example:

"Information will be kept totally private – it's privileged information."

"Secret…not allowed out, secret, no matter what."

The Census Bureau uses the terms privacy and confidentiality to mean two different things, (privacy referring to non-intrusiveness in data collection and confidentiality referring to protections after the data is collected.) However, this distinction is not made by our respondents.

When they expand on these definitions, respondents tend to rely on images which are drawn from the realm of interpersonal communication. Thus, they say, "confidential is if I tell you something, it remains just between us" or "you don't tell it out anywhere." The basic model for the respondents' understandings of confidentiality is a verbal communication between two persons.

3.3. 'Confidentiality' is Unconditional

Another feature of their use of the term is important for this discussion: Respondents explicitly or implicitly treat a promise of confidentiality as unlimited and unconditional. It may be for this reason that respondents in the previous study (Gerber, 2001) often didn't like the term "strictly confidential" – because it implies gradations in a concept that they tend to see as an absolute. Respondents in the current study explicitly express the same assumption of unqualified confidentiality:

"People are very very very very concerned about their privacy…the information would be made confidential under all circumstances no matter what."

"means that it should be kept between you and me and you're not to repeat it to anyone else under any circumstances."

3.4. Extending "Confidentiality" to a Large Group of People

The basic image of confidentiality is about information that never passes outside of a dyad.

As a result, extending the concept to a large federal agency poses a certain cognitive challenge. Respondents who are familiar with surveys understand that the data they provide is not intended for just one other person, but promising confidentiality implicitly restricts access to the information. So who will see the information they give?

Respondents are aware that the Census Bureau is a large Federal agency, with many employees. They generally expand the range of confidentiality to apply to at least some of the Census Bureau's personnel: "the people who read my form" or "the ones that need the data." More commonly respondents extend the range of confidentiality to include the whole agency: "The Census Bureau, and nobody outside." They see this, in general, as the promise they've been given when the letter says, "You answers are confidential."

3.5. Supporting the Idea of Confidentiality: Law

Several of the ideas presented in the letter were intended as supports for the idea of confidentiality. One was the citation of the privacy law.

Three version of the citation language were used:

- "Federal law protects your privacy (Title 13, Sections 9 and 214)."
- "Strong federal laws require the Census Bureau to keep your responses on this form confidential (Title 13, Sections 9 and 214)."
- "Your answers are confidential, by law (Title 13, Sections 9 and 214)."

The mention of legal protections was generally popular with respondents. Between the three alternatives, the most effective was the first. It is possible that this results from mentioning protection of privacy. Also, some respondents found that the adjective "strong" (in the second version) was, as they put it, "over the top." That is, they seemed to see it as an exaggeration.

The legal citation (Title 13, Sections 9 and 214) was required information, so our research question was where to place it for best effect. When respondents were faced with the citation on the front of the letter (as in the third version above,) they found it bureaucratic and difficult to understand. However, when it was on the back, (as in the first two versions above,) they responded positively. The back of the letter interpreted as being for was for "the details," and the legal citation was clearly a detail for them. The back of the letter also provided respondents with an Internet address they assumed they could use to follow up the citation. This seemed to relieve them of the annoyance of not understanding what Title 13 was or where it could be found.

However, it should be noted that respondents did not understand that the law being cited was special to the Census Bureau. They sometimes thought that this was "The Privacy Act" or told us that they thought it must cover all government agencies.

3.6. Supports for the Confidentiality Concept: Nonidentification

Another support for the promise of confidentiality was to tell respondents that they would not be personally identified in the data that we release. In the past, this was expressed in the reassurance that the data would be used for "statistical purposes only." In the previous study, (Gerber 2000) this phrase was found to be difficult to understand. To some extent this was because respondents were unfamiliar with statistics, and did not automatically assume that it meant the removal of identifiers. We attempted to revise this language to make the intent clearer.

- "Your survey answers will only be used to produce statistics, and for no other purpose."
- "Your answers are confidential. That means the Census Bureau cannot give out information that identifies you or your household. The numbers we publish will not contain names or addresses."

The phrase "only be used to produce statistics" functioned rather better than "statistical purposes" had previously done. The concept of statistics is somewhat familiar to respondents, although they generally cannot provide anything like an abstract definition. The definitions involved counting or categorizing data. For example, one respondent described statistics as "putting things in buckets" like age and family size. Others gave rambling accounts of how someone might want to know how many people there were of certain ethnicities or how many

people live in a particular jurisdiction, etc. It seemed to us that the term "statistics" by itself is relatively nonproblematic, as long as the term is not intended to convey complex analysis or mathematical procedures beyond counting.

The relative success of the term statistics may also result from the current context of the term. It seems likely that the previously used phrase "statistical purposes" was confusing to respondents primarily because they could not understand what a *statistical* purpose would be, as opposed to any other sort of purpose. When we discussed the second part of the phrase with the respondents ("and for no other purpose") they sometimes expressed brief confusion. What else could you with data but to count it up? In the current formulation, respondents easily dismiss this concern, however, because those *other* purposes have been clearly delineated as events that are not going to happen.

While the phrase "used to produce statistics" partially explained what we intended to do with the data, it did not serve as a support for the idea of confidentiality. Some respondents have the sophistication to understand that statistical data sets remove identifiers, but others do not have this idea. As a result, counting things up is not conceptually related to confidentiality. Reminding respondents that their data will be compiled and used served more as support for the importance of the survey than it served as a reassurance of confidentiality.

However, the sentence "That means the Census Bureau cannot give out information that identifies you or your household" was seen as a definition of what we meant by "confidentiality." It clearly conveyed to them that names, addresses, and other identifiers such as phone numbers and social security numbers[3], would not be divulged. It proved very popular with our respondents. Since the respondents have to process the idea of confidentiality to understand what a big government agency might mean by it, it was useful to have this definition. It did not appear to be connected in respondents' minds to the idea of "statistics," however.

The second sentence in this "definition" was "The numbers we publish will not contain names or addresses." This did not appear as useful as the sentence that preceded it. Respondents were confused about what the "numbers" were, and it appears that this term is vaguer than "statistics." Respondents also indicated that they did not know what "publish" meant, under these circumstances. This sentence was not helpful, and was dropped.

3.7. Informing Respondents: 72 year ConfidentialityLlimit

After 72 years, census data becomes available to the public for genealogical, historical, and other uses. This information was conveyed in the following phrases:

- "Your Census remains confidential for 72 years."
- "Census information must remain confidential for 72 years."

There was some difficulty in understanding these statements, since it was not clear what would happen after the specified time limit. Some respondents thought the data would become public, while others thought that it would simply be destroyed. Some respondents were surprised that data is kept that long, as they could not see any possible use for such old data.

The most salient reaction to this feature was that "72 years" was considered a strange, arbitrary number. Respondents sometimes laughed aloud when they first saw this feature, and speculated on how such a number could have been derived. Besides that, it troubled only a few respondents, who were unhappy to think that there was any limit to confidentiality at all. Most respondents took the approach that they would probably be dead in 72 years, and wouldn't care who saw their information afterwards.

3.8. Informing Respondents: Administrative RecordsUse

The Census Bureau receives data from other federal agencies to augment analysis of the census. These administrative records come under the protection of Title 13 once the data has been

[3] In fact we do not ask for social security number in the decennial questionnaire. Respondents all had an opportunity to examine the questionnaire before we began the interview. Many, however, came to believe during the interview that they must have seen a request for social security number. This may have been the result of the stress we were putting on confidentiality: if the data is not sensitive, this does not make sense.

received. The descriptions of administrative records use were as follows:

- "Other government agencies may give us additional information about your household. We might combine this information with your answers to improve census results. The same legal protections apply to any information we receive from other agencies.
- "To improve census results, other government agencies may give us additional information about your household. The additional information we receive is legally protected, just like your census answers."

The placement of these statements was important in determining respondents' reactions to these statements. Both statements were used on the back of two versions of the letter. The second statement was also used on the front of the third version. This information was better received when it occurred on the back of the letter, after respondents had a chance to process the simple confidentiality message. The back of the letter was the text that the respondents had already decided was "details."

Two aspects of these statements must be examined: respondents' understanding of what the statements meant, and the salience of the statements. We intended to communicate the following concepts in these phrases: 1. that we might receive administrative records data from other agencies, 2. that the exchange was only one way, and 3. all such data would be treated as Title 13 once it was received.

Respondents generally understood that we were receiving data from other agencies. The "legal protections" statement was also understood as intended: respondents were able to connect these phrases back to the earlier citation of the law.

The rest of the communication was not successful. The problems were that respondents saw the exchange of data as two-way, and that they didn't understand the purposes to which the administrative records data would be put.

Many respondents indicated that they thought that, if we received data. we must also be giving it out. To some extent, this reaction is rooted in a common belief that all government agencies share data. In our previous studies, (Gerber, 2001, 2002) respondents frequently assumed that "all computers are connected." Their primary concern was that people with know-how could get information about anyone by hacking or other ruses. Many of our current respondents share that belief, but the emphasis seemed to be on official data sharing. They were convinced that agencies, or at least people "above a certain level" in the agencies, could have access to their data if they really wanted it. The administrative records statement serves to reinforce this belief.

The sharing of data among government agencies is not only assumed, it is often considered highly legitimate. Respondents often make the assumption that data is shared not just among Federal agencies, but with state and local governments as well. "Government" in this view is all one thing, and of course data will be given to those who need to use it. This is perhaps the widest extension of the confidentiality concept: "confidential with anyone in government with a need to know." But it still posits restrictions – respondents are often very clear that the data cannot and should not be available to commercial interests. Such respondents often explicitly tell us that they trust the government, either because of personal experience (such as being in the military) or as a matter of patriotism.

Belief in a two-way data exchange was also rooted in respondents' understandings of the purpose of the exchange. (The purpose is left unspecified in the letters.) Many respondents assumed that the reason for the data exchange was to check the accuracy of their answers. In order to do this, they assume, the data has to be specifically connected to individuals or households. In this logic, in order for us to perform the check, we must have given at least a name, or name and address, to the other agency. They conclude that the exchange is per se a breach of confidentiality because it must involve releasing identities.

Respondents were also concerned about the confidentiality promises that may have been made by the other agencies, which also appeared to have been breached. Another cause for concern was that the data held by these other agencies might well be wrong, so our "check" could result in faulty data about them. This might have vague potential repercussions down the line. For example, one respondent mentioned an elderly parent who had recently left her household to live with another sibling. She speculated that her tax form might include a

different number of dependents than she would be reporting on the Census, and that this might cause unspecified difficulties with tax and social service agencies. Respondents also told us that if we actually needed more data (although they couldn't understand why we might) we should properly get it directly from them.

The assumption of two way data transfer had much less effect on respondents' opinions than we might have anticipated. It destroyed belief in the promise of confidentiality only for a minority of respondents. One reason for this was the common assumption that wide data sharing is a normal part of the way government operates. Thus, information about administrative records could be seen as an "admission," or "being up front about it." In the eyes, of such respondents, this information, no matter how it was misinterpreted, actually lent us some credibility.

Others thought that the administrative records sentences were there primarily to protect the Census Bureau. It was seen as "the stuff you have to say" to cover the agency "in case something bad happens." That way, they felt, the agency could not later be blamed for concealing risks. Some respondents indicated that they are used to seeing similar "fine print" in privacy statements sent to them by banks and credit cards, which they think serves this purpose.

Another reason some respondents were not disturbed by the assumption of two-way data flow was that they did not think they had anything to loose. That is, there was nothing in their personal records that could cause trouble for them. ("I'm not doing anything.") However, they also thought that the assumed data exchange might have a chilling effect on others providing data. The most common examples of people who might be discouraged from respondent were undocumented immigrants to the country. Also mentioned were people who might have larger numbers in their households than are allowed by landlords or social agencies; and people who might have some involvement with courts or the police.

Although the discussion of administrative records use was troubling to some respondents after they processed it, it was not very salient to them on first reading. They often did not focus on the administrative records statement until we pointed it out to them. This may be accounted for by the immediate context of the statement on the back of the letter. The text on the back of the letter contained information that was designed to explain Title 13 protections, and it appeared that for some respondents, this was more salient than the administrative records language.

The more welcome information about Title 13 protections used two different approaches. One stressed agencies legally prevented from obtaining the data. The other stressed penalties for breaking Title 13 regulations.

- "The answers you give on the census form cannot be obtained by law enforcement, by immigration, or by tax collection agencies. Your answers cannot be used in court. They cannot be obtained with a Freedom of Information Act (FOIA) request."

- "Any Census employee who releases your information can be imprisoned up to 5 years or fined up to $250,000 or both."

Specific information about who could not get the data was popular with respondents, and was apparently much more noticeable to them than the administrative records statement that immediately followed in the next paragraph. It is possible that attention patterns in reading account for this. Respondents may have simply "skimmed" the mention of "other agencies," assuming it was an addition to the list of agencies which could not demand Title 13 data that they had just read. Thus, in this placement the administrative records language was far less salient than in its other placements.

Detailing legal penalties for breaches of Title 13 was not universally popular, but it also appeared to be more salient than the administrative records statements on first reading.

3.9 Continuity and Change

In many ways the respondents in this research resemble our previous respondents in 2000. Both sets of respondents tend to accept our pledges of confidentiality as being inherently conditional. They assume, and largely accept, the idea of government data sharing. They may not see it as a breach of the confidentiality pledges given to them. Now as in 2000, they are willing to provide data, but believe that

anyone whose answers might bring trouble with the government will not respond.

The general message that data provided to the census is legally protected is fairly clear, and is understood by most respondents. Respondents still tend to see participating in the census as a worthwhile activity, and tend to see the Census Bureau in a relatively positive light, in comparison with other government agencies. (Gerber, 2001.)

Despite these continuities, however, some changes in these respondents' approach to issues of privacy and confidentiality were apparent. There are three elements in this change: first, familiarity with privacy statements from many sources, second, changes in the legal context, and third, a greater awareness of lapses in protection of data.

3.9.1 Familiarity with privacy statements

Respondents reported to us that they now receive many privacy statements from all sorts of sources: banks, insurance companies, credit cards, etc. These statements have become familiar, and respondents tend to assimilate our language to these other sources. It is seen as "typical" and "what you have to say." They sometimes told us that expect privacy statements to be in difficult, bureaucratic language. When asked to describe the content of all these statements, one respondent summed it up as "blah blah blah." Respondents often report that they do not read or pay much attention to these statements. The confidentiality messages we intend them to receive may thus not be experienced as interesting or particularly noticeable.

Respondents also are used to seeing a list of exceptions, exclusions and conditions in these statements, which they often describe as "the fine print." (Often this is literally true.) This may overall reduce the salience of information like the administrative records statements, which some interpret as a limit on our pledge of confidentiality. Thus, respondents may be beginning to expect conditions in the promises of confidentiality from large bureaucratic organizations.

3.9.2 Changes in the legal context

Although most respondents liked the idea of having a law to protect their answers, some did not experience that information as particularly convincing. A few pointed out that laws can, and have changed. One skeptical respondent mentioned "The Patriot Act" and said that as far as he knew, that trumped all other privacy legislation. He thought that the effect of this law was that, if an official wanted information about him, they could get it. This sense of change in the legal situation is new to the current research. When we looked at these ideas n 2000, it was never mentioned. Other legal changes may have impinged on this. Since the previous research, respondents have become familiar with the privacy rules that govern HIPAA (Health Insurance Portability and Accountabilty Act.) Several of our respondents mentioned this, because they worked in the health care field. While health industry employees have been trained about data security, it does not appear to give them any particular sense of confidence. HIPAA was several times mentioned to us as an example of how rules and regulations can be ignored or subverted by employees. For example, one respondent described hospital training that required employees not to discuss the condition of a patient with another employee who was a friend of that patient (but not part of the treatment team.) The respondent thought that the policy would be immediately breached because "people talk;" it is simply human nature. Respondents who work in other bureaucratic settings also mentioned the "people talk" limitation to privacy policy, however rigorously the policy is conceived.

3.9.3 Lapses in data protections

Belief in our pledges of confidentiality is increasingly affected by the very common awareness of respondents of data that has been inadvertently released or lost by government and private organizations. Most frequently mentioned was the loss of a laptop with data about present and ex-military personnel from the VA (Veterans Administration.) The opportunity to interview in locations remote from the DC metro area, where the VA laptop was local news, was useful in this regard. Both in Hawaii and St. Louis, these events were highly salient. While the details of this data spill were fresh some 6 months after the events, respondents were also aware in a less detailed way of news stories about other data problems in banks, stores and other private interests. These events led some

respondents to express what they felt was a healthy skepticism about our promises. They told us that although they gave us credit for having good policies, they were uncertain that we would be able to carry them out.

Once again, this skepticism is rooted in the fallibility of human behavior: people not only "talk," they make errors and lose things. There was no sense that organizations have the means to prevent such breaches. Even severe penalties, such as those mentioned in our letter, are not thought to prevent such errors. The penalties themselves are sometimes the source of skepticism. A few respondents told us that they understood how large agencies operate, and thought that employees would be punished by "a slap on the wrist" rather than making a data breach public.

Thus, attention seems to have shifted since the research we did in 2000. In the prior research, the primary causes for concern about privacy focused on deliberate attacks on data privacy. Respondents were worried about the capability of clever and determined hackers to undermine computer systems. While this concern persists in the current context, it is less salient in comparison with a new concern. In the current research, skepticism about confidentiality has shifted to perceived bureaucratic incompetence and the perceived tendency of large organizations to cover up lapses in protection.

4. Conclusions

This paper examines respondent understanding and acceptance of language intended to convey confidentiality protections to respondents in a cover letter for the decennial census mail-out questionnaire. These cognitive interviews lead us to conclude that the main message should be kept as simple as possible, but that respondents like the opportunity to see additional details on the reverse side of the letter.

Respondents modify a dyadic concept of confidentiality to accommodate data sharing within a large organization, or even within government as a whole. However, the belief in wide government data sharing interferes with the attempt to describe a secure one-way flow of data. If respondents posit an exchange data as a means of checking their answers with other agencies, this leads them to think that confidentiality is being broken. However, many see data sharing within government as highly legitimate, and are not concerned.

There is evidence in these interviews of some new features of respondents' reaction to privacy and confidentiality language. Changes in the laws surrounding privacy lead some respondents to see privacy pledges as historically conditional: they have been changed in the recent past and therefore can be unpredictably changed in the future. Because of increased exposure, respondents may now be spending less attention on privacy language. They may experience the many statements they see as highly bureaucratic and designed to primarily to protect the organization. They have come to expect exclusions and conditions in these statements. An additional change is the increasing awareness of respondents of the gap between an organization's privacy policy and its ability or willingness to enforce it. As a result, the privacy policies, in themselves, have a somewhat limited effect on respondents' belief in pledges of confidentiality.

References

Gerber, E. (2001) "The Privacy Context of Survey Response: An Ethnographic Account" in Doyle, P., Lane, J., Theeuwes, J., and Zayatz, L., eds, *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies.* New York: North Holland, pp. 371-394.

Gerber, E. (2002) Privacy Schemas and data Collection: An Ethnographic Account. *US Census Bureau,* Census 2000 Testing and Experimentation Report.

Gerber, E. (2003) "Respondents' Understanding of Confidentiality Language," *Proceedings of the American Statistical Association, Survey Methods Section [CD-ROM],* Alexandria VA: American Statistical Association.

Landreth, A. (2001) "SIPP Advance Letter Research: Cognitive interview Results: Implications and Letter Recommendations." *US Census Bureau*, Center for Survey Methods Research, Statistical Research Division's Study Series, Survey Methodology #2003-11.