Aiming at a More Cost-Efficient Census Via Online Data Collection: Privacy Trade-Offs of Geo-Location

Research in progress

In an effort to reduce costs associated with data collection, in 2010 the Census Bureau announced that the 2020 Census of the United States would take place via two channels: the classical paper questionnaire and the online questionnaire. In addition to an adequate awareness campaign, the success of the initiative may depend on addressing individuals' possible security and privacy concerns and providing adequate incentives for a sizable portion of US citizens to transition from the offline to the online format. In this research in progress we analyze a specific incentive that the Census is considering: pre-populating location information in the form, so as to reduce completion time and effort. We study how individuals may react to this initiative, given the privacy concerns that pre-populating may arise. We describe a series of online experiments of geo-location that investigate the impact of awareness of being geo-tracked on willingness to provide, among other pieces of information, Census-related personal information.

1. Introduction

The most recent Census of the United States, Census 2010, cost approximately $13 billion dollars. The costs have roughly doubled in nominal terms each decade since 1970 (Economist, 2011) – a cost escalation that was not accompanied by a tantamount increase in accuracy of the count (Edmonston & Schultze, 1995). Reasons for the rise in costs include population growth, growth in the number of housing units, and a decline in the response rate to the mailed questionnaire (which triggers costly physical visits of Census personnel to non-responding households). Given its mounting costs, a congressional panel recently questioned the feasibility of the 2020 Census, expressing concerns over the US government's ability to afford it (Mervis, 2011).

Cost inefficiencies have been at the center of various critiques to the Census Bureau, together with the perception of privacy intrusiveness of the questionnaire itself, especially in its long form (USA Today, 2012). (Unlike the short form, the long form includes questions about sensitive socio-economic information such as marital status, education, citizenship, and disabilities.)[1] In response, the Census has implemented a number of changes (for instance, more effective advertising campaigns and the exclusive use of the short form questionnaire), lowering operational costs of the 2010 Census below the $7 billion budget (Census Bureau, 2010a), and contributing to a higher than expected response rate (72%) to the mailed questionnaire (Census Bureau, 2010b).

These efforts notwithstanding, Census data collection as performed in its old format (mailed questionnaires followed up by interviews by phone or in person in case of non-response) remains costly. Given the cost-efficiencies and the remarkable enhancement of the connection between citizens and governmental institutions made possible by IT,[2] an option the Census has considered so as to make the data collection process more efficient is the completion of the questionnaire via the Internet. Such option will be provided to U.S. residents for the first time in 2020.[3] However, can larger participation be guaranteed in the online versus the paper format of the questionnaire? Several issues need to be addressed in order for this initiative to generate the desired outcome: groups or areas where Internet penetration or computer literacy may be low should be reached; each household should only be able to submit one form (either online or offline); furthermore, privacy and security concerns of transmitting sensitive information through the Internet should be assuaged. In this paper we focus on the privacy trade-offs involved in the collection of data through the Internet. Specifically, we investigate reactions to geo-location, or the ability to identify the geographic location from which people are connecting to fill the online form.

One possible incentive that could be provided to U.S. residents in order to be convinced to use the online form, as opposed to the standard paper form, is a lower time and effort necessary for

---

[1] For a detailed comparison between the short and long form, see: http://www.census.gov/dmd/www/pdf/d3239a.pdf.

[2] For a much broader initiative (the Aadhaar project) regarding digital identities for Indian citizens, see Bapna and Sundararajan (2010). Information on a 2011 survey for the Aadhaar project is available at: http://www.stern.nyu.edu/experience-stern/faculty-research/sundararajan-uid-results.

[3] See http://voices.washingtonpost.com/federal-eye/2010/05/2020_census_will_have_an_onlin.html. In 2000, only a small proportion of US households were offered an option to submit the Census form online, but since the operation was not successful, the Bureau returned to standard mail delivery for the 2010 Census: see http://www.cbsnews.com/8301-201_162-57559640/u.s-census-will-offer-online-option/.

completion. Title 26 of the US Code allows the Census Bureau to receive certain information about citizens from other entities. Such information could be used to pre-populate specific fields in the Census form, thus reducing the amount of time necessary to complete it. Similarly, the use of certain technological tools, such as GPS or IP-capture, could allow for location specific information to be provided automatically. Setting aside issues relating to technical feasibility and accuracy, geo-location may have unintended consequences, including the perception of being tracked by the government, therefore potentially raising among Census respondents the impression of a privacy intrusion. At a time when data collection programs carried out by the US government have been publicly exposed, and private companies' tracking and behavioral advertising practices have been subject to scrutiny and critique, US citizens may have become particularly sensitive to the possibility of being watched, monitored, or surveilled. Under such a scenario, a Census' initiative focusing on pre-populating geo-location data in online Census forms may thus backfire and reduce, rather than increase, residents' response rates.

In a series of experiments we analyze individuals' reaction to geo-location (their location being identified) in terms of their willingness to disclose personal information in an online questionnaire. In a first online experiment, we manipulate participants' awareness of their location being identified, and measure their willingness to reveal two types of personal information: engagement in unethical or somewhat compromising behaviors, and Census-related information (demographics and living arrangements). In a second online experiment (yet to be started), we test whether geo-location related privacy concerns are higher for government institutions as compared to other entities, and whether this is due to people feeling surveilled and monitored by the government.

2. Related literature

Reactions to location identification have been analyzed in different streams of research. The literature on ubiquitous computing acknowledges various possible privacy issues associated with environment-aware systems and location tracking technologies (Agre & Rotenberg, 1997; Barkhuus & Dey, 2003; Doheny-Farina, 1994; Garfinkel, 2001; Harper, 1995; Junglas & Spitzmuller, 2005; Kaasinen, 2003; Minch, 2004), but it doesn't always provide practical recommendations on how to address such concerns (Schilit et al., 2003).

One of the suggested solutions is user control: the possibility to hide one's location from certain people (or entities), or for specific places or points in time (Gruteser & Liu, 2004), or, similarly, interruptions in the transmission of one's location data (Huang, Matsuura, Yamane & Sezaki, 2005). A potential problem with control-based solutions is that the more sophisticated the control algorithms, the higher the probability that the average person may find them difficult to use, and the lower her confidence that her chosen settings reflect indeed her privacy preferences. An even less user-friendly approach may consist in location obfuscation techniques, such as mixing personal data with data related to other users or adding noise to one's location (Brush, Krumm & Scott, 2010). More generally, perceived self-efficacy, or the judgment of one's capability to execute certain tasks properly, has been shown to significantly affect IT adoption (Agarwal, Sambamurthy & Stair, 2000; Chau, 2001; Hong, Thong, Wong & Tam, 2001; Johnson & Marakas, 2000; Luarn & Lin, 2005). Moreover, a system that is dynamic but that interrogates the user on her location preferences at different places or points in time may become annoying or intrusive itself, thus affecting adoption (Myles, Friday & Davies, 2003; Weiser & Brown, 1996).

While much of this literature analyzes the effects of privacy concerns on the adoption of location-based services (e.g., Zweig & Webster, 2002), we attempt to reverse this relationship, focusing on what are the privacy concerns raised by geo-location, and what are the consequent reactions. Related studies on the consequences of privacy concerns (more broadly defined, rather than strictly related to geo-location) in the online world have shown that people may respond to privacy invasions in a similar way as they respond to unfair transactions (Ashworth & Free, 2006). Moreover, individuals may be more willing to complete a privacy-sensitive transaction (such as the purchase of somewhat sensitive items, such as condoms) if the website is more privacy-friendly (Gideon, Cranor, Egelman & Acquisti, 2006). People may also be willing to pay a premium for protecting their privacy if privacy-related information is explained in a simple, accessible way (Tsai, Egelman, Cranor & Acquisti, 2011). The exact time in which privacy-relevant information is made salient was also shown to affect consumer choices, and specifically, willingness to pay for privacy (Egelman, Tsai, Cranor & Acquisti, 2009). Targeted advertisements have been shown to backfire if they are perceived as too privacy intrusive or "too targeted" (White, Zahay, Thorbjørnsen & Shavitt, 2008). On the other hand, if consumers have high privacy controls, personalized ads seem more effective (Tucker, 2011).

A stream of studies have specifically investigated privacy concerns raised by location tracking. Theoretical models have been proposed to account for the effect of various variables, such as the purpose of location data collection, the possible use by third parties, possible errors in storing the information, trust towards the requester, and usefulness of location data for the task at hand (Junglas & Spitzmuller, 2005). Some studies have analyzed the level of acceptance of location tracking by governmental institutions, but only in the special case of emergencies (Aloudat, Michael, Chen & Al-Debei, 2013). Particularly relevant to our case is work by Thomas et al. (2013), which focused on specific sub-populations, such as older adults, who may be less prone to the understanding and use of such technologies. Research has also shown that location privacy is valuable to individuals (Cvrcek, Kumpost, Matyas & Danezis, 2006), and that the need for it varies according to many determinants, such as time and specific place, who requests the information, how often it is requested, how populated is the place one is at when the request reaches them, and so on (Sadeh et al., 2009; Toch et al., 2010).

3. Experiments

We designed several experiments to investigate individuals' perceptions of the privacy trade-offs raised by geo-location technologies, and their consequences in terms of willingness to disclose further personal information. We investigated how people may react to geo-location in the context of online surveys, and whether they harbor specific reluctance to a government institution performing geo-location.

3.1 Experiment 1

In an effort to analyze possible differences in willingness to disclose personal information, for this experiment, we addressed two different populations: users (workers) of Amazon Mechanical Turk (MTurk) and participants from a pool of students and non-students managed by a North Eastern American University. We refer to the MTurk experiment as Experiment 1A, and to the University experiment as Experiment 1B. In order to investigate the mechanism behind the results, we also ran a follow-up study, which we refer to as Experiment 1C, for which participants were recruited from MTurk.

3.1.1 Experiment 1A

Methods. Experiment 1A was a between-subjects, three-condition randomized experiment, in which we manipulated participants' awareness of their location being identified. In a Control condition, we geo-located participants but we did not mention it. In a Geo-Located condition, at the beginning of the study (ostensibly a survey on ethical behaviors), participants were shown their current location (city, state, country, zip code) and asked to confirm that their location was correct and complete before proceeding. In a Requested Location condition, participants were explicitly asked to provide their location before proceeding. The remainder of the survey was identical for all participants, who were asked seven Census related questions and 16 privacy-sensitive questions regarding engagement in a series of unethical behaviors (a dependent variable that has been used in the disclosure literature to measure willingness to disclose and therefore, indirectly, privacy concerns; see Acquisti, John & Loewenstein, 2012; Brandimarte, Acquisti & Loewenstein, 2013; John, Acquisti & Loewenstein, 2011). Responses to this second set of questions were on a scale from 0 (Never) to 4 (Often), the last option (5) being "I prefer not to say." The order in which these two different types of personal information were asked was counterbalanced. We could not include Census questions that constituted personally identifiable information (PII), such as first name, last name and phone number, due to the terms of service of the platform we used to recruit participants. General questions about feeling tracked or monitored and about related privacy concerns concluded the survey (see the Appendix for the full list of questions).

We expected participants in the Control condition to have lower privacy concerns as compared to the other two conditions. Moreover, in line with the results in Brandimarte, Acquisti & Loewenstein (2013), where feeling of control over disclosure of personal information were shown to decrease privacy concerns, we predicted higher willingness to disclose and lower sense of being monitored in the Requested Location condition as compared to the Geo-Located condition, where participants may experience discomfort from the fact that someone knows where they are without them explicitly revealing it (people may feel monitored even way beyond what they truly are). If the data supported this hypothesis, they would suggest that time and effort reduction made possible by geo-location is actually outweighed by the perceived intrusiveness of the geo-location technology, and that simply presenting people with their location information being pre-populated in a Census form may not be an effective strategy to incentivize online participation.

Results. We invited 403 participants (37% female, M_{age} = 29.8, SD = 9.4) from MTurk to participate in a "Survey on Ethical Behaviors," and paid them $0.3. A small percentage of participants preferred not to respond to at least one of the questions on sensitive behaviors (8.7%, 8.2% and 6.8% in the Control, Geo-Located and Requested Location conditions, respectively), and this percentage did not differ significantly across conditions (all p-values larger than .1), suggesting that the inclusion of the option "Never" as a possible response to these questions soothed concerns arising from admitting certain sensitive behaviors. Responses to the 16 sensitive behavior questions were averaged to an overall willingness to disclose score (Cronbach's α = .802) and compared across conditions.

Supporting our predictions, participants were more likely to answer questions on sensitive behaviors in the Control condition ($M_{disclosure}$ = 1.09, SD = .60) as compared to the Geo-Located ($M_{disclosure}$ = .96, SD = .59, p = .03) and the Requested Location ($M_{disclosure}$ = .97, SD = .61, p =

.04) conditions. However, contrary to our expectations, willingness to disclose was the same between the latter two conditions ($p > .1$). We suspect this could be due to some error in the detection of the location of participants. The percentage of correct zip codes as reported by participants in the Geo-Located condition is probably an over-estimation, since these participants had no incentive in truly reporting their location: out of the 132 participants in this condition, 40% reported that the captured zip code was correct. However, based on the correspondence between the captured zip code and the one provided by participants in the Requested Location condition, the software we used (geo-location tool provided by the Qualtrics company, a leader in online survey development) only captured the correct 5-digit zip code in 25 out of 134 cases, was off by the last digit in 15 cases, and by the last two digits in 26 other cases. This implies that, for approximately half of our sample, the last three digits of the captured zip code were incorrect, which may have decreased the feeling of being tracked, reducing privacy concerns.[4]

The main result is robust to other specifications. In particular, we analyzed the data not only by calculating disclosure scores, but also as a panel, where the single respondent represented the grouping variable, and the list of 16 sensitive questions constituted the panel dimension. The dependent variable (*Disclosure*, or how participants responded to each sensitive question) was coded as a dummy, equal to 0 if the answer was "Never" and 1 otherwise. We then estimated a panel probit model with this binary dependent variable as a function of the condition in which participants were randomized (obviously, one of the 3 conditions has to be omitted for multi-collinearity; the results are similar whether it is the Geo-Located or the Requested Location condition to be dropped, so we only report the results for the model where the Geo-Located condition is dropped) and demographics (age, gender and race; due to the vast majority of whites in the sample and the sparse representation for all other races, race was coded as 1 for whites and 0 otherwise). The results are reported in Table 2. Confirming our previous analysis, participants in the Control condition were on average 13.8% more likely to admit to sensitive behaviors than participants in the Geo-Located condition, while participants in the Requested Location condition were just as willing to admit as participants in the Geo-Located condition. Gender and race also showed statistically significant effects, with males disclosing more than females, and whites disclosing more than the participants of other races. A panel logit estimation gives essentially the same results.

As far as the Census-related questions go, we observed a ceiling effect: all participants answered all of them. We suspect that this result could be due to the perceived low sensitivity of these questions. In order to overcome this effect, in a future iteration of the experiment we may try to "make" these questions more intrusive. One way to do this could be to manipulate the identity of the requester: certain questions may be perceived as tame if taken in the context of a research study, run by a research institution, but they may be perceived as quite intrusive if requested by the Government, or by other entities (advertising companies, insurance companies, and so on). We test this conjecture in Experiment 1C.

In a future iteration of the study, we may also include different measures of privacy concerns (such as the General Privacy Concern Index; Westin & Louis, 1991) since the four items we used did not explain the effect we found on the sensitive behaviors questions (p-values > .1 for pair-

[4] In order to test this conjecture, in another experiment we tested the effectiveness of showing only the first two digits of the zip code, which are usually correct, followed by three asterisks, and emphasized that this was done in order to preserve the privacy of participants (see Experiment 2 below).

wise comparisons across all items). Namely, participants did not report different propensity to provide false responses, concerns about self-incrimination, concerns about data privacy or confidentiality. In hindsight, we realize that the last two questions may have been interpreted as a request to evaluate the trustworthiness of the researchers (in general, quite highly rated), rather than to report the discomfort due to the experimental manipulation.

3.1.2 Experiment 1B

Methods. Experiment 1B was identical to Experiment 1A, but we collected data from a different population – namely, a pool of participants managed by a North Eastern American University.

Results. We recruited 186 participants (60% female, M_{age} = 26.8, SD = 10.8) from a pool of students and non-students managed by a North Eastern American University. They were invited to take the same "Survey on Ethical Behaviors" used in Experiment 1A, and compensated with participation in a lottery for a $50 Amazon gift card. A minority of participants preferred not to respond to one or more of the questions on sensitive behaviors (15.9%, 9.4% and 9.8% in the Control, Geo-Located and Requested Location conditions, respectively). Although, perhaps surprisingly, slightly more participants preferred not to answer in the Control condition, this percentage did not differ significantly across conditions (all p-values larger than .1). Responses to the 16 sensitive behaviors questions were averaged to an overall willingness to disclose score (Cronbach's α = .819) and compared across conditions.

Again, confirming our hypothesis, participants were more likely to answer questions on sensitive behaviors in the Control condition ($M_{disclosure}$ = 1.21, SD = .75) as compared to the Geo-Located ($M_{disclosure}$ = 1.00, SD = .69, p = .05) and the Requested Location ($M_{disclosure}$ = .91, SD = .49, p < .01) conditions. Unexpectedly, willingness to disclose seemed slightly higher in the Geo-Located condition as compared to the Requested Location condition, but the difference was not statistically significant (p > .1). Consistent results were obtained estimating probit or logit panel models.

Interestingly, participants in the Geo-Located condition who reported that their estimated location (down to zip code) was correctly identified, were less willing to disclose ($M_{disclosure}$ = 0.86, SD = .44) than those who reported their location to be incorrect ($M_{disclosure}$ = 1.29, SD = .97, p = .018). No difference was found in the Requested Location condition between participants (unaware that their location was identified independently of them providing it) who reported to be in the same location (down to zip code) as the one estimated by the geo-location tool. This finding would lend some support to our explanation for the insignificant difference in willingness to disclose between the Requested Location and the Geo-Located conditions: if the geo-location procedure is successful, people do respond by revealing less, but if the procedure is not accurate, then the manipulation loses its effectiveness. Notice, in fact, that the mean disclosure score in this latter case is no different than the one in the Control condition. It should be noted that this result was not found with the MTurk sample (Experiment 1A): it is possible that participants in that experiment were overall less truthful in reporting whether the estimated location was correct or not, or it could be that the geo-location tool was more accurate for the university participant pool, as most of them connected from a location on or close to campus, for which the zip code was indeed captured correctly.

The same ceiling effect found in Experiment 1A for Census-related questions was observed here, with all but three participants answering all of them. Furthermore, and again replicating the results in Experiment 1A, participants across the various conditions did not report different propensity to provide false responses, concerns about self-incrimination, concerns about data privacy or confidentiality.

3.1.3 Experiment 1C

Methods. In order to test whether the ceiling effect we observed for the Census-related questions in Experiments 1A and 1B was due to the low perceived sensitivity of such questions, we recruited 402 participants from MTurk to evaluate, on a scale from 1 (Not at all) to 7 (Very much), the level of intrusiveness of nine of the sensitive behaviors questions we used in the previous experiments, all the Census-related questions therein, and other PII requested in the 2010 Census form (see the Appendix for the full list of questions). We randomly assigned participants to one of four conditions, differing by the entity requesting the data (researchers, the U.S. Government, the Census Bureau specifically, and advertisers). Based on the results in Experiments 1A and 1B, we expected Census-related questions to be, in general, perceived as less intrusive than questions on sensitive behavior. Furthermore, due to recent media attention on surveillance programs by the U.S. Government, we hypothesized that Census questions, and specifically PII, would be perceived as more intrusive if asked by governmental institutions or advertisers than by researchers.

We made it clear to participants that they were not to respond to the questions, but just to evaluate how sensitive they evaluated them. Therefore, in order to avoid confusion, we did not ask for age and gender, which were indeed present in the 2010 Census forms. Hence, we cannot report the demographics for this particular sample, but we have no grounds to believe that they should be significantly different from the demographics of the MTurk sample used in Experiment 1A, since recruiting procedures were kept identical.

Results. We constructed two scores by averaging the sensitive behaviors questions (Cronbach's $\alpha = .906$) and the Census-related questions (Cronbach's $\alpha = .932$), respectively. Although we did find that Census-related questions are perceived as less intrusive ($M_{intrusive} = 3.12$, $SD = 1.31$) than sensitive behaviors questions ($M_{intrusive} = 4.74$, $SD = 1.66$, paired $t(401) = -15.971$, $p < .001$), the pattern for perceived intrusiveness as a function of the entity requesting the data was opposite to the one we predicted. Census-related questions were, on average, perceived as more intrusive if requested by researchers (or advertisers) than if requested by governmental institutions (irrespective of whether the requester was the Government in general or, more specifically, the Census Bureau; see Table 2 for descriptive statistics and Figure 1 for a graphical representation). Sensitive behaviors questions, on the other hand, were perceived as more intrusive if the requester was a governmental institution than if it was a research institution (see Table 1 for descriptive statistics and Figure 2 for a graphical representation).

Discussion. Experiments 1A and 1B showed that willingness to disclose sensitive information decreases if people are aware that their location is being tracked. Although the experiments did not show an effect specifically for Census-related questions, this was possibly due to the perceived low sensitivity of the questions presented to participants, which may generate a ceiling effect with everybody willingly responding to all questions. While Experiment 1C suggested that Census-related questions are not perceived as privacy intrusive, and are even perceived as

relatively less intrusive when requested by governmental institutions as compared to research institutions or advertisers, this study was hypothetical, and it could not rule out the possibility that, when presented in real life with this set of questions coming from the Census Bureau, people may indeed find them intrusive (and therefore, subject to the effect found in Experiments 1A and 1B for sensitive behaviors questions). In fact, the result of Experiment 1C could be explained by the fact that the terms of service of the platform we used to recruit respondents does not allow researchers to request PII, which may have caused MTurk workers to rate PII as more sensitive if requested by researchers or advertisers than by governmental institutions. Future iterations of the experiment should address the ceiling effect seen for Census-related questions, for instance manipulating purpose of data collection and use of collected data; methods used to extract geo-location data; accuracy of geo-location data.

3.2 Experiment 2

Experiment 2 tackled some of the problems encountered in Experiments 1A and 1B, and also expanded on Experiment 1C by measuring actual willingness to disclose. Specifically, it tested the hypotheses that: 1) by geo-locating participants correctly, there would be significant pairwise differences in disclosure rates between the three conditions used in Experiments 1A and 1B, with disclosure rates being highest in the Control condition, lowest in the Geo-Located condition, and the Requested Location condition being in between; 2) willingness to disclose varies depending on the institution requesting the information (Census Bureau, generic Government institution or Researchers); and 3) there is an interaction of geo-location and institution type, with the strongest effect on willingness to disclose being observed for Government institutions.

Method. We recruited MTurk Workers for a 5-minute "Study on Past behaviors and habits" and paid them $0.50. The study used a 3x3 between-subjects design where we manipulated institution requesting for information as in Experiment 1C (but dropping the Advertisers group), and geo-location as we did in Experiments 1A and 1B. The only difference was that, rather than showing a (possibly incorrect) 5-digit zip code, we only showed (together with city, State and Country) the first two digits, followed by 3 asterisks. We also mentioned in the instructions that, for their privacy, participants were only shown the first two digits of their zip code, in order to increase salience of geo-location concerns. As dependent variables, we used the same questions as in Experiments 1A and 1B, but we also added some other Census-related questions (education, occupation, income, living in a rural area, and children living in the house) in order to increase the perceived intrusiveness of the overall set of questions.

Results. 694 MTurk Workers (41.2% female, M_{age} = 31.1, SD = 10.6) participated in this study (Workers who participated in Experiment 1A could not take this second study). A 2x2 factorial ANOVA on the sensitive behaviors questions, with age, gender and race as covariates, revealed a main effect of geo-location ($F(2,682) = 4.165$, $p < .05$) and type of institution requesting the data ($F(2,682) = 4.493$, $p < .05$), but there was no significant interaction. Moreover, replicating the main result of Experiment 1A, planned contrasts revealed that willingness to disclose was higher in the Control conditions that in the two geo-location conditions ($t(685) = 3.22$, $p = .001$), but that the Requested Location and the Geo-Located condition did not differ from each other ($t(685) = .48$, $p > .10$). Notice that, in this Experiment, 165 out of 229 (72%) participants reported that the tracked location was correct down to the zip code level. This suggests that accuracy of geo-location may not be the main driver of the effect: simply mentioning location (whether it's displayed or explicitly requested) is enough to trigger privacy concerns and reduce willingness to

disclose sensitive information. Planned contrasts also revealed that willingness to disclose was higher if Researchers requested it as compared to Census or Government ($t(685) = 3.40$, $p = .001$). Interestingly, willingness to disclose looked higher for Government than for Census in the Control condition (see Figure 3), and the opposite seemed to be occurring for the two location treatments, but the differences were not statistically significant (hence the lack of significance in the interaction term).

Finally, the privacy concern score showed a symmetric trend to the average disclosure of sensitive information. A 2x2 ANOVA, with the same covariates used for the previous dependent variable, showed a marginally significant main effect of geo-location ($F(2,677) = 2.36$, $p < .10$), a main effect of the institution requesting the data ($F(2,677) = 3.70$, $p < .05$), but no significant interaction.

Discussion. Experiment 2 showed that the Institution requesting the data affects not only perceived intrusiveness of hypothetical questions, but also actual disclosure. In particular, there seems to be higher concerns and lower willingness to disclose sensitive information to Governmental institutions. Still, lower perceived sensitivity of Census-related questions caused a ceiling effect, by which almost everybody (627 out of 694 participants) provided all the requested information (31 skipped one question, 18 skipped two, and 18 skipped three or more), irrespective of requesting institution and geo-location.

3.3 Experiment 3

In a third experiment, we tested whether the results of Experiment 1C could be altered by priming participants to think about surveillance. We expected to see that, with no priming, Census-related questions may be perceived as less intrusive when requested by governmental institutions than researchers, whereas surveillance priming would act as an alarm bell, increase the perceived level of intrusiveness of Census questions, and possibly flip the relationship observed in Experiment 1C.

Methods. Similarly to Experiment 1C, we recruited 603 MTurk Workers to rate the level of intrusiveness of Census-related and sensitive questions, and paid them $0.50. Again, since we made it clear to participants that demographic questions were not to be answered, but only rated in terms of perceived intrusiveness, we did not collect demographics, but have no reason to believe that they should be different than Experiment 1A. We used the same measures as in Experiment 1C and assigned participants to one of six conditions, in a 3x2 between-subjects design where we manipulated the alleged entity requesting the data (Census, Government, or Researchers) and the presence of surveillance priming. The priming was implemented as an anagram to be solved by the participant before starting to answer questions about perceived intrusiveness. In the priming and no-priming conditions, we provided, in random order, the 7 letters of the names “Snowden” and “Clinton,” respectively, and asked participant to solve the anagram by identifying the name of a famous person that used all 7 letters. Notice that we used names with the same length, associated one way or another with the US Government, and arguably controversial figures for different reasons.

Results. Most participants were able to reconstruct the two last names: 247 out 296 (83.4%) participants in the Surveillance Priming condition answered “Snowden,” and 270 out of 307 (87.95%) participants in the No Surveillance Priming condition answered “Clinton.” The

difference in correct guesses was not statistically significant, but we included a binary variable (equal to 1 if participants solved the anagram correctly) as a covariate in our analysis, since participants who were not able to identify Snowden's name were not effectively primed.

Similarly to Experiment 1C, we constructed two scores by averaging the sensitive behaviors questions (Cronbach's $\alpha = .919$) and the Census-related questions (Cronbach's $\alpha = .931$), respectively. We replicated the result from Experiment 1C by which Census-related questions are perceived as less intrusive ($M_{intrusive} = 3.06$, $SD = 1.38$) than sensitive behaviors questions ($M_{intrusive} = 4.70$, $SD = 1.69$, paired $t(602) = -19.197$, $p < .001$).

Using the binary variable for whether participants correctly solved the anagram as a covariate, a 3 (Institution) x 2 (Surveillance Priming) MANOVA revealed a main effect of the requesting institution on perceived intrusiveness of the Census questions ($F(2, 596) = 5.476$, $p < .01$), but the surveillance priming we used showed no significant effect ($F(1, 596) = .673$, $p > .05$). The interaction effect was also not significant ($F(2, 596) = .492$, $p > .05$). For the sensitive behaviors questions, on the other hand, we observed both a main effect of requesting institution ($F(2, 596) = 15.721$, $p < .001$) and surveillance priming ($F(2, 596) = 4.327$, $p < .05$), but no interaction effect ($F(2, 596) = .753$, $p > .05$).

As post-estimation tests revealed, we replicated the result from Experiment 1C by which Census-related questions were, on average, perceived as more intrusive if requested by researchers ($M_{intrusive} = 3.25$, $SD = 1.15$) than if requested by the Census Bureau ($M_{intrusive} = 2.80$, $SD = 1.42$, $t(399) = -3.494$, $p < .01$). Moreover, perhaps surprisingly and unfortunately for the Census Bureau, Census-related questions were, on average, perceived as more intrusive if requested by the Bureau itself than if requested by the Government ($M_{intrusive} = 3.12$, $SD = 1.52$, $t(399) = -2.154$, $p < .05$). No statistically significant difference was found between the level of intrusiveness of Census questions if requested by researchers or the Government in general ($t(402) = -.989$, $p > .05$; see Table 3 for descriptive statistics).

Sensitive behaviors questions, on the other hand, were perceived as more intrusive if the requester was the Census Bureau ($M_{intrusive} = 5.08$, $SD = 1.73$, $t(399) = 5.443$, $p < .001$) or Government ($M_{intrusive} = 4.84$, $SD = 1.68$, $t(402) = 4.070$, $p < .001$) than if it was a research institution ($M_{intrusive} = 4.19$, $SD = 1.53$; see Table 3 for descriptive statistics). There was no statistical difference in perceived intrusiveness between the Census Bureau and the Government conditions ($t(399) = -1.386$, $p > .05$).

Discussion. Although the priming method we used showed a significant effect on the sensitive behaviors questions, it was not effective in increasing the perceived level of intrusiveness of Census-related questions, and thus, was not strong enough to cause a reversal in the effect found in Experiment 1C. For future iterations, we may consider stronger priming stimuli, such as reading a passage about the US Government surveillance programs recently brought to the attention of the media, or videos of Edward Snowden's revelations.

3.4 Experiment 4

In a final experiment, we brought together the features of both Experiment 1A and Experiment 3 in order to test for different willingness to disclose to various institutions as a function of a surveillance prime.

Methods. We recruited 601 MTurk Workers (43% female, M_{age} = 31.2, SD = 10.3) for a Study on "Past behaviors and habits," and paid them $0.50. We used the same dependent variables (sensitive questions, census-related questions, and questions about feeling monitored) and the same design to introduce geo-location as the one used in the Geo-Located condition of Experiment 2. Furthermore, we used the same surveillance primes used in Experiment 3. The questions were the same as the ones used in Experiment 2.

We randomly assigned participants to one of six conditions, in a 3x2 design where we manipulated the alleged entity requesting the data (Census Bureau, US Government or Researchers institution) and the presence of the surveillance prime used in Experiment 3. Consistent with the "chilling" effect found in recent work (Marthews & Tucker, 2014), we expected participants to be less willing to disclose to governmental institutions if a surveillance prime was provided, since, especially at this point in time, when the U.S. Government is at the center of media attention for data collection programs, such as PRISM, surveillance primes may heighten the feeling of being monitored.

Results. Most participants were able to reconstruct the two last names: 224 out of 300 (74.7%) participants in the Surveillance Priming condition answered "Snowden," and 264 out of 301 (87.7%) participants in the No Surveillance Priming condition answered "Clinton." The difference was statistically significant ($p < .001$), so, as we did for Experiment 3, we included the binary variable for guessing the right last name as a covariate in our analysis.

We were still unable to eliminate the ceiling effect observed for Census-related questions, with a negligible fraction of participants preferring not to answer (537 participants answered all questions, 38 skipped one, 10 skipped two, 16 skipping 3 or more). As for the sensitive behavior questions, a 3x2 ANOVA on the willingness to disclose score, with age, gender, and the binary variable for exposure to surveillance priming as covariates, revealed a main effect of institution ($F(2,592) = 3.93$, $p < .05$), but no effect of priming and no significant interaction (see Figure 6). Planned contrasts showed that participants were more likely to disclose to researchers (M = 1.01, SD = .57, $t(598) = 2.397$, $p < .05$) than to Census or Government (M = .88, SD = .59). Interestingly, the overall level of concern for the Census Bureau or the Government, as measured by the average of our 4-item scale (Cronbach's $\alpha = .82$) was slightly lower with surveillance priming (M = 2.51, SD = 1.11) than without (M = 2.55, SD = 1.07), but the difference was not statistically significant ($t(298) = -.54$, $p > .1$).

Discussion. Experiment 4 was designed to test the hypothesis that participants would be more concerned about their privacy towards Governmental institutions as compared to researchers in the presence of surveillance priming. Instead, consistent with the results found in Experiment 3, where we did not directly measure disclosure (but only perceived intrusiveness of questions), Experiment 4 showed that participants are not significantly affected by surveillance primes when answering Census-related questions or sensitive behavior questions, but they are less willing to disclose sensitive information to Governmental institutions than to researchers.

4. Conclusions

Experiment 1 suggests that willingness to disclose what is perceived to be sensitive information depends on other information, namely geo-location, being already available to the requester. Specifically, willingness to provide sensitive information decreases if one is aware that one's

location is being tracked and showed on the screen, or explicitly requested, as compared to a control condition in which location is not tracked. Census related information, on the other hand, may not be perceived as sensitive, and therefore one might conclude that pre-population of location information in Census forms could effectively affect online participation rate. Actual Census forms, though, did receive some criticism for their privacy intrusive nature, and because, as shown in Experiments 1A and 1B, sensitive information is subject to the effects of geo-location technologies, the Census Bureau should take into account citizens' privacy concerns when implementing geo-location tools.

Experiments 1C and 2 complement the previous experiments by showing that not only geo-location, but also the type of institution requesting the data affect perceived sensitivity and willingness to disclose. Although geo-location concerns seem to affect these DVs equally across institutions, it seems to be the case that participants are less willing to provide sensitive information to governmental institutions than to researchers. We should point out though, that this effect was not found for questions that are included in the Census forms.

Experiments 3 and 4 suggest that, while surveillance primes may not affect perceived intrusiveness of, and willingness to respond to Census questions, they may increase the perceived intrusiveness, and reduce willingness to provide sensitive information, which reinforces the idea that, if governmental institutions, which have been at the center of media attention for surveillance programs and invasion of citizens' privacy, plan to collect sensitive information, they may face reluctant responders. This may suggest that, if actual information requested in Census forms is perceived as intrusive, pre-population of location information in Census forms may not create an effective incentive to obtain a high participation rate to the online version of the Census, as time and effort saved in completion of the form would be overridden by privacy concerns arising from location tracking.

Different strategies may be considered in order for the online Census initiative to become successful, such as campaigns focusing on the completion of the form as a duty, something that, one way or the other (online or offline) must be done, so the choice would be between a paper format that cannot come pre-populated in any field, or an online format, that could potentially be quicker to complete. Campaigns may also emphasize that geo-location is very different from location tracking: with geo-location, the government would be able to identify the location of a citizen at a specific point in time (when she is filling the online Census form), not to track her over time, with the two activities exposing the citizen to very different risks (Shokri, Theodorakopoulos, Danezis, Hubaux & Le Boudec, 2011). Another aspect to emphasize could be the reason why geo-location is used: citizens should have it very clear in their minds that the only purpose for the government to geo-locate them is to save the Government (and thus, ultimately, citizens themselves) money, and to save citizens time while filling a form. Moreover, it should be emphasized that such information will not be used for other purposes.

Besides location information, which could be easily obtained both in the case of completion of the Census form on a laptop (by capturing the IP address) or on a mobile device (using GPS data), other types of information could be pre-populated once the citizen identified herself with first name and last name. Through Title 26 of the US Code, the Census Bureau has access to administrative data, such as data from the Department of Motor Vehicles (DMV) of each State. DMV has data on date of birth (and therefore age) of anyone possessing a driver's license, thus allowing for these fields to be automatically filled. This may provide an additional incentive for

citizens to fill the Census form online as opposed to paper, but it may also be perceived as privacy intrusive if it is unexpected, and if it is not clear to individuals how the Census Bureau may have obtained that information.

ACKNOWLEDGMENTS

This research was supported by NSF Research grant NCRN-MN #1130706, “Data Integration, Online Data Collection, and Privacy Protection for Census 2020.”

References

Acquisti, A., John, L. K. and Loewenstein, G. (2012). The Impact of Relative Standards on the Propensity to Disclose. *Journal of Marketing Research*, 49(2), 160-174.

Agarwal, R., Sambamurthy, V. and Stair, R. M. (2000). Research report: The evolving relationship between general and specific computer self-efficacy – An empirical assessment. *Information Systems Research*, 11(4), 418-430.

Agre, P.E. and Rotenberg, M. (1997). Technology and Privacy: The New Landscape. Cambridge MA: MIT Press.

Aloudat, A., Michael, K., Chen, X. and Al-Debei, M. M. (2013). Social Acceptance of Location-Based Mobile Government Services for Emergency Management. *Telematics and Informatics*, 31(1), 153-171.

Ashworth, L. and Free, C. (2006). Marketing dataveillance and digital privacy: Using theories of justice to understand consumers' online privacy concerns. *Journal of Business Ethics*, 67(2), 107-123.

Bapna, R and Sundararajan, A. (2010). Building institutions through identity. The Wall Street Journal, September 29. Available at: http://www.livemint.com/Opinion/hR9gaKUrvwgXbasgH210GP/Building-institutions-through-identity.html.

Barkhuus, L. and Dey, A. K. (2003). Location-based services for mobile telephony: a study of users' privacy concerns. In *Proceedings of the 9th International Conference on Human-Computer Interaction*.

Brandimarte, L., Acquisti, A. and Loewenstein, G. (2013). Misplaced Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science*, 4(3): 340-347.

Brush, A. J., Krumm, J. and Scott, J. (2010). Exploring end user preferences for location obfuscation, location-based services, and the value of location. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, ACM, pp. 95-104.

Census Bureau (2010a). Costs of the Census: Some good news. Director's Blog. Available at: http://directorsblog.blogs.census.gov/2010/08/10/costs-of-the-census-some-good-news/.

Census Bureau (2010b). The Numbers Are In: 72 Percent of Nation's Households Mail Back 2010 Census Forms. Available at: http://www.census.gov/2010census/news/releases/operations/the-numbers-are-in.html.

Chau, P. Y. K. (2001). Influence of computer attitude and self-efficacy on IT usage behavior. *Journal of End User Computing*, 13(1), 26-33.

Cvrcek, D., Kumpost, M., Matyas, V., & Danezis, G. (2006). A study on the value of location privacy. In *Proceedings of the 5th ACM workshop on Privacy in electronic society*, ACM, pp. 109-118.

Doheny-Farina, S. (1994). The Last Link: Default = Offline, Or Why Ubicomp Scares Me. *Computer-mediated Communication*, 1(6), 18-20.

Economist, The (2011). Costing the count. June 2. Available at: http://www.economist.com/node/18772674?story_id=18772674&CFID=165420949&CFTOKEN=32425086

Edmonston, B. and Schultze, C. (1995). Modernizing the U.S. Census. The National Academies Press.

Egelman, S., Tsai, J., Cranor, L. F. and Acquisti, A. (2009). Timing is everything?: The effects of timing and placement of online privacy indicators. In *Proceedings of the 27th international conference on Human factors in computing systems*, ACM, pp. 319-328.

Garfinkel, S. (2001). Database Nation: The Death of Privacy in the 21st Century. O'Reilly & Associates.

Gideon, J., Cranor, L., Egelman, S. and Acquisti, A. (2006). Power strips, prophylactics, and privacy, oh my! In *Proceedings of the second symposium on Usable privacy and security*, pp. 133-144.

Gruteser, M. and Liu, X. (2004). Protecting privacy, in continuous location-tracking applications. *Security & Privacy, IEEE*, 2(2), 28-34.

Harper, R. H. (1995). Why people do and don't wear active badges: A case study. *Computer Supported Cooperative Work* (CSCW), 4(4), 297-318.

Hong, W., Thong, J. Y. L., Wong, W. M. and Tam, K. Y. (2001). Determinants of user acceptance of digital libraries: An empirical examination of individual differences and system characteristics. *Journal of Management Information Systems*, 18(3), 97-124.

Huang, L., Matsuura, K., Yamane, H. and Sezaki, K. (2005). Enhancing wireless location privacy using silent period. In *Proceedings of the Wireless Communications and Networking Conference, IEEE*, Vol. 2, pp. 1187-1192.

John, L. K, Acquisti, A. and Loewenstein, G. (2011). Strangers on a plane: Context-dependent willingness to divulge sensitive information. *Journal of consumer research*, 37(5), 858-873.

Johnson, R. D., & Marakas, G. M. (2000). Research report: The role of behavior modeling in computer skills acquisition – Toward refinement of the model. *Information Systems Research*, 11(4), 402-417.

Joinson, A. N. (2009). Privacy concerns, trust in government and attitudes to identity cards in the United Kingdom. In *Proceedings of the 42nd Hawaii International Conference on System Sciences*, IEEE, pp. 1-10.

Junglas, I. A. and Spitzmuller, C. (2005). A research model for studying privacy concerns pertaining to location-based services. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, IEEE.

Kaasinen, E. (2003). User Needs for Location-aware Mobile Services. *Personal and Ubiquitous Computing*, 7(1), 70-79.

Luarn, P. and Lin, H. H. (2005). Toward an understanding of the behavioral intention to use mobile banking. *Computers in Human Behavior*, 21(6), 873-891.

Marthews, A. and Tucker, C. (2014). Government Surveillance and Internet Search Behavior. Available at SSRN: http://ssrn.com/abstract=2412564 or http://dx.doi.org/10.2139/ssrn.2412564.

Mervis, J. (2011). Can the US afford the next Census? *Science Insider*, available at http://news.sciencemag.org/2011/04/can-u.s.-afford-next-census.

Minch, R. P. (2004). Privacy issues in location-aware mobile devices. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, IEEE.

Myles, G., Friday, A. and Davies, N. (2003). Preserving privacy in environments with location-based applications. *Pervasive Computing, IEEE*, 2(1), 56-64.

Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M. and Rao, J. (2009). Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing*, 13(6), 401-412.

Schilit, B. N., LaMarca, A., Borriello, G., Griswold, W. G., McDonald, D., Lazowska, E., Balachandran, A., Hong, J. and Iverson, V. (2003). Challenge: Ubiquitous location-aware computing and the "Place Lab" initiative. In *Proceedings of the 1st ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*, ACM, pp. 29-35.

Shokri, R., Theodorakopoulos, G., Danezis, G., Hubaux, J. P. and Le Boudec, J. Y. (2011). Quantifying location privacy: The case of sporadic location exposure. In *Privacy Enhancing Technologies*, Springer Berlin Heidelberg, pp. 57-76.

Thomas, L., Little, L., Briggs, P., McInnes, L., Jones, E. & Nicholson, J. (2013). Location tracking: views from the older adult population. *Age and ageing*, 42(6), 758-763.

Toch, E., Cranshaw, J., Drielsma, P.H., Tsai, J. Y., Kelley, P. G., Springfield, J., Cranor, L. F., Hong, J. and Sadeh, N. (2010). Bridging the gap between physical location and online social networks. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, ACM, pp. 119-128.

Tsai, J. Y., Egelman, S., Cranor, L. and Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, 22(2), 254-268.

Tucker, C. (2011). Social Networks, Personalized Advertising, and Privacy Controls. In *Workshop of Economics and Information Security*.

USA Today (2012). Opposing view: Census survey intrusive and expensive. July 15. Available at: http://usatoday30.usatoday.com/news/opinion/story/2012-07-15/Census-American-Community-Survey/56241350/1.

Weiser, M. and Brown, J. S. (1996). Designing calm technology. *PowerGrid Journal*, 1(1), 75-85.

Westin, A. and Louis, H. & Associates (1991). Harris-Equifax Consumer Privacy Survey. Tech. rep. Conducted for Equifax Inc.

White, T. B., Zahay, D. L., Thorbjørnsen, H. and Shavitt, S. (2008). Getting too personal: Reactance to highly personalized email solicitations. *Marketing Letters*, 19(1), 39-50.

Zweig, J. and Webster, J. (2002). Where is the line between benign and invasive? An examination of psychological barriers to the acceptance of monitoring systems. *Journal of Organizational Behavior*, 23, 605-633.

Tables and Figures

Table 1. Marginal effects for panel probit estimation – Experiment 1A. Standard errors in brackets. ** indicates significance at the 1% level; * at 5% level.

	Disclosure
Control	.138* (.070)
Requested_Location	.006 (.070)
Male	.194** (.061)
Age	-.003 (.003)
White	.213** (.065)
	Number of groups = 403 Wald χ^2 (5) = 26.78

Table 2. Descriptive Statistics for Experiment 1C

	Institution	Mean	Std. Deviation	N
Census_scale	Researchers	3.292011	1.2671673	99
	Government	2.875223	1.2935118	102
	Census	2.931129	1.4425579	99
	Advertisers	3.387701	1.1733705	102
	Total	3.121664	1.3110152	402
Sensitive_scale	Researchers	4.205387	1.5799324	99
	Government	5.071895	1.6026691	102
	Census	4.890011	1.7526926	99
	Advertisers	4.767974	1.5968970	102
	Total	4.736595	1.6596210	402

Table 3. Descriptive Statistics for Experiment 2

	Institution	Surveillance	Mean	Std. Deviation	N
Census_scale	Census	0	2.7186	1.34300	105
		1	2.8956	1.49527	94
		Total	2.8022	1.41611	199
	Government	0	3.1518	1.55002	100
		1	3.0865	1.50679	102
		Total	3.1188	1.52488	202
	Researchers	0	3.1595	1.15754	102
		1	3.3455	1.13487	100
		Total	3.2516	1.14731	202
	Total	0	3.0062	1.36956	307
		1	3.1133	1.39572	296
		Total	3.0588	1.38235	603
Sensitive_scale	Census	0	4.9132	1.78936	105
		1	5.2589	1.65731	94
		Total	5.0765	1.73257	199
	Government	0	4.6233	1.83799	100
		1	5.0523	1.49764	102
		Total	4.8399	1.68439	202
	Researchers	0	4.1906	1.41951	102
		1	4.1856	1.64396	100
		Total	4.1881	1.53090	202
	Total	0	4.5787	1.71332	307
		1	4.8251	1.66050	296
		Total	4.6996	1.69070	603

Figure 1. Estimated mean for Census-related questions score – Experiment 1C

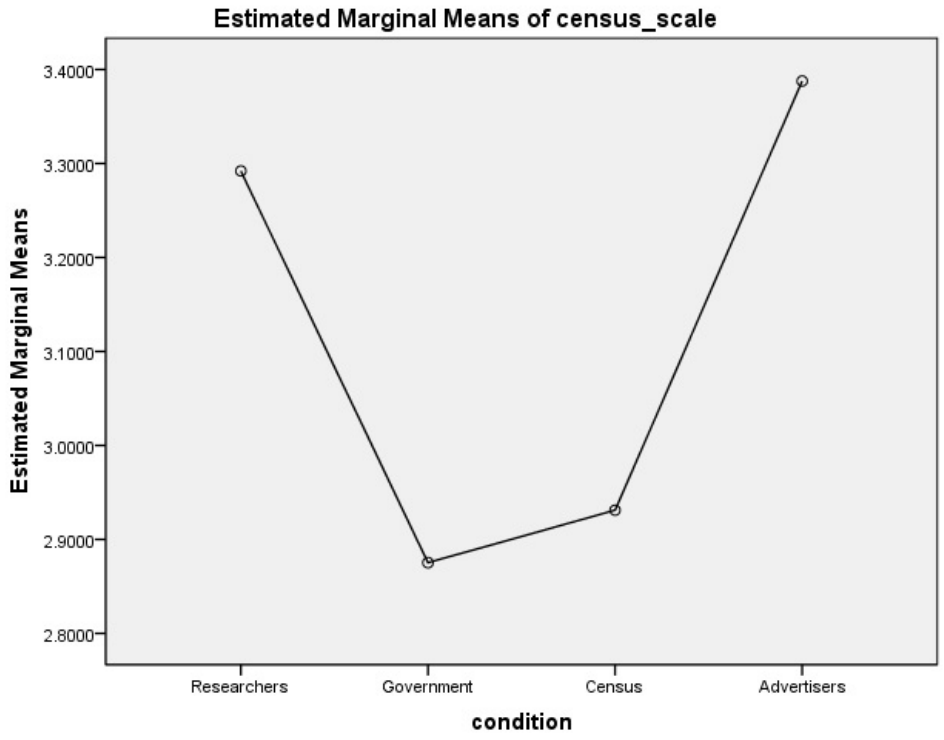


Figure 2. Estimated mean for sensitive behaviors questions score – Experiment 1C

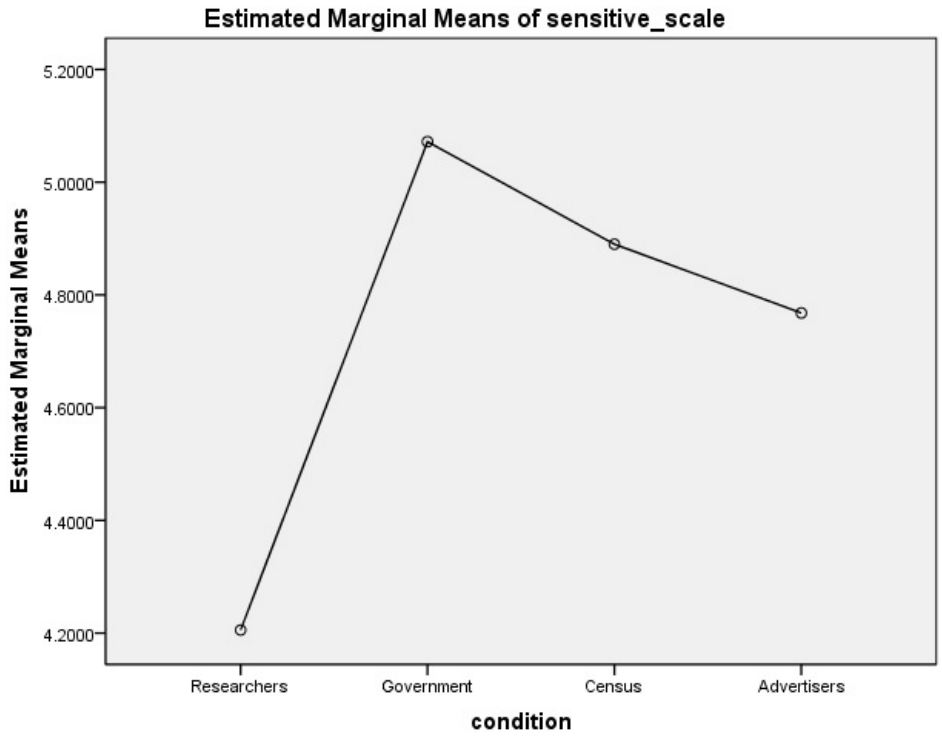
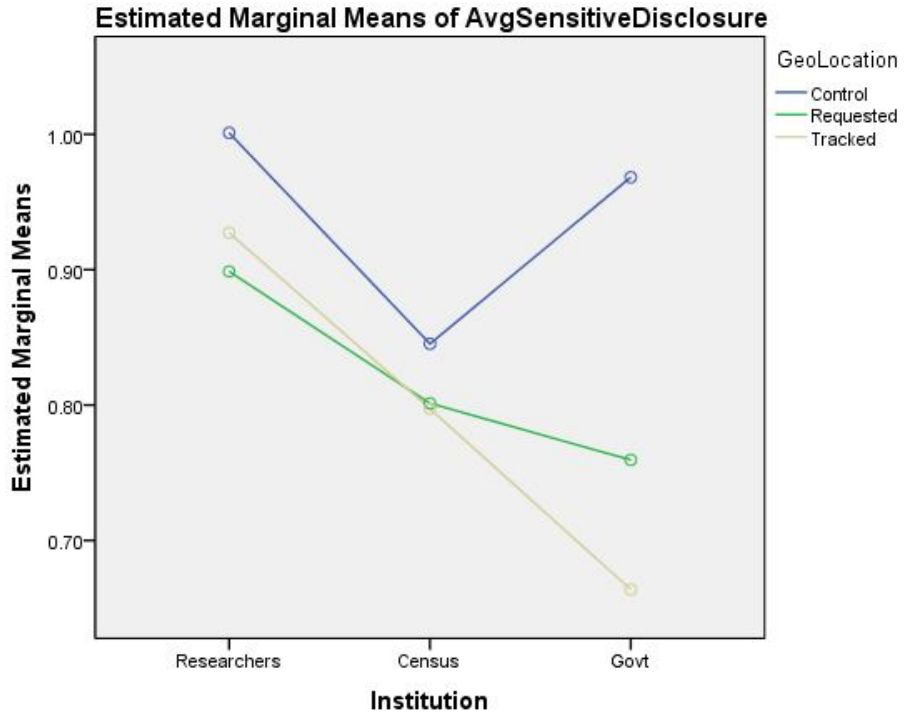


Figure 3. Estimated mean for sensitive behaviors questions score – Experiment 2



Covariates appearing in the model are evaluated at the following values: Gender = 1.4121, Age = 31.1311, White = .7666

Figure 4. Estimated mean for Census-related questions score – Experiment 3

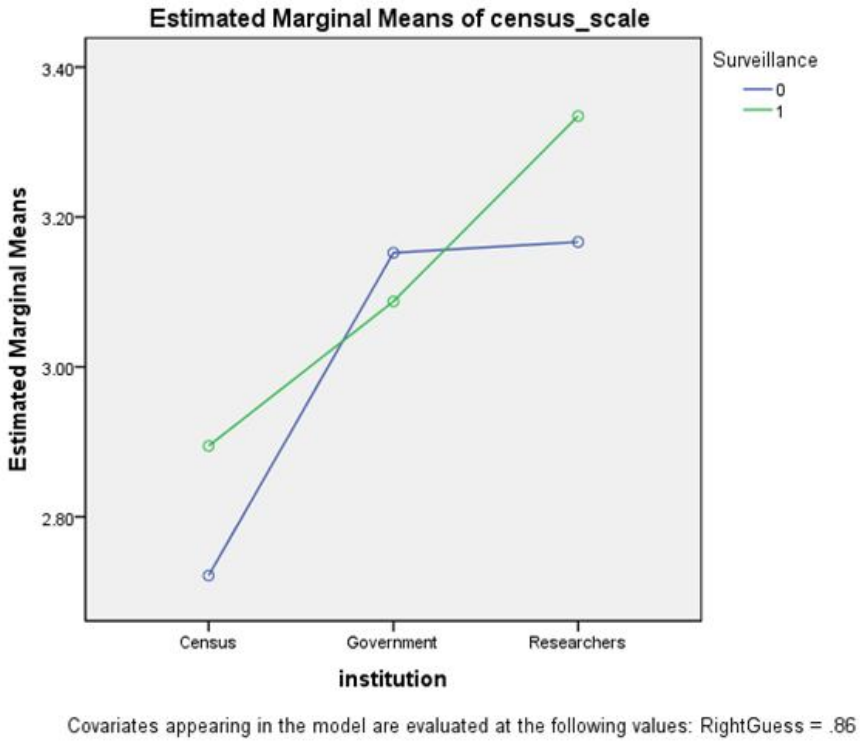
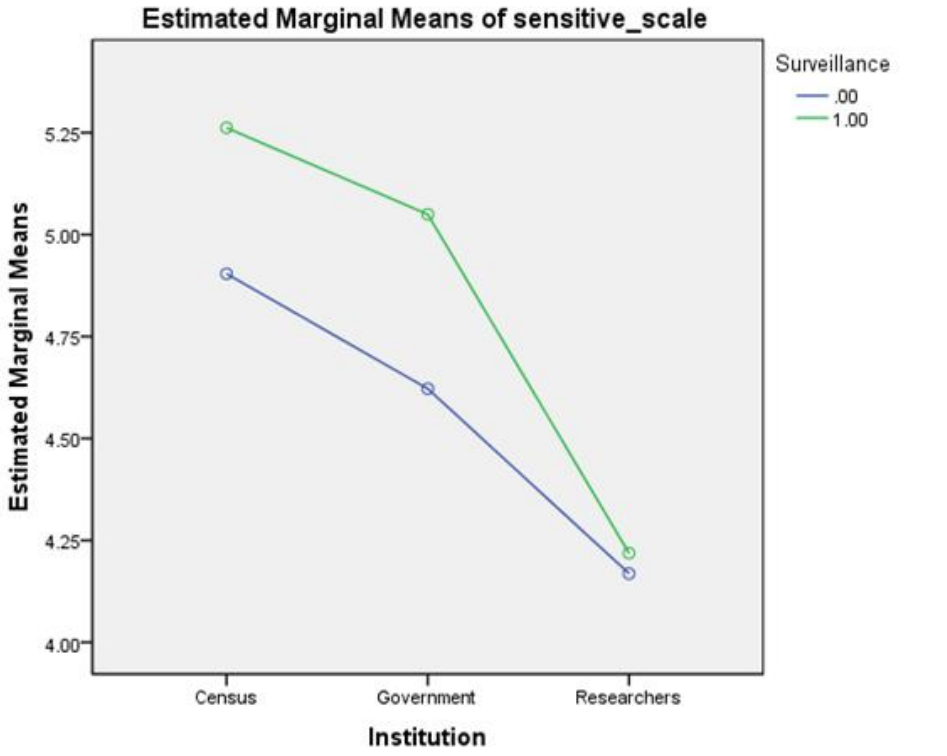
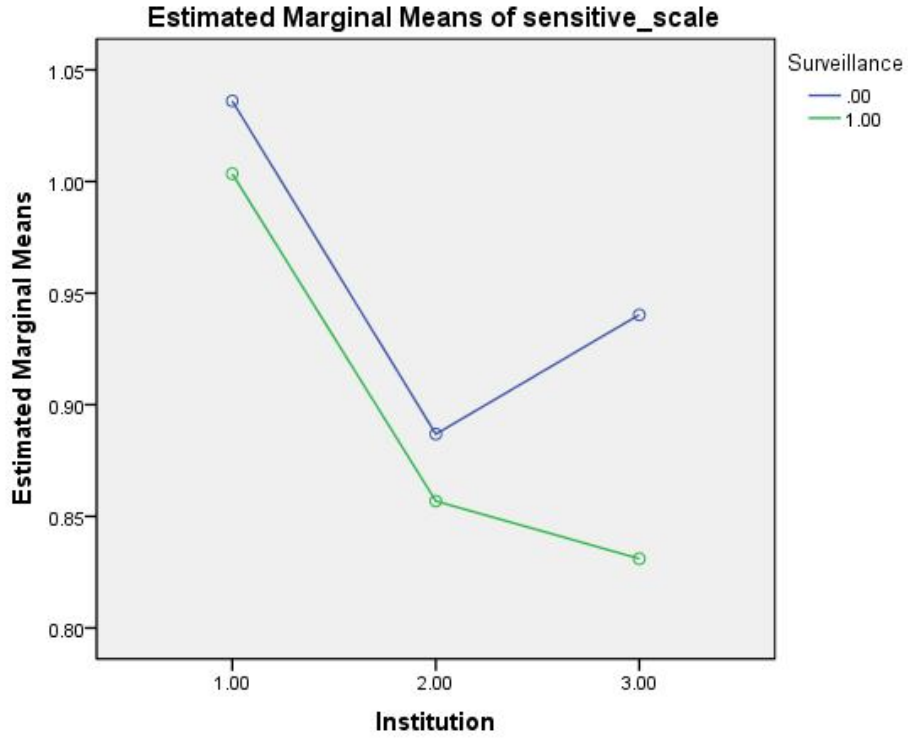


Figure 5. Estimated mean for sensitive behaviors questions score – Experiment 3



Covariates appearing in the model are evaluated at the following values: RightGuess = .8574

Figure 5. Estimated mean for sensitive behaviors questions score – Experiment 4



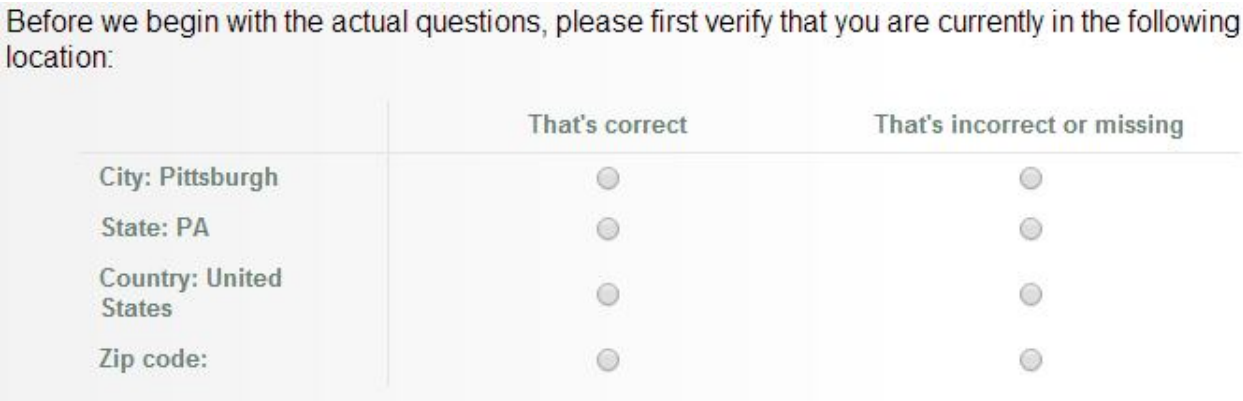
Covariates appearing in the model are evaluated at the following values: gender = 1.4293, age = 31.1597, GuessRight = .8120

Appendix

Geo-location manipulation used in Experiments 1A and 1B – screenshots.

Control condition: no indication of location tracking

Geo-Located condition:



Requested Location condition:

All conditions: the order was randomized for the block of sensitive behaviors questions. Also, the order of the two blocks was randomized. The possible responses were: Never, Once, Twice, Sometimes, Often, I prefer not to say.

Sensitive behaviors ((*) indicates that the question also appeared in Experiment 1C):

"In the next page, you will be shown a list of various behaviors that people sometimes engage in. We ask that you read each of those and indicate how often (if at all) you personally did each of the described behaviors.

For each of the following, please indicate how often (if at all) you personally did the described action. Have you ever:

(*) Had sex with the current partner of a friend? (1)

(*) Lied about your age? (2)

Cheated at sports or games? (3)

(*) Tried to gain access to someone else's phone or email without their knowledge or consent? (4)

(*) Made up a serious excuse, such as grave illness or death in the family, to get out of doing something? (5)

Called in sick when you were not sick? (6)

Stolen anything worth more than $25? (7)

Let a friend drive after s/he had too had much to drink or had used drugs? (8)

Masturbated in a public space, such as a public restroom? (9)

Looked at pornographic material? (10)

(*) Cheated on your tax return? (11)

(*) Fantasized about doing something terrible (e.g., torturing) to someone? (12)

(*) Encouraged someone to drink when you were trying to seduce them? (13)

(*) Smoked marijuana or an illegal drug? (14)

(*) Made a false insurance claim? (15)

Taken nude pictures of yourself or your partner? (16)

Census-related questions: (they all appeared in Experiment 1C, together with first name, last name, date of birth, and phone number)

What is your gender? (17)

What is your age? (18)

What is your race? (19)

Where do you currently live? (20) (House, Apartment, Mobile home, Other - please specify)

Do you currently live with other people? (21)

Do you own the place where you live? (22) (Yes, No - I'm renting, No - I'm living there without paying rent)

Do you have a mortgage or loan? (23)

Final questions:

Thank you for your responses. Now we would like to ask about your thoughts and feelings as you answered. Please indicate to what extent you personally agree or disagree with each of the following statements: (Likert scale from Strongly Disagree (1) to Strongly Agree (5))

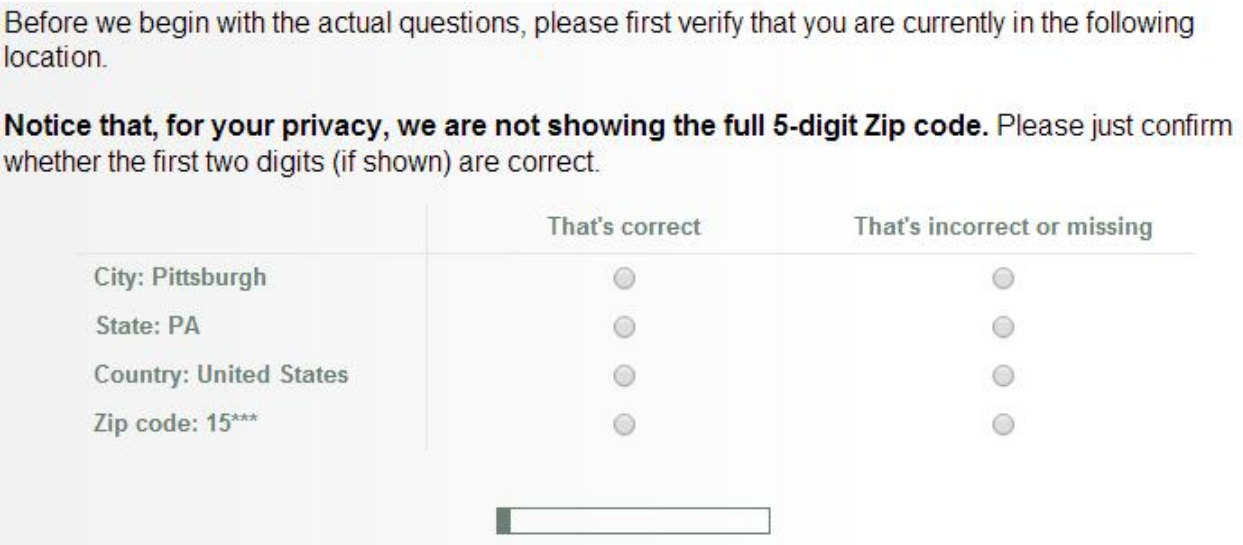
Some of the responses I gave are actually not true (24)

I was concerned about incriminating myself (25)

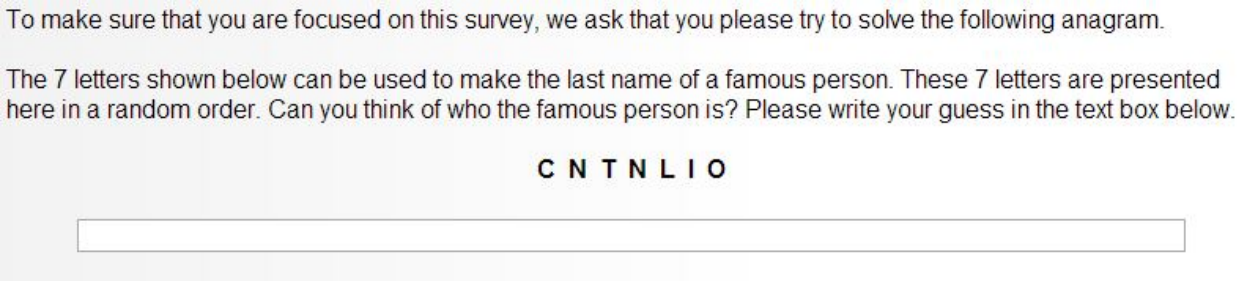
I was concerned about whether my responses would truly be private (26)

I was concerned about whether my responses would be kept confidential (27)

Geo-location instrument used in Experiments 2 and 4 – screenshot.



Surveillance priming instrument used in Experiments 3 and 4 – screenshots.



To make sure that you are focused on this survey, we ask that you please try to solve the following anagram.

The 7 letters shown below can be used to make the last name of a famous person. These 7 letters are presented here in a random order. Can you think of who the famous person is? Please write your guess in the text box below.

S W D N N O E