A Comparison of Six Sample Providers Regarding Online Privacy Benchmarks

Sebastian Schnorf
Google
Zurich, Switzerland
sebschnorf@google.com

Aaron Sedley
Google
Mountain View, CA
asedley@google.com

Martin Ortlieb
Google
Zurich, Switzerland
mortlieb@google.com

Allison Woodruff
Google
Mountain View, CA
woodruff@acm.org

ABSTRACT

Researchers increasingly utilize online tools to gather insights. We show how privacy comfort as measured by questionnaires differs across various survey sample providers. To investigate potential differences depending on provider, we fielded a small set of privacy-related benchmark questions regarding past experience, present and future concerns to six major US survey providers. We found substantial differences depending on privacy benchmark and provider population, illustrating that privacy-related research may yield different insights depending on provider choice.

1. INTRODUCTION

In research, online tools such as Amazon's Mechanical Turk and Google Consumer Surveys (GCS) are increasingly employed because they allow quick and convenient gathering of insights. However, there is an ongoing debate about how the populations of these providers differ regarding key characteristics [1, 2, 11, 12, 13; see 2 for a comprehensive review of data sources]. Biases, for example towards more heavy Internet users, pose a challenge to research because we might underrepresent or not be aware of valid concerns of a general population segment. In the case of product-related research, we might for instance fail to identify barriers to more widespread product usage. Accordingly, as privacy researchers increasingly utilize these tools to measure attitudes and behaviors, it is important to assess underlying variances for privacy-related dimensions to create a better understanding of various survey provider populations. The arrival of GCS as an alternative to Mechanical Turk in comparison and contrast to other online survey panels prompted us to compare the landscape of the offerings more comprehensively.

GCS [4] presents a small set of questions to users, providing them access to content that is behind a pay-wall and might not be otherwise available to them for free. GCS is easy to deploy, and several comparison studies [5, 8] found GCS to be more accurate than both probability and non-probability based Internet panels. For example, Pew Research states that GCS "…appears to conform closely to the demographic composition of the overall Internet population... In addition, there is little evidence so far that the GCS sample is biased toward heavy Internet users" [13].

The contribution of this paper is an assessment of the variances in response to privacy-related questions across sample providers. In particular, we were interested in how GCS performs versus other platforms. We provide a better understanding of different provider populations and by that hope to inform future privacy projects.

Symposium on Usable Privacy and Security (SOUPS) 2014, July 9-11, 2014, Menlo Park, CA.

2. METHOD

Measuring privacy attitudes and behavior is a challenging endeavor. In order to keep this project at a manageable size, we decided to field a small set of basic privacy questions regarding past, present and hypothetical experiences. In this approach we were well aware that we might miss many concepts for measuring privacy such as those explored by the Westin questions [14, 15]. In addition to the privacy questions, we also asked about standard socio- as well as technographic information (Internet use frequency, technology optimism and adoption behavior). For the complete questionnaire, see the Appendix.

We fielded our questions to six major US survey providers in December 2013. Note that collection time and response rate may greatly vary depending on fielding time. Furthermore the cost per response ranged from as low as $1 to as high as $12.

Table 1. Overview of survey providers

	GCS	MTU	GFK	PSR	USA	SSI
Name	Google Consumer Surveys	Amazon Mechanical Turk	GFK Knowledge Panel	Princeton Survey Res.	uSamp Online Market Res.	Survey Sampling Interntl
Respondents(N)	1101	1112	1031	835	1100	1115
Time to collect	24h	5.5h	3d	4d	8d	8d
Response	18.1%	n/a	~30%	11.8%	n/a	9.1%
Sampling	River sample	Opt-in	Prob-based (ABS)	Prob-based (RDD)	Opt-in panel	Opt-in panel
Recruiting	Pay-wall	Qualif. Mturker	Online Comm	Phone	Online Comm	Online Comm
Demographics	Inferred	Excluded	Included	Included	Included	Included
Weighting Ref.	CPS [4]	No	CPS [4]	CPS [4]	CPS [4]	CPS [4]
Representation	US pop.	US Mturker	US pop.	US pop.	US pop.	US pop.

GCS, as introduced earlier, is a new method for performing Internet surveying. In GCS, response data is supplemented with inferred demographics based on IP and site traffic information. This demographic information, such as age and gender, is then used to weight individual responses to represent the general US population [4]. In our study, we fielded both a short version of GCS with one or two individual questions, as well as a longer version that currently allows up to ten questions.

Mechanical Turk (MTU) is an online crowdsourcing marketplace that enables individuals or businesses to co-ordinate the use of human intelligence to perform tasks that computers are currently unable to do. The MTU platform can also be used to field surveys to its workers. GFK [6] and PSR [16] are both international market research companies working with probability-based sampling. In the case of GFK, they randomly select individuals from a large pool of 97% of US household addresses, including people without Internet access [6]. Because GFK's and PSR's recruiting includes people who do not use the Internet, from these sources we screened for people who use the Internet for personal purposes. USA [19] and SSI [21] are both opt-in panels, meaning that these providers recruit people from online platforms and communities, and then, based on some screening criteria and demographic reference [4], select respondents for certain surveys.

3. RESULTS

We present our findings for present, hypothetical and past privacy experiences. We start by giving a descriptive overview of our data and then proceed to more detailed regressions models.

Table 2. Overview of privacy data (means / weighted means)

Survey Question	GCS[1]	MTU	GFK	PSR	USA	SSI
4. Privacy Discomfort General [1,5]	3.46 /3.42	3.35	3.60 /3.63	3.74 /3.70	2.64	2.79
5. Privacy Discomfort Organis. [1,5]	3.56 /3.55	3.14	3.50 /3.53	3.65 /3.60	2.54	2.71
6. Hypothetical Privacy Concern [1,5]	3.05 /3.06	2.85	2.91 /2.99	2.76 /2.77	3.00	3.00
7. Privacy Incident Experience (share)	0.23 /0.24	0.16	0.15 /0.14	0.20 /0.22	0.17	0.16

Table 2 shows an overview of our response data across the six survey providers (for the actual question posed, see the Appendix). Note again, that for GFK and PSR, we screened for people who use the Internet for personal purposes. In Table 2, we show means for our four key privacy questions. For the first two questions, we recoded the bipolar answer scale to values ranging from one to five, so a value of three relates to a neutral rating. The third question about hypothetical privacy concerns is on a uniform scale from one to five. The last question reports the share of respondents who reported having experienced a privacy incident.

Table 2 demonstrates variability in the response means across survey providers. However, for those providers working with a weighting factor, the results are not substantially different between weighted and un-weighted results. Note again that in case of GCS, weighting is applied to inferred demographics, whereas GFK and PSR employ weighting based on participants' responses to demographic questions. Due to only marginal differences between weighted and un-weighted figures (see Table 2) we decided to proceed with our more detailed analysis using un-weighted data only.

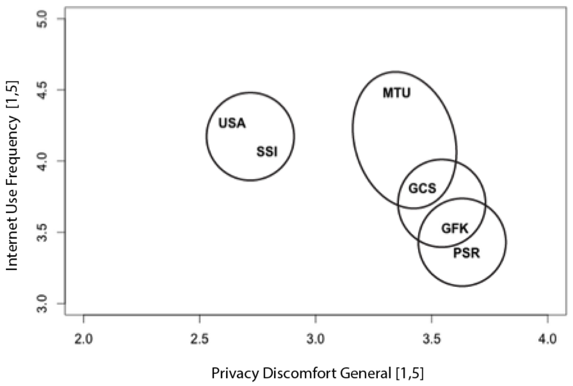


Figure 1. Internet Use vs. Privacy Discomfort (means)

Figure 1 plots the mean of current general privacy discomfort on the x-axis and one of our technographic variables, Internet use frequency, on the y-axis (for other technographic values see Table 6 in the Appendix). The circles indicate homogeneous groups for the privacy variable according to the Kruskal-Wallis test with p levels adjusted to the number of pairwise comparisons ($p < 0.05 / 15$). The two dimensional plot yields a typical pattern across our data: GCS does not differ substantially regarding current privacy discomfort from GFK and PSR, which use random sampling and are deemed most representative of the general population [6].

Table 3. Different response depending on baseline and question

Baseline Provider	Privacy Comfort General	Privacy Comfort Organis.	Hypothetical Privacy Concern	Privacy Incident Experience
GCS	PSR, USA, SSI	MTU, USA, SSI	MTU, PSR	GFK
MTU	GFK, PSR, USA & SSI	GCS, GFK, PSR, USA, SSI	GCS	
GFK	MTU, USA, SSI	MTU, USA, SSI		GCS
PSR	GCS, MTU, USA, SSI	MTU, USA, SSI	MTU	
USA	GCS, MTU, GFK, PSR	GCS, MTU, GFK, PSR	PSR	
SSI	GCS, MTU, GFK, PSR	GCS, MTU, GFK, PSR	PSR	

[1] For GCS we display figures of the longer survey version. We found differences to the short GCS version; in particular, there is much lower incident experience of .14 (N=18K). The difference is most probably due to the drop off rate of around 50% in the longer GCS version (for further comparisons see Table 6 in the Appendix).

In Table 3, we list all significantly different populations (p <0.05/15) depending on baseline provider and question asked. For instance, we observed significantly different responses for hypothetical privacy concern in GCS versus MTU. Note that in this table, random samplers GFK and PSR do not yield different results from each other.

Further below we elaborate on the differences in more detail. We use GCS as our baseline and control for socio- and technographic variables.

3.1 Privacy Comfort General

In the question about current privacy comfort, we asked respondents how comfortable or uncomfortable they are with online information about themselves that anyone can find and see (question 4 in the Appendix). To avoid priming users by directly asking about concerns, we asked about "privacy comfort" and used a bipolar answer scale (see Figure 2, upper right).

As shown previously, there is quite some variability in the overall mean of the response to this question. Figure 2 shows the full distribution of responses. The figure yields again the following response patterns (means): rather skeptical ratings for GFK (3.60) and PSR (3.74), more optimistic ratings for USA (2.64) and SSI (2.79), and in between are GCS (3.46) and MTU (3.35).

In the following analysis, we investigate how strongly the privacy bias persists if we control for gender, Internet use frequency, technology optimism and adoption behavior. We conducted this analysis using the GCS responses as our baseline.

According to Table 3, from a GCS perspective, we would expect different responses from PSR, USA and SSI. As can be seen in Table 3, the differences with PSR persist, however, at a rather low level of .13 +/- .06. The differences to USA and SSI are much stronger, suggesting a lower score of approximately - .40, which is quite substantial on a five-point scale. With MTU, a new effect emerges in the other direction (+ .41). So even though Mechanical Turkers are a population of heavy users, earlier adopters and tech optimists (see Table 6 in the Appendix), they tend to have more current privacy discomfort when compared with GCS respondents.

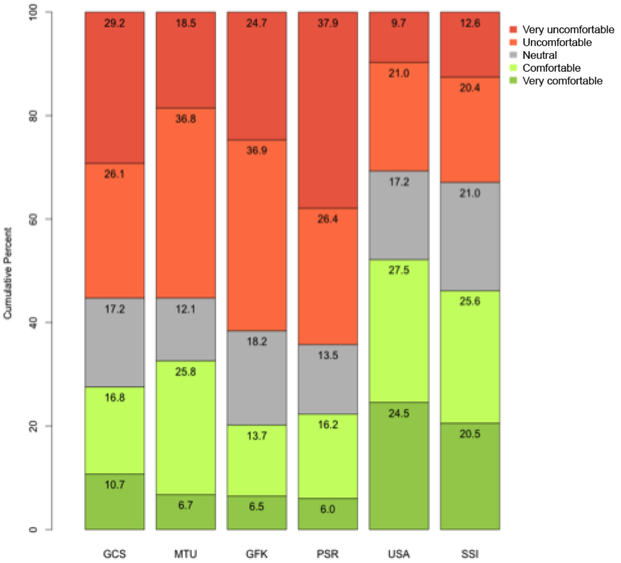
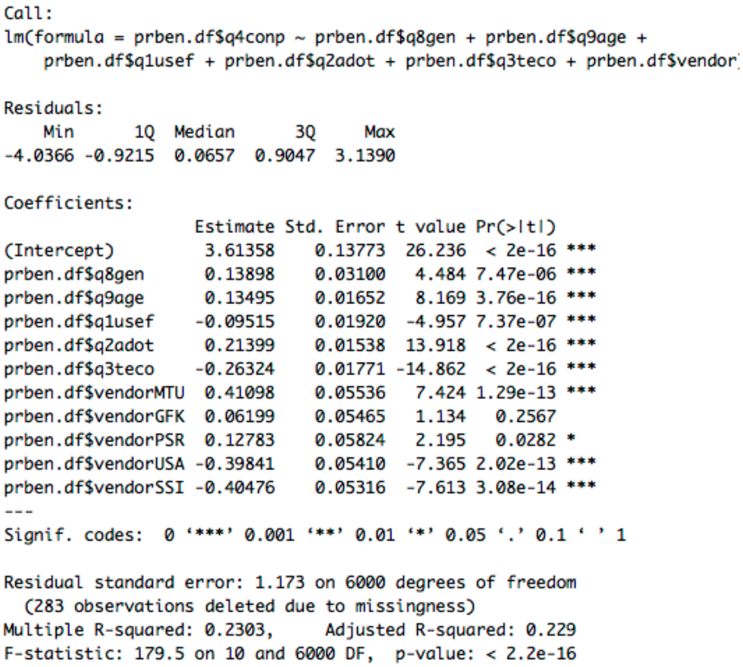


Figure 2. General privacy comfort response distribution

The small difference between multiple and adjusted R-square in the lower section of Table 4 shows that there is not much correlation between our control variables, which is a positive sign suggesting they cover different aspects of privacy attitudes.

Table 4. Regression with current privacy discomfort as outcome



```
Call:
lm(formula = prben.df$q4conp ~ prben.df$q8gen + prben.df$q9age +
    prben.df$q1usef + prben.df$q2adot + prben.df$q3teco + prben.df$vendor

Residuals:
    Min      1Q  Median      3Q     Max 
-4.0366 -0.9215  0.0657  0.9047  3.1390 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)         3.61358    0.13773  26.236  < 2e-16 ***
prben.df$q8gen      0.13898    0.03100   4.484 7.47e-06 ***
prben.df$q9age      0.13495    0.01652   8.169 3.76e-16 ***
prben.df$q1usef    -0.09515    0.01920  -4.957 7.37e-07 ***
prben.df$q2adot     0.21399    0.01538  13.918  < 2e-16 ***
prben.df$q3teco    -0.26324    0.01771 -14.862  < 2e-16 ***
prben.df$vendorMTU  0.41098    0.05536   7.424 1.29e-13 ***
prben.df$vendorGFK  0.06199    0.05465   1.134   0.2567    
prben.df$vendorPSR  0.12783    0.05824   2.195   0.0282 *  
prben.df$vendorUSA -0.39841    0.05410  -7.365 2.02e-13 ***
prben.df$vendorSSI -0.40476    0.05316  -7.613 3.08e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.173 on 6000 degrees of freedom
  (283 observations deleted due to missingness)
Multiple R-squared: 0.2303,	Adjusted R-squared: 0.229 
F-statistic: 179.5 on 10 and 6000 DF,  p-value: < 2.2e-16
```

Our interpretation of the results are that MTurkers may be more tech savvy than the SSI and USA panelists, and that their greater knowledge of technical issues may make them more sensitive to the potential for hacking and misuse of data online. They may be more afraid based on knowledge, whereas the GFK and PSR samples may be more afraid based on the unknown. SSI and USA panelists probably recognize there is some danger, but they may not care as much about their own personal privacy as the average person, considering they have opted in to answer surveys about themselves.

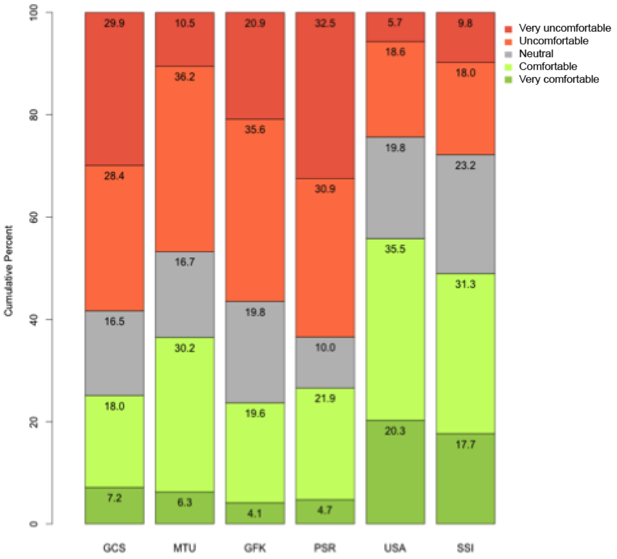


Figure 3. Privacy comfort re organizations response distribution

3.2 Privacy Comfort Related to Organizations

Similar to the previous question, we wanted to investigate current privacy comfort, but this time regarding a business or organization as these more specific concerns may be different (see question 5 in the Appendix).

Running the regression analysis as described in the previous chapter, but with data about organizational comfort, confirmed the differences noted in Table 3 for MTU, USA, and SSI. The significant effect with MTU was small (.11), however, here we saw much less privacy concerns with respondents from USA (-.61) and SSI (- .57). These two opt-in survey panels attracted relative heavy users and had substantially less privacy discomfort when compared to GCS, especially with organizations.

Another observation is that when comparing comfort on a general level, most of the populations we surveyed tended to have a higher level of general discomfort as compared to discomfort about organizations. This finding seconds insights of other studies on users perception of privacy [10, 17]. Interestingly, for the GCS population, the opposite is true: GCS respondents have a higher level of discomfort regarding organizations and a lower level of general discomfort. This finding may be related to the GCS survey administration method in which users are asked a question in order to access content; this method may confuse or concern some users.

3.3 Hypothetical Privacy Concerns

To gauge near future or imagined privacy concerns, we asked respondents to what extent they think information about themselves on the Internet, that is available to another person, business or organization, might cause negative experiences (see question 6 in the Appendix).

As shown in Table 2, we see not much variability across providers for this question, with means ranging from 2.76 to 3.05. GCS is marginally leading this list and from a GCS perspective (Table 3), the regression model predicts lower results with the MTU (- .30) and PSR populations (- .24). In the case of MTU, this difference is interesting because the MTU population is in the lead regarding current privacy discomfort.

3.4 Privacy Incident Experience

We also asked respondents if they had any negative privacy experience (see question 7 in the Appendix). As previously shown, the share of respondents with privacy incident experience in our study ranges from 15% to 23%. With the weighted data, GCS is leading this list with 24% of respondents with incident experience. Does this mean that for finding people with incident experience, GCS is a good option? Based on our project, this turns out to be only the case for the longer version of GCS, we had a much lower share of 15% in the short version of GCS (see Table 6 in the Appendix). One explanation for this difference is that at this stage of the longer questionnaire we have a drop-off rate of around 50%. Researchers using GCS to gauge incident experience need therefore carefully take into account the specific setup of their surveys.

The only difference we could confirm through the Kruskal-Wallis test was between GCS and GFK (see the last column in Table 3). Interestingly, when we run the regression, in this case a binominal model due to the dichotomous outcome variable, we see quite strong negative effects for all providers, except for PSR which is the only one using phone recruitment (see Table 5).

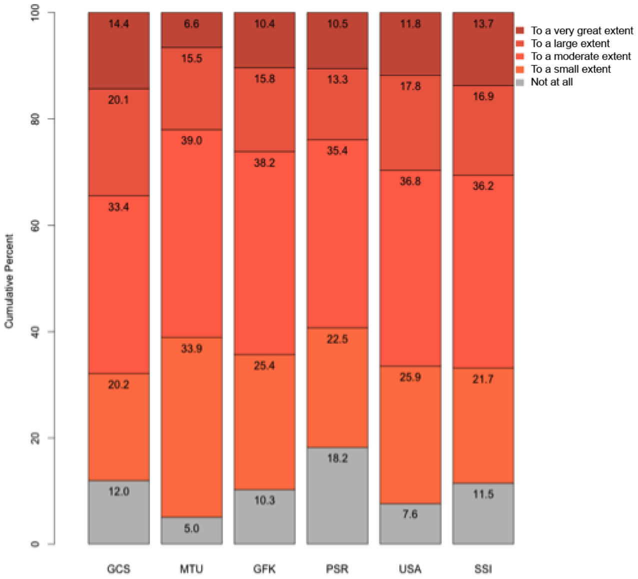
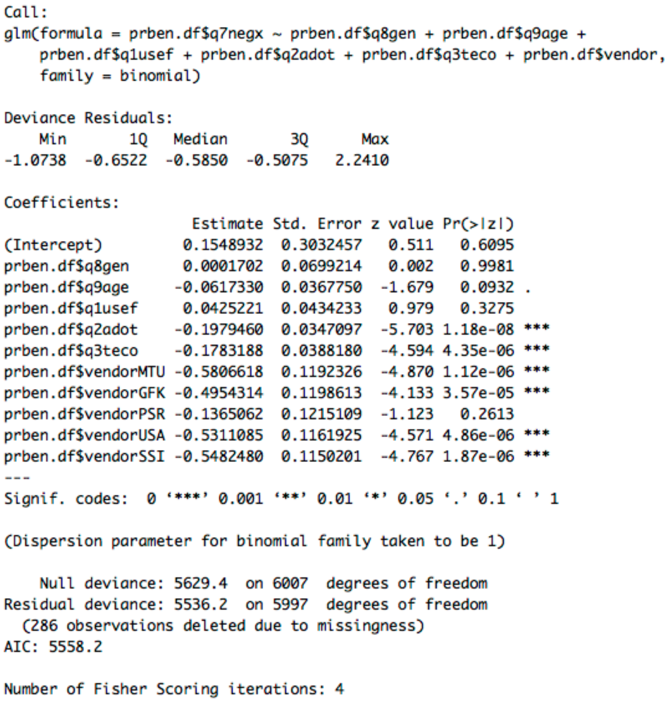


Figure 4. Hypothetical privacy concern response distribution

In general, self-administered surveys like those online show higher reports of sensitive behaviors (e.g. drug use) and lower social desirability effects (e.g. saying you voted in the last election even though you did not) [20]. It may be that online panelists are less sensitive to privacy incidents than average internet users, since they have agreed to join a panel and answer questions about themselves.

Table 5. Regression with privacy incidents as outcome



```
Call:
glm(formula = prben.df$q7negx ~ prben.df$q8gen + prben.df$q9age +
    prben.df$q1usef + prben.df$q2adot + prben.df$q3teco + prben.df$vendor,
    family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0738  -0.6522  -0.5850  -0.5075   2.2410

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)         0.1548932  0.3032457   0.511   0.6095
prben.df$q8gen      0.0001702  0.0699214   0.002   0.9981
prben.df$q9age     -0.0617330  0.0367750  -1.679   0.0932 .
prben.df$q1usef     0.0425221  0.0434233   0.979   0.3275
prben.df$q2adot    -0.1979460  0.0347097  -5.703 1.18e-08 ***
prben.df$q3teco    -0.1783188  0.0388180  -4.594 4.35e-06 ***
prben.df$vendorMTU -0.5806618  0.1192326  -4.870 1.12e-06 ***
prben.df$vendorGFK -0.4954314  0.1198613  -4.133 3.57e-05 ***
prben.df$vendorPSR -0.1365062  0.1215109  -1.123   0.2613
prben.df$vendorUSA -0.5311085  0.1161925  -4.571 4.86e-06 ***
prben.df$vendorSSI -0.5482480  0.1150201  -4.767 1.87e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5629.4  on 6007  degrees of freedom
Residual deviance: 5536.2  on 5997  degrees of freedom
  (286 observations deleted due to missingness)
AIC: 5558.2

Number of Fisher Scoring iterations: 4
```

One of the two following questions for those respondents with incident experiences was about the severity of the most negative experience (question 7a in the Appendix). The response mean was lowest with SSI (2.92) and highest with GFK (3.46). The GFK population, even though the lowest share of people with overall incident experience (15%), features the most severely “privacy-burnt” respondents.

The second follow-up question asked about the type of the most negative privacy incident (question 7b in the Appendix). Figure 5 shows that, with the exception of SSI, unwanted ads and spam are the most salient privacy issue. Reputation and financial harm are other prominent types of privacy incidents across the providers (the margin of error with a 95% confidence interval is around +/-3%.). Please note again that participants may well have had other incidents that were less serious and not reported (and different populations may assess “most negative” differently).

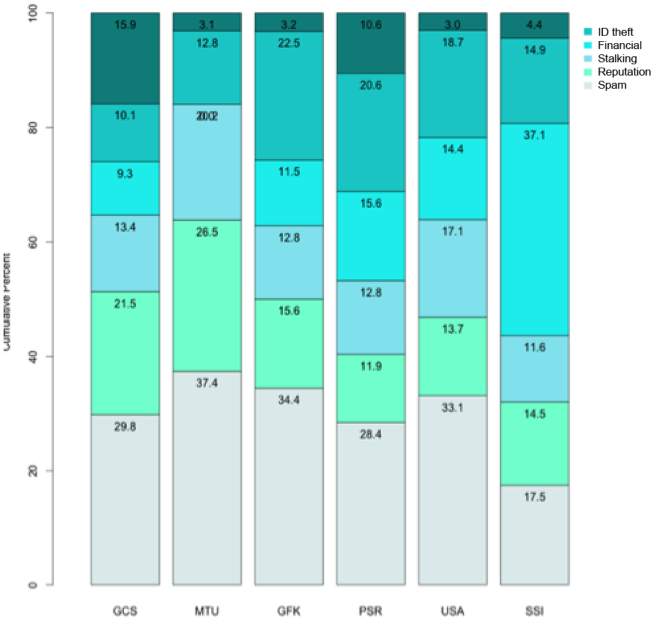


Figure 5. Type of most negative privacy incidents

Mentions of “Other” incidents were rare. Although we did not analyze them quantitatively, we briefly examined the free text. Most items appeared to describe duplicates of existing items (e.g., reputation damage), or unauthorized credit card use or account hijacking.

4. DISCUSSION

In this paper we identified variance in responses to privacy-related questions across six survey sample populations. Overall, the variations we have found suggest that how providers source participants, such as opt-in or probability-based sampling, meaningfully affect the population estimates. The major contribution of this research is a better understanding of each survey provider’s sample frame and its implications for privacy-related research:

- GFK and PSR, whose sampling methods are designed to maximize general population representativeness, show higher average privacy discomfort than other survey populations. GCS does not differ substantially from these random samples regarding current discomfort. We therefore tentatively conclude that for studies focusing on current privacy discomfort of people who are regular users of the Internet, GCS might be a good alternative to probability-based samples.
- Mechanical Turkers, even though a population of heavy users, early adopters and technology optimists, tend to have a higher level of current privacy discomfort. This finding shows that privacy researchers can explore privacy attitudes of higher technical complexity with MTU, but the generalizability of their findings to a broader audience might be limited due to issues such as feature comprehension.
- The two opt-in samples from USA and SSI both attract heavier Internet users that have substantially lower current privacy discomfort. Based on our findings and related research [3] we tend to conclude that researchers working with USA and SSI might underestimate privacy concerns.
- People in most of our survey provider populations tend to have a higher level of general privacy discomfort as compared to discomfort related to organizations. Interestingly, this is not the case for GCS. This might be a result of the GCS setup on third party websites. For researchers this means that with GCS, they might be able get more skeptical privacy voices regarding organizations such as for brand comparisons.
- There is not much variability among the six survey providers regarding hypothetical privacy concern. However, when taking GCS as a baseline, the MTU and PSR sample has lower hypothetical privacy concerns. So even though more privacy concerned at present, Mechanical Turkers are not more pessimistic regarding the near future.
- Among all the surveys fielded, 15% to 24% of people have experienced privacy incidents. The highest proportion, among GCS respondents, might be a result of the specific setup in this study (multi-question, which yielded a high break-off rate) and point to the importance for researchers to monitor response rates. Regarding the nature of the most negative experience, unwanted ads and spam are most salient privacy issues experienced across most panels.

In this paper we studied differences across six sample providers with a few basic privacy questions. Further research could be conducted using an extended and refined set of questions. Furthermore, we expect that over time providers will look for new channels to source study participants, therefore respondent populations might change. Repeat comparative studies will help elicit these differences.

5. ACKNOWLEDGMENTS

Our thanks go to Mario Callegaro, Vasyl Pihur and Jessica Staddon for thoughtful comments at various stages of this project.

6. REFERENCES

[1] Berinsky, A., Huber, G. A. and Lenz, G. A. 2012. Evaluating Online Labor Markets for Experimental Research: Amazon’s Mechanical Turk http://dl.dropboxusercontent.com/u/7536991/Mechanical_Turk.pdf

[2] Callegaro, M. et al. (2014). A Critical Review of Studies Investigating the Quality of Data Obtained with Online Panels Based on Probability and Nonprobability Samples. Callegaro, M., et al. (Eds), *Online Panel Research, A Data Quality Perspective.* (pp 23-53) Chichester: Wiley.

[3] Couper, M. and Singer, E. 2013. Informed Consent for Web Paradata Use. *Survey Research Methods* 7(1), 57-67.

[4] CPS, Current Population Survey, http://www.bls.gov/cps/

[5] GCS, Google Consumer Surveys, Google Inc. 2013. http://www.google.com/insights/consumersurveys/

[6] GFK knowledge panel: http://www.gfk.com/Documents/GfK-KnowledgePanel.pdf

[7] Lee Rainie, Sara Kiesler, Ruogu Kang, and Mary Madden. Anonymity, Privacy, and Security Online. Pew Internet & American Life Project, September 2013.

[8] McDonald, P., Mohebbi, M. and Slatkin, B. 2012.Comparing Google Consumer Surveys to Existing Probability and Non-Probability Based Internet Surveys. Google Whitepaper. http://www.google.com/insights/consumersurveys/static/357791852901773780/consumer_surveys_whitepaper.pdf

[9] MTU, Amazon's Mechnical Turk. https://www.mturk.com/mturk/welcome

[10] Ortlieb, M. (in press) The Anthropologist's View on Privacy. *IEEE Security & Privacy*, 12 (3), 2-4.

[11] Paolacci, G., Chandler, J., and Ipeirotis, P. (2010). Running Experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411-419.

[12] Peer, E., Vosgerau, J., and Acquisti, A. (in press). Reputation as a Sufficient Condition for Data Quality on Amazon Mechanical Turk. *Behavior Research Methods*.

[13] Pew Research Center 2012. A Comparison of Results from Surveys by the Pew Research Center and Google Consumer Surveys.

[14] Ponnurangam K. and Cranor L. Privacy Indexes: A Survey of Westin's Studies. Tech. Report CMU-ISRI-5-138, Institute for Software Research International (ISRI), Carnegie Mellon University, December 2005.

[15] J.D. Power and SSI. Consumer Concerns About Data Privacy Rising: What Can Business Do? October 2013.

[16] PSR http://www.psrai.com/pdf/psrai_telephone_omnibus.pdf

[17] Rosenberg, A. 2010. 5 Essential Facebook Privacy Tips. http://mashable.com/2010/05/18/facebook-privacy-tips

[18] Sosik, V.S., Burzstein, E., Consolvo, S., Huffaker, D., Kossinets, G., Liao, K., McDonald, P., and Sedley, A. 2014 Online Micro-Surveys for User Experience Research. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, 889-892.

[19] SSI: Survey Sampling International http://www.surveysampling.com/

[20] Tourangeau, R. Yan, T. 2007. Sensitive Questions in Surveys. *Psychological Bulletin* 133 (5), 859-883.

[21] USA, uSamp (United Sample Inc.) http://www.usamp.com/panel.html

7. APPENDIX

Questionnaire:

1. For personal purposes, how often do you normally use the Internet?

Every hour or more often
Every few hours
Once or twice a day
Multiple times per week
Once per week or less often

2. Which of the following best describes when you buy or try out new technology?

Among the first people
Before most people, but not among the first
Once many people are using it
Once most people are using it
I don't usually buy or try out new technology

3. In general, how would you rate technology's impact on people's lives?

Very positive
Somewhat positive
Neither positive nor negative
Somewhat negative
Very negative

4. How comfortable or uncomfortable are you with information about yourself on the Internet that anyone can find and see?

Very comfortable
Somewhat comfortable
Neither comfortable nor uncomfortable
Somewhat uncomfortable
Very uncomfortable

5. How comfortable or uncomfortable are you providing information about yourself online to a business or organization?

Very comfortable
Somewhat comfortable
Neither comfortable nor uncomfortable
Somewhat uncomfortable
Very uncomfortable

6. To what extent do you think information about yourself on the Internet, that is available to another person, business or organization, might cause you negative experiences?

To a very great extent
To a large extent
To a moderate extent
To a small extent
Not at all

7. Have you had any negative experiences because information about yourself on the Internet was available to another person, business or organization?
[separate GCS with this screener question]

Yes (--> 7a & 7b)
No

7a. Recall the most negative experience you had due to information about yourself on the Internet. What consequences were there?
[multi-select]
[randomize, for MTurk order below]

Unwanted commercial offers or spam
Reputation damage or embarrassing situation
Stalking or harassment
Financial loss
Identity theft
Other: (please specify)
[for GCS, None of the above]

7b. Recall again the negative experience you had due to information about yourself on the Internet. How severe were the consequences?

Extremely severe
Very severe
Moderately severe
Slightly severe
Not at all severe

Additional Questions for MTurk & GCS

8. What is your gender?
- Male
- Female
- I prefer not to answer

9. What is your age?
- 18-24 years old
- 25-34
- 35-44 [For GCS, there is one bucket for 35-54]
- 45-54 [For GCS, there is one bucket for 35-54]
- 55 or older
- I prefer not to answer

10. Which of the following best describes where you live?
- in a city of at least 250,000 people
- within 50 miles of a 250,000+ city
- within 50 miles of a 50,000+ city
- further than 50 miles from a 50,000+ city

Table 6. Overview of all data (means / weighted means)

Survey Question	**GCS micro**	**GCS long**	**MTU**	**GFK**	**PSR**	**USA**	**SSI**
1. Usage frequency [1,5]	3.67 /3.69	3.81 /3.85	4.48	3.53 /3.46	3.36 /3.49	4.27	4.07
2. Adopter category [1,5]	3.45 /3.44	3.27 /3.26	2.62	3.58 /3.65	3.65 /3.59	2.44	2.77
3. Technology optimism [1,5]	3.69 /3.72	3.46 /3.79	4.33	2.68 /3.69	3.68 /3.63	4.34	4.09
4. Discomfort General [1,5]	3.62 /3.64	3.46 /3.42	3.35	3.60 /3.63	3.74 /3.70	2.64	2.79
5. Discomfort Organis. [1,5]	3.81 /3.78	3.56 /3.55	3.14	3.50 /3.53	3.65 /3.60	2.54	2.71
6. Hypothetical Concern [1,5]	2.78 /2.82	3.05 /3.06	2.85	2.91 /2.99	2.76 /2.77	3.00	3.00
7. Incident Experience	0.14 /0.13	0.23 /0.24	0.16	0.15 /0.14	0.20 /0.22	0.17	0.16
7a Incident Outcome	-	-	-	-	-	-	-
7b Incident Severity [1,5]	3.08 /3.12	-	3.34	3.46 /3.39	3.17 /3.09	3.03	2.92
8. Gender share female	-	0.41 /0.46	0.38	0.46 /0.51	0.49 /0.51	0.52	0.52
9. Age group [1,4]	-	3.00 /2.89	2.08	2.90 /2.95	3.09 /2.79	2.80	2.80