How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples

Elissa M. Redmiles*, Sean Kross†, and Michelle L. Mazurek*
*University of Maryland
{eredmiles, mmazurek}@cs.umd.edu
†University of California San Diego
seankross@ucsd.edu

***Abstract*—Security and privacy researchers often rely on data collected from Amazon Mechanical Turk (MTurk) to evaluate security tools, to understand users' privacy preferences and to measure online behavior. Yet, little is known about how well Turkers' survey responses and performance on security- and privacy-related tasks generalizes to a broader population. This paper takes a first step toward understanding the generalizability of security and privacy user studies by comparing users' self-reports of their security and privacy knowledge, past experiences, advice sources, and behavior across samples collected using MTurk (n=480), a census-representative web-panel (n=428), and a probabilistic telephone sample (n=3,000) statistically weighted to be accurate within 2.7% of the true prevalence in the U.S.**

Surprisingly, the results suggest that: (1) MTurk responses regarding security and privacy experiences, advice sources, and knowledge are more representative of the U.S. population than are responses from the census-representative panel; (2) MTurk and general population reports of security and privacy experiences, knowledge, and advice sources are quite similar for respondents who are younger than 50 or who have some college education; and (3) respondents' answers to the survey questions we ask are stable over time and robust to relevant, broadly-reported news events. Further, differences in responses cannot be ameliorated with simple demographic weighting, possibly because MTurk and panel participants have more internet experience compared to their demographic peers. Together, these findings lend tempered support for the generalizability of prior crowdsourced security and privacy user studies; provide context to more accurately interpret the results of such studies; and suggest rich directions for future work to mitigate experience- rather than demographic-related sample biases.

I. Introduction

A number of recent security and privacy studies have used data collected on Amazon Mechanical Turk (MTurk) to evaluate new tools and report on users' behavior [1]–[5]. While work from the social sciences has resulted in mixed findings about the validity of MTurk study results related to topics such as health behavior and politics [6], [7], little work in our field has examined the validity of security- and privacy-specific information collected on MTurk.

Due in part to concerns about the generalizability of MTurk responses, security and privacy researchers have begun to turn to near-census-representative but non-probabilistic web panels to sample users who better represent the demographics of the U.S. population [8]–[11]. These web panels are thought to be a relatively low cost, more representative alternative to MTurk. Again, however, no prior work has compared security and privacy research done using such panels, and related work in the social sciences has obtained mixed results [12]–[15].

The referenced validation papers from the social sciences provide important context. We argue, however, that further validation specific to our field is necessary, not only because of existing mixed results, but because security and privacy tool evaluations and surveys differ importantly from studies in other fields in at least three ways:

- Asking questions about online behavior on the internet is inherently different than asking questions about other behaviors (e.g., smoking). Questions about online behavior, including security and privacy behavior, may vary significantly depending on the internet skill of the respondents [16], [17], which may in turn vary depending on the platform used for data collection.
- Prior work offers limited evidence that demographics may not necessarily covary with responses about security and privacy topics [18], potentially differing from other social science topics previously measured in survey generalizability studies [19], [20].
- Security and privacy topics are rarely, if ever, queried in broad, general surveys (such as those conducted by government agencies), and thus prior work offers little insight into sample-related differences in users' responses about these topics.

Thus far, only two studies in our field have closely examined the quality of data collected using MTurk or other web panels [21], [22]. Both studies looked only at the results of privacy research: Kang et al. [22] compared MTurk survey results about privacy topics to responses from a probabilistic sample, and Schnorf et al. compared the results of four privacy questions deployed on six non-probabilistic, near-probabilistic, and probability-based web-panels, including MTurk, to each other [21]. There is further room for study, however, in two key respects: examining security more broadly (not just privacy) and evaluating more questions for both privacy and security, and comparing to a truly probabilistic, low-error-margin sample, which is currently believed to be the best available way to closely approximate generalizability to the general population [23].

In this paper, we take a first step toward filling this gap: we compare users' self-reports about their security and privacy behavior, knowledge, past experiences, and advice sources across surveys conducted on MTurk (n=480), using a nearly census-representative web-panel (n=428), and via a probabilistic telephone survey (n=3,000). Our work is: a) the first to study the generalizability of MTurk surveys about both privacy and security behavior, knowledge, experiences and advice sources, as compared to a probabilistic sample weighted to be representative of the entire U.S. population; b) the first to compare security and privacy data collected using a census-representative panel with other samples; and c) the first privacy or security study to explore the impact of weighting MTurk data to improve the generalizability of results. Unlike any prior work, we compare these samples not only at the macro level—comparing entire samples—but also by demographic subset. For example, we consider whether MTurk participants (hereafter, Turkers) who are 18 to 29 respond in line with 18-to-29-year-olds in the probabilistic sample.

We find that, surprisingly, MTurk responses to the security and privacy questions that we asked were more similar to responses from the weighted probabilistic sample than are the responses of the census-representative panel. In general, MTurk respondents tend to mirror the probabilistic sample – for those who are under the age of 50 – with regard to their reported advice sources (although they do report seeking out advice about security and privacy from websites with more frequency); negative experiences (they do report higher frequency of a few security-related experiences); and confidence in their knowledge about the majority of security and privacy topics. However, MTurk respondents of all ages universally report higher internet activity than the probabilistic sample; and panel respondents are more representative of the general population for older (50+) adults on all metrics.

To investigate what factors influence differences between MTurk responses and those from the probabilistic sample, we implemented a simple demographic-based statistical weighting. Such approaches have been successful in other fields to correct sample-related differences [7], [24]. This weighting did not significantly reduce the differences between the probabilistic and MTurk data, suggesting that computer-security surveys may differ from surveys about other behaviors in important and previously unexplored ways. For example, higher levels of online activity in the MTurk population—potentially an indication of internet skill and/or early technology adoption [25]—may be the root cause of response differences for digital security and privacy surveys, rather than demographic bias.

We also explored the effect of time on responses, finding that Turkers' responses to the specific questions we ask are consistent over time, despite important, relevant, and widely reported news events. This suggests that people are able to consistently, and presumably accurately, report on their perceptions of their own digital security and privacy experiences.

Taken together, our findings offer tempered support for the validity of self-report data for security and privacy perception measurements, and specifically for the generalizability of prior survey results from Turkers with respect to users' digital security and privacy behavior, knowledge, and experiences.

In addition, our results suggest that differences between MTurk survey results and probabilistic results will not simply be solved by better demographic recruitment, underscoring the unique challenges of user studies about digital security and privacy. Future work is needed to better understand the origin of these unique biases and develop techniques and novel measurement approaches to account for them.

II. Related Work

Representative samples ensure accurate and generalizable research results [26]–[28]. Below, we describe three different sampling methods that have been used in previous security and privacy studies: probabilistic samples, web panels, and MTurk. We also provide a review of related work evaluating these different sampling approaches, and contextualize our study within this body of research.

A. Probabilistic Samples

Probabilistic samples statistically guarantee that every person in a given population (e.g., the U.S.) has a non-zero chance of taking a given survey. Probabilistic samples allow researchers to extrapolate true population prevalences using statistical weighting techniques [29]–[32]. Such samples may be collected in person via face-to-face surveys administered by an interviewer, via mail, or via the telephone (households without a telephone will be contacted by mail and provided with the necessary resources to participate) [33]. Prior work in the survey methodology field has shown the results of telephone, mail, and face-to-face surveys to be relatively equivalent [34], [35]; as such, phone surveys are most often conducted due to the fact that they are cheaper and have higher response rates. Probabilistic surveys are rarely conducted in security and privacy [18], [36], [37], likely due to the fact that they are extremely expensive ($15-$30/response). Thus, in this paper we examine in what cases, and for what demographics, other, less expensive, sampling techniques can serve as reasonable alternatives to probabilistic sampling.

B. Web Panels

Web-panel samples can be obtained by hiring a panel company (e.g., Survey Sampling International, Forsa, Qualtrics) to administer your survey to a set number of their panel participants [38]. Panel participants are potential respondents who are recruited by the panel company via mailings, frequent flyer programs, web advertisements, and other techniques. These panel participants receive invitations to complete different surveys, based on whether they satisfy the demographic criteria for each survey. Respondents are compensated with various incentives including charity donations, frequent flyer miles, and gift cards; responses typically cost the researcher $2-$5 each [38], [39].

While these panels allow researchers to specify demographic requirements (e.g., request a sample that matches the demographic makeup of the U.S.), there is significant bias in which people become part of the panel and respond to which surveys. Prior work shows that over 90% of panel members who are invited to take a survey do not respond, and the effects of this non-reponse bias on data quality are not yet fully understood [38].

A significant body of work has been devoted to better understanding how panel responses differ from traditional probabilistic responses, beyond non-response rates. Heeren et al. compared panel and probabilistic telephone survey responses to a questionnaire about alcohol behavior, and found that panel respondents tended to report socially-undesirable behaviors somewhat more frequently with few reporting differences on other behaviors [12]. Similarly, in a survey on road safety administered face-to-face and via a panel, Goldenbeld and de Craen found only small differences between responses, but also noted the tendency of panel respondents to more frequently report socially-undesirable behaviors [13]. Fricker et al. observed lower item non-response in panel as compared to telephone respondents, but also that panel respondents tended to offer less differentiated answers to opinion scales [14]. Yeager et al. also compared a telephone and web survey conducted with probabilistic and non-probabilistic samples, and found that sampling bias from the non-probabilistic sampling method rather than mode effects (i.e., differences in responses related to use of telephone, web, or paper) tend to be the largest hindrance in the use of online surveys [15].

C. Crowdsourcing and Mechanical Turk

MTurk is a crowdsourcing platform that allows researchers to post HITs (tasks) that workers registered on the site can complete for compensation [40]. MTurk, and to a lesser extent alternatives such as Crowdflower and Prolific, have been used extensively to conduct both experimental and survey research in security and privacy, political science, economics, and psychology. The crowdsourcing nature of MTurk allows researchers to reach a far more diverse subject pool than may be locally accessible, provides an efficient means of collecting large numbers of responses quickly, and is far less expensive than other sampling methods (responses cost $0.75-$1.50 each) [40], [41].

Paolacci et al. as well as Ross et al. analyzed the demographics of MTurk and found that MTurk users tend to be more highly educated and younger than the general population [40], [42]. Additionally, Goodman et al. found that MTurk users may also hold different values and possess different personality characteristics than their peers [43].

Significant work in other fields, such as psychology, survey methodology, and political science, has been done to evaluate sample bias and compare MTurk samples with other types of samples. Behrend et al. found that MTurk respondents were significantly more diverse than were respondents collected through convenience sampling (e.g., recruiting at a university) [44]. Turkers also answered the psychology questionnaires administered in that study more reliably. Relatedly, Hauser and Schwartz found Turkers to be significantly more attentive than college students recruited with convenience sampling, leading to higher-validity results. However, Goodman et al. found the opposite: Turkers were less attentive in their study than convenience-sampled college students [43].

A smaller body of work has compared MTurk to non-convenience samples. Bartneck et al. found a significant, but very small, difference between survey responses from Turkers and web panel respondents on a survey about image features [45]. Berinsky et al. on the other hand, found that MTurk users were less representative of the U.S. population than were panel and probabilistic sample respondents [6]. Finally, Simons and Chabris compared results from MTurk and a traditional probabilistic phone survey for a questionnaire about memory [7]. They found that, with statistical weighting, the MTurk results could generalize to the U.S. population with little difference in responses.

D. Security and Privacy Sample Comparison

All of the aforementioned work has been conducted in the fields of psychology, survey methodology, economics, and political science. While results from these studies are relevant, security and privacy may differ with regard to question sensitivity, topic complexity, and relevance to survey mode (e.g., asking questions about internet use on the internet).

Early work in usability and security studied the use of MTurk for experimental studies, focusing on potential pitfalls of using the platform and best practices for recruiting respondents [46], [47]. These studies touch on potential concerns regarding the sample bias inherent to using Turkers as participants [47], but include no experiments to validate or alleviate these concerns.

More recently, Kang et al. compared MTurk and Pew survey responses, finding significant differences in privacy values and beliefs [22]. Although this comparison was not made using weighted Pew data, and thus was not fully representative of the U.S., it does illustrate important ways in which privacy research results drawn from MTurk may not match the broader U.S. population. Additionally, Schnorf et al. conducted a comparison of privacy-survey results administered on six different web-panel and crowdsourcing platforms [21]. Their work also identifies inconsistencies in privacy survey results across survey platforms. We expand on this work by directly comparing MTurk, a demographically representative web panel, and a probabilistic survey, using questions about *both* security and privacy, specifically examining responses around behavior, experiences, knowledge, and advice sources.

III. Methodology

In this section we provide details on the questions we compare, as well as each of the datasets used in our analysis, including the survey development and sampling procedure for each. We also detail our statistical analysis and the limitations of our work.

To which of the following have you turned to for advice about how to protect your personal information online? [Multiple selection]

- Friend or Peer
- Family Member
- Co-worker
- Librarian or resource at library
- Government website
- Website run by a private organization
- Teacher

As far as you know, have you ever... (Answer choices: Yes, No, Do Not Know)

- Had important personal information stolen such as your
- Social Security Number, your credit card, or bank account
- information?
- Had inaccurate information show up in your credit report?
- Had an email or social networking account of yours compromised or taken over without your permission by someone else?
- Been the victim of an online scam and lost money?
- Experienced persistent and unwanted contact from someone online?
- Lost a job opportunity or educational opportunity because of something that was posted online?
- Experienced trouble in a relationship or friendship because of something that was posted online?
- Had someone post something about you online that you didn't want shared?

Do you ever use the internet to... (Answer choices: Yes, No, Do Not Know)

- Use social media, such as Facebook, Twitter or Instagram?
- Apply for a job?
- Apply for government benefits or assistance?
- Apply for a loan or cash advance?
- Search for sensitive health information?
- Buy a product, such as books, toys, music or clothing?

Do you feel as though you already know enough about...
(Answer choices: Already know enough, Would like to learn more,
Doesn't apply, Do not know)

- Choosing strong passwords to protect your online accounts?
- Managing the privacy settings for the information you share online?
- Understanding the privacy policies of the websites and applications you use?
- Protecting the security of your devices when using public WiFi networks?
- Protecting your computer or mobile devices from viruses and malware?
- Protecting your computer or mobile devices from viruses and malware?
- Avoiding online scams and fraudulent requests for your personal information?

Fig. 1: Survey questions asked to all respondents.

A. *Questions*

The survey questions in our datasets query respondents' security and privacy experiences, advice sources, confidence in their knowledge about security and privacy topics, and internet behaviors (see Figure 1), among other topics. Many of the survey questions are drawn from existing pre-tested questions used by Pew and Reason-Rupe [48]–[51], the survey was extensively pre-tested before deployment to ensure validity. Additionally, the question order was randomized and demographic questions were administered at the end of the questionnaire to prevent bias [33], [52].

It is important to note that the questions we use were determined by the ones available in the probabilistic dataset, which we did not collect ourselves but rather obtained via a Data Access Grant, as detailed below.

B. *Datasets*

In our analysis we use four datasets: a dataset obtained through a probabilistic telephone sample, two datasets obtained using MTurk, and a nearly census-representative dataset obtained using a web panel.

1) Probabilistic Telephone Sample: We received the probabilistic survey data through a Data Access Grant from Data&Society, an internet think tank. [1] Data&Society contracted Princeton Survey Research Associates International (PSRAI) to collect the data. PSRAI collected 3,000 responses to this survey using a computer-assisted-telephone-interview (CATI), random digit dial (RDD) methodology from November 28 to December 23, 2015. To maximize the recruitment of a representative sample, the survey was administered by professionally trained interviewers in both English and Spanish, and interviews were conducted on multiple days of the week and at multiple times of day. As this was a probabilistic survey, the survey data was weighted to balance demographics to match the U.S. population. The data in this dataset is statistically estimated to be accurate within 2.7% of the true prevalence in the population. See Appendix VII for additional details on weighting.

2) Census-Representative Web-Panel Sample: We collected our census-representative web-panel sample from Survey Sampling International. The dataset (n=428) was collected in January 2017. We imported the questions (Figure 1) into Qualtrics and included all response options (including "prefer not to answer" and "don't know") that were included in the original telephone-interview scripts. Question order was randomized, and demographic questions were asked at the end of the survey to prevent bias. Quota sampling was used to ensure that the demographics of the respondents closely matched the U.S. Census for age, race, gender, and income. Survey Sampling International respondents are provided with benefits such as gift cards, airline frequent flyer miles, and donations to charities of their choice. The survey and data collection were approved by our Institutional Review Board.

3) MTurk Web Sample: We collected one dataset from MTurk in January 2017 and one in March 2018 [2]. These datasets were collected using the same survey questions. We recruited 480 MTurk users to complete our 2017 survey and 493 MTurk users to complete our 2018 survey. Both sets of MTurk users had a 95% approval rating or above and reside in the U.S. (Prior work has shown that MTurk users with 95% approval ratings produce high-quality data and do not require attention checks [55].) Respondents were compensated with $1 for their participation. This survey and the data collection were approved by our Institutional Review Board.

4) Time Differences Among Samples: There is a one-year time difference between the collection of our telephone sample and the census-representative and 2017 MTurk samples, due in part to constraints related to receiving access to the probabilistic dataset. To understand how and whether this time difference would confound our results, we conducted two subsequent analyses to evaluate the level of time-stability of the survey questions that we asked.

[1] The survey development and deployment for this survey was approved by Chesapeake IRB [53].

[2] Specifically, at the end of March 2018, after the Cambridge Analytica scandal was broadly reported [54].

In our first analysis, we compared our telephone survey data to users' responses to the same questions asked by Pew, using another probabilistic telephone survey in July 2013 [56]. The Pew sample contains only a subset of our questions: five of eight security and privacy experiences (having information stolen, having had an email or social network account compromised, having been the victim of an online scam and lost money, having experienced trouble in a relationship, and having lost a job or other opportunity because of something posted online). We conducted X^2 proportion tests to compare responses between the samples, with Bonferroni correction to reduce Type I error [57] introduced by conducting multiple question-by-question comparisons.

We find only one significant difference between our December 2015 sample and the July 2013 sample on these questions: significantly more respondents in 2015 reported having had information stolen than in 2013 (full results tables can be found in Appendix VIII-A. Additionally, we compare the two internet behavior responses for which we could locate identical questions in the Pew archives—use of social media and use of the internet to buy a product—finding no significant differences. Specifically, the proportion of respondents in our December 2015 sample reporting use of social media was not significantly different from a September 2013 Pew survey [58] (X^2 test $p-value$ = 0.609), and the proportion reporting use of the internet to buy a product was not significantly different from a May 2011 Pew survey [59] (X^2 test $p-value$ = 0.619).

In our second analysis, we compared responses to each of the questions in our 2018 MTurk survey dataset to the 2017 MTurk survey dataset (all comparisons, together with a description of the MTurk sample demographics, are shown in Appendix VIII-B). We found only one significant difference among the 28 items asked: a significantly lower proportion of 2018 respondents (94%) reported purchasing products on the internet, as compared to the 2017 respondents (99%). The confidence interval for the difference between these proportions is [3% - 8%]. The effect size by Cohen's H is 0.306 (small). We hypothesize that this change may be due to a growing shift away from product purchases and toward experiential purchases among Millennials, who are most heavily represented in the MTurk sample. Further, the proportion of MTurk 2018 respondents (94%) reporting internet product purchases is not significantly different (X^2 test $p-value$ = 0.803) from the proportion of census-representative web-panel respondents (90%) [60].

In sum, we find only two differences in our time-based comparisons. This is perhaps surprising, as the samples we compare spanned different relevant and widely reported news events (e.g., the Snowden surveillance revelations in 2014, the 2016 U.S. presidential election, reporting on the Cambridge Analytica scandal in 2018) and substantial time gaps (14 months and 2 years). Given the lack of observed differences, particularly for security- and privacy-related items in the more recently compared samples, we believe that any differences between the samples in our full analysis are most likely related to the samples themselves, rather than potential time confounds. As such, in the remainder of our paper we use the MTurk 2017 dataset only, and we refer to this sample simply as MTurk (we use this older sample for consistency as it was collected at the same time as our SSI sample).

Further, these results offer initial evidence that people's perceptions of their digital security and privacy experiences – at least measured through our particular survey questions [3] – are time stable and robust to major news events. This suggests that our respondents were able to consistently, and presumably accurately, answer questions about their digital security and privacy knowledge and advice sources as well as questions about experiences they have ever had and behaviors they have ever done. This consistency lends support for the continued use of survey studies to assess such experiences and use them to benchmark the need for new technologies and protective tools.

C. Sample Comparison Analysis

We made question-by-question comparisons between the samples. As all of the questions were binary (don't know responses were grouped as non-response, given that respondents were required to provide answers to each question), we used X^2 proportion tests to compare responses. In addition to comparing total response proportions per question from each sample, we also compared the responses by age subset (18-29, 30-49, and 50+) and by educational attainment subset (less than high school, graduated from high school, completed some college, and hold a bachelors or above). We first conducted omnibus tests to compare all three samples, and subsets from all three samples; the results of these comparisons are in Appendix X. For every variable with a significant omnibus result, we conducted pairwise proportion tests comparing the panel and MTurk samples each to the probabilistic sample.

As in our time-analyses, we reduce Type I error by applying a Bonferroni correction to each p-value. Bonferroni tends to be conservative (higher chance of a Type II error, or failing to identify a meaningful difference) compared to other multiple-hypothesis-testing correction methods. In comparing sampling methods, it is not clear which kind of error is more detrimental to our understanding. We chose the Bonferroni correction because its effects on our conclusions are clear (each p-value is multiplied by the number of tests performed, in this case 28) and it decreases the chance of a Type I error. The analysis code is in Appendix IX.

D. Limitations

Self-report studies have a number of limitations, including over- and under-reporting, sample bias, and social-desirability bias. However, while our study utilizes self-report data, our main claims are not about the accuracy of respondents' answers to a given question, but rather about whether and how responses from different samples resemble each other. For the purpose of our analysis we consider the probabilistic sample responses to

[3] We might expect attitudinal questions (e.g., is it acceptable for Facebook to advertise you based on attribute A) to be more susceptible to time and news events.

be the baseline, as they are the most representative self-report data we have about U.S. users' security and privacy behaviors, experiences, and knowledge. We do not make any claims about the validity of respondents' reports, aside from noting that the prevalence statistics observed in the probabilistic sample agree with prior samples collected by Pew, which utilized similar questions (see Appendix VIII-A). That said, it is possible that respondents' answers to the probabilistic telephone survey are less reflective of their true behaviors or experiences than respondents' answers to the MTurk or web surveys. Prior work, however, suggests that this is unlikely, and probabilistic surveys have been accepted as the baseline of self-reported "truth" since the early 1990s [28], [61].

There are a number of other limitations specific to our study, in addition to the time difference discussed in Section III-B4 above. First, two of our samples were collected via the web (both using the same questions and interface) while the third was collected via phone. This may introduce mode effects [33]; however, prior work shows that phone- and web-survey responses are reasonably equivalent and that respondents in both groups tend to exhibit similar levels of attentiveness, while respondents may be more likely to share sensitive information via web surveys due to lack of fear of judgement from an interviewer [62], [63]. Second, our research only addresses the responses of U.S. internet users, and thus we can offer no insight into the generalizability of results for international security and privacy studies. Third and finally, our work does not evaluate usability assessment questions. We received the probabilistic dataset through a data grant, and thus we were restricted to reusing the questions in the dataset we received. Further, tool use questions are difficult to ask using a phone survey. Future work wishing to compare usability assessment questions across samples could potentially explore nearly-probabilistic web sampling methods like GfK KnowledgePanel, which uses probabilistic methods to select participants for the sample.

IV. Results

Below, we present our comparison of users' negative security experiences, advice sources, security knowledge, and internet behavior across our three core datasets: probabilistic, census-representative web panel, and MTurk. First, we present the demographics of our three samples. Next, we compare the overall results of the three survey samples, followed by comparisons by age and educational subset. Finally, we compare the generalizability of the statistically weighted MTurk responses.

A. Demographics

The demographics of respondents in the probabilistic sample were nearly representative of the United States prior to being weighted to account for non-response, and after weighting they are, within a small error margin, representative of the demographics of the United States. The demographics of respondents in the panel sample were nearly representative of the United States, although these respondents differed slightly in age and were slightly more educated than the general

Sample Demographics

	Metric(%)	**MTurk**	**Panel**	**Prob.W**	**Prob.UW**	**Census**
Sex	Male	50	49	49	52	48
	Female	48	51	51	48	52
Race/Ethn.	Caucasian	84	69	63	58	66
	Hispanic	4	12	16	19	15
	African American	10	14	12	14	11
	Other	5	7	7	7	8
Education	LT H.S.	0.4	3	13	13	13
	High School	12	31	28	27	28
	Some college	41	34	30	24	31
	B.S. or above	46	31	29	35	28
Age	18-29 years	20	27	20	16	21
	30-49 years	58	23	33	24	35
	50+ years	22	49	44	56	44
Income	<$30k	25	28	*NA*∗	34	32
	$30k-$50k	24	23.5	*NA*∗	15	19
	$50k-$75k	26	19	*NA*∗	11	18
	$75k-$100k	12	13	*NA*∗	9	11
	$100k-$150k	8	10	*NA*∗	8	12
	$150k+	3	5	*NA*∗	7	10

TABLE I: Demographics for our three samples and the U.S. [64]. Values may not add to 100% due to non-response. UW for unweighted, W for weighted. *Income was the unweighted metric of interest.

population [64]. Finally, the MTurk sample was more educated, younger, more white, and less wealthy than the U.S. population. See Table I for a comparison of the demographics in these three samples to the U.S. census [64].

B. Overall Comparison

We first compared the results of the three samples for all respondents (Table II). Perhaps surprisingly, we find that, overall, MTurk provides a more generalizable set of results than does the panel. That is, MTurk responses more closely match those of the U.S. (i.e. the responses from the probabilistic survey) than do the answers of the census-representative panel respondents. See Figure 2 for a summary of the results.

Advice Sources. MTurk respondents reported seeking advice from co-workers, friends, and librarians with nearly the same frequency as the general population. However, Turkers were significantly more likely (58%) to report seeking out digital security advice from a website than were respondents in the general population (21%), and less likely (3% vs. 7%) to report seeking out advice from teachers. Panel respondents were also more likely (30% vs. 21%) to report seeking out digital security advice from websites, and less likely (3% vs. 7%) to report seeking out teachers, although the latter result was not significant. Panel respondents also reported seeking out friends as an advice source more often (48%) than the U.S. (39.0%). Overall, respondents in both samples are more likely than the general population to report that they would seek out advice on digital security in general, and report using a wider variety of sources. It is interesting to note that there were no significant differences in the frequency with which all three samples reported consulting co-workers and librarians, two

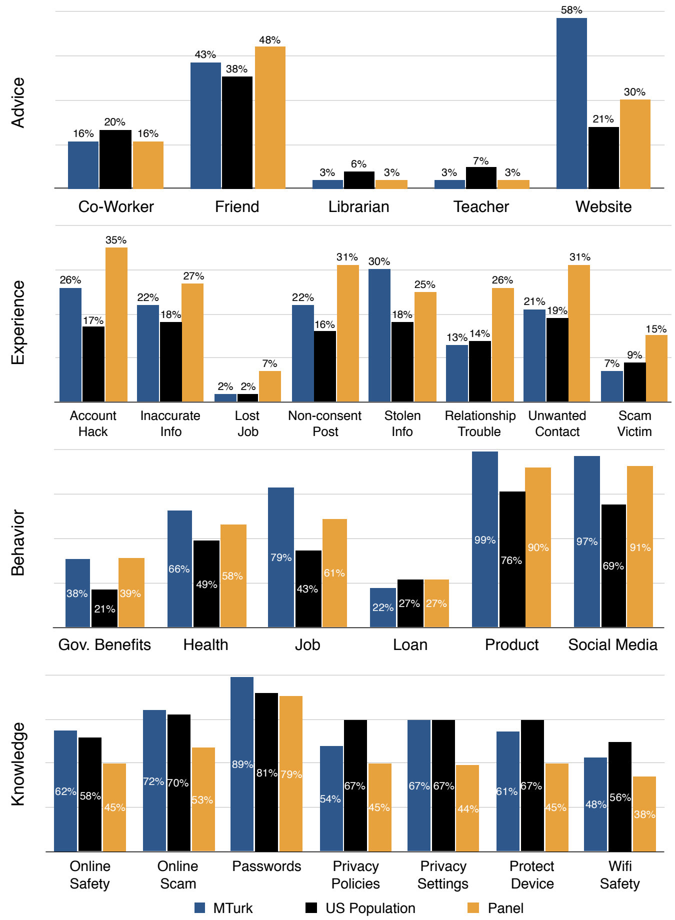

Fig. 2: Comparison of the overall proportion of responses to each question for the three populations.

Overall Sample Comparison on 28 Measured Variables

	Metric (%)	**MTurk**	**Panel**	**Prob**	**p-value** Prob vs. MTurk	Panel
Advice	Co-worker	15.6	16.4	20.2	–	–
	Friend	43.1	47.9	38.6	1.0	0.01*
	Librarian	2.7	3.3	5.4	–	–
	Teacher	2.9	3.3	6.9	0.034*	0.155
	Website	57.7	30.4	21.2	< 0.001*	0.001*
Experience	Account Hack	25.7	35.3	18.1	0.005*	< 0.001*
	Inaccurate Info	21.9	26.9	17.8	0.899	< 0.001*
	Lost Job	2.1	6.5	1.9	1.0	< 0.001*
	Non-consent Post	22.1	31.3	18.2	1.0	< 0.001*
	Stolen Info	30.5	24.8	17.8	< 0.001*	0.017*
	Relation Trouble	13.2	25.9	16.2	1.0	< 0.001*
	Unwanted Contact	20.9	31.3	19.4	1.0	< 0.001*
	Scam Victim	7.5	15.0	7.5	1.0	< 0.001*
Behavior	Gov. Benefits	38.1	39.0	22.9	< 0.001*	< 0.001*
	Heath	66.1	58.4	50.3	< 0.001*	0.06
	Job	78.9	61.2	50.3	< 0.001*	< 0.001*
	Loan	22.4	27.3	14.7	0.001*	< 0.001*
	Product	99.4	90.2	78.2	< 0.001*	< 0.001*
	Social Media	96.7	90.7	73.7	< 0.001*	< 0.001*
Knowledge	Online Safety	62.1	44.6	61.3	1.0	< 0.001*
	Online Scam	71.8	53.3	72.7	1.0	< 0.001*
	Passwords	88.9	78.5	84.0	0.193	0.179
	Privacy Policies	53.8	44.9	70.2	< 0.001*	< 0.001*
	Privacy Settings	66.9	44.4	70.9	1.0	< 0.001*
	Protect Device	61.3	45.3	70.5	0.003*	< 0.001*
	Safety on Wifi	47.9	37.6	59.3	< 0.001*	< 0.001*

TABLE II: Pairwise comparison of the proportion of responses to each question in the three samples (MTurk, panel, and probabilistic weighted to represent the U.S.) and results of the proportion tests comparing responses to each question from MTurk to those in the probabilistic survey and from the panel survey to the probabilistic survey. Proportions highlighted in blue are significantly greater than the probabilistic proportion, while those in orange are significantly less. *p*-values for variables for which the omnibus test result was null are indicated with –. *p*-values corrected for the number of comparisons performed.

advice sources not typically considered in security studies [8], [65].

Negative Experiences. 30% of MTurk respondents reported having had information stolen and 26% reported having had their email accounts compromised as compared to 18% (stolen information) and 17% (email compromised) of respondents in the general population. That said, MTurk respondents and the U.S. population reported similar frequencies of falling victim to an online scam, having something posted about them online without their consent, experiencing relationship trouble or unwanted contact as a result of something online, and losing a job or other opportunity as a result of something they posted online. Panel respondents, on the other hand, reported higher levels of victimization for all of the negative incidents, as shown in Figure 2.

Internet Behavior. This higher reporting of negative experiences in the online survey samples may result from being more active online: in general, respondents from both the MTurk and the panel samples tended to report higher rates of all internet behaviors. For example, 97% of MTurk respondents, and 91% of panel respondents, report using the internet for social media as compared to 74% of the U.S. population. Similarly, 22% of MTurk respondents and 27% of panel respondents report using the internet to apply for a loan, while only 15.0% of the U.S. reports doing so. This finding seems reasonably intuitive, as those who participate in MTurk or in panels are likely to be more comfortable online.

Security & Privacy Knowledge. Finally, respondents in both the MTurk and panel samples were less likely than the U.S. population to report feeling like they already knew enough about the security and privacy topics queried. More specifically, 54% of MTurk respondents vs. 70% of the U.S. population felt they knew enough about privacy policies, 61% vs. 71% felt they knew enough about how to protect their devices from viruses and malware, and 48% vs. 59% felt they new enough about how to protect their devices while using public wifi. Similarly, panel respondents were less likely to report feeling

Age Comparison: 18-29 Years

	Metric (%)	MTurk	Panel	Prob	p-value Prob vs. MTurk	p-value Prob vs. Panel
Advice	Co-worker	15.1	18.6	19.1	–	–
	Friend	48.4	61.0	50.0	–	–
	Librarian	2.2	1.7	7.1	–	–
	Teacher	7.5	8.5	11.0	–	–
	Website	63.4	33.9	24.2	< 0.001*	1.0
Experience	Account Hack	23.7	46.6	23.3	1.0	< 0.001*
	Inaccurate Info	9.7	22.0	9.5	1.0	0.008*
	Lost Job	2.2	11.9	2.9	1.0	0.003*
	Non-consent Post	31.2	44.9	31.1	–	–
	Stolen Info	24.7	24.6	9.5	0.002*	< 0.001*
	Relationship Trouble	14.0	36.4	30.1	0.059	1.0
	Unwanted Contact	21.5	41.5	30.1	–	–
	Scam Victim	10.8	23.7	7.3	1.0	< 0.001*
Behavior	Gov. Benefits	33.3	43.2	30.3	–	–
	Health	66.7	62.7	48.2	0.043*	0.17
	Job	89.2	83.1	77.0	–	–
	Loan	24.7	37.3	18.9	1.0	< 0.001*
	Product	98.9	94.1	83.0	0.003*	0.099
	Social Media	97.8	95.8	87.6	0.165	0.447
Knowledge	Online Safety	65.6	44.9	67.4	1.0	< 0.001*
	Online Scam	75.3	52.5	79.8	1.0	< 0.001*
	Passwords	87.1	83.9	92.1	–	–
	Privacy Policies	51.6	51.7	75.4	< 0.001*	< 0.001*
	Privacy Settings	75.3	47.5	84.0	1.0	< 0.001*
	Protect Device	64.5	45.8	76.2	0.665	< 0.001*
	Safety on Wifi	48.4	36.4	68.3	0.009*	< 0.001*

TABLE III: Comparison of the three samples for the subset of respondents who are 18-29 years old (see Table II caption).

Age Comparison: 30-49 Years

	Metric (%)	MTurk	Panel	Prob	p-value Prob vs. MTurk	p-value Prob vs. Panel
Advice	Co-worker	14.2	18.2	24.7	–	–
	Friend	45.9	58.6	38.6	1.0	0.006*
	Librarian	2.7	8.1	5.8	–	–
	Teacher	3.3	2.0	6.5	–	–
	Website	52.5	31.3	23.5	< 0.001*	1.0
Experience	Account Hack	29.0	35.4	21.2	–	–
	Inaccurate Info	21.9	29.3	23.5	–	–
	Lost Job	1.1	7.1	2.4	–	–
	Non-consent Post	23.0	34.3	22.5	–	–
	Stolen Info	29.0	21.2	23.5	–	–
	Relationship Trouble	14.8	35.4	18.4	1.0	0.005*
	Unwanted Contact	19.7	30.3	19.4	–	–
	Scam Victim	8.2	15.2	8.3	–	–
Behavior	Gov. Benefits	40.4	44.4	20.4	< 0.001*	< 0.001*
	Health	66.1	64.6	51.9	0.022*	0.614
	Job	88.5	72.7	59.8	< 0.001*	0.486
	Loan	25.1	38.4	18.3	1.0	< 0.001*
	Product	99.5	90.9	79.4	< 0.001*	0.262
	Social Media	96.7	91.9	80.0	< 0.001*	0.172
	Online Safety	66.7	50.5	63.8	–	–
Knowledge	Online Scam	74.9	56.6	73.1	1.0	0.028*
	Passwords	91.3	72.7	83.7	0.373	0.301
	Privacy Policies	56.3	41.4	69.3	0.033*	< 0.001*
	Privacy Settings	68.9	44.4	71.6	1.0	< 0.001*
	Protect Device	64.5	51.5	71.7	1.0	0.002*
	Safety on Wifi	50.3	44.4	65.7	0.005*	0.002*

TABLE IV: Comparison of the three samples for the subset of respondents who are 30-49 years old (see Table II caption).

like they knew enough about all of the privacy and security topics queried except passwords (see proportions in Table II). This is surprising, and perhaps indicates that these respondents, who are more active online, also have a better sense of the breadth of information available about security and privacy.

C. By Age

Next, in order to understand which samples are most representative for different demographics, we divided the responses from each sample by age, comparing respondents who were 18-29 years, 30-49 years, and over 50 years of age.

Age: 18-29 years. Similar to the results in the overall comparison, considering only those respondents who were 18-29 years old, the responses from MTurk sample more closely matched the U.S. population than did the panel responses, as shown in Table III. In fact, the MTurk and U.S. population responses for this age group were very closely matched (6 significant differences out of 26 variables), more so than the MTurk responses overall (14.0 significant differences). For those 18-29 years old, a higher proportion of Turkers reported that they would seek out advice from a website (63% for MTurk respondents vs. 24% in the general population); a higher proportion reported having information stolen online (25% vs. 10%); and a higher proportion reported engaging in two internet behaviors: searching for health information (67% vs. 48%) and purchasing products online (99% vs. 83%). Finally, a lower proportion of MTurk respondents reported feeling like they knew enough about privacy policies (52% vs. 75%) and protecting their devices when using public wifi (48% vs. 68%). We hypothesize that MTurk responses for the 18-29 year old age group very closely match the general population because younger users tend to be early adopters [66]–[68], and thus, there may not be a large difference between 18-29 year olds who use MTurk and those who do not use MTurk.

The responses from the panel sample for those aged 18-29 differed from the general population responses on 12 variables. These differences were primarily in the higher reporting of negative experiences and lower reporting of feeling knowledgable about privacy and security topics, as shown in Table III.

Age: 30-49 years. The MTurk and panel results for respondents aged 30-49 years were nearly equal in their similarity to the probabilistic sample: MTurk respondents' reports differed from the general population on 8 variables, while panel respondents' reports differed on 9 variables. Turkers' reports differed with regard to websites as an advice source (53% vs. 24%) and also differed for all of the internet behaviors except applying for loans online (see Table IV). MTurk responses also differed in prevalence from the probabilistic sample with regard to feelings of knowledge about privacy policies (56% vs. 69%) and protecting their devices when on public wifi (50%

Age Comparison: 50 Years and Older

	Metric (%)	MTurk	Panel	Prob	p-value Prob vs. MTurk	p-value Prob vs. Panel
Advice	Co-worker	17.0	14.2	16.6	–	–
	Friend	39.0	35.5	32.0	–	–
	Librarian	3.0	1.9	4.1	–	–
	Teacher	0.5	0.9	5.0	0.36	0.184
	Website	59.5	28.0	17.4	< 0.001*	0.011*
Experience	Account Hack	28.9	23.5	12.2	< 0.001*	< 0.001*
	Inaccurate Info	28.4	27.5	17.4	0.003*	0.017*
	Lost Job	3.0	3.3	0.9	–	–
	Non-consent Post	17.5	22.3	6.7	< 0.001*	< 0.001*
	Stolen Info	26.5	34.0	17.4	0.04*	< 0.001*
	Relation Trouble	15.6	11.5	6.0	< 0.001*	0.174
	Unwanted Contact	26.1	22.0	13.2	< 0.001*	0.038*
	Scam Victim	5.5	10.0	6.9	–	–
Behavior	Gov. Benefits	38.5	34.1	20.8	< 0.001*	< 0.001*
	Heath	53.1	66.0	50.0	1.0	< 0.001*
	Job	65.0	43.6	25.7	< 0.001*	< 0.001*
	Loan	19.0	16.6	9.0	< 0.001*	0.031*
	Product	99.5	87.7	74.3	< 0.001*	< 0.001*
	Social Media	96.0	87.2	59.6	< 0.001*	< 0.001*
Knowledge	Online Safety	41.7	56.0	55.5	0.007*	1.0
	Online Scam	52.1	67.0	68.1	< 0.001*	1.0
	Passwords	87.5	78.2	79.4	–	–
	Privacy Policies	52.0	42.7	68.0	< 0.001*	< 0.001*
	Privacy Settings	42.7	61.0	62.5	< 0.001*	1.0
	Protect Device	42.2	56.5	65.9	< 0.001*	0.327
	Safety on Wifi	35.1	45.0	48.2	0.014*	1.0

TABLE V: Comparison of the three samples for the subset of respondents over 50 years old (see Table II caption).

Education Comparison: No College Degree

	Metric (%)	MTurk	Panel	Prob	p-value Prob vs. MTurk	p-value Prob vs. Panel
Advice	Co-worker	11.8	11.9	16.6	–	–
	Friend	40.9	44.0	35.4	–	–
	Librarian	2.0	2.4	5.6	–	–
	Teacher	2.0	2.4	6.4	0.221	0.295
	Website	54.3	26.6	17.3	< 0.001*	0.008*
Experience	Account Hack	24.8	30.0	16.7	0.075	< 0.001*
	Inaccurate Info	20.5	22.9	16.2	–	–
	Lost Job	2.0	4.4	2.6	–	–
	Non-consent Post	24.4	29.4	17.8	0.431	< 0.001*
	Stolen Info	26.4	20.1	16.2	0.002*	1.0
	Relationship Trouble	13.4	23.9	17.8	–	–
	Unwanted Contact	21.7	29.7	21.3	–	–
	Scam Victim	6.7	14.0	6.7	1.0	0.001*
Behavior	Gov. Benefits	35.4	39.2	23.9	0.004*	< 0.001*
	Health	64.6	54.9	48.2	< 0.001*	1.0
	Job	78.7	54.3	49.4	< 0.001*	1.0
	Loan	19.7	23.9	14.2	0.852	0.002*
	Product	99.6	87.7	71.6	< 0.001*	< 0.001*
	Social Media	96.5	89.1	73.7	< 0.001*	< 0.001*
Knowledge	Online Safety	63.8	44.7	61.2	1.0	< 0.001*
	Online Scam	71.3	54.3	69.4	1.0	< 0.001*
	Passwords	88.6	77.1	80.6	0.084	1.0
	Privacy Policies	53.1	45.7	69.6	< 0.001*	< 0.001*
	Privacy Settings	68.1	46.1	67.4	1.0	< 0.001*
	Protect Device	62.2	44.7	68.6	1.0	< 0.001*
	Safety on Wifi	47.2	37.2	57.3	0.095	< 0.001*

TABLE VI: Comparison of the three samples for the subset of respondents with less than a B.S. (see Table II caption).

vs. 66%).

Panel respondents were also less likely than the general population to report feeling like they knew enough about privacy policies (41% vs. 69%) and protecting their devices when on public wifi (44% vs. 66%), as well as about privacy settings (44% vs. 72%), how to protect their computers from viruses and malware (52% vs. 72%), and how to protect themselves from online scams (57% vs. 73%). In addition to knowledge-related differences, panel respondents differed from the U.S. population in their more frequent use of the internet to apply for loans (38% vs. 18%) and government benefits (44% vs. 20%); their experiences with relationship trouble as a result of online posts (35% vs. 18%) and their experiences with having their email compromised (35% vs. 21%); and their more frequent consultation of friends for security and privacy advice (59% vs. 39.0%). These results suggest that MTurk and panel samples may be nearly equally as generalizable to the U.S. population for those aged 30-49, with MTurk responses differing primarily for internet behavior and panel responses differing primarily for knowledge about security and privacy topics.

Age: Over 50 years. In contrast to the other age subsets, for those over the age of 50, the panel responses more closely matched the responses of the general population (13 differences) than did the responses from MTurk (18 differences), as shown in Table V. This higher degree of similarity between the panel and the U.S. is largely due to more similarity in panel and U.S. respondents' desire to learn more about various security topics. The differences between the general population and panel were primarily related to negative experiences and internet behaviors. Panel respondents reported higher rates of victimization for five of the seven negative experiences: having an email account compromised (24% vs. 12%), having inaccurate information about themselves show up in a credit report (28% vs. 17%), having something posted about them without their consent (22% vs. 7%), having information stolen (34% vs. 17%), and having unwanted contact online (22% vs. 13%). They also reported higher rates of all the internet behaviors queried. Finally, they more frequently reported consulting websites for security and privacy advice than the general population (28% vs. 17%), and fewer panel respondents felt that they knew enough about privacy polices (43% vs. 68%).

MTurk respondents, on the other hand, were less likely than the general population to feel that they knew enough about all but one of the privacy and security topics queried (passwords). They were more likely to report doing all of the online activities other than searching for health information, were more likely to report all of the negative experiences except losing a job or opportunity due to a social media post and falling victim to an

Education Comparison: College Degree or Higher

	Metric (%)	MTurk	Panel	Prob	p-value Prob vs. MTurk	p-value Prob vs. Panel
Advice	Co-worker	20.3	25.4	27.5	0.95	1.0
	Friend	46.5	56.7	44.5	1.0	0.272
	Librarian	3.7	5.2	4.8	1.0	1.0
	Teacher	3.7	5.2	7.7	1.0	1.0
	Website	61.3	38.8	28.7	< 0.001*	0.562
Experience	Account Hack	27.2	47.0	21.1	1.0	< 0.001*
	Inaccurate Info	24.0	35.8	21.6	1.0	0.01*
	Lost Job	2.3	11.2	0.6	1.0	< 0.001*
	Non-consent Post	19.8	35.8	19.5	1.0	< 0.001*
	Stolen Info	35.0	34.3	21.6	0.001*	0.04*
	Relationship Trouble	12.9	30.6	13.5	1.0	< 0.001*
	Unwanted Contact	20.3	35.1	16.2	1.0	< 0.001*
	Scam Victim	8.8	17.2	8.7	1.0	0.087
Behavior	Gov. Benefits	42.4	38.8	21.4	< 0.001*	< 0.001*
	Health	69.1	65.7	54.0	0.002*	0.365
	Job	79.3	76.9	52.0	< 0.001*	< 0.001*
	Loan	26.3	35.1	15.9	0.012*	< 0.001*
	Product	99.1	96.3	91.3	0.003*	1.0
	Social Media	97.2	94.0	74.0	< 0.001*	< 0.001*
Knowledge	Online Safety	59.4	44.0	61.4	1.0	0.005*
	Online Scam	71.9	50.7	79.1	0.675	< 0.001*
	Passwords	88.9	82.1	90.8	1.0	0.077
	Privacy Policies	53.5	43.3	71.6	< 0.001*	< 0.001*
	Privacy Settings	65.0	41.0	77.5	0.004*	< 0.001*
	Protect Device	59.4	46.3	74.3	< 0.001*	< 0.001*
	Safety on Wifi	47.5	38.8	62.9	< 0.001*	< 0.001*

TABLE VII: Comparison of the three samples for respondents with a B.S. or above (see Table II caption).

Education Comparison: High School Degree or Lower

	Metric (%)	Panel	Prob	p-value
Advice	Co-worker	7.5	13.7	1.0
	Friend	43.8	28.9	0.014*
	Librarian	2.7	5.3	1.0
	Teacher	1.4	8.2	0.147
	Website	19.9	12.3	0.539
Experience	Account Hack	32.2	13.8	< 0.001*
	Inaccurate Info	19.2	12.7	1.0
	Lost Job	5.5	3.3	1.0
	Non-consent Post	31.5	17.0	0.002*
	Stolen Info	11.6	12.7	1.0
	Relationship Trouble	26.0	18.1	0.907
	Unwanted Contact	28.1	19.6	0.761
	Scam Victim	15.8	6.6	0.01*
Behavior	Gov. Benefits	34.9	20.9	0.01*
	Health	50.0	45.3	1.0
	Job	49.3	49.7	1.0
	Loan	17.1	12.1	1.0
	Product	80.8	65.9	0.014*
	Social Media	85.6	73.7	0.079
Knowledge	Online Safety	45.9	60.1	0.051
	Online Scam	58.2	66.8	1.0
	Passwords	75.3	76.7	1.0
	Privacy Policies	47.9	68.9	< 0.001*
	Privacy Settings	47.3	66.7	< 0.001*
	Protect Device	50.0	64.8	0.028*
	Safety on Wifi	41.1	57.2	0.013*

TABLE VIII: Comparison of panel and probabilistic samples for respondents who hold no more than a high school diploma (see Table II caption).

online scam, and were also more likely to report seeking out advice from websites. We hypothesize that the panel sample may be more representative of the general population in this case because older adults are more familiar with the concept of survey panels, even if they were formerly familiar with telephone panels, and thus are more likely to participate in web panels. Further, there were significantly more adults over the age of 50 (49%) in the panel sample than in the MTurk sample (22%). Consequently, there may be less selection bias in which older adults chose to participate in web panels than in those who chose to use MTurk, a relatively new technology (founded in 2005).

D. By Education

In addition to comparing responses by age, we also compared responses by education. We initially attempted to compare those with a high school degree or less, those with some college credit, and those with a bachelor's degree or higher. However for several metrics the expected cell counts for the MTurk respondents in the first category (n = 57) did not satisfy the assumptions of the X^2 proportional test. As such, we were unable to compare the samples subdivided into these three educational subsets. Instead, we compared sample responses across two educational subsets: those who had not earned a bachelor's degree and those with a bachelor's degree or additional higher education. [4]

For both subsets, MTurk responses were more representative than the panel, largely due to the fact that Turkers and U.S. users reported similar levels of interest in learning more about various security topics. For a comparison of the three samples for both education subsets, see Tables VI- VII. We also compared the panel sample with the general population for those who did not hold more than a high school diploma, as there were sufficient respondents in this category from the panel sample. We find the panel sample to be somewhat representative of this population (Table VIII). The differences between the panel and the U.S. in this education subset center around knowledge about digital security and privacy topics (4 differences) and negative experiences (3 differences).

E. Demographic Weighting of MTurk

Finally, to account for demographic bias in the MTurk sample, we applied survey raking (i.e., weighting) to balance

[4] Using Fisher's Exact Test as an alternative is not appropriate in this situation, since the marginal totals of the contingency table are not fixed. FET assumes that they are fixed [69]. In addition, low counts for certain metrics would cause the simulated estimates of these counts to not be meaningful. Both of these considerations factored into our decision to adjust the education groupings instead.

MTurk Demographic Weighting

	Metric	MTurk W	MTurk UW	Prob	p-value Prob vs. MTurk W	p-value Prob vs. MTurk UW
Advice	Co-worker	12.1	15.6	20.2	0.001*	0.661
	Friend	42.5	43.1	38.6	1.0	1.0
	Librarian	2.4	2.7	5.4	0.2	0.458
	Teacher	1.9	2.9	6.9	0.001*	0.034*
	Website	51.5	57.7	21.2	< 0.001*	< 0.001*
Experience	Account Hack	22.6	25.7	18.1	0.649	0.005*
	Inaccurate Info	18.2	21.9	17.8	1.0	0.899
	Lost Job	2.6	2.1	1.9	1.0	1.0
	Non-consent Post	23.4	22.1	18.2	0.289	1.0
	Stolen Info	29.7	30.5	17.8	< 0.001*	< 0.001*
	Relationship Trouble	12.5	13.2	16.2	1.0	1.0
	Unwanted Contact	18.8	20.9	19.4	1.0	1.0
	Scam Victim	6.6	7.5	7.5	1.0	1.0
Behavior	Gov. Benefits	35.5	38.1	22.9	< 0.001*	< 0.001*
	Health	63.1	66.1	50.3	< 0.001*	< 0.001*
	Job	74.3	78.9	50.3	< 0.001*	< 0.001*
	Loan	19.3	22.4	14.7	0.388	0.001*
	Product	99.6	99.4	78.2	< 0.001*	< 0.001*
	Social Media	96.1	96.7	73.7	< 0.001*	< 0.001*
Knowledge	Online Safety	65.0	62.1	61.3	1.0	1.0
	Online Scam	74.9	71.8	72.7	1.0	1.0
	Passwords	89.3	88.9	84.0	0.09	0.193
	Privacy Policies	56.5	53.8	70.2	< 0.001*	< 0.001*
	Privacy Settings	67.1	66.9	70.9	1.0	1.0
	Protect Device	63.2	61.3	70.5	0.057	0.003*
	Safety on Wifi	47.4	47.9	59.3	< 0.001*	< 0.001*

TABLE IX: Comparison of weighted (W) and non-weighted (UW) MTurk data and the U.S. (see Table II caption).

the MTurk sample demographics to be more representative of the U.S.

Survey raking is a commonly used technique in survey methodology and election polling that has also been applied successfully to improve the generalizability of MTurk survey data from other fields [7], [24]. Survey raking involves computing weights for each response based on the demographics of the respondent, the proportion of respondents with the same demographics in the sample, and the proportion of respondents with those demographics in the census. Each weight is a fraction: the proportion of respondents with a given set of demographics in the population of interest (in this case the U.S. Census) divided by the proportion of respondents with those demographics in the sample. We completed this weighting process based on three age subsets (18-29, 30-49, and 50+) and three education subsets (H.S. or less, some college, bachelors or more), using the anesrake R package [70].

We find that this weighting improves the generalizability of MTurk responses only slightly, reducing the number of significant differences between the MTurk responses and those of general U.S. users from 13 to 11 (Table IX): the difference between the proportion of respondents who report knowing enough about how to protect their devices in the MTurk sample and the probabilistic sample becomes insignificant with weighting, and the difference between the proportion of respondents who report having had their accounts hacked also becomes insignificant. We hypothesize that the overall lack of improvement is due to the fact that responses to security and privacy surveys covary with internet experience, and Turkers (even those who are older or less educated) tend to be more tech-savvy than their peers. Further, while there were differences observed for all age groups, the most differences between the MTurk and probabilistic samples were observed for those over the age of 50. Weighting simply amplifies the responses of the 22% of MTurk respondents who were over 50; thus, if these respondents are not representative (e.g., they are highly tech-savvy or security- and privacy-sensitive), having more of their non-representative responses in the dataset will not reduce bias.

V. Discussion

In this section we discuss the impact of our findings on the future deployment of security and privacy studies that collect self-report data. We also provide a set of suggested guidelines for using different types of samples based on our results (see Figure 3 for a summary). Finally, we conclude by synthesizing our work with the results of prior sample comparisons focused on privacy, as well as with a brief set of suggestions for future work.

A. The Forgotten 40%

Overall, we find that MTurk responses to the security and privacy questions we asked were more representative of the U.S. than were responses from a census-representative web-panel, except for respondents over the age of 50 or those with a high school education or lower. While it is promising that results from MTurk relatively closely represent the general population for those aged 18-49 years, it is important to remember that nearly half (44%) of the U.S. population is 50 years of age or older and 40% of the population holds no more than a high school diploma. Given the heavy use of MTurk and college-aged convenience samples for the collection of security and privacy survey data, and our finding that MTurk was not as representative for those over the age of 50 years or with less education, the results of many prior security and privacy studies may not generalize to these users. Even with demographic weighting, MTurk did not improve greatly in generalizability, implying that Turkers who are older or less educated are not very similar to their peers. While the panel sample was somewhat representative of these older and less educated populations, there were still a number of significant differences related to confidence in knowledge about privacy and security topics and internet behaviors.

Security is a collective behavior; the security of every user, including the most recent adopter, impacts the entire community. Further, prior work has found education-related differences in users' advice sources and security outcomes [8], [18], [71]. As such, we argue that extending security and privacy research to include these populations can sometimes be critical. To this end, we present in the next section a set of suggested guidelines and considerations for selecting an appropriate survey sample for security and privacy research.

B. Picking a Sample

Selecting a sampling method for any tool evaluation or survey involves a number of considerations, including resources, the

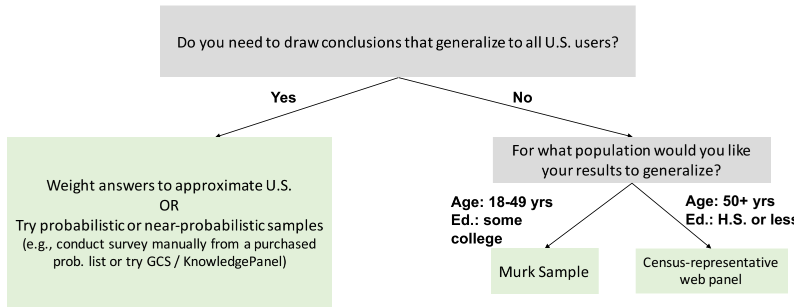

Fig. 3: Decision chart for selecting a security and privacy survey sample based on the results of our analysis.

desired population for which the results should generalize, and the appropriate mode of deployment (e.g. telephone, web). Figure 3 summarizes the discussion below in an attempt to provide an easy decision-making tool for security and privacy researchers.

Based on our results, we suggest that researchers seeking to generalize their study of security and privacy topics to those 18-29 years of age need look no further than MTurk. This suggestion matches with studies from other fields showing that MTurk provides high quality data for this age range [44].

Our results also suggest that those wishing to generalize their studies to those aged 18 through 49 years may use MTurk, while bearing in mind that Turkers' heavy internet use may skew results. On the other hand, researchers seeking to study security and privacy constructs on those aged 30 and over, may find a web panel to be the least expensive option. As with MTurk results for those aged 30-49 years, researchers should be careful to bear in mind that panel respondents also reported heavier internet use than the general population. As security behaviors have been shown to relate to internet skill [17], [72]–[75], which in turn has been shown to correlate with internet use [76]–[79]), researchers must be careful to interpret results from these samples in context.

To improve generalizability, researchers might consider alternatives such as using a nearly-probabilistic sample like Google Consumer Surveys (GCS) [80] or a probability-based web-panel such as GFK KnowledgePanel. GCS presents survey questions to users as an alternative to a paywall and thus limits the amount of questions that can be asked to 10, including demographics. We did not evaluate GCS in our work, as the question limit would not accommodate even our short survey, and thus we cannot comment on its generalizability; however, we suspect this question limit may limit the applicability of GCS for security and privacy research anyway. GFK KnowledgePanel, on the other hand, offers unlimited questions, but is fairly expensive ($8-$12/response). Panels such as this use probabilistic techniques to invite respondents (e.g., statistically sampling people to whom they want to mail panel invitations) to join the panel, but they suffer from high nonresponse rates and significant, unbalanced bias among those who do respond. This difference in willingness to participate in a web panel may significantly relate to constructs such as internet skill that covary with constructs measured in security and privacy studies [81]. As we were unable to evaluate this sampling method in this work, we cannot comment on whether these samples would perform better than the less-expensive panel sample that we analyzed.

C. Comparison with Prior Work

Using data collected between 2013 and 2014 and a set of seven yes/no questions, Kang et al. found that their MTurk sample responses differed from those of their probabilistic telephone survey with regard to two of three behaviors related to managing personal information online – MTurk respondents were more likely to report having tried to hide their identity and having tried to hide online content – and two of four privacy and security attitudes: MTurk workers reported more worry about information available about them on the internet and were more likely to report thinking that people should have the ability to be anonymous online [22]. They do not compare response differences by demographic sub-groups, but do use regression models to show significant covariance by age and educational attainment. Our results complement their findings: we find the most differences between the MTurk and probabilistic sample for internet behaviors, and Kang et al. find significant differences for information management behaviors, suggesting that behavior may be an area with especially significant sample-related bias.

Schnorf et al. collected data in December 2013 [21]. They examined differences in response rates to a privacy experiences question asking if respondents had experienced one of five privacy incidents (unwanted commercial offers or spam, reputation damage or embarrassing situation, stalking or harassment, financial loss, identity theft) and differences in response rates to two yes/no questions about privacy concerns. They compared responses to this question from six sample providers, including Princeton Survey Research International (the provider that was used to collect our probabilistic telephone data), SSI (the provider used to collect our census-representative web panel) and MTurk. Similar to our results, they found no significant differences among the six samples they compare (including the three that are shared with our work) and reports of incidents. However, they do find differences between the probabilistic survey and MTurk responses about privacy concerns – similar to Kang et al.'s findings about privacy beliefs. Schnorf et al. hypothesize that these differences in concern may be due to higher tech-savvyness among Turkers. We do not examine beliefs in our work, while Kang et al. do not examine knowledge, experiences, or advice sources. As Kang et al. and Schnorf et al.'s work explore only privacy-related beliefs, and a very small set of such beliefs, future work may wish to further explore security-belief biases and a broader set of privacy-belief biases.

In sum, our results and those of Kang et al. and Schnorf et al. suggest that samples used in security and privacy surveys may make a significant difference when querying behavior and beliefs, while having a smaller but still noticeable impact on self-reported confidence or knowledge, and little impact on reports of negative experiences or advice sources. Our findings in particular indicate that these impacts are most pronounced when seeking to understand older and less educated populations.

D. Future Work

Our findings suggest two main directions for future work.

First, our results showed common trends in the differences between responses from Turkers and panel participants. This indicates that it may be possible to develop a set of statistical weights to balance the results obtained from these populations to better reflect the entire U.S. population. We implemented the simplest such weighting schemes—survey raking, or demographic balancing of the responses—and found that this approach yielded only small adjustments to the raw data. This was surprising given that demographic weighting has resolved issues of sample bias for surveys on other, non-computer-security topics. This suggests that surveys about computer security and privacy may be importantly different from surveys about offline experiences or behavior, perhaps because the key difference in samples lies in metrics such as internet skill rather than demographics. New approaches involving weighting based on known values (e.g., based on the results of a probabilistic survey such as that analyzed in this work) are being explored in the survey methodology field [15], [82]. Thus, weighting MTurk responses to balance with a known distribution of U.S. users' internet skills may be a promising direction for future work.

Second, our evaluation was limited strictly to self-report survey questions relating to users' security and privacy experiences, advice sources, knowledge, and behaviors; prior work has focused primarily on privacy beliefs, experiences, and behaviors [21], [22]. Security and privacy user studies, however, often seek to evaluate behavior on real tasks. We did not evaluate whether the security and privacy behaviors of MTurk or panel participants on real tasks matched the behavior of the U.S. population. Such an evaluation may be difficult, or even impossible, as the companies that provide probabilistic survey samples are not designed to ask users to complete tasks (and these surveys are always conducted via telephone, face-to-face with an interviewer, or with a mailed paper survey). A potential mechanism for observing biases in task data is comparison of MTurk task performance to real-world log data observations. However, such log data can be difficult or very expensive to obtain. As such, survey-based comparisons such as ours can provide an initial understanding of sample biases, which can be used to develop bias-correcting techniques; once validated for survey questions, these corrective measures can then be evaluated for possible efficacy in the more difficult environment of task-based studies.

VI. Conclusion

In this work we examined whether the results of surveys about security and privacy administered on MTurk or on census-representative web panels generalize to the U.S. population. Perhaps surprisingly, we find that MTurk responses are more representative of the U.S. than are the census-representative panel responses, except for respondents aged 50 and older or for respondents with no more than a high school education. Both MTurk and panel respondents tend to report higher levels of all internet behaviors that were measured, tend to report seeking out security and privacy advice from websites more than the general population, and tend to less frequently report feeling like they know enough about certain security and privacy topics, especially privacy policies. Overall, our findings are encouraging for the continued use of MTurk for low-cost, convenient samples in security and privacy research, as long as the potential pitfalls we identify are carefully managed. It is important to note, however, that our results still show significant differences between MTurk and panel results and results from the general population, especially when measuring behavior. The populations for whom these differences are most stark—older and less educated users—are traditionally underrepresented in security and privacy research, and improving security and privacy research may be especially critical for these groups.

References

[1] B. Ur, P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer *et al.*, "How does your password measure up? the effect of strength meters on password creation." in *USENIX Security Symposium*, 2012, pp. 65–80.

[2] J. Bonneau and S. E. Schechter, "Towards reliable storage of 56-bit secrets in human memory." in *USENIX Security*, vol. 2014, 2014, pp. 607–623.

[3] K. Fawaz, H. Feng, and K. G. Shin, "Anatomization and protection of mobile apps' location privacy threats," in *24th USENIX Security Symposium (USENIX Security 15)*. Washington, D.C.: USENIX Association, 2015, pp. 753–768. [Online]. Available: https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/fawaz

[4] S. Dechand, D. Schürmann, K. Busse, Y. Acar, S. Fahl, and M. Smith, "An empirical study of textual key-fingerprint representations," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, 2016, pp. 193–208. [Online]. Available: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/dechand

[5] I. Ion, R. Reeder, and S. Consolvo, ""... no one can hack my mind": Comparing expert and non-expert security practices," in *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, 2015, pp. 327–346.

[6] A. J. Berinsky, G. A. Huber, and G. S. Lenz, "Evaluating online labor markets for experimental research: Amazon. com's mechanical turk," *Political Analysis*, vol. 20, no. 3, pp. 351–368, 2012.

[7] D. J. Simons and C. F. Chabris, "Common (mis) beliefs about memory: A replication and comparison of telephone and mechanical turk survey methods," *PloS one*, vol. 7, no. 12, p. e51876, 2012.

[8] E. Redmiles, S. Kross, and M. L. Mazurek, "How I Learned to be Secure: a Census-Representative Survey of Security Advice Sources and Behavior," in *CCS*, 2016. [Online]. Available: http://dl.acm.org/citation.cfm?id=2978307

[9] R. Wash and E. Rader, "Too much knowledge? security beliefs and protective behaviors among united states internet users," in *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, 2015, pp. 309–325.

[10] A. P. Felt, A. Ainslie, R. W. Reeder, S. Consolvo, S. Thyagaraja, A. Bettes, H. Harris, and J. Grimes, "Improving ssl warnings: Comprehension and adherence," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 2893–2902.

[11] S. Egelman, S. Jain, R. S. Portnoff, K. Liao, S. Consolvo, and D. Wagner, "Are you ready to lock?" in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2014, pp. 750–761.

[12] T. Heeren, E. M. Edwards, J. M. Dennis, S. Rodkin, R. W. Hingson, and D. L. Rosenbloom, "A comparison of results from an alcohol survey of a prerecruited internet panel and the national epidemiologic survey on alcohol and related conditions," *Alcoholism: Clinical and Experimental Research*, vol. 32, no. 2, pp. 222–229, 2008.

[13] C. Goldenbeld and S. De Craen, "The comparison of road safety survey answers between web-panel and face-to-face; dutch results of sartre-4 survey," *Journal of safety research*, vol. 46, pp. 13–20, 2013.

[14] S. Fricker, M. Galesic, R. Tourangeau, and T. Yan, "An experimental comparison of web and telephone surveys," *Public Opinion Quarterly*, vol. 69, no. 3, pp. 370–392, 2005.

[15] D. S. Yeager, J. A. Krosnick, L. Chang, H. S. Javitz, M. S. Levendusky, A. Simpser, and R. Wang, "Comparing the accuracy of rdd telephone surveys and internet surveys conducted with probability and non-probability samples," *Public opinion quarterly*, p. nfr020, 2011.

[16] E. Hargittai and E. Litt, "New strategies for employment? internet skills and online privacy practices during people's job search," *IEEE security & privacy*, vol. 11, no. 3, pp. 38–45, 2013.

[17] E. Hargittai *et al.*, "Facebook privacy settings: Who cares?" *First Monday*, vol. 15, no. 8, 2010.

[18] E. M. Redmiles, S. Kross, and M. L. Mazurek, "Where is the Digital Divide? Examining the Impact of Socioeconomics on Security and Privacy Outcomes," http://drum.lib.umd.edu/handle/1903/18867, 2016.

[19] M. Denton, S. Prus, and V. Walters, "Gender differences in health: a canadian study of the psychosocial, structural and behavioural determinants of health," *Social science & medicine*, vol. 58, no. 12, pp. 2585–2600, 2004.

[20] D. R. Kinder and L. M. Sanders, *Divided by color: Racial politics and democratic ideals*. University of Chicago Press, 1996.

[21] S. Schnorf, A. Sedley, M. Ortlieb, and A. Woodruff, "A comparison of six sample providers regarding online privacy benchmarks," in *SOUPS Workshop on Privacy Personas and Segmentation*, 2014.

[22] R. Kang, S. Brown, L. Dabbish, and S. B. Kiesler, "Privacy attitudes of mechanical turk workers and the us public." in *SOUPS*, 2014, pp. 37–49.

[23] L. Chang and J. A. Krosnick, "National surveys via rdd telephone interviewing versus the internet: Comparing sample representativeness and response quality," *Public Opinion Quarterly*, vol. 73, no. 4, pp. 641–678, 2009.

[24] J.-C. Deville, C.-E. Särndal, and O. Sautory, "Generalized raking procedures in survey sampling," *Journal of the American statistical Association*, vol. 88, no. 423, pp. 1013–1020, 1993.

[25] E. Hargittai, "Survey measures of web-oriented digital literacy," *Social science computer review*, vol. 23, no. 3, pp. 371–379, 2005.

[26] L. Kish, "Survey sampling," 1965.

[27] R. M. Groves, F. J. Fowler Jr, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau, *Survey methodology*. John Wiley & Sons, 2009, vol. 561.

[28] J. A. Krosnick, "Survey research," *Annual review of psychology*, vol. 50, no. 1, pp. 537–567, 1999.

[29] B. Kitchenham and S. L. Pfleeger, "Principles of survey research: part 5: populations and samples," *ACM SIGSOFT Software Engineering Notes*, vol. 27, no. 5, pp. 17–20, 2002.

[30] R. Tourangeau, L. J. Rips, and K. Rasinski, *The psychology of survey response*. Cambridge University Press, 2000.

[31] P. V. Marsden and J. D. Wright, Eds., *Handbook of survey research*, 2nd ed. Bingley, UK: Emerald, 2010. [Online]. Available: http://www.worldcat.org/search?qt=worldcat_org_all&q=9781848552241

[32] G. Kalton and D. Kasprzyk, *Treatment of missing survey data*. Department of Biostatistics, University of Michigan, 1986.

[33] J. A. Krosnick, *Handbook of Survey Research*, 2010. [Online]. Available: http://www.sciencedirect.com/science/book/9780125982269

[34] A. L. Holbrook, M. C. Green, and J. A. Krosnick, "Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias," *Public opinion quarterly*, vol. 67, no. 1, pp. 79–125, 2003.

[35] E. D. de Leeuw, *Data quality in mail, telephone and face to face surveys.* ERIC, 1992.

[36] L. Rainie, S. Kiesler, R. Kang, M. Madden, M. Duggan, S. Brown, and L. Dabbish, "Anonymity, privacy, and security online," *Pew Research Center*, vol. 5, 2013.

[37] J. Staddon, D. Huffaker, L. Brown, and A. Sedley, "Are privacy concerns a turn-off?: engagement and privacy in social networks," in *Proceedings of the eighth symposium on usable privacy and security*. ACM, 2012, p. 10.

[38] R. Baker, S. J. Blumberg, J. M. Brick, M. P. Couper, M. Courtright, J. M. Dennis, D. Dillman, M. R. Frankel, P. Garland, R. M. Groves *et al.*, "Research synthesis aapor report on online panels," *Public Opinion Quarterly*, vol. 74, no. 4, pp. 711–781, 2010.

[39] M. P. Couper and P. V. Miller, "Web survey methods introduction," *Public Opinion Quarterly*, vol. 72, no. 5, pp. 831–835, 2008.

[40] G. Paolacci, J. Chandler, and P. G. Ipeirotis, "Running experiments on amazon mechanical turk," *Judgment and Decision making*, vol. 5, no. 5, pp. 411–419, 2010.

[41] K. Casler, L. Bickel, and E. Hackett, "Separate but equal? a comparison of participants and data gathered via amazon's mturk, social media, and face-to-face behavioral testing," *Computers in Human Behavior*, vol. 29, no. 6, pp. 2156–2160, 2013.

[42] J. Ross, L. Irani, M. Silberman, A. Zaldivar, and B. Tomlinson, "Who are the crowdworkers?: shifting demographics in mechanical turk," in *CHI'10 extended abstracts on Human factors in computing systems*. ACM, 2010, pp. 2863–2872.

[43] J. K. Goodman, C. E. Cryder, and A. Cheema, "Data collection in a flat world: The strengths and weaknesses of mechanical turk samples," *Journal of Behavioral Decision Making*, vol. 26, no. 3, pp. 213–224, 2013.

[44] T. S. Behrend, D. J. Sharek, A. W. Meade, and E. N. Wiebe, "The viability of crowdsourcing for survey research," *Behavior research methods*, vol. 43, no. 3, pp. 800–813, 2011.

[45] C. Bartneck, A. Duenser, E. Moltchanova, and K. Zawieska, "Comparing the similarity of responses received from studies in amazon's mechanical turk to studies conducted online and with direct recruitment," *PloS one*, vol. 10, no. 4, p. e0121595, 2015.

[46] A. Kittur, E. H. Chi, and B. Suh, "Crowdsourcing user studies with mechanical turk," in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2008, pp. 453–456.

[47] P. G. Kelley, "Conducting usable privacy & security studies with amazon's mechanical turk," in *Symposium on Usable Privacy and Security (SOUPS)(Redmond, WA*, 2010.

[48] "Pew American Trends Panel," http://www.pewresearch.org/methodology/u-s-survey-research/american-trends-panel/, 2016.

[49] "Pew Internet and American Life Project," http://www.pewinternet.org/, 2016.

[50] "National Cybersecurity Alliance," https://staysafeonline.org/, 2015.

[51] "Reason-Rupe Surveys," http://reason.com/poll, 2016.

[52] N. Schaeffer and S. Presser, "The science of asking questions," *Annual Review of Sociology*, 2003. [Online]. Available: http://dx.doi.org/10.1146/annurev.soc.29.110702.110112

[53] "Chesapeake irb," https://www.chesapeakeirb.com/, 2016.

[54] T. Guardian, "The cambridge analytica files: A year-long investigation into facebook, data, and influencing elections in the digital age," 2018. [Online]. Available: https://www.theguardian.com/news/series/cambridge-analytica-files

[55] E. Peer, J. Vosgerau, and A. Acquisti, "Reputation as a sufficient condition for data quality on amazon mechanical turk," *Behavior research methods*, vol. 46, no. 4, pp. 1023–1031, 2014.

[56] "Anonymity Omnibus Dataset," http://www.pewinternet.org/datasets/july-2013-anonymity-omnibus/, 2013.

[57] O. J. Dunn, "Estimation of the medians for dependent variables," *The Annals of Mathematical Statistics*, pp. 192–197, 1959.

[58] P. R. Center, "Social media update, 2013," 2013.

[59] K. Zickuhr and A. Smith, "Digital differences," 2012.

[60] L. Greene, "The future of experience: Why brands and retailers need to look at trends and innovation in experience culture, as consumer appetites

for experiences reach critical mass," *Journal of Brand Strategy*, vol. 5, no. 3, pp. 238–244, 2016.
[61] P. Salant and D. A. Dillman, *How to conduct your own survey*. Wiley, 1994.
[62] F. Kreuter, S. Presser, and R. Tourangeau, "Social desirability bias in cati, ivr, and web surveys the effects of mode and question sensitivity," *Public Opinion Quarterly*, vol. 72, no. 5, pp. 847–865, 2008.
[63] K. L. Manfreda and V. Vehovar, "Internet surveys," *International handbook of survey methodology*, pp. 264–284, 2008.
[64] "American community survey 5-year estimates," http://www.census.gov/programs-surveys/acs/news/data-releases/2015/release.html, 2015.
[65] E. Rader and R. Wash, "Identifying patterns in informal sources of security information," *Journal of Cybersecurity*, vol. 1, no. 1, pp. 121–144, 2015.
[66] E. M. Rogers, "Elements of diffusion," *Diffusion of innovations*, vol. 5, pp. 1–38, 2003.
[67] T. Laukkanen and M. Pasanen, "Mobile banking innovators and early adopters: How they differ from other online users?" *Journal of Financial Services Marketing*, vol. 13, no. 2, pp. 86–94, 2008.
[68] E. Martínez and Y. Polo, "Adopter categories in the acceptance process for consumer durables," *Journal of Product & Brand Management*, vol. 5, no. 3, pp. 34–47, 1996.
[69] W. J. Conover and W. J. Conover, "Practical nonparametric statistics," 1980.
[70] J. Pasek, "anesrake: Anes raking implementation," *Comprehensive R Archive Network. Version 0.4.< http://cran. r-project. org/web/packages/anesrake/index. html>. Accessed July*, vol. 12, p. 2010, 2010.
[71] J. Warshaw, N. Taft, and A. Woodruff, "Intuitions, analytics, and killing ants: Inference literacy of high school-educated adults in the us," in *Symposium on Usable Privacy and Security (SOUPS)*, 2016.
[72] E. Hargittai and E. Litt, "New strategies for employment? internet skills and online privacy practices during people's job search," *IEEE S&P*, 2013.
[73] E. Litt and E. Hargittai, "Smile, snap, and share? a nuanced approach to privacy and online photo-sharing," *Poetics*, vol. 42, pp. 1–21, 2014.
[74] E. Redmiles, S. Silverstein, W. Bai, and M. Mazurek, "More skilled internet users behave (a little) more securely," *SOUPS*, 2016. [Online]. Available: https://www.usenix.org/sites/default/files/soups16poster20-redmiles.pdf
[75] M. Tan and T. S. Teo, "Factors influencing the adoption of internet banking," *Journal of the AIS*, vol. 1, no. 1es, p. 5, 2000.
[76] M. S. Eastin and R. LaRose, "Internet self-efficacy and the psychology of the digital divide," *Journal of Computer-Mediated Communication*, vol. 6, no. 1, pp. 0–0, 2000.
[77] E. Hargittai and A. Hinnant, "Digital inequality: Differences in young adults' use of the internet," *Communication Research*, 2008. [Online]. Available: http://crx.sagepub.com/content/35/5/602.abstract
[78] S. Livingstone and E. Helsper, "Balancing opportunities and risks in teenagers' use of the internet: The role of online skills and internet self-efficacy," *New media & society*, vol. 12, no. 2, pp. 309–329, 2010.
[79] E. Hargittai, "Digital na(t)ives? variation in internet skills and uses among members of the "net generation"," *Sociological inquiry*, vol. 80, no. 1, pp. 92–113, 2010.
[80] P. McDonald, M. Mohebbi, and B. Slatkin, "Comparing google consumer surveys to existing probability and non-probability based internet surveys," *Google Whitepaper*, 2012.
[81] R. D. Hays, H. Liu, and A. Kapteyn, "Use of internet panels to conduct surveys," *Behavior research methods*, vol. 47, no. 3, pp. 685–690, 2015.
[82] C. DiSogra, C. Cobb, E. Chan, and J. M. Dennis, "Calibrating non-probability internet samples with probability samples using early adopter characteristics," in *Joint Statistical Meetings (JSM), Survey Research Methods*, 2011, pp. 4501–4515.

APPENDIX

VII. PROBABILISTIC SURVEY WEIGHTING

The weighting information below was provided by PSRAI in their survey report. For full weighting information, please visit http://bit.ly/2kIsa1D.

"Weighting is generally used in survey analysis to adjust for effects of the sample design and to compensate for patterns of nonresponse that might bias results. The weighting was accomplished in multiple stages to account for the disproportionately-stratified samples, the overlapping landline and cell sample frames, household composition, and differential non-response associated with sample demographics. The weights correct for differential non-response that is related to particular demographic characteristics of the sample. The weight ensures that the demographic characteristics of the sample closely approximate the demographic characteristics of the target population.

In addition to demographic weighting, sampling design weights were also calculated and applied. Specialized sampling designs and post-data collection statistical adjustments require analysis procedures that reflect departures from simple random sampling. PSRAI calculates the effects of these design features so that an appropriate adjustment can be incorporated into tests of statistical significance when using these data. The so-called "design effect" or deff represents the loss in statistical efficiency that results from a disproportionate sample design and systematic non-response.

The survey's margin of error is the largest 95% confidence interval for any estimated proportion based on the total sample. For example, the margin of error for the total sample in this survey is 2.7 percentage points. This means that in 95 out every 100 samples using the same methodology, estimated proportions based on the entire sample will be no more than 2.7 percentage points away from their true values in the population. It is important to remember that sampling fluctuations are only one possible source of error in a survey estimate. Other sources, such as measurement error, may contribute additional error of greater or lesser magnitude."

VIII. TIME COMPARISON

A. Our 2015 Probabilistic Telephone Survey vs. Pew 2013 Probabilistic Telephone Survey

The table below compares the responses of respondents in our probabilistic sample to responses from a Pew Research Center survey using the same questions from 2013 (n=1,002) [56].

Experience	**Our Sample**	**2013 Pew**	**p-value**
Stolen Info.	18%	10%	<0.001
Account compromised	18%	21%	0.12
Scam Victim	8%	6%	0.47
Relationship Trouble	16%	13%	0.063
Lost Job	2%	1%	0.83

TABLE X: Comparison of reports of security and privacy experiences in our sample vs. Pew Research Center 2013 survey [56]

B. MTurk 2018 Sample vs. Mturk 2017 Sample

We collected two MTurk samples using identical methodology in March 2018 and in January 2017. Table XI compares the demographics of the two samples and Table XVIII compares

	Metric (%)	MTurk 2017	MTurk 2018	Census
Sex	Male	50	58	48
	Female	48	41	52
Race/Ethn.	Caucasian	84	81	66
	Hispanic	4	11	15
	African American	10	10	11
	Other	5	8	8
Education	LT H.S.	0.4	0	13
	High School	12	10	28
	Some college	41	39	31
	B.S. or above	46	52	28
Age	18-29 years	20	36	21
	30-49 years	58	48	35
	50+ years	22	17	44
Income	<$30k	25	25	32
	$30k-$50k	24	3–	19
	$50k-$75k	26	23	18
	$75k-$100k	12	11	11
	$100k-$150k	8	7	12
	$150k+	3	2	10

TABLE XI: Demographics for our two MTurk samples and the U.S. [64]. Values may not add to 100% due to non-response.

the proportion of respondents who reported each advice source, experience, knowledge, or internet use per sample.

	Metric	MTurk 2017	MTturk 2018	p-value
Advice	Co-worker	0.16	0.16	1.00
	Friend	0.43	0.45	1.00
	Librarian	0.03	0.03	1.00
	Teacher	0.03	0.02	1.00
	Website	0.58	0.52	1.00
Experience	Compromised Email	0.26	0.25	1.00
	Inaccurate Info	0.22	0.23	1.00
	Lost Job	0.02	0.05	0.46
	Post	0.22	0.25	1.00
	Stolen Info	0.30	0.31	1.00
	Relationship Trouble	0.13	0.14	1.00
	Unwanted Contact	0.21	0.25	1.00
	Scam Victim	0.08	0.11	1.00
Internet	Gov. Benefits	0.38	0.38	1.00
	Health	0.66	0.70	1.00
	Job	0.79	0.80	1.00
	Loan	0.22	0.26	1.00
	Product	0.99	0.94	< 0.001*
	Social Media	0.97	0.92	0.06
Knowledge	Online Protect	0.62	0.59	1.00
	Online Scam	0.72	0.67	1.00
	Passwords	0.89	0.83	0.17
	Privacy Policies	0.54	0.51	1.00
	Privacy Settings	0.67	0.66	1.00
	Protect Comp	0.61	0.61	1.00
	Wifi Protection	0.48	0.53	1.00

TABLE XII: X^2 comparison of MTurk results in 2017 and 2018.

IX. Analysis Code

Here: https://bit.ly/2I2QIBr we provide the code used in our statistical analysis. The datasets will be released pending approval from our institutional review board (for the MTurk and panel datasets) and approval from Data&Society, the think tank that awarded us the probabilistic dataset.

X. Omnibus Test Results

Tables XIII- XVIII show the results of our omnibus comparisons for each question, both overall and among age and education subsets. Only when the omnibus test was significant did we conduct the pairwise tests whose results are given in Section IV.

	Metric	Prob	Panel	Mturk	Statistic	p-value
Advice	Co-worker	0.20	0.16	0.16	7.61	0.578
	Friend	0.38	0.43	0.48	14.49	0.019*
	Librarian	0.06	0.03	0.03	8.81	0.318
	Teacher	0.07	0.03	0.03	17.70	0.004*
	Website	0.21	0.58	0.30	266.20	< 0.001*
Experience	Stolen Info	0.17	0.26	0.35	69.16	< 0.001*
	Inaccurate Info	0.18	0.22	0.27	22.35	< 0.001*
	Lost Job	0.02	0.02	0.07	31.53	< 0.001*
	Non-consent Post	0.16	0.22	0.31	38.81	< 0.001*
	Stolen Info	0.18	0.30	0.25	48.04	< 0.001*
	Relation Trouble	0.14	0.13	0.26	30.18	< 0.001*
	Unwanted Contact	0.19	0.21	0.31	30.53	< 0.001*
	Scam Victim	0.09	0.07	0.15	26.69	< 0.001*
Behavior	Gov. Benefits	0.21	0.38	0.39	80.67	< 0.001*
	Health	0.49	0.66	0.58	44.57	< 0.001*
	Job	0.43	0.79	0.61	137.56	< 0.001*
	Loan	0.12	0.22	0.27	49.14	< 0.001*
	Product	0.76	0.99	0.90	143.88	< 0.001*
	Social Media	0.69	0.97	0.91	166.70	< 0.001*
Knowledge	Online Safety	0.58	0.62	0.45	43.77	< 0.001*
	Online Scam	0.70	0.72	0.53	65.85	< 0.001*
	Passwords	0.81	0.89	0.79	18.17	0.003*
	Privacy Policies	0.67	0.54	0.45	129.92	< 0.001*
	Privacy Settings	0.67	0.67	0.44	114.59	< 0.001*
	Protect Device	0.67	0.61	0.45	106.67	< 0.001*
	Safety on Wifi	0.56	0.48	0.38	79.64	< 0.001*

TABLE XIII: Omnibus X^2 proportion test comparing the three samples.

	Metric	Prob	Panel	Mturk	Statistic	p-value
Advice	Co-worker	0.21	0.15	0.19	0.87	1
	Friend	0.50	0.48	0.61	5.04	1
	Librarian	0.07	0.02	0.02	7.57	0.0591
	Teacher	0.13	0.07	0.09	1.42	1
	Website	0.28	0.63	0.34	56.04	< 0.001*
Experience	Compromised email	0.24	0.24	0.47	26.55	< 0.001*
	Inaccurate Info	0.11	0.10	0.22	14.80	0.016*
	Lost Job	0.03	0.02	0.12	19.29	0.002*
	Post	0.30	0.31	0.45	8.30	0.409
	Stolen Info	0.11	0.25	0.25	27.77	< 0.001*
	Relationship Trouble	0.30	0.14	0.36	13.63	< 0.028*
	Unwanted Contact	0.29	0.21	0.41	10.23	0.156
	Scam Victim	0.10	0.11	0.24	27.02	< 0.001*
Internet	Gov. Benefits	0.29	0.33	0.43	7.16	0.726
	Health	0.47	0.67	0.63	15.82	0.01*
	Job	0.75	0.89	0.83	8.24	0.422
	Loan	0.17	0.25	0.37	18.23	0.003*
	Product	0.83	0.99	0.94	23.82	< 0.001*
	Social Media	0.88	0.98	0.96	14.04	0.023*
Knowledge	Online Protect	0.66	0.66	0.45	20.76	< 0.001*
	Online Scam	0.78	0.75	0.53	36.85	< 0.001*
	Passwords	0.90	0.87	0.84	8.00	0.476
	Privacy Policies	0.75	0.52	0.52	38.02	< 0.001*
	Privacy Settings	0.80	0.75	0.47	70.03	< 0.001*
	Protect Comp	0.75	0.65	0.46	42.45	< 0.001*
	Wifi Protection	0.68	0.48	0.36	46.17	< 0.001*

TABLE XIV: Omnibus X^2 proportion test for 18-29 year old respondents.

	Metric	Prob	Panel	Mturk	Statistic	p-value
Advice	Co-worker	0.26	0.14	0.18	10.24	0.155
	Friend	0.42	0.46	0.59	15.61	0.011*
	Librarian	0.06	0.03	0.08	4.10	1
	Teacher	0.06	0.03	0.02	5.34	1
	Website	0.24	0.53	0.31	57.95	< 0.001*
Experience	Compromised email	0.21	0.29	0.35	12.25	0.057
	Inaccurate Info	0.23	0.22	0.29	2.07	1
	Lost Job	0.02	0.01	0.07	9.57	0.217
	Post	0.21	0.23	0.34	6.73	0.898
	Stolen Info	0.23	0.29	0.21	2.96	1
	Relationship Trouble	0.20	0.15	0.35	18.97	0.002**
	Unwanted Contact	0.21	0.20	0.30	6.37	1
	Victim Scam	0.09	0.08	0.15	5.17	1
Internet	Gov. Benefits	0.20	0.40	0.44	47.27	< 0.001*
	Health	0.54	0.66	0.65	15.16	0.013*
	Job	0.59	0.89	0.73	55.15	< 0.001*
	Loan	0.17	0.25	0.38	22.24	< 0.001*
	Product	0.77	0.99	0.91	47.41	< 0.001*
	Social Media	0.79	0.97	0.92	35.35	< 0.001*
	Online Protect	0.60	0.67	0.51	7.87	0.507
Knowledge	Online Scam	0.71	0.75	0.57	12.73	0.045*
	Passwords	0.82	0.91	0.73	16.58	0.007*
	Privacy Policies	0.66	0.56	0.41	34.84	< 0.001*
	Privacy Settings	0.69	0.69	0.44	29.49	< 0.001*
	Protect Comp	0.67	0.65	0.52	17.68	0.004*
	Wifi Protection	0.62	0.50	0.44	26.27	< 0.001*

TABLE XV: Omnibus X^2 proportion test for 30-49 year old respondents.

	Metric	Prob	Panel	Mturk	Statistic	p-value
Advice	Co-worker	0.16	0.17	0.14	0.83	1
	Friend	0.32	0.39	0.35	4.38	1
	Librarian	0.05	0.03	0.02	2.80	1
	Teacher	0.05	0.01	0.01	14.86	0.015*
	Website	0.17	0.59	0.28	169.15	< 0.001*
Experience	Compromised Email	0.13	0.23	0.29	47.91	< 0.001*
	Inaccurate Info	0.18	0.28	0.28	24.18	< 0.001*
	Lost Job	0.01	0.03	0.03	10.54	0.0134
	Post	0.08	0.17	0.22	62.18	< 0.001*
	Stolen Info	0.18	0.34	0.27	38.14	< 0.001*
	Relationship Trouble	0.05	0.12	0.16	26.76	< 0.001*
	Unwanted Contact	0.14	0.22	0.26	28.52	< 0.001*
	Scam Victim	0.08	0.06	0.10	3.44	1
Internet	Gov. benefits	0.18	0.39	0.34	40.39	< 0.001*
	Health	0.47	0.66	0.53	17.68	0.004*
	Job	0.22	0.65	0.44	132.25	< 0.001*
	Loan	0.07	0.19	0.17	24.10	< 0.001*
	Product	0.72	0.99	0.88	76.74	< 0.001*
	Social Media	0.56	0.96	0.87	146.07	< 0.001*
Knowledge	Online Protect	0.53	0.56	0.42	14.18	0.022*
	Online Scam	0.66	0.67	0.52	20.56	< 0.001*
	Passwords	0.77	0.88	0.78	7.72	0.547
	Privacy Policies	0.64	0.52	0.43	60.52	< 0.001*
	Privacy Settings	0.61	0.61	0.43	29.41	< 0.001*
	Protect Comp	0.63	0.56	0.42	45.30	< 0.001*
	Wifi Protection	0.49	0.45	0.35	12.61	0.047*

TABLE XVI: Omnibus X^2 proportion test for respondents over the age of 50.

	Metric	Prob	Panel	Mturk	Statistic	p-value
Advice	Co-worker	0.15	0.12	0.12	6.82	0.857
	Friend	0.33	0.41	0.44	9.14	0.269
	Librarian	0.06	0.02	0.02	10.13	0.164
	Teacher	0.07	0.02	0.02	13.89	0.025*
	Website	0.17	0.54	0.27	164.60	< 0.001*
Experience	Compromised Email	0.16	0.25	0.30	31.28	< 0.001*
	Inaccurate Info	0.16	0.20	0.23	9.80	0.194
	Lost Job	0.03	0.02	0.04	3.81	1
	Post	0.17	0.24	0.29	22.72	< 0.001*
	Stolen Info	0.16	0.26	0.20	17.44	0.004*
	Relationship Trouble	0.16	0.13	0.24	10.50	0.137
	Unwanted Contact	0.21	0.22	0.30	9.78	0.195
	Scam Victim	0.09	0.07	0.14	18.17	0.003*
Internet	Gov. Benefits	0.22	0.35	0.39	37.01	< 0.001*
	Health	0.47	0.65	0.55	24.53	< 0.001*
	Job	0.43	0.79	0.54	74.26	< 0.001*
	Loan	0.11	0.20	0.24	18.63	0.002*
	Product	0.67	1.00	0.88	116.94	< 0.001*
	Social Media	0.70	0.96	0.89	88.09	< 0.001*
Knowledge	Online Protect	0.57	0.64	0.45	29.87	< 0.001*
	Online Scam	0.66	0.71	0.54	27.06	< 0.001*
	Passwords	0.77	0.89	0.77	12.53	0.049*
	Privacy Policies	0.65	0.53	0.46	73.42	< 0.001*
	Privacy Settings	0.64	0.68	0.46	49.82	< 0.001*
	Protect Comp	0.65	0.62	0.45	59.95	< 0.001*
	Wifi Protection	0.54	0.47	0.37	43.03	< 0.001*

TABLE XVII: Omnibus X^2 proportion test for respondents with less than a bachelors degree.

	Metric	Prob	Panel	Mturk	Statistic	p-value
Advice	Co-worker	0.28	0.20	0.25	4.78	1
	Friend	0.46	0.47	0.57	7.06	0.763
	Librarian	0.06	0.04	0.05	0.59	1
	Teacher	0.08	0.04	0.05	5.19	1
	Website	0.28	0.61	0.39	83.51	< 0.001*
Experience	Compromised Email	0.20	0.27	0.47	43.40	< 0.001*
	Inaccurate Info	0.22	0.24	0.36	13.36	0.033*
	Lost Job	0.01	0.02	0.11	68.67	< 0.001*
	Post	0.16	0.20	0.36	19.14	0.002*
	Stolen Info	0.22	0.35	0.34	24.09	< 0.001*
	Relationship Trouble	0.12	0.13	0.31	27.81	< 0.001*
	Unwanted Contact	0.17	0.20	0.35	27.95	< 0.001*
	Scam Victim	0.08	0.09	0.17	9.89	0.185
Internet	Gov. Benefits	0.19	0.42	0.39	51.24	< 0.001*
	Health	0.52	0.69	0.66	20.52	< 0.001*
	Job	0.44	0.79	0.77	73.97	< 0.001*
	Loan	0.14	0.26	0.35	34.74	< 0.001*
	Product	0.88	0.99	0.96	19.03	0.002*
	Social Media	0.67	0.97	0.94	78.30	< 0.001*
Knowledge	Online Protect	0.58	0.59	0.44	14.73	0.016*
	Online Scam	0.75	0.72	0.51	51.94	< 0.001*
	Passwords	0.87	0.89	0.82	9.78	0.195
	Privacy Policies	0.69	0.54	0.43	59.00	< 0.001*
	Privacy Settings	0.70	0.65	0.41	82.87	< 0.001*
	Protect Comp	0.70	0.59	0.46	53.97	< 0.001*
	Wifi Protection	0.59	0.47	0.39	39.71	< 0.001*

TABLE XVIII: Omnibus X^2 proportion test for respondents with a bachelors or above.