Towards Understanding Differential Privacy: When Do People Trust Randomized Response Technique?

Professor Evan Peck, Bucknell University
Professor Darakhshan Mir, Bucknell University
Brooke Bullek, Bucknell University
Stephanie Garboski, Bucknell University

1. Goals and Purpose

Please describe the goal(s) and purpose of this project. What open research problem or question does the project address?

Our research synthesizes two important fields of CS (data privacy and human-computer interaction) and postulates important questions regarding the usability of an up-and-coming privacy protocol known as differential privacy (DP). DP has made headlines recently as the mechanism underlying Apple iOS 10's ability to safeguard its users' staggering amounts of data while using this data to facilitate machine learning and "smart suggestions" technology. However, as is perhaps expected, DP and other algorithmically complex privacy protocols are implemented by domain experts and thus do not lend themselves to being easily understood by the average user.

This presents a desire for a multidisciplinary approach, one that studies the "cognitive and behavioral analysis of privacy decisions." Reports facilitated by workshops organized by the White House and the President's Council of Advisors on Science and Technology (PCAST) have stressed the need for such an investigation, and rightfully so: nearly every aspect of our lives, from our medical records to music tastes to shopping habits, is intertwined with a culture of data overcollection. As convenient and mutually beneficial as this culture often is, the statistics show that 2.5 exabytes of data are produced every day (and that's an outdated source!) It's natural to wonder not only how all of this data is being generated, but also which variables can influence user-facilitated disclosure of this data.

Our main research question was the following: Would most people agree that their data was more exposed (and thus less safe) with a privacy mechanism that added less obfuscating "noise" to their data? Or would the technique dissuade them from choosing the more anonymous privacy mechanism out of a sense of pride ("I have nothing to hide") or responsibility ("I need to tell the truth, and anonymity is deceptive")?

2. Related Work

Include a scholarly review of related prior work. Include citations to the relevant literature.

The increasing collection and analysis of personal data has created a complex, data-based ecosystem. In the course of constant human-data interactions (source), inhabitants of this data-based ecosystem are implicitly or explicitly making decisions about their privacy. However, privacy-preserving mechanisms, like the ones that may exist on your phone, typically remain nonexistent or otherwise hidden from users despite mounting concern for privacy. Similarly, standard anonymizing techniques are often insufficient; they may rely on removing identifiers

such as full names, but fail in the sense that the remaining "anonymized" information can be synthesized to reconstruct an identity (Latanya Sweeney proved this by cross referencing voting records and medical data (source)). However, even if well-intentioned, these mechanisms, and the rationale for using them, are often hidden to the people who contribute their data. How might the transparency of these mechanisms influence the privacy decisions of people?

We can consider a specific scenario in which one might be asked to donate their sensitive data—for instance, cell phone records. It's been found that the analysis of phone records in a developing country can help researchers study the spread of malaria (source); however, releasing this collected data could compromise the privacy of those who donated their phone records. Do you donate and risk your privacy, or opt out and potentially hinder medical research? To allay these privacy concerns, data curators can use mechanisms such as differential privacy (DP) that promise individuals that their data will only be used in a manner that does not compromise their privacy, thereby incentivizing participation (sources 1, 2). DP provides a knob through which the relationship between the obfuscation and utility of these aggregate statistics is mediated, often by adding mathematically calibrated noise to them. In this way, the privacy of individual contributors is protected, but the information released in aggregate maintains statistical credibility.

Other terminology that frequently pops up in these discussions is the notion of a **local** versus **global** differential privacy model (source). We can envision three distinct parties: the user, the data aggregator, and the data scientist. In the local model, the user applies noise to their own data using some mechanism we can currently abstract as a "randomizing device," and the aggregator is essentially a middleman who is responsible for analyzing the data from all users at once. The aggregator applies some computations and is able to remove, or at least minimize, the noise from the randomizing device before sending the clean and accurate statistics to the data scientist. The downside to the local model is the sheer amount of data that must be collected for the data aggregator to puzzle out accurate statistics with any reasonable degree of certainty. Fuzzy microdata, while reassuringly anonymous, is not so reassuring from a utilitarian perspective. To offset this, the data from many, many users needs to be collected in aggregate; this is why critics deem differential privacy "privacy in the land of plenty" (source). In either case, to incentivize individuals to contribute potentially sensitive information, DP “promises” them that by allowing their data to be used for such aggregations, they are unlikely to be affected (adversely or otherwise) (source).

RRT is a version of the local DP model that is widely used to facilitate higher rates of sensitive disclosure (sources 1, 2, 3, 4). Suppose an administrator is trying to gauge the number of college students that have cheated on an exam. Rather than asking directly, which would likely provoke negative (non-incriminating) responses, the administrator may opt to use RRT. Each student is given a randomizing device, like a coin, and the administrator poses the question, “Have you ever cheated on a college exam?” Rather than answering outright, students only answer honestly if the coin lands on “heads,” and if the coin lands on “tails” they must respond. Thus, a “Yes” response cannot be interpreted as an admission of guilt. Still, because the probability of landing on either side is known to be 50/50, the administrator can statistically deduce the approximate frequency of the sensitive behavior (cheating on a college exam) by examining the results in aggregate (source).

Boeije and Lensvelt-Mulders were among the first to administer sensitive questionnaires using RRT via a computer-assisted self-interviewing environment (source). Terms such as

“cheating" and “trust," defined by the theoretical framework, were eclipsed by “luck" and “forced dishonesty" from the experimental data, highlighting the importance of the meaning of honesty for respondents (source). A common frustration emerges when respondents are forced to supply false positive responses and admit within an impersonal interface that they had committed unlawful activities. Furthermore, participants in older studies reported that the “computer does not encourage telling the truth" (source). This observation of depersonalization and lack of “encouragement" for “truthful answers on sensitive questions," reinforces the need to examine understanding and trust as metrics when administering RRT or other protocols.

Although privacy research often investigates social media platforms, occasionally, these experiments have been funneled into fabricated, controlled environments meant to directly assess the role played by a system’s UI and overall presentation (source 1, source 2). Privacy “nudges” include adjusting interface colors, buttons, font-weight, and additional verification steps that dissuade users from posting irresponsibly (source). These nudges counteract asymmetric information by placing emphasis on users and illuminating their privacy settings. Consequently, they prompt people to exhibit greater caution and awareness when sharing personal information (source). This indicates that the transparency of privacy protocols are as important as reliability and efficiency where usable privacy is concerned.

3. **Process**

Please describe your approach and/or strategy in solving the problem.

Our main focus of this academic year was in creating a cohesive report on our prior research in order to submit to various conferences to attempt for publication. We were able to take our research and form around it a body of background information and conclusions drawn that would best explain our research to a broader audience. After being accepted for publication and presentation at ACM CHI 2017, we worked on further broadening our audience by forming blog posts that describe both our research and the basics of differential privacy.

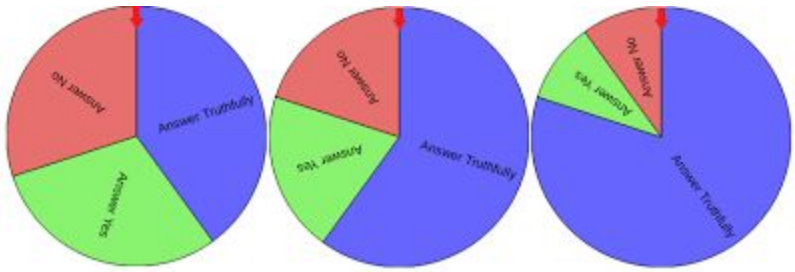
A secondary focus was in developing groundwork for future work that extends our current research, in the form of a secondary experiment. This experiment would attempt to answer the questions outlined in Future Work (5). In order to design this secondary experiment, we researched further how users make privacy decisions, and what influences them to increase their awareness of and for privacy. To approach this, we started broadly, by researching across the body of available publications in online privacy, and then we narrowed down our experiment to specific scenarios that would work for our experimental questions. We decided that an in-person experiment would reach our goals most effectively, and that we wished to manipulate how we described the research group conducting the experiment; would the difference between a “good” corporation and a “bad” corporation collecting sensitive data affect a participant’s privacy decisions? One of our research problems was to explore further how we can design an interface that empowers users to understand DP and make privacy decisions. To contextualize this problem, we debated varying interfaces for DP that we thought might best represent the protocols at hand. Although we considered creating a “Chrome” extension that utilized a more opaque version of differential privacy, we settled on an in-person RRT method, that would hopefully utilize the spinners from our previous successful research. The “Chrome” extension would have likely limited our database of participants to Mechanical Turkers again (which we

didn’t want) and didn’t allow us to fully explore the question of how context impacts privacy decisions and understanding of differential privacy.

4. Results and Discussion

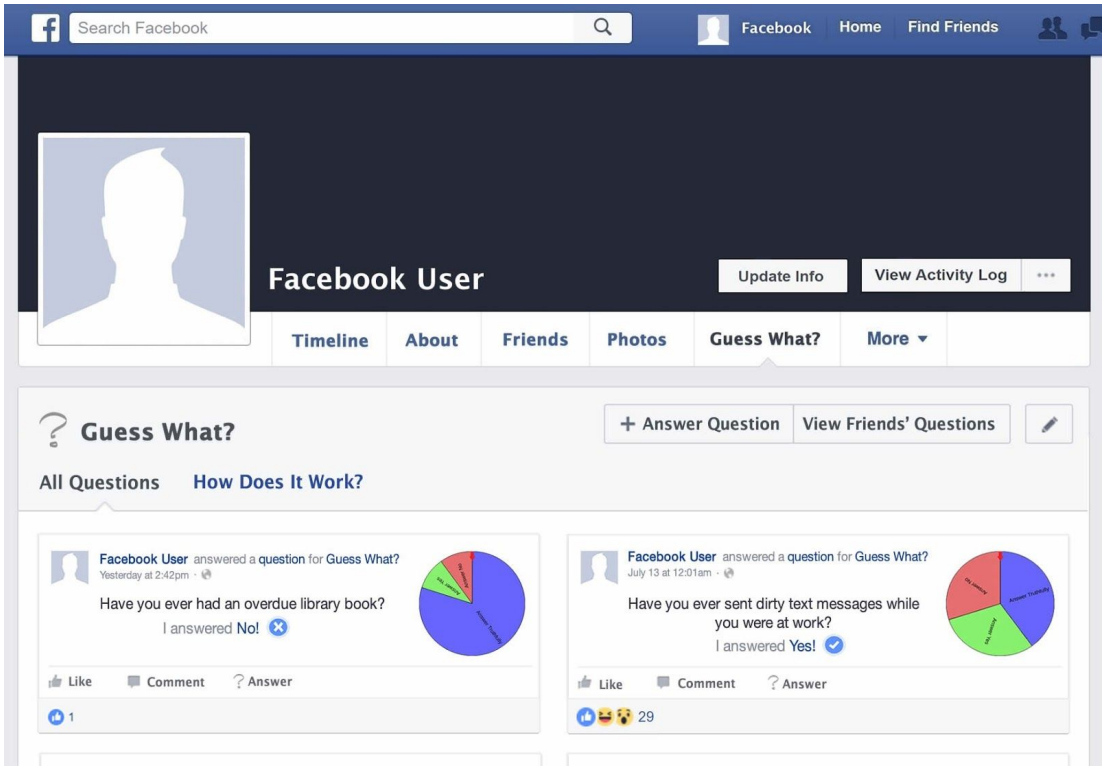
Please list or describe the conclusions of the work and the results achieved. Provide empirical and/or theoretical justification of the results. Include any limitations of the results. Feel free to include any screenshots or diagrams in this section.

We collected data from 228 participants, aiming to gauge their personal comfort, understanding, and trust in Randomized Response Technique (RRT) as a function of random noise. We used a "symmetric" RRT randomizing device whose probabilities of answering yes, no, or honestly varied between spinners; the probability of landing on the "answer honestly" segment was known as the spinner bias.



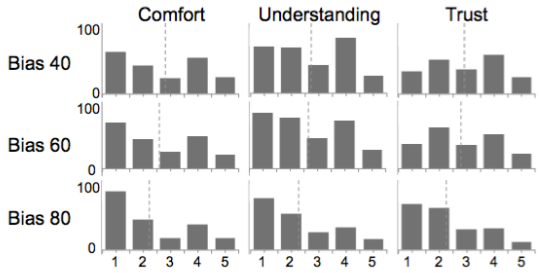
Bias 40 **Bias 60** **Bias 80**

We developed two frameworks in which we presented the application of RRT. One was a completely isolated, ambiguous "database" where we gave our word as researchers that this data would be protected (albeit still susceptible to being compromised by hackers). Another environment tried to simulate a Facebook feature that would publicly display participants' (post-RRT) anonymized responses on their social media profiles. The latter context became necessary in light of the excessive anonymity felt by Mechanical Turk participants that compromised the nature of an experiment rooted in how people felt about privacy online.

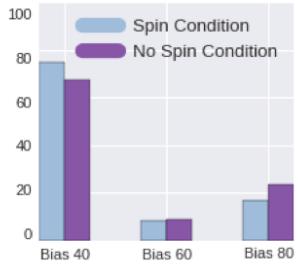


Despite identical spinners and RRT methodologies, participants felt vastly less comfortable with the Facebook framing. This conclusion was garnered from analysis of the collective five-point Likert scale responses from a number of questions comprising user sentiment towards each spinner, including comfort, understanding, and trust. Other factors, such as those gauged with attitudinal trust questions, showed that participants who were innately more trusting or optimistic in their daily affairs tended to avoid the most anonymous (40% biased) spinner. What could this mean in the broader scope of anthropocentric privacy experiments? Variables ranging from the participant's mood that day to more ingrained psychosocial traits and quirks all have potential to swing results in favor of one extreme (anonymity/concern) to the other (honesty/apathy) when analyzing privacy-related decisions.

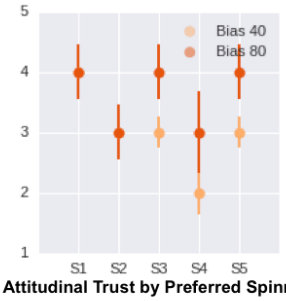
Below are the graphs used to support our claims in our paper, "Towards Understanding Differential Privacy: When Do People Trust Randomized Response Technique?" Figure a. consists of a set of histograms that demonstrate the Likert scale responses from participants along the comfort, understanding, and trust (CUT) metrics for each spinner. Figure b. plots the number of times each spinner was selected to answer the final sensitive question. Figure c. shows disparities in attitudinal trust between the privacy-favoring (Bias 40) and honesty-favoring (Bias 80) groups, where these groups refer to which spinner was ultimately favored to answer the final sensitive question.



Comfort, Understanding, and Trust Sentiment
a



Preferred Spinner
b



Attitudinal Trust by Preferred Spinner
c

5. **Future Work**

Please list or describe your plans for future work on this project or new questions that have resulted from the progress you have made on the research problem.

Future work should further address the following questions:

- How can we design an interface that empowers users to understand DP and make privacy decisions?
- What does anonymity mean?
- How does context impact privacy decisions?

Further studies we'd like to conduct or encourage the next student researchers to conduct would consider the multiplicity of cognitive biases and behavioral influences that can be tapped into when developing privacy-preserving algorithms intended for usability purposes. We also believe it's worthwhile to conduct field experiments that may benefit from being revisited in a different context (for example, a small town/suburb as opposed to a large city, or a survey of rural/largely privacy illiterate participants as opposed to tech-savvy Mechanical Turk users). Finally, we'd like to consider a more pronounced mechanism of adding randomizing noise to data, such that participants could toy with this noise themselves and explore fleshed-out data visualizations. Being a relatively new field/intersection, our privacy + HCI research has potential to shed light on how users consider the devices and mechanisms that collect our data on a daily basis.

6. **Web Links**

Please provide links to your project web pages. If there is more than one, please list each link on a separate line.

https://sites.google.com/a/bucknell.edu/hci_privacy/creu-research-journals

ACM entry to our paper: http://dl.acm.org/citation.cfm?id=3025698

7. **Presentations and Publications**

Please list the titles and venues of any presentations or publications resulting from this project. Please list each on a separate line.

- **The Comfort Quandary: Do People Really Trust Algorithms That Preserve Their Privacy Online?** - Susquehanna Valley Undergraduate Research Symposium
- **Towards Understanding Differential Privacy: When Do People Trust Randomized Response Technique?** - CHI '17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems