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

Brooke Bullek  Stephanic Garboski  Darakhshan J. Mir  Evan M. Peck
Bucknell University  Bucknell University  Bucknell University  Bucknell University
Lewisburg, PA. USA  Lewisburg, PA. USA  Lewisburg, PA. USA  Lewisburg, PA. USA
btb004@bucknell.edu  sag033@bucknell.edu  d.mir@bucknell.edu  evan.peck@bucknell.edu

ABSTRACT
As a consequence of living in a data ecosystem, we often relinquish personal information to be used in contexts in which we have no control. In this paper, we begin to examine the usability of differential privacy, a mechanism that proposes to promise privacy with a mathematical “proof” to the data donor. Do people trust this promise and adjust their privacy decisions if the interfaces through which they interact make differential privacy less opaque? In a study with 228 participants, we measured comfort, understanding, and trust using a variant of differential privacy known as Randomized Response Technique (RRT). We found that allowing individuals to see the amount of obfuscation applied to their responses increased their trust in the privacy-protecting mechanism. However, participants who associated obfuscating privacy mechanisms with deception did not make the “safest” privacy decisions, even as they demonstrated an understanding of RRT. We demonstrate that prudent privacy-related decisions can be cultivated with simple explanations of usable privacy.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation (e.g. HCI): User Interfaces; K.4.1 Public Policy Issues: Privacy

Author Keywords
randomized response; user-centered differential privacy;

INTRODUCTION
The increasing collection and analysis of personal data (ranging from our shopping behavior to our mental and physical health) has created a complex, data-based ecosystem. In the course of constant human-data interactions [11], inhabitants of this data-based ecosystem are implicitly or explicitly making decisions about their privacy. Yet the data collection (and analysis) components are not always designed to raise awareness of privacy or empower individuals to take control of their privacy decisions. When the components of this ecosystem do consider privacy, they may turn to privacy-preserving principles and mechanisms embedded in the design of the system itself – the so-called “Privacy by Design” approach [5]. 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?

Consider the following scenario: suppose analysts wish to examine detailed data collected from many individuals to study aggregate statistics about a population. However, releasing accurate aggregate data or even anonymized data may compromise the privacy of people [13, 20]. To allay these privacy concerns, data curators often seek to 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 [8, 10]. 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. This noise is introduced by individuals who perturb their data themselves (local differential privacy model) or by a supposedly trusted data curator who adds noise to the aggregate statistics before publicly releasing them (global model) [15]. 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) [9].

We study the circumstances under which people are able to trust the “promise of” differential privacy under a local model – Randomized Response Technique (RRT) [26]. Using RRT, individuals are responsible for adding noise to their own data before releasing it. By requesting information on a probability basis, RRT has been shown to elevate rates of self-disclosure and mitigate respondent discomfort [23, 27]. We use RRT in order to achieve control over experimental conditions despite DP’s complex (and diverse) domain, and to focus on whether an individual is likely to understand the implications of the “promise of differential privacy”. Unlike the global model of DP, individuals using RRT don’t need to reason whether they should trust 3rd parties to faithfully implement DP. Further, RRT encompasses an important subset of algorithms that satisfy DP with a level of privacy that depends on a parameter of the technique known as the bias, which we will revisit shortly.

This work extends existing research examining the relationship between users and RRT designs [7, 17, 18]. However, we
concentrate on user experience rather than statistical accountability of the collected, obfuscated data. Our methodology asks the core user-centered question that accompanies RRT—as privacy protections are manipulated via different levels of noise, how do users’ attitudes and behavior reflect that change? By studying the responses of individuals using RRT with varying levels of self-introduced noise, we can begin to investigate the conditions in which people trust the promise of differential privacy. We administered a sensitive questionnaire to 228 participants in order to investigate the following questions.

- **How does the amount of noise that a user adds to their data impact their trust in the technique?** When noise manipulations were made transparent, the amount of mathematical noise (privacy) was directly related to participants’ comfort, understanding, and trust in the security of the database.

- **Do individual differences in attitudes or demographics impact levels of trust and comfort with RRT?** We found that participants who were more trusting in their day-to-day lives relayed concerns that “privacy equals lying,” resulting in the selection of less safe privacy mechanisms.

- **Does altering the interface of RRT’s randomizing device (in this case, animation) impact user trust in the device?** We found no significant difference between representations. It is inconclusive whether animations communicate increased transparency in randomizing devices.

**BACKGROUND**

RRT is widely used to facilitate higher rates of sensitive disclosure [3, 22, 23, 26, 27]. 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 (e.g. a spinner like the one in Fig. 2), and the administrator poses the question, “Have you ever cheated on a college exam?” Rather than answering outright, students only answer honestly if the spinner lands on “Answer Truthfully.” If the spinner lands on “Answer Yes” or “Answer no”, they must respond “Yes” or “No,” respectively. Thus, a “Yes” response cannot be interpreted as an admission of guilt. Still, because the probability of landing any of the three segments is a known property of the spinner, the administrator can statistically deduce the approximate frequency of the sensitive behavior (cheating on a college exam) by examining the results in aggregate [19].

Since its conception, researchers have devised improvements to RRT to facilitate a higher rate of compliance, which correlates with a better understanding of the forced-response technique [17]. Boeije and Lensvelt-Mulders were among the first to administer sensitive questionnaires using RRT via a computer-assisted self-interviewing environment [4]. 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 [4].

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 that they had not done [7]. Furthermore, participants in older studies reported that the “computer does not encourage telling the truth” [4]. 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.

**Communicating Privacy**

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 [1, 14]. Field trials go so far as to introduce authentic additions (e.g. with browser extensions) to existing online platforms (e.g. Facebook) to study real privacy decisions [24, 25]. Privacy “nudges” include adjusting interface colors, buttons, font-weight, and additional verification steps that dissuade users from posting irresponsibly [24]. 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 [25]. This indicates that the transparency of privacy protocols are as important as reliability and efficiency where usable privacy is concerned. To gauge effectiveness of RRT, we used a hybrid of these two approaches; namely, a sand-boxed environment that borrows inspiration from the popular social media platform Facebook.

**METHODS**

235 United States residents were recruited using Amazon’s Mechanical Turk (MTurk), an online platform for workers that has been widely used in privacy studies [16]. Seven participants were discarded after choosing not to give consent of their data after the experiment. Of the 228 remaining participants (105 identified as male, 121 female, 2 other), most (51.3%) fell in the range of 25 to 34 years. Participants were compensated $1.30 for approximately 10 minutes to exceed federal minimum wage requirements.

**Experiment Materials**

Each participant answered 13 sensitive privacy questions to “beta test” a hypothetical new Facebook feature that would publish their responses on their profiles (see Fig. 1). Participants were instructed that this hypothetical feature would permit them to share risqué information online while “keeping...
their Facebook friends guessing.” This framing, while unconventional for RRT studies, was devised to heighten sensitivity and compel participants to remain wary of disclosing private information, even in the presence of a trusted research group with IRB approval. In short, we took advantage of a social context in which participants felt they had something at stake to test RRT in the otherwise anonymous environment of Mechanical Turk. The questions (three of which are listed below) were chosen based on use in prior studies, and were presented to each participant in a random order [2].

- Have you used recreational drugs in the past 6 months?
- Have you engaged in unprotected sex in the past 6 months?
- Have you ever cheated while in a relationship?

For the “randomizing device” that RRT uses to perturb participants’ responses, we used a virtual colored spinner (Fig. 2). Each spinner presented three options for answering a sensitive question – “Answer Yes,” “Answer No,” and “Answer Truthfully” – where the “Yes” and “No” segments were identical in size. For each question, participants clicked a button to spin the wheel and help guide their answers. This is known as a symmetric, forced response randomizing device and has been found to foster the most trust in RRT surveys [3, 21].

**Design and Measures**

We used a 3 by 2 design in which the independent variable tested within participants was the degree of noise (or bias) in the spinner. After a training session to ensure an understanding of the spinner’s randomization, all participants were asked sensitive questions in blocks of four (Fig. 2), each with a different probability that the participant must “Answer Truthfully.” We refer to this weighting as the spinner’s bias (Fig. 3). Before viewing a final highly sensitive question, participants were tested in the spinner most recently used. Using a 5-point Likert scale (“Strongly Agree” to “Strongly Disagree”), participants ranked their comfort, understanding, and trust of each spinner.

- [Comfort] By anonymizing my own answers with this spinner, I would feel comfortable sharing answers to sensitive questions with my Facebook friends.
- [Understanding] It is clear that the technique guarantees secrecy about someone’s activities in real life.
- [Trust] Given access to my responses and this spinner, someone else is unlikely to guess my real answers.

**Attitudinal Trust** was collected to interpret the underlying motivators that influenced each participant’s preferred spinner. Below are the statements whose true/false values were individually ranked by participants with a five-point Likert scale.

- S1 I don’t mind giving out some minor personal information (such as birthday and gender) when registering for accounts on questionable websites.
- S2 Generally speaking, I think that most people have the best intentions.
- S3 I don’t mind lending money to my friends.
- S4 Of the following, I have at least one account with largely unrestricted (i.e. public) visibility settings: Facebook, Twitter, Google+.
- S5 I believe crime statistics in the media accurately reflect what is happening in the U.S.

**RESULTS**

Our primary comparisons in this study involve repeated measures of ordinal data (survey responses of different spinners). As a result, we apply an exact Wilcoxon-Pratt Signed-Rank test for significance and calculate effect sizes. We respond to each of our research questions posed in the introduction.

<table>
<thead>
<tr>
<th>Bias</th>
<th>Comfort</th>
<th>Understand</th>
<th>Trust (all 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40-60</td>
<td>2.74 .13</td>
<td>2.54 .12</td>
<td>2.21 .10</td>
</tr>
<tr>
<td>60-80</td>
<td>4.89 .23</td>
<td>4.60 .22</td>
<td>7.02 .33</td>
</tr>
<tr>
<td>80-40</td>
<td>5.41 .25</td>
<td>5.47 .26</td>
<td>7.31 .34</td>
</tr>
</tbody>
</table>

Table 1. Using Wilcoxon-Pratt Signed-Rank on survey responses for each spinner, we found significant differences between all pairings. All comparisons of the bias 40 and bias 60 spinner were significant ($p < .05$) but with a trivial effect size. Other spinner comparisons (60 v. 80, 40 vs. 80) were also significant ($p < .0001$) with larger effect sizes.
How does manipulating RRT’s anonymizing noise impact a user’s trust in the strategy? Across all spinners, participants were uncomfortable with the idea of sharing sensitive information on Facebook (comfort: MD = 2). However, we found that participant responses were significantly different across all spinner (bias) pairings (see Table 1). The CUT survey (Fig. 4.a) revealed similar trends – negative responses for the most honest (and least private) spinner and significantly more positive responses for the most anonymous spinner. The “middle-of-the-road” bias 60 spinner also showed significant differences between the other two spinners. However, the effect size between the bias 40 and bias 60 spinner was trivial (around .1). This suggests that most people do not distinguish between the change in noise between the bias 40 and bias 60 spinner. However, of the participants that do, they correctly perceive the bias 60 spinner as offering less protection.

When forced to choose a spinner at the end of the experiment, participants vastly preferred the most anonymous spinner (bias 40) (see Fig. 4.b). Surprisingly, the spinner that forced the highest degree of honesty (bias 80) was chosen more than twice as often as the bias 60 spinner.

Do individual differences in attitudes or demographics impact levels of trust and comfort with RRT? The attitudinal trust survey lends insight to the users who preferred the most truthful spinner. Fig. 4.c plots the median Likert sentiment for each of the 5 attitudinal trust statements of bias 40 and bias 80 choosers. 3 out of the 5 statements revealed higher attitudinal trust in participants who preferred the low-privacy, bias 80 spinner. In the open-ended response portion of the survey, several participants from the low-privacy group explained that despite the higher perceived risk of the bias 80 spinner, it was still preferable to being forced to give a false positive response. Their responses included statements such as “I like to be truthful and not deceive or lie to someone”, “I don’t like to lie”, and “Because you can almost always answer truthfully, which means that you have less chance of being forced to lie” The evidence suggests there is an interesting perspective among privacy-lenient users that equates the anonymizing noise provided by RRT with “lying.”

Does altering the interface of RRT’s randomizing device (in this case, animation) impact user trust in the device? We found that animation had little (or no) effect on participants’ CUT metrics or preferred spinners (Fig. 4.b).

LIMITATIONS
While Mechanical Turk provided a large pool of participants, their inflated trust (in both Amazon and academic studies) may not be representative of the average user on the internet [6]. In addition, our Facebook framing may have increased user desires to preserve privacy over traditional contexts, and may not be accurately extrapolated to user experiences. A confounding factor may be participants’ concern with how Facebook friends who lack understanding of or exposure to RRT may misinterpret their responses. Thus, in this public context, participants will likely be more cautious than if their responses were only seen by the researcher who understands RRT.

DISCUSSION AND CONCLUSION
We used RRT to examine one aspect of DP in a clear way—how do users respond to privacy being protected through data perturbations which they can see and understand? Our approach considered the impact of transparent privacy protocols on comfort, understanding, and trust in sensitive online questionnaires. Using virtual spinners as randomizing devices that obfuscate participants’ answers, we found that participants vastly preferred the most “anonymous” spinner.

Still, there emerged a distinctive group of participants who preferred the most “truthful” spinner because it minimized the questionable ethical consequences of lying in their eyes. High self-reported attitudinal trust strongly correlated with this “low-privacy” sentiment, marking a dichotomous preference for anonymity versus honesty among these groups. For future work, this warrants the investigation of innate human tendencies to equate privacy to “lying” if an individual considers data-obfuscation to be unethical. It also raises the question of whether other cognitive biases can sway privacy-related decisions, particularly when using RRT as a proxy for DP.

Further steps could be taken to safeguard users against their own implicit cognitive biases in other unsafe circumstances, such as e-commerce and mobile apps that unnecessarily collect location data. Usable privacy, particularly via elegant, responsive interfaces that offer intuitive explanations of privacy protocols, is an integral part of achieving this undertaking.
REFERENCES


