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

Brooke Bullek
Bucknell University
Lewisburg, PA. USA
btb004@bucknell.edu

Stephanie Garboski
Bucknell University
Lewisburg, PA. USA
sag033@bucknell.edu

Darakhshan J. Mir
Bucknell University
Lewisburg, PA. USA
d.mir@bucknell.edu

Evan M. Peck
Bucknell University
Lewisburg, PA. USA
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

CHI 2017, May 06–11, 2017, Denver, CO, USA.
Copyright © 2017 ACM. ISBN 978-1-4503-4655-9/17/05 $15.00
DOI: http://dx.doi.org/10.1145/3025453.3025698

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.

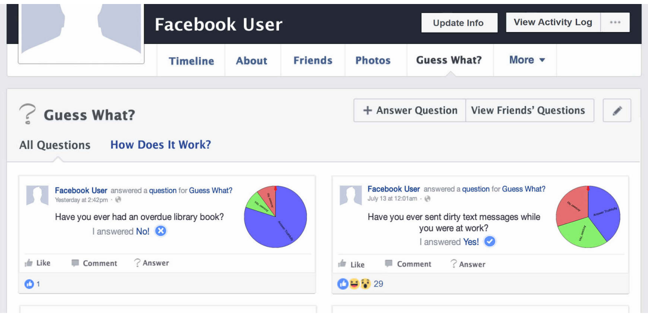
- *How does the amount of noise that a user adds to their data impact their trust in the technique?* When noise manipulations were made transparent, the amount of mathematical noise (privacy) was directly related to participants' comfort, understanding, and trust in the security of the database.
- *Do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



Facebook User answered a question for Guess What?
Yesterday at 2:42pm
Have you ever had an overdue library book?
I answered No!

Facebook User answered a question for Guess What?
July 13 at 12:01am
Have you ever sent dirty text messages while you were at work?
I answered Yes!

Figure 1. The screenshot to set the stage for the Facebook framing that would take place throughout the experiment. The hypothetical "Guess What?" feature allowed participants to envision their responses in a public, high-risk environment.

Communicating Privacy

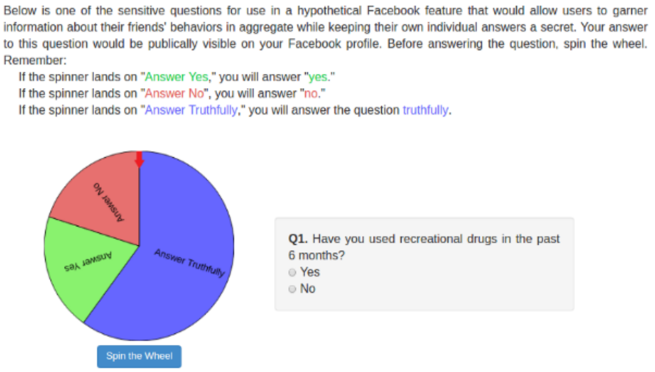
Although privacy research often investigates social media platforms, occasionally, these experiments have been funneled into fabricated, controlled environments meant to directly assess the role played by a system's UI and overall presentation [1, 14]. Field trials go so far as to introduce authentic 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



Below is one of the sensitive questions for use in a hypothetical Facebook feature that would allow users to garner information about their friends' behaviors in aggregate while keeping their own individual answers a secret. Your answer to this question would be publically visible on your Facebook profile. Before answering the question, spin the wheel. Remember:

If the spinner lands on "Answer Yes," you will answer "yes."
If the spinner lands on "Answer No", you will answer "no."
If the spinner lands on "Answer Truthfully," you will answer the question truthfully.

Q1. Have you used recreational drugs in the past 6 months?
○ Yes
○ No

Figure 2. One of 12 screens containing a sensitive question to be answered by the participant using RRT.

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].

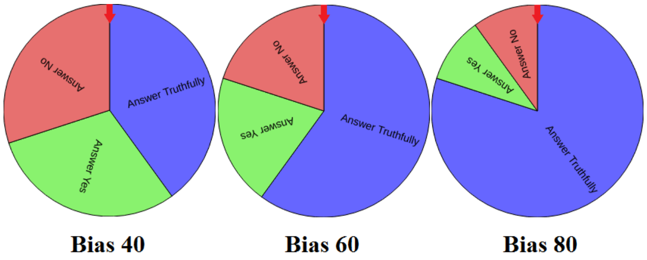
- Have you used recreational drugs in the past 6 months?
- Have you engaged in unprotected sex in the past 6 months?
- Have you ever cheated while in a relationship?

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 asked to select which spinner bias to use in order to conceal their answer. The spinner chosen in this step was labeled the participant's *preferred spinner*.

The independent variable *between* participant groups was the presence of animation in the spinner. Participants were randomly assigned to either *Spin* or *No Spin* groups. While the Spin group experienced fluid animation of the spinner, the No Spin group witnessed their spinners statically "jump" to a location. This manipulation is motivated by research in information visualization suggesting that animated transitions can help people understand changes in statistical data [12]. In this study, we investigate whether it impacts participants' perception of a randomizing mechanism.



Bias 40 Bias 60 Bias 80

Figure 3. The three spinners (randomizing devices) that each participant used to answer sensitive questions. We refer to "bias" as the probability the participant must answer truthfully (represented as blue segments).

Comfort, Understanding, and Trust (CUT) scales were presented to participants after each set of 4 sensitive questions. This provided a direct method of gathering feedback about 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.

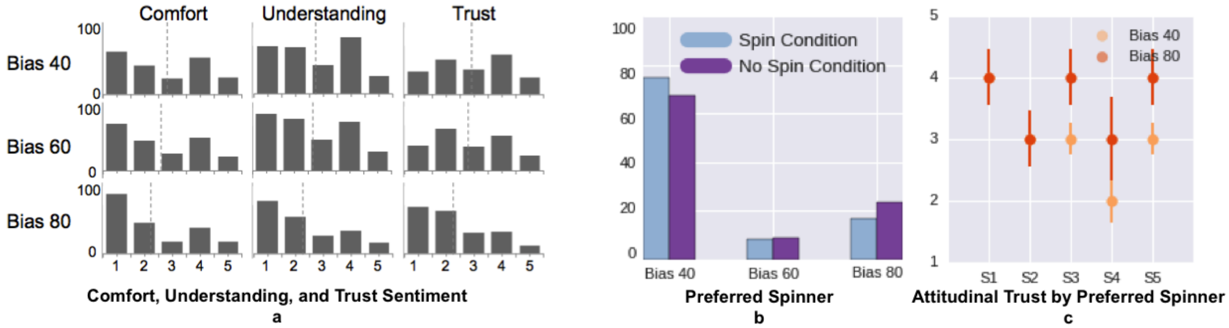
S1 I don't mind giving out some minor personal information (such as birthday and gender) when registering for accounts on questionable websites.
S2 Generally speaking, I think that most people have the best intentions.
S3 I don't mind lending money to my friends.
S4 Of the following, I have at least 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 to test for significance and calculate effect sizes. We respond to each of our research questions posed in the introduction.

	Comfort		**Understand**		**Trust**		**(all 3)**
Bias	Z	r	Z	r	Z	r	p
40-60	2.74	.13	2.54	.12	2.21	.10	$< .05$
60-80	4.89	.23	4.60	.22	7.02	.33	$< .0001$
40-80	5.41	.25	5.47	.26	7.31	.34	$< .0001$

Table 1. Using Wilcoxon-Pratt Signed-Rank on survey responses for each spinners, 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.



Comfort, Understanding, and Trust Sentiment
a

Preferred Spinner
b

Attitudinal Trust by Preferred Spinner
c

Figure 4. (a) We show the distribution of responses to our 5-point Likert scale CUT survey. A dotted line designates the mean. The majority of participants' ratings indicated higher comfort, understanding, and trust metrics in the most anonymous (bias 40) spinner. (b) The percent of participants that chose each bias as their preferred spinner. Given a choice of spinner to answer the final sensitive question, over three quarters of participants selected bias 40. (c) The group that preferred the bias 80 spinner may be accounted for in part by examining attitudinal trust; a disparity in preferences towards "lending money to friends" and "believing crime statistics in the media" surfaces in bias 40 vs. bias 80 choosers.

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 other use cases. 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 obfuscated 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

1. Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and Human Behavior in the Age of Information. *Science* 347, 6221 (2015), 509–515.

2. Alessandro Acquisti, Leslie K. John, and George Loewenstein. 2011. Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information. *Journal of Consumer Research* 37 (2011), 858–873.

3. Graeme Blair, Kosuke Imai, and Yang-Yang Zhou. 2015. Design and Analysis of the Randomized Response Technique. *J. Amer. Statist. Assoc.* 110, 511 (2015), 1304–1319.

4. Hennie Boeije and Gerty Lensvelt-Mulders. 2002. Honest by Chance: A Qualitative Interview Study to Clarify Respondents' (Non-)Compliance with Computer-Assisted Randomized Response. *Bulletin of Sociological Methodology* 75 (2002), 24–39.

5. Ann Cavoukian. 2010. Privacy by Design: The Definitive Workshop. A foreword by Ann Cavoukian, Ph.D. *Identity in the Information Society* 3, 2 (2010), 247–251.

6. Rena Coen, Jennifer King, and Richmond Wong. 2016. The Privacy Policy Paradox. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016).*

7. Elisabeth Coutts and Ben Jann. 2011. Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods & Research* 40, 1 (2011), 169–193.

8. Cynthia Dwork. 2006. Differential Privacy. In *Proceedings of International Colloquium on Automata, Languages and Programming (ICALP 2006)*, Vol. 4052. Springer Verlag, 1–12.

9. Cynthia Dwork. 2011. The Promise of Differential Privacy. A Tutorial on Algorithmic Techniques. In *Proceedings of IEEE Symposium on Foundations of Computer Science (FOCS '11)*. 1–2.

10. Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *Third Theory of Cryptography Conference (TCC 2006)*. Springer Berlin Heidelberg, 265–284.

11. Hamed Haddadi, Richard Mortier, Derek Mcauley, and Jon Crowcroft. 2013. *Human-data interaction*. Technical Report. University of Cambridge.

12. Jeffrey Heer and George G. Robertson. 2007. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics* (2007), 1240–1247.

13. Nils Homer, Szabolcs Szelinger, Margot Redman, David Duggan, Waibhav Tembe, Jill Muehling, John Pearson, Dietrich Stephan, Stanley Nelson, and David Craig. 2008. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLOS Genet* 4, 8 (2008).

14. Thomas Hughes-Roberts and Elahe Kani-Zabihi. 2014. On-Line Privacy Behavior: Using User Interfaces for Salient Factors. *Journal of Computer and Communications* (2014), 220–231.

15. Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal Mechanisms for Local Differential Privacy. In *Advances in Neural Information Processing Systems 27 (NIPS 2014)*. 1–9.

16. Patrick Gage Kelley. 2010. Conducting Usable Privacy & Security Studies with Amazon's Mechanical Turk. *Symposium on Usable Privacy and Security (SOUPS 2010)* (2010).

17. Johannes a Landsheer, Peter Van Der Heijden, and Ger Van Gils. 1999. Trust and Understanding, Two Psychological Aspects of Randomized Response. *Quality & Quantity* 33 (1999), 1–12.

18. Gerty Lensvelt-Mulders, Joop Hox, and Peter Van Der Heijden. 2005. How to Improve the Efficiency of Randomised Response Designs. (2005), 253–265.

19. Peng Tu Liu, Lien Ping Chow, and Wiley Henry Mosley. 1975. Use of the randomized response technique with a new randomizing device. *J. Amer. Statist. Assoc.* 70, 350 (1975), 329–332.

20. Arvind Narayanan and Vitaly Shmatikov. 2008. Robust de-anonymization of large sparse datasets. *IEEE Symposium on Security and Privacy* (2008), 111–125.

21. Martin Ostapczuk, Morten Moshagen, Zengmei Zhao, and Jochen Musch. 2009. Assessing Sensitive Attributes Using the Randomized Response Technique: Evidence for the Importance of Response Symmetry. *Journal of Educational and Behavioral Statistics* 34, 2 (2009), 267–287.

22. Karen Soeken and George Macready. 1982. Respondents' Perceived Protection When Using Randomized Response. *Psychological Bulletin* 92, 2 (1982), 487–489.

23. Uchila Umesh and Robert Peterson. 1996. A Critical Evaluation of the Randomized Response Method. *Sage Publications* 20, 1 (1996), 104–138.

24. Yang Wang, Pedro Leon, Xiaoxuan Chen, Saranga Komanduri, and Gregory Norcie. 2013. From Facebook Regrets to Facebook Privacy Nudges. *Ohio State Law Journal* (2013), 1307–1335.

25. Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A Field Trial of Privacy Nudges for Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. 2367–2376.

26. Stanley Warner. 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *J. Amer. Statist. Assoc.* 60, 309 (1965), 63–69.

27. Felix Wolter and Peter Preisendörfer. 2013. Asking Sensitive Questions: An Evaluation of the Randomized Response Technique Versus Direct Questioning Using Individual Validation Data. *Sociological Methods & Research* 42, 3 (2013), 321–353.