Proceedings of the Human Factors and Ergonomics Society Annual Meeting
2023, Vol. 67(1) 226–232

DOI: 10.1177/21695067231195006
journals.sagepub.com/home/pro
Sage

Exploring Use of Explanative Illustrations to Communicate Differential Privacy Models

Aiping Xiong[1], Chuhao Wu[1], Tianhao Wang[2], Robert W. Proctor[3], Jeremiah Blocki[3], Ninghui Li[3], and Somesh Jha[4]

Abstract

Proper communication is key to the adoption and implementation of differential privacy (DP). In this work, we designed explanative illustrations of three DP models (Central DP, Local DP, Shuffler DP) to help laypeople conceptualize how random noise is added to protect individuals' privacy and preserve group utility. Following a pilot survey and an interview, we conducted an online experiment ($N = 300$) exploring participants' comprehension, privacy and utility perception, and data-sharing decisions across the three DP models. We obtained empirical evidence showing participants' acceptance of the Shuffler DP model for data privacy protection. We discuss the implications of our findings.

Keywords

Differential Privacy, User-centered Deployment, Usable Privacy

Introduction

Differential Privacy (DP, also called Central DP) is a promising approach for preserving privacy with a quantifiable protection guarantee and acceptable utility in the context of statistical information disclosure (Dwork, 2006). Specifically, it adds noise to the aggregated-level results such that an individual's information disclosure is bounded. The US Census Bureau has implemented Central DP to protect the privacy of each participant of the 2020 Census (Abowd, 2018). In recent years, Local DP (Duchi et al., 2013) has become popular because of its deployment in companies such as Google, Apple, and Microsoft. Local DP differs from Central DP in that random noise is added by each user before sending the data to a central party. Thus, users do not need to rely on a trusted third party. Nevertheless, removing the trusted central party comes at the cost of utility. Since every user adds some independently generated noise, the effect of noise adds up when aggregating the result.

More recently, Shuffler DP has been introduced in academia (Cheu et al., 2019) and industry (Bittau et al., 2017), which achieves a middle ground between Central DP and Local DP. Shuffler DP involves an auxiliary party called the shuffler. Users send their perturbed data to the shuffler that shuffles the users' data, and then sends data to the server, thus removing the linkage between users and their reports. Due to this anonymity property, users can add less noise while achieving the same level of privacy. A drawback of Shuffler DP is that it requires that the shuffler should not collude with the server. Otherwise, the user obtains protection only corresponding to the Local DP noise without benefit of shuffling.

With the increasing deployment of DP and its variants, research has been conducted to understand DP from human aspects, such as whether individuals can understand these techniques and consequently increase their willingness to share data (Bullek et al., 2017; Xiong et al., 2020) and their expectations (Cummings et al., 2021). Yet, previous studies have mainly focused on Central DP and Local DP. Little work has been conducted regarding Shuffler DP. Another shortcoming in those previous studies is that they have mostly focused on communicating the privacy benefit of DP. In real-world scenarios, users make the data-sharing decisions by evaluating more than one attribute that may influence the final decision (Krause & Horvitz, 2008). Besides privacy benefit, DP introduces utility cost. Algorithms that follow the concept of DP have a privacy parameter ϵ that

[1]Pennsylvania State University, University Park, USA
[2]University of Virginia, Charlottesville, VA, USA
[3]Purdue University, West Lafayette, IN, USA
[4]University of Wisconsin-Madison, Madison, WI, USA

Corresponding Author:
Aiping Xiong, Pennsylvania State University, Innovation Center, University Park, 16802, USA.
Email: axx29@psu.edu

determines the tradeoff between privacy and utility for a request. Given DP, there is a natural tradeoff between information loss and privacy.

Visualization of such tradeoff using statistics (e.g., confidence intervals) has been proposed and evaluated (Nanayakkara et al., 2022). Results showed that the visualization improved researcher users' comprehension of differentially private data release. Considering that the visualization of those statistics is not accessible to laypersons, we propose to use illustrations communicating the privacy-utility tradeoff of DP models. Prior studies found that learning from illustrated texts produced better performance than learning from texts alone in various educational settings (Levie & Lentz, 1982). Dual coding theory (Clark & Paivio, 1991) also indicates that conveying information in both verbal and non-verbal (e.g., pictorial codes) representations provides double routes for the processing, encoding, and retrieval of the presented information.

We designed explanative illustrations (Mayer & Gallini, 1990) for each model, in which verbal information in natural language and symbolic graphics are presented to promote the comprehension and consideration of the privacy-utility tradeoff across the three DP models. Moreover, with techniques such as Central DP, a company can still collect raw data from individuals, indicating the risk of compromise about which individuals were most concerned (Xiong et al., 2020). Thus, a simple and transparent illustration of such implications is also proposed to help individuals have a complete understanding of DP.

Experiment

We conducted an online survey ($N = 300$) examining the effects of illustrations on participants' comprehension of the DP models, their perceived utility and privacy protection, and data-sharing decisions. A pilot survey and an interview study were conducted before the experiment, validating the initial illustration design and survey questions. Findings in these studies led to improved illustrations and survey questions that were examined in the experiment. We measured participants' data-sharing decisions in two types of scenarios (public good and commercial interests). To contextualize the corresponding decision making, we asked participants to imagine that they were one of the users in the described scenario.

Recruitment and ethics

Both the pilot and formal study were conducted on Amazon Mechanical Turk (MTurk). The human intelligence tasks (HITs) were posted with restrictions to US workers with at least 95% approval rate and 100 or more approved HITs. We made these restrictions in the studies to accurately represent sample restrictions of the recent MTurk research (Hauser & Schwarz, 2016). Participants of the interview study were recruited through convenience sampling by selecting acquaintances who had limited knowledge or prior experience with DP. All studies complied with the American Psychological Association Code of Ethics and were approved by the Institutional Review Board at the authors' institutes. Informed consent was obtained for each participant.

Pilot study

We conducted an online pilot survey on Amazon MTurk ($n = 30$) and then more detailed online interviews of lay users ($n = 6$) to check that the survey questions could be understood and identify any potential problems in the initial illustrations (e.g., text descriptions and data flow diagrams) in advance, such that the methodology could be fine-tuned before launching into the main study.

The online survey participants were typical MTurk workers: mostly White (76.7%), slightly more male (56.7%), most in the age range of 25-44 years (80.0%), about 73.3% having a Bachelor's degree or above, and about 56.7% working in the field of computer or information technology. The survey took 7.5 min (median) to complete, and the payment was $1.50 for each participant. In the online interview, six Ph.D. students with diverse backgrounds (e.g., industrial engineering, education, and agriculture), gender (two female), and ethnicity (e.g., white, East Asian, Latino, African American) were recruited as participants. They completed the online survey before joining the interview session. The interview protocol followed a semi-structured design. Guiding questions were asked based on their survey responses, including 1) "Is everything in the diagram clear to you? Or did you feel confused about anything in the figure?"; 2) "Did you notice any color difference in the diagram? What did (do) you think it means?"; and 3) "Considering the diagram and the above text description altogether, do you feel like the diagram is telling you new information? Any suggestion to improve both of them?" Each participant was compensated with a $20 Amazon gift card after the interview.

We improved the illustration of each model and the survey based on the observations from the pilot survey and the interview study. The distribution of survey time (e.g., the 15-s median viewing time of the illustration of each model) and "Did not read it carefully" theme in the interview indicate that participants tended to skim through the pages and omit illustration details. To address the problem, we set a minimum viewing time (60 s) for each illustration to prevent participants from proceeding too quickly. We also divided the data flow diagram into multiple steps and presented the whole flow with animation to increase users' attention to the details of each step (Hong et al., 2004). We expected the reduced information on each page would help participants better comprehend the data perturbation mechanisms.

The "Local DP vs. Shuffler DP" theme in the interview revealed that participants had difficulty understanding the implications of different data perturbation processes. Thus,

besides clearly presenting technical details, we emphasized the data perturbation implications on security and privacy (e.g., "Because Local DP is applied before the app collects users' data, users' personal information will not be leaked even if the app's database is compromised or hacked by attackers."). Regardless of participants' performance, we presented the illustration or animation again to facilitate further comprehension of each model. We then asked the comprehension question again to assess the effect of feedback and repetition. We also considered that a direct comparison across models might assist users in understanding the differences and impact their data-sharing decisions.

Methodology

Participants. We recruited 400 participants from MTurk. The median survey completion time was about 20 min. The payment was $3.50 for each participant. Four duplicate survey responses were removed. We further filtered participants by the survey duration. Since it took 10 min to watch all videos, we used 15 min as the lower threshold. We plotted the distribution of completion time and cut off responses longer than an hour. As a result, we included 300 participants in the data analysis, with 160 in the illustration condition and 140 in the animation condition. Among those 300 participants, 147 of them viewed the Central DP at first and 153 of them viewed the Local DP at the very beginning. Participants were mostly White (75.3%), slightly more male (57.0%) than female, and most in the age range of 25 − 44 years (76.3%). About 79.6% of the participants had a Bachelor's degree, Medical degree, or worked in a computer or information technology field.

Stimuli. To come up with the illustrations of Central DP, Local DP, and Shuffler DP, we started from the data flow descriptions evaluated by Xiong et al. (2020), which showed the best comprehension results from end users. To make the three DP trust models comparable, all the illustrations followed the same style and logic: We first presented a text description, which was followed by the corresponding data flow diagram. We expected that the text descriptions would help participants conceptualize DP when viewing the data flow diagrams. A utility heatmap showing the utility cost at the aggregated level was presented at the end.

After designing the illustrations, we conducted multiple rounds of internal discussion and review of the illustrations. In the discussions, we involved DP experts to ensure that our illustration of each model was technically accurate. We also conducted pilot studies to validate the illustrations with laypeople. The illustrations for all models were improved in a similar way. Next, we describe the descriptions and illustrations using the Shuffler DP model as an example.

Text description. Besides describing the DP data flow, we made the implication of the DP model explicit in the text description (e.g., collusion between the shuffler and the server of the shuffler DP results in little benefit of shuffling). Based on the pilot survey and the interview study findings, we added a legend listing the set of icons used in the diagram and described the meaning for each of them. Key icons were also embedded in the text description to help participants associate the text and the diagram. In addition, we improved the wording and emphasized the data perturbation processes and implications for privacy protection.

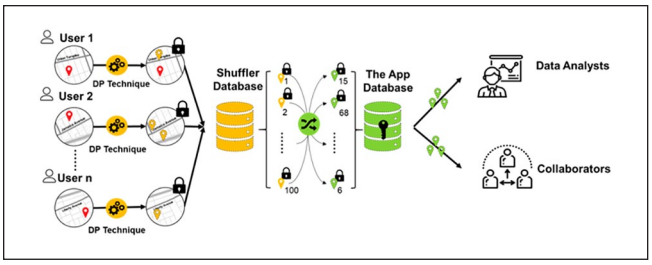

Figure 1. Processes of data perturbation and shuffling using Shuffler DP.

Data flow diagram. The data flow diagram starts from the data collection of individual users (see Figure 1). A snapshot of the map includes a red pin, indicating the actual location of a user. A gear icon represents the DP technique. After the processing of the DP technique, a user's actual location is blurred with some noise (e.g., it becomes a yellow pin somewhere else) such that the user's presence at the location becomes uncertain. We vary the noise perturbation across the users. For example, while a single noise obscured user 1's actual location, user *n*'s actual location was replaced by another noise. Then the shuffler is introduced. A security lock is used to indicate an extra layer of security added to the perturbed data in the shuffler database. Data shuffling (e.g., data of user 1 assigned to user 6) is presented afterward. A green shuffle icon is also presented to indicate the break of the linkage between the users and their data. After data shuffling, an encryption key is used in the App database to indicate the unlock of the security protection for data publishing to data analysts or collaborators.

We also improved the diagram based on the results of the pilot survey and the interview. To increase the contrast in color coding, we used the yellow color referring to the minimized privacy risk with DP, and the green color shows the shuffled data after another layer of security protection (Mayhorn et al., 2004). For color deficient participants, we added a dashed line to code the perturbed data (see Figure 1).

Utility heatmap. We also proposed illustrations showing how the DP model impacts the utility of the collected data at an aggregated level by comparing it to actual data before data perturbation (see Figure 2). To provide spatial context, we used a map with dots that represent a user's location. We set the opacity of a dot as X and allowed occlusion so that the data flow diagram, the red dots indicate users' actual locations. The yellow

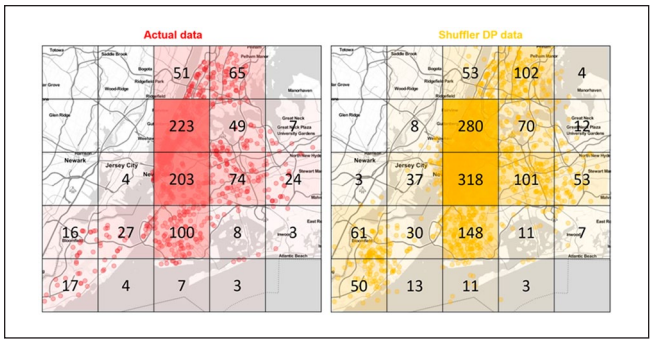

Figure 2. Illustrations of the utility cost for the Shuffler DP model.

dots represent the perturbed location information, indicating relatively low risks. To enhance the illustration of the utility implication, we added a layered heatmap to the original data visualization and labeled the number of data points in each cell. We used positive noise rather than unbiased noise in DP on purpose to make utility cost easily understandable to laypeople.

Procedure. After informed consent, participants were randomly assigned to the illustration condition or the animation condition. For both conditions, the survey started with a description of the location data collection and use. We then introduced one DP model to address the re-identification of anonymized location data. There was a 60-s minimum viewing time for each illustration in the illustration condition. The animation of each model was automatically played. When each animation ended, participants were directed to the next survey page automatically.

After answering one comprehension question following each model, participants received feedback about their performance and were instructed to read/watch the illustration or animation for a second time. Then, a 7-point Likert Scale was used to evaluate the perceived utility and privacy protection of the DP model. Following that, we asked about participants' data-sharing decisions, which were situated in two scenarios: collecting data for research in disease control and prevention (i.e., public good) and for companies to make relevant commercial recommendations (i.e., commercial interests). The 7-point Likert Scale was also used to indicate participants' agreement on sharing data with the DP model. We counterbalanced the order of the two scenarios between participants. The three DP models were presented in a randomized order for each participant except that the Local DP was always presented before the Shuffler DP since the former serves as the basis for the latter. At the end of the survey, participants filled out their demographic information.

Results

Correct answer rate of the comprehension question collapsed across participants (see Table 1) were entered into 3 (model: Central, Local, Shuffler) × 2 (presentation: once, twice) × 2 (order: Central DP first, Local DP first) Chi-squared tests ($\alpha = .05$). Post-hoc tests with Bonferroni corrections were performed, testing all pairwise comparisons with corrected p-values for possible inflation. Participants' average ratings for perceived privacy protection and utility were analyzed with ANOVA using the same three factors as the Chi-squared tests, respectively. We labeled participants' data-sharing decisions as "Yes" if they gave a scale rating of "5" or above. Data-sharing decision rates collapsed across participants were entered into 3 (model: Central, Local, Shuffler) × 2 (scenario: public good, commercial interests) × 2 (order: Central DP first, Local DP first) Chi-squared test. Post-hoc tests were also performed for both perception and decision-making measures. Results between the animation and the illustration conditions did not show any statistically significant differences. Thus, we presented the descriptive results in the tables and figures but omitted them in the reported statistical analyses.

Table 1. Correct answer rate for the comprehension question of each DP model. Numbers in the parentheses indicate the number of participants in each condition.

	Central DP		Local DP		Shuffler DP	
	Once	Twice	Once	Twice	Once	Twice
Illustration (160)	72.5%	79.4%	45.6%	47.5%	71.9%	80.0%
Animation (140)	64.3%	78.6%	56.4%	47.1%	75.7%	79.3%

Comprehension. Table 1 shows the correct answer rate for comprehension questions across the three models after viewing the illustrations or animations once and twice. Compared with the Central DP and the Shuffler DP, the Local DP had an overall lower correct rate (= 39.83, $p < .001$). The question for Local DP asked about the privacy implication of data sharing with a third party, which was not directly explained in the illustration. The low correct rate suggests that participants may only have grasped information explicitly expressed for the model (i.e., privacy implication of a breached or hacked database).

Comparing the results of watching animations/ illustrations once versus twice, we only found that the Central DP showed a significant increase (= 7.76, $p = .005$), and such pattern was more evident for the animation condition than for the illustration condition ($p = .002$). We further examined whether the order of model presentation had an effect. When comprehension questions were asked for the first time, there was no significant difference between the two presentation orders. However, for the second time, the correct rate for the Central DP was significantly higher when the Central DP was presented last than when it was presented first (= 9.00, $p = .003$). Thus, the increased correct rate of the Central DP could be attributed to the order of presentation instead of increased understanding of the model after viewing the illustration again. When Central DP was presented last, participants could have a clearer memory of the corresponding

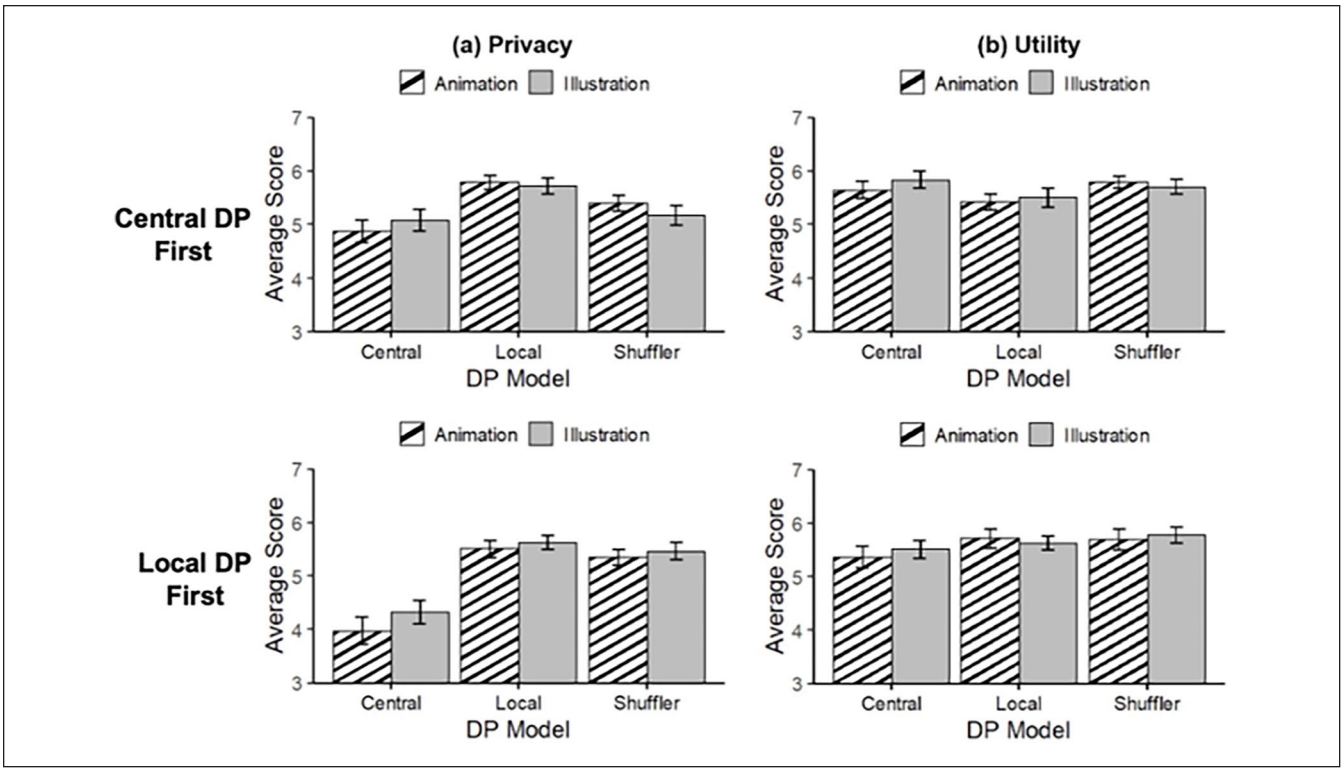

Figure 3. Results of privacy and utility perception.

information. Based on the results of comprehension questions, we further filtered the data for the analysis hereafter. Specifically, only participants with at least two correct answers in either the first or the second time were kept, resulting in 261 responses (138 in the illustration condition, 123 in the animation condition, 129 with Central DP presented first, and 132 with Local DP presented first).

Privacy and utility perception. Figure 3 shows participants' average rating of perceived privacy and utility for the three DP models. For the perceived privacy protection, the main effects of model ($F(2,771) = 40.75$, $p < .001$, $= .10$), presentation order ($F(1,771) = 8.49$, $p = .004$, $= .01$), and their interaction ($F(2,771) = 7.65$, $p < .001$, $= .02$), were all significant. Post-hoc comparisons revealed that participants gave higher rating for the Shuffler DP (5.35) than that of the Central DP (4.56, $< .001$), but the rating of Shuffler DP was lower than that of the Local DP (5.66, $= 0.04$). Such results are in agreement with the correct understanding of privacy implications across the three DP models. Post-hoc comparisons also showed that the ordering effect was only evident for the Central DP. Specifically, when the Local DP was presented first, the average rating for the Central DP was lower (4.14) than that when the Central DP was presented initially (4.98, $< .001$). Thus, the presentation of Local DP and Shuffler DP could have impacted people's perceived privacy of the Central DP, but not vice versa.

In terms of the perceived utility, the ANOVA showed no significant effects at all. Given the heatmap and numbers (see Figure 2), the utility implications of DP models should not be difficult for participants to understand. A possible explanation for the obtained results is that the reduced accuracy of the three models were all acceptable for the participants. Alternatively, participants might have fewer concerns about the utility compared to the privacy.

Data-sharing decision. Participants' data-sharing decisions (see Figure 4) were 58% for the Central DP, 73.6% for the Local DP, and 70.7% for the Shuffler DP. The main effect of DP model was significant ($= 32.40$, $p < .001$). Post-hoc pairwise comparisons revealed that the differences were mainly due to the decision rate of the Central DP being smaller than that of the Local DP and the Shuffler DP ($< .001$). Also, participants showed more willingness to share data in the public-good scenario (71.9%) than in the commercial-interests scenario (63%, $= 13.80$, $p < .001$). The main effect of presentation order ($= 8.80$, $p < .01$) and its interaction with DP model ($= 9.01$, $p = .002$) were significant. Specifically, participants were more likely to share data when the Central DP was presented first (71.1%) than when the Local DP was presented first (63.9%, $= .001$), which again implies the impact of the Local DP and the Shuffler DP on participants' evaluation of the Central DP.

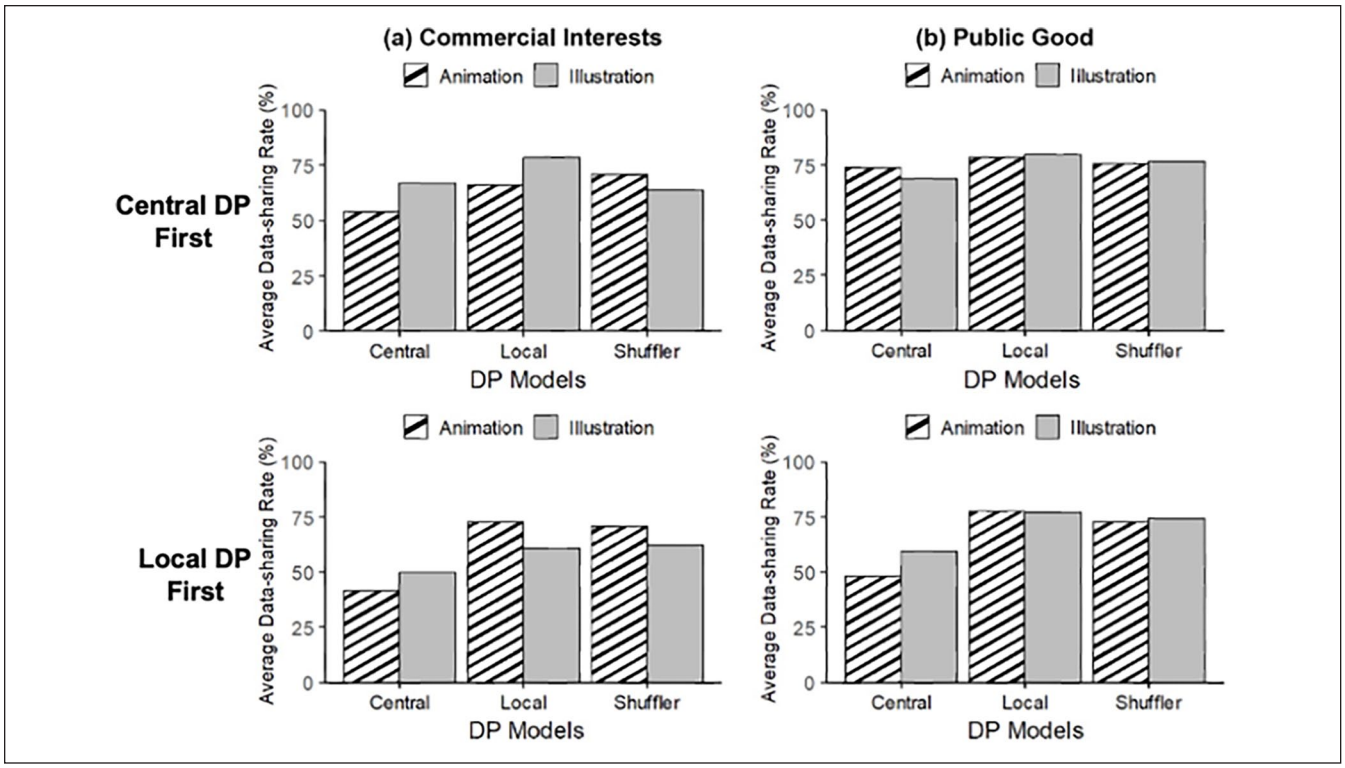

Figure 4. Results of data-sharing decision.

Discussion

We found that participants' perceived privacy protection but not their perceived utility varied across the DP models. Specifically, they perceived the most privacy protection of the Local DP, followed by the Shuffler DP and the Central DP, indicating that they understood the key differences across the DP models. Such a pattern was more evident when the Central DP was presented after the other two models than when it was presented first. Regardless of the scenarios, participants preferred stronger privacy protection (i.e., more selection of the Local DP or the Shuffler DP than the Central DP) when asked to share location data. The effect of DP model was also more evident when the Central DP was presented later. Such order effect of the Central DP can be interpreted as an effect of reference frames (Kahneman & Tversky, 2013), suggesting that it is essential to have a relative base (or frame) for end users' practical evaluation of perceived privacy protection.

Limitations

Our work has limitations. *First*, we recruited MTurk workers in the study. Thus, participants are younger, more technical, and more privacy-sensitive than the overall U.S. population (Kang et al., 2014). This is evident in our results, which demonstrate a large percentage of participants have experience in the fields of computer or information technology. We believe these limitations are acceptable, as the public has limited knowledge of differential privacy in general. *Second*, we only asked participants' data-sharing decisions on two data usage scenarios (i.e., public good and commercial interests), which we considered being reasonably representative but not comprehensive. *Third*, we did not obtain any differences in perceived utility across the three models. Future work could consider asking open-ended questions or conducting an interview study to understand the reasons behind participants' selections.

Implications to DP Community and Beyond

To design and deploy privacy-enhancing technologies that effectively protect end users, we need to understand how end users perceive privacy protection, the associated cost, and what influences their decision to adopt (or not adopt) privacy-enhancing tools. A key takeaway from our work is that explanative illustrations can be effective in communicating DP models to end users. Given an adequate understanding of the DP models, end users' perceived privacy protection matches the protection offered by each model. End users also prefer stronger privacy protection for their disclosure decisions. To this end, we argue that organizations and companies should consider being transparent on the details of DP deployment (e.g., trust model and the choice of privacy parameters) that can impact end users' data-sharing decisions.

Our work can benefit the DP community and beyond in various aspects. *First*, users' awareness and comprehension of DP implementation can foster their trust and confidence in organizations/companies (i.e., data users) for data collection and use. *Second*, we demonstrate how to design explanative illustrations to communicate DP and use various measures to examine the effectiveness of such communication. A similar approach can be used by organizations and companies for communicating extra DP models and other privacy-enhancing tools. *Third*, our findings can offer insights to organizations and companies for communicating DP in practice. For example, a comparison to existing privacy-enhancing tools might help users better understand and accept DP models.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1931441.

References

Abowd, J. M. (2018). Protecting the confidentiality of America's statistics: Adopting modern disclosure avoidance methods at the census bureau,https://www.census.gov/newsroom/blogs/research-matters/2018/08/protectingtheconfi.html.

Bittau, A., Erlingsson, A., Maniatis, P., Mironov, I., Raghunathan, A., Lie, D., Rudominer, M., Kode, U., Tinnes, J., & Seefeld, B. (2017). Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*, pp. 441–459.

Bullek, B., Garboski, S., Mir, D. J., & Peck, E. M. (2017). Towards understanding differential privacy: When do people trust randomized response technique?" In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 3833–3837. ACM.

Cheu, A., Smith, A., Ullman, J., Zeber, D., & Zhilyaev, M. (2019). Distributed differential privacy via shuffling. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer*, pp. 375–403.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*, 149–210.

Cummings, R., Kaptchuk, G., & Redmiles, E. M. (2021). I need a better description: An investigation into user expectations for differential privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pp. 3037–3052.

Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2013). Local privacy and statistical minimax rates. In *IEEE 54th Annual Symposium on Foundations of Computer Science*, pp. 429–438. IEEE.

Dwork, C. (2006). Differential privacy. In *International Colloquium on Automata, Languages, and Programming*, pp. 1–12, Springer.

Hauser, D. J., & Schwarz, N. (2016). Attentive turkers: Mturk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*, 400–407.

Hong, W., Thong, J. Y., & Tam, K. Y. (2004). Does animation attract online users' attention? the effects of flash on information search performance and perceptions. *Information Systems Research*, *15*, 60–86.

Kang, R., Brown, S., Dabbish, L., & Kiesler, S. (2014). Privacy attitudes of mechanical turk workers and the us public. In *Tenth Symposium on Usable Privacy and Security*, pp. 37–49.

Krause, A., & Horvitz, E. (2008). A utility-theoretic approach to privacy and personalization. In *Proceedings of the 23rd National Conference on Artificial Intelligence*, pp. 1181–1188.

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *ECTJ*, *30*, 195–232.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, *82*, 715–726.

Mayhorn, C. B., Wogalter, M. S., Bell, J. L., & Shaver, E. F. (2004). What does code red mean? *Ergonomics in Design*, *12*, pp. 12–14.

Nanayakkara, P., Bater, J., He, X., Hullman, J., & Rogers, J. (2022). Visualizing privacy-utility trade-offs in differentially private data releases. *Proceedings on Privacy Enhancing Technologies*, *2*, 601-618.

Roth, E., Zhang, H., Haeberlen, A., & Pierce, B. C. (2020). Orchard: Differ-entially private analytics at scale. In *14th USENIX Symposium on Oper-ating Systems Design and Implementation (OSDI 20),* pp.1065–1081.

Xiong, A., Wang, T., Li, N., & Jha, S. (2020). Towards effective differential privacy communication for users' data sharing decision and comprehension. In *2020 IEEE Symposium on Security and Privacy (SP)*, pp. 392–410, IEEE.