The Influence of Explanation Designs on User Understanding Differential Privacy and Making Data-sharing Decision

Zikai Alex Wen[a], Jingyu Jia[b,*], Hongyang Yan[d], Yaxing Yao[c], Zheli Liu[b], Changyu Dong[d,*]

[a]*Computational Media and Arts Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China*
[b]*College of Computer Science, Nankai University, Tianjin, China*
[c]*Department of Information Systems, University of Maryland, Baltimore County, Baltimore, United States*
[d]*Institute of Artificial Intelligence and Blockchain, Guangzhou University, Guangzhou, China*

Abstract

Differential privacy (DP) technologies are being promoted by organizations to encourage data sharing, but without a proper understanding of how these technologies work, individuals may make incorrect data-sharing decisions. A design gap exists in effectively communicating the workings of DP technologies, such as Local DP, to users. Our research aimed to fill this gap through the use of an explanatory illustration. We conducted an online survey with 228 participants to assess the impact of different explanation designs on understanding DP and data-sharing decisions. Our study found that the visual explanatory illustration was more effective in assisting individuals to comprehend Local DP's privacy protection against large organizations as compared to the textual description, with the illustration group exhibiting an increase of 51.4% in their comprehension. The study also found that improved knowledge of privacy-enhancing technologies does not guarantee willingness to share protected data. To prevent misinformed decisions, future research could focus on designing a more effective way of communicating the privacy protections of these technologies to users, building on the insights gained from our study.

1. Introduction

The analysis of big data has the potential to bring positive benefits to society, but if it is conducted without proper data privacy protection, it poses a significant risk to the privacy and security of individuals [1]. To address this issue, organizations such as government agencies [2, 3] and companies [4, 5] have been advertising that they adopted differential privacy (DP) technologies to protect everyone's privacy in big data analyses. However, studies [6, 7, 8] have revealed that many regular users may have misunderstandings about the way privacy-enhancing technologies operate and their ability to effectively protect privacy. For instance, many people may mistakenly think that using private browsing mode will conceal their physical location from the websites they visit [6].

In an effort to increase the understanding of DP technologies among the general public, a growing number of research projects [9, 10, 11, 12, 13] have sought to enhance the explanation

[*] Jingyu Jia and Changyu Dong are the corresponding authors.

of DP technology. Some researchers have utilized text-based descriptions [9, 10] to explain DP technologies, while others [11, 12, 13] have utilized visual encoding. However, no existing design effectively allows users to grasp the protection of numerical private data offered by DP technology in a quick and simple manner.

Our study addressed the need by formulating a design space of DP explanation using the design space exploration research method [14]. The design space consists of four forms of explaining DP technology: textual descriptions, tables, charts, and explanatory illustrations. Describing the data perturbation process in Local DP using text is challenging for people to understand [9]. To address this challenge, we focused on designing visual explanations, including a table, a chart, and an explanatory illustration. We conducted a preliminary study with five participants to assess the clarity and impact of the visual designs on people's decisions about sharing data. The results showed that the explanatory illustration had the most potential to help users comprehend how DP technology protects numerical private data effectively, compared to the other two visual explanations.

To determine the impact of the different design elements used in textual descriptions and visual illustrations on users' understanding of DP as a privacy-enhancing technology and their willingness to share private data, we conducted a large-scale survey. The survey aimed to address the following three research questions.

RQ1 Can our explanatory illustration design improve people's understanding of Local DP compared to a text description?

RQ2 If RQ1 holds, what design elements contribute to its effectiveness?

RQ3 If RQ1 holds, does increased knowledge acquisition through the illustration design affect people's decisions to share personal data? If so, what is the reason behind this?

We carried out an online survey with 228 participants. The survey was designed as an A/B test, where participants were randomly assigned to either read a textual description of Local DP or an explanatory illustration. We asked three types of closed-ended questions: questions to test participants' knowledge, questions to gauge the extent of their privacy concerns, and questions to assess their willingness to share data. The same closed-ended questions were repeated before and after the DP explanation was provided. Finally, participants were given their ratings for their privacy concerns and their willingness to share data and asked to explain any changes or similarities in their ratings.

The results of our survey provided answers to our three research questions. Firstly, both the textual description and the explanatory illustration helped participants understand Local DP to some extent. However, the improvement in the understanding of how Local DP protects data privacy from organizations was much greater in the group that read the explanatory illustration, with a 51.4% increase in the rate of correct answers, compared to the group that read the textual description, where the rate barely changed. This suggests that the explanatory illustration was more effective in conveying the concept of Local DP. We attribute this to the clear explanation of the mathematical concept of probability using an analogy of a lottery draw. Lastly, the survey results indicated that even though participants learned more about a privacy-enhancing technology, their willingness to share data under its protection did not necessarily increase.

Our findings also revealed certain limitations in the current explanatory illustration design. Some participants had additional questions that were not addressed, leading to confusion regarding Local DP. To address this issue, future research could incorporate natural language dialog techniques along with our explanatory illustration design to create a more effective way of communicating privacy protections to users. This approach has the potential to better inform individuals when making decisions about sharing private data.

2. Background and Related Work

In this chapter, we start by giving an overview of DP by discussing its background information. This includes explaining the meaning and relationship between various privacy parameters in the mathematical definition of DP. Then, we describe how DP has been embraced by leading IT companies and how they have used it to assure users of the protection of their data privacy. Additionally, we link the research on explaining DP to the research on explaining privacy-enhancing technologies in general. Lastly, we provide a summary of prior design work aimed at explaining DP to the general public and how the limitations of these designs motivated our research.

2.1. Background on Differential Privacy

DP is the state-of-the-art privacy definition for privacy-preserving data analysis, which protects personal privacy by limiting the impact of an individual record on the statistical analysis results. DP assumes that a trusted third-party data curator collects users' authentic data and adds random noise to the analysis results so that malicious parties cannot reverse-engineer the details of individual records through the analysis results. The formal definition of DP is as follows.

Definition 1. *(Differential Privacy) [15] Let $\varepsilon \geq 0$ and $\delta \in [0,1)$. A randomized mechanism $M : \mathcal{X}^n \rightarrow \mathcal{Y}$ satisfies $(\varepsilon, \delta)-$differential privacy if and only if for any two datasets $X, X' \in \mathcal{X}^n$ that differs in only one record, and any $Y \subseteq \mathcal{Y}$, we have*

$$Pr[M(X) \in Y] \leq e^{\varepsilon} Pr[M(X') \in Y] + \delta$$

As shown in Definition 1, the impact of changes in individual user records on the database analysis results is limited by a pair of parameters: ε and δ. One uses the privacy parameter ε to adjust the level of privacy protection of the DP: a small value of ε indicates a high level of privacy protection. Conversely, a larger value of ε indicates a higher level of privacy leakage. The parameter δ is used to limit the variation of the output distribution of the analysis results beyond e^{ε}.

In practice, ε is usually fixed to a reasonably safe value to ensure that the output distribution of the analysis results is essentially the same. δ is usually a negligible value to ensure that the differences in output distributions are bounded by e^{ε} in the vast majority of cases.

Local DP is a variant of DP that requires a weaker trust assumption than DP. Local DP does not need a trusted data curator because the curator can no longer accurately infer a user's authentic data after the user perturbs their data locally. The formal definition of Local DP is as follows.

Definition 2. *(Local Differential Privacy) [16] Let $\varepsilon \geq 0$ and $\delta \in [0,1)$. A local randomized mechanism $M : \mathcal{X} \rightarrow \mathcal{Y}$ satisfies $(\varepsilon, \delta)-$local differential privacy if and only if for any two individual records $x, x' \in \mathcal{X}$ and any $Y \subseteq \mathcal{Y}$, we have*

$$Pr[M(x) \in Y] \leq e^{\varepsilon} Pr[M(x') \in Y] + \delta$$

By comparing the difference between Definition 1 and Definition 2, we can tell that the only difference between Local DP and DP is the number of data records that need to be obfuscated. The M mechanism in Local DP adds random noise to two individual data records to ensure that no one can infer from the data analysis output which data record produced the analysis result. In other words, Local DP can be interpreted as a special case of DP to protect two datasets of size 1, so its privacy protection capability is required to be higher than that of DP.

2.2. Using Differential Privacy in Practice

IT companies [4, 5, 17] and government agencies [18, 3, 2] have implemented Local DP to safeguard user privacy. For instance, Google created RAPPOR [4] to add noise to string input data on a user's device, while Apple employed a Local DP algorithm to protect user input data and website visit history [5]. As these DP technologies become increasingly prevalent, our research focuses on explaining their benefits and limitations to the general public.

2.3. Explanation of Privacy-Enhancing Technologies

DP technology is designed to enhance privacy. Other privacy-enhancing technologies include anonymous web browsing [19], geo-blocking [20], and end-to-end encrypted secret chats [21], etc. Computer systems utilize these technologies, including DP, to let users feel more at ease when sharing personal information with certain parties. It is important to clearly explain the strengths and limitations of privacy-enhancing technologies to users to prevent any unintended harm from misunderstanding these technologies' effectiveness.

Studies aimed at effectively communicating the privacy risks [22, 23, 6, 24, 25, 26] and benefits [27, 28, 7, 8] of privacy-enhancing technologies have been conducted for many years. Despite this, research [22, 26] has shown that a significant portion of these explanations fail to help individuals make informed decisions about sharing their personal information. In some instances, these explanations may even be misleading [6, 7, 8], such as leading users to believe that private browsing fully conceals their physical location from websites [6].

Like other privacy-enhancing technologies, the current explanations of DP may also be inadequate or misleading to ordinary users, who have typically learned about DP through text-based descriptions. DP has a unique feature compared to other privacy-enhancing technologies, as its protection is probabilistic, meaning there is a degree of privacy leakage. As a result, researchers [29, 10, 11, 12, 13, 30, 31] have begun to investigate the impact of existing DP explanations on users' data-sharing decisions.

2.4. Designs of Differential Privacy Explanation

Cummings et al. [29] discovered that traditional written explanations of Local DP do not effectively guide users in making informed decisions about sharing personal data. The researchers provided six different explanations of Local DP to the survey participants to help them comprehend Local DP. After the participants had been given the explanations, they were asked both knowledge-based and voluntary questions related to private data sharing. The results revealed that fewer than half of the participants answered the knowledge questions correctly, leading many to have misconceptions about Local DP's ability to secure the data they upload to the cloud. This finding has prompted a growing number of researchers [10, 11, 12, 13] to investigate ways to better communicate Local DP's data protection abilities to the general public.

The existing approaches to enhance the explanation of DP can be classified into two categories: text description design [9, 10, 31] and visual illustration design [11, 12, 13, 30].

The text description design is the first category that we will discuss. Xiong et al.[9] attempted to design various texts that explain DP and Local DP. However, their user study showed that people had difficulty comprehending the working principle of Local DP from the text, particularly in understanding the data perturbation process. Franzen et al.[10] utilized the risk communication format from the medical field to design an explanation of DP privacy guarantees, but their user study revealed that their explanation design did not significantly improve users' understanding of DP knowledge or increase their confidence levels. Smart et al. [31] elaborated on how the algorithm parameters can impact users' safety when sharing more browser history data. The results from their knowledge test indicated that users were overconfident in their understanding. In general, the current research on text description design has not been effective in helping users grasp DP knowledge.

The second category is the visual illustration design, which aims to overcome the limitations of text descriptions by incorporating visual aids. The dual coding theory [32] supports this design approach, as it suggests that combining images and language is more conducive to information processing, comprehension, and recall. Based on this theory, Nanayakkara et al. [13] designed an interactive visualization tool for data managers who are unfamiliar with DP, to help them provide correct DP protection for user data. Xiong et al. [12] created explanatory illustrations to demonstrate how DP protects geolocation data privacy for ordinary users. Their user studies showed that the illustration designs helped people better understand DP, but it still required extensive reading and comprehension. Additionally, their design cannot be applied directly to explain DP protection for numerical data, such as income levels and medical records, which is a high-demand scenario for ordinary users who want to quickly understand DP's protective effect on their numerical data before deciding whether to share it. Our research goal is to address such high-demand scenarios through visual design.

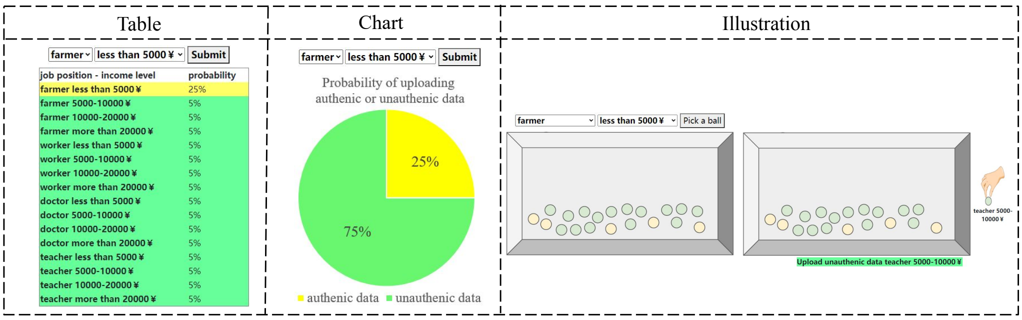
Previous studies [13, 12, 9] have conducted controlled experiments to assess the impact of various DP explanations on users' understanding of DP and data sharing decisions. However, these experiments only evaluated different content designs using the same improvement method. While illustrations have been used to explain DP, there has been no comparison between illustrated explanations and the most effective textual explanations. Textual explanations are more prevalent in current DP explanations than illustrated ones. Thus, our research aims to compare the effectiveness of visual and textual designs in improving users' understanding of Local DP.

3. Design Space Exploration

Due to difficulties in comprehending Local DP through text descriptions [9], we aim to explore alternative explanation designs to communicate the data perturbation process better. Before developing these designs, analyzing the existing textual content designs used to describe Local DP is necessary.

Previous designs on textual DP explanations [9, 29] have used the word 'randomly' to describe the perturbation of user data in Local DP. For example, Xiong et al. [9] wrote, "The app will modify your data *randomly* before sending it to the app server." Cummings et al. [29] described, "In the local DP model, users perturb their information *randomly* with the help of a collection mechanism, such as their device, before sending it to the curator for analysis." These general descriptions do not provide a clear picture of the randomized mechanism.

Describing the mechanism in detail may introduce unintelligible technical jargon to ordinary users. Thus, we considered using visuals to explain the mechanism more intuitively. We used the design space exploration method [14] to systematically evaluate three design categories: Table,



job position - income level	probability
farmer less than 5000￥	25%
farmer 5000-10000￥	5%
farmer 10000-20000￥	5%
farmer more than 20000￥	5%
worker less than 5000￥	5%
worker 5000-10000￥	5%
worker 10000-20000￥	5%
worker more than 20000￥	5%
doctor less than 5000￥	5%
doctor 5000-10000￥	5%
doctor 10000-20000￥	5%
doctor more than 20000￥	5%
teacher less than 5000￥	5%
teacher 5000-10000￥	5%
teacher 10000-20000￥	5%
teacher more than 20000￥	5%

Figure 1: Three non-textual design prototypes for communicating the data perturbation process of Local DP: Table (a probability distribution), Chart (a pie chart), and Illustration (a lottery game).

Chart, and Illustration, as shown in Figure 1. The data table displays the probability distribution for each authentic or fake data that may be uploaded online. The chart depicts the expected proportion of authentic or fake data in the uploaded data. The illustration uses a series of schematics or pictures to show the process of adding random noise to authentic data. Of the three designs, the illustration is the most effective in communicating the process and the most challenging to create.

The following paragraphs elaborate on how the different design prototypes illustrate the data perturbation procedure. These prototypes have been created based on typical user scenarios. We have adopted a usage scenario described by Cummings et al. [29] where users are asked to provide information about their job title and income to aid in advancing social justice. The users can choose from four job titles and four income levels to upload.

Our prototypes will demonstrate a fundamental data perturbation process that protects numerical data. This process involves a randomized response and sets the privacy budget, ε, to ln(3). It means that the user's actual data has a 25% chance of being uploaded. In contrast, each of the remaining fake data options has a 5% probability of being uploaded instead of the actual data.

(1) Table Design: The table design shows all the available data upload options and ranks them based on their upload probability. Once the user inputs mock data, the table updates the upload probability and highlights the user's input row in yellow and the remaining options in green. Suppose the user understands the concept of probability. In that case, they can comprehend that there is a 75% chance of the input data being transformed into fake data and a 5% chance of each fake data replacing the authentic data.

(2) Chart Design: The pie chart is the best way to visually represent the perturbation probability of the user's input data. In our design, a quarter of the chart is filled with yellow to represent the 25% chance of the authentic data being sent to the server. The remaining three-quarters of the chart is green, indicating the non-authentic data. The pie chart is less misleading than the table design as its background color proportion corresponds to the perturbation probability. However, it sacrifices data details, as the user cannot see the specific fake data or their replacement probability.

(3) Illustration Design: The illustration design is inspired by lottery events used to teach probability. The authentic data is represented as yellow balls in a raffle box, while the fake data is represented as green balls. The data perturbation process of Local DP is analogized to randomly picking a ball from the raffle box. The small balls symbolize a 5% probability, the minimum upload probability in the user scenario. Thus, the illustration shows five yellow and 15 green balls in the raffle box.

When the user inputs data and clicks the "Pick a Ball" button, they will see an animated illustration of a hand picking a ball from a gray box. As depicted in Figure 1, on this occasion, a green ball was selected, representing an altered data record for a teacher with an income between 5,000 and 10,000, as opposed to the user's original input of a farmer with an income less than 5,000. The illustration design combines the benefits of the previous two designs and eliminates their drawbacks. Users can click the button repeatedly to see how probability impacts the data perturbation process.

4. Iterative Design and Evaluation Method

Our iterative design and evaluation process consisted of three steps, as shown in Figure 2. Each of these steps is described in the following three sections.

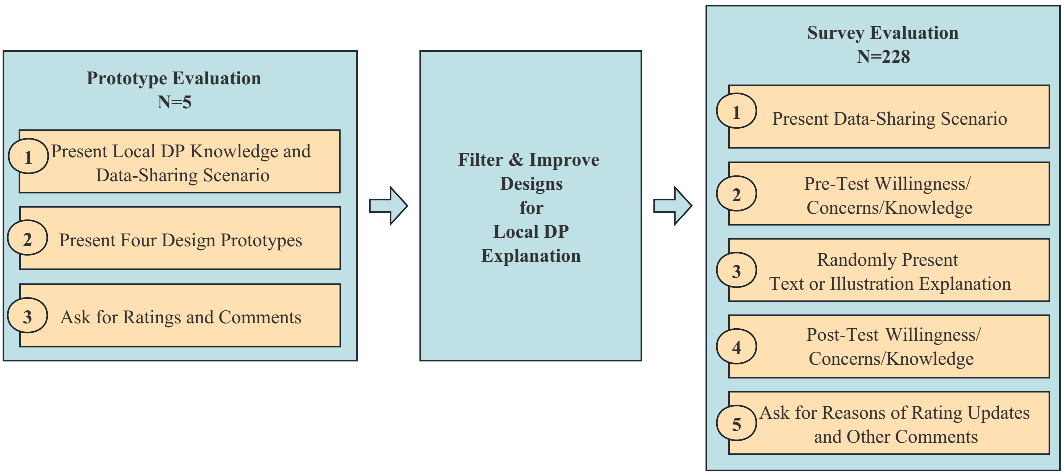

Prototype Evaluation
N=5
1 Present Local DP Knowledge and Data-Sharing Scenario
2 Present Four Design Prototypes
3 Ask for Ratings and Comments
Filter & Improve Designs for Local DP Explanation
Survey Evaluation
N=228
1 Present Data-Sharing Scenario
2 Pre-Test Willingness/ Concerns/Knowledge
3 Randomly Present Text or Illustration Explanation
4 Post-Test Willingness/ Concerns/Knowledge
5 Ask for Reasons of Rating Updates and Other Comments

Figure 2: The frame diagram of our iterative design and evaluation method.

The first step was a small pilot study with five participants, where we aimed to identify the most effective prototype among the three visual designs in explaining the data perturbation process. We also gathered feedback and suggestions for improvement. In the second step, the chosen prototype was improved based on the feedback from the pilot study. Finally, the most effective visual design was compared to the textual design in a large-scale online survey with 228 participants to answer three research questions. These questions were: which design helped people better understand local DP (**RQ1**)? Were there differences in knowledge test scores between the text and illustration groups, and why (**RQ2**)? Did the differences in knowledge acquisition affect people's data-sharing decisions, and why (**RQ3**)?

5. Prototype Evaluation

We needed to determine which of the non-textual prototypes for explaining the data perturbation process was most effective for ordinary users, as each prototype has its advantages and disadvantages. To do so, we carried out a small pilot user study with approval from Guangzhou University. The study aimed to identify which prototypes had the potential to effectively explain the data perturbation process and which prototypes had severe limitations, and the feedback from the users would be used to improve the preliminary designs.

5.1. Method

The study was conducted either in-person or remotely via Zoom, with each participant receiving a cash reward of $30 US dollars. The audio and screen recordings were transcribed and analyzed using an open-coding technique [33], with common themes identified across the studies. In the key findings, we include quotes from participants identified by ID numbers following the letter P for "participant" (e.g., P1).

Participants. Five participants were recruited, including four non-experts and 1 DP expert, with ages ranging from 19 to 57 years old and varying levels of education and Internet experience. Two got high school degrees, and another two participants got bachelor's degrees. The participants had all faced privacy breaches and consequences such as harassment and phishing.

Procedure. Before evaluating the prototypes, the participants received training on Local DP to familiarize themselves with the necessary background information. This training was not part of the final design. After completing the training, the participants were shown each of the design prototypes. The prototypes were based on an imaginary scenario where people were asked to share their data for the purpose of income justice research, which was adapted from a previous DP explanation design study [29]. The participants were asked to rate each prototype on two five-point Likert scale questions: (1) Does this prototype effectively explain the data perturbation process? and (2) Does this prototype alleviate my concerns, making me comfortable with sharing my data? Their evaluations were taken into consideration, and they were encouraged to provide suggestions for improvement. The entire study took approximately one and a half hours.

5.2. Key Findings and Discussion

Table 1 showed that all participants either agreed or strongly agreed that the text description was clear. Out of the five participants, four preferred the clarity of the explanatory illustration, with only one favoring the text description. The clarity scores of the data table and the pie chart were not as high, with participants showing a divided opinion. To gain a deeper understanding of these scores, we conducted a qualitative coding of the participants' comments about the prototypes.

Participants	Text		Table		Chart		Illustration	
	Clear	Share	Clear	Share	Clear	Share	Clear	Share
P1	4	4	2	2	4	4	4	4
P2	4	4	5	4	3	3	5	4
P3	4	4	4	4	3	3	5	4
P4	4	4	4	4	3	3	3	3
P5	5	4	2	2	5	4	4	3

Table 1: Ratings of Local DP explanation design clarity and people's willingness to share data after reading the explanations.

Our qualitative analysis revealed a split in participants' evaluations of the clarity of the data table and pie chart. P1 and P5 who disagreed that the data table was clear cited their dislike of math and numbers while favoring the pie chart as more intuitive. Conversely, the other participants thought the pie chart was insufficient in presenting information, while the data table was clear as it displayed all the altered data values.

This outcome indicates the presence of two groups: one that struggles to understand the data table and another that finds the pie chart insufficient. In contrast, our analysis of participant comments on the illustration prototype did not reveal such a divide, leading us to conclude that it would be safe to omit the data table and pie chart from future formal studies and solely compare the explanatory illustration and text description.

The evaluation of the illustration design was positively received by four of the study participants, who appreciated the engaging lottery analogy used to explain the data perturbation process. Additionally, P4 and P5 recommended incorporating further illustrations to convey why privacy risks they were concerned about were not present after using Local DP. This idea was supported by the research of Xiong et al. [12] which found that explaining how Local DP eliminates privacy risks without delving into technical details would aid users in understanding its privacy protection capabilities. In response, the next iteration of the illustration design will emphasize addressing this user requirement by providing more elaborate explanations. The improved design will be discussed in the following section.

Our participants also recommended changes to the hypothetical data-sharing scenario design as they needed clarification on the implications of sharing income data for scientific research. They suggested using a more familiar scenario, such as sharing medical data. The scenario of sharing income data was initially taken from Cummings et al. [29]. However, Cummings et al. also created a medical scenario in their work, and other researchers [9] have used a similar scenario for user testing. Therefore, we modified the hypothetical data-sharing scenario in our formal survey to ask participants to imagine sharing their medical records with a non-profit organization for research purposes.

6. Explanatory Illustration Design

According to the initial user study results, it is suggested that the illustration be enhanced to clarify how Local DP secures user privacy against potential privacy violations by hackers, governments, and organizations. The illustration should demonstrate the data perturbation process and then show how Local DP protects against these adversary types, which are listed in detail in Table 2. Cummings et al. [29] have emphasized that these three types of adversaries are the primary concerns for users, and thus, the new illustration content should effectively address these issues.

No.	Abbreviation	Description
1	Hack	The hacker attacks the database to access your authentic data.
2	Law	The government compulsorily acquires your authentic data.
3	Organization	The organization directly uses or shares your authentic data.

Table 2: The three types of privacy concerns that Local DP can eliminate.

However, explaining all potential privacy attacks to users is not practical as they want a quick overview of Local DP. As users may have varying levels of concern about different adversaries, we designed a customization mechanism. The system prompts users to prioritize their concerns about privacy threats, then only explains how the highest concern won't happen. The other explanations are accessible through a menu for those interested in learning more. This targeted approach may effectively address user concerns.

The design of this new communication flow and how the default illustration is chosen based on the user's response will be explained in the following paragraphs, followed by a description of the added content in the illustration design.

New Communication Flow. We suggest altering the conventional method of seeking user consent to share data by first asking them about their privacy fears. Our new approach involves a communication flow that begins with assessing users' privacy concerns and then providing a tailored illustration to address their top concerns. Users are prompted to rate their worry regarding hacker attacks, government requests, and organizational violations on a scoring scale from 1 to 5, where 1 indicates "not concerned at all," 2 stands for "slightly concerned," 3 represents "somewhat concerned," 4 means "moderately concerned," and 5 signifies "very concerned."

Users can express their level of concern about hacker attacks by assigning a high score to the *Hack* concern. They can show distrust towards the government by assigning a high score to the *Law* concern. We revised the definition of the *Organization* concern in Cummings et al. [29] so users will indicate that they are worried about their data being stolen by organizations if they assign a high score to this concern.

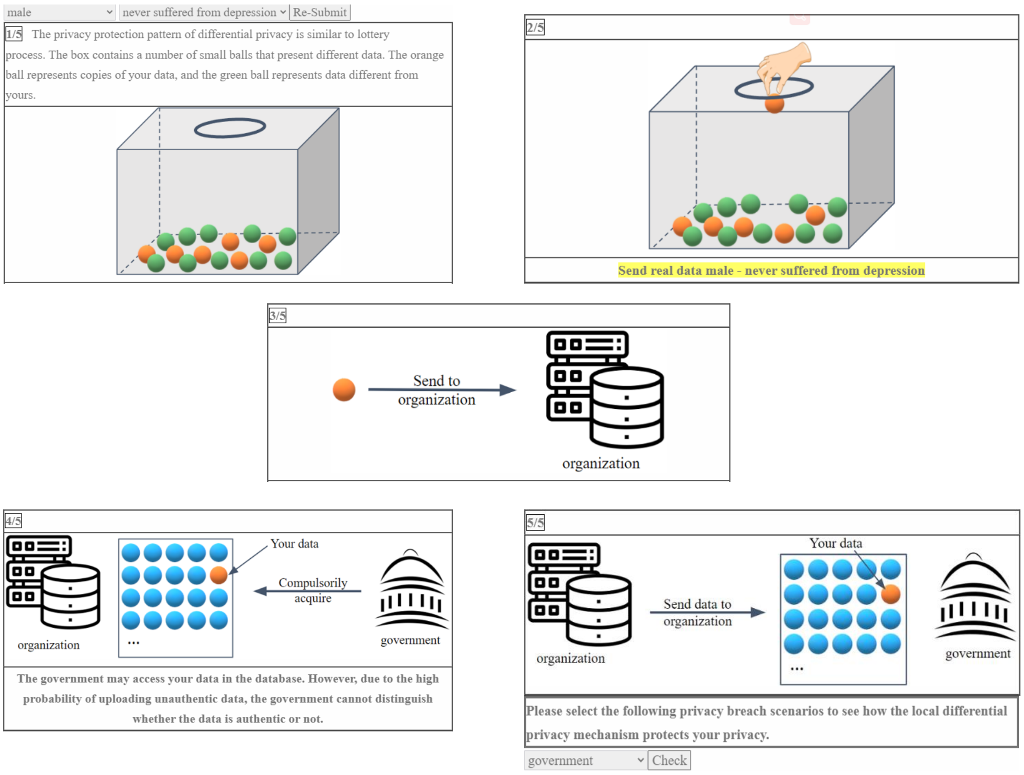
After the user rates their privacy concerns, the illustration presents information to alleviate their top concern. If there is a tie for the highest score, the system randomly selects one of the concerns to address. The improved illustration design will be described in the next section.

Illustration Design Refinement. The illustration shown in Figure 3 depicts only one scenario of privacy concerns - in this case, government access to user data. The illustration has five frames, with the first two frames depicting the data perturbation process, which is the same as in the original design. The remaining three frames are newly added content and show how, in the case of government access, the government cannot access real user information. The illustration demonstrates that while the government has access to user-uploaded data, it cannot determine the authenticity of the data due to the protection provided by the data perturbation process through Local DP.

We have developed a template to showcase how Local DP safeguards against potential privacy breaches from three sources: the government, hackers, and organizations. For each adversary, two explanations are provided: one for when users upload real data, and another for when they upload modified data. Figure 3 is an example of how the template explains how Local DP protects user information from being accessed by the government. The image and description in the illustration must be adjusted to fit each individual privacy threat scenario.

Table 3 provides a list of the icons and text descriptions used in all privacy breach scenarios. The three symbols represent the privacy hazards from the government, hackers, and analysts. The design provides two distinct perspectives, as the algorithm simulates either uploading authentic data or altered data. When uploading authentic data, the design must inform users that the source of their top privacy concern has access to their actual information. However, they can still deny the authenticity of any data that has been compromised. When uploading altered data, the design directly informs users that the adversaries do not have access to their authentic information. In this case, there is no need to have separate explanations for different subjects.

Our final DP explanation design presents two novel features in comparison to the standard DP explanation methods currently in use. Firstly, it elicits users' levels of concern regarding privacy attacks and then assuages their worst fears. Secondly, our design is more in tune with



male ⌄ never suffered from depression ⌄ Re-Submit

1/5 The privacy protection pattern of differential privacy is similar to lottery process. The box contains a number of small balls that present different data. The orange ball represents copies of your data, and the green ball represents data different from yours.

2/5

Send real data male - never suffered from depression

3/5

4/5

The government may access your data in the database. However, due to the high probability of uploading unauthentic data, the government cannot distinguish whether the data is authentic or not.

5/5

Please select the following privacy breach scenarios to see how the local differential privacy mechanism protects your privacy.

government ⌄ Check

Figure 3: The default illustration explaining how Local DP protects user privacy from the government.

people's cognitive tendencies to process information through a blend of text and visual elements. In theory, a personalized multimedia-based explanation should be more appealing and effective than a rigid text-only explanation [32]. Nonetheless, further research through a formal survey is necessary to verify this and determine the reasons behind it.

7. Survey Evaluation

Our illustration design for explaining Local DP features novel elements, such as visualizing the data perturbation process as a lottery and customizing the content for specific privacy concerns. Through a large-scale online survey, we aim to answer the following research questions:

1. Which design, our illustration or the traditional text-based explanation, is better at helping people understand Local DP? (RQ1)
2. If there is a difference in knowledge test performance between the text group and the illustration group, what design elements are responsible? (RQ2)
3. Are the differences in knowledge acquisition reflected in people's data-sharing decisions, and why? (RQ3)

Adversary	Upload Data	Additional Explanation
	Authentic	**The government** may access your data in the database. However, due to the high probability of uploading unauthentic data, the government cannot distinguish whether the data is authentic or not.
	Unauthentic	**The government** can only get access to the unauthentic data you sent.
	Authentic	**The hacker** may get your real data. But since many people also uploaded unauthentic data, hackers cannot tell whether the data is authentic or not.
	Unauthentic	**The hacker** can only get access to the unauthentic data you sent.
	Authentic	**The analyst** is unlikely to analyze your authentic data from the analysis results given by the organization, because the probability of uploading unauthentic data is high.
	Unauthentic	**The analyst** can only get access to the unauthentic data you sent.

Table 3: The icons and text descriptions for explaining three privacy attack scenarios.

7.1. Method

To conduct the online surveys, we utilized the reliable crowdsourcing platform Prolific [1]. Our study was designed as an A/B test, randomly dividing participants into two groups. One group read the text description of Local DP, while the other group read the explanatory illustration. The survey lasted approximately 20 minutes, and participants received £0.15 (US$0.17) per minute for their time. The participants answered questions on our custom website, which recorded both their objective and subjective responses and the time spent on each page. To analyze the collected data, we employed mixed methods [34], which allowed us to incorporate both quantitative and qualitative results to answer our research questions.

Participants. We recruited 228 participants to complete our survey, 117 in the text group and 111 in the illustration group. We required participants to be native English speakers to eliminate the influence of language ability on reading the survey material.

The participants' demographic information is summarized in Table 4. There was an equal representation of males and females in both groups. The age range of participants was from 18 to over 55, with a majority of 24-44 years old (58.1% in the text group and 57.6% in the illustration group). Both groups had a higher education level, with a majority being higher education students or university graduates. The majority of participants in both groups were from the United Kingdom (66.7% in the text group and 57.7% in the illustration group), while the remaining participants were from predominantly English-speaking countries such as the United States, Canada, and South Africa.

Procedure. The survey first gathered the participants' demographic information and had them take a DP knowledge test. After completing the pre-test, they were presented with a medical scenario in which they were asked if they would like to share their medical records with a non-profit organization for research purposes. Participants rated their willingness to share the data on

[1] https://www.prolific.co/

Gender	**Male**	**Female**	**Other**	**N/A**		
Text	47.0%	53.0%	0.0%	0.0%		
Illustration	53.2%	45.9%	0.0%	0.9%		
Age	**18-24**	**25-34**	**35-44**	**45-54**	**55^{+}**	**N/A**
Text	10.3%	37.6%	20.5%	17.9%	13.7%	0.0%
Illustration	13.5%	27.9%	29.7%	9.9%	18.0%	0.9%
Education	**Middle or Lower**	**High School**	**College**	**Bachelor's**	**Graduate**	**N/A**
Text	1.7%	13.7%	20.5%	43.6%	20.5%	0.0%
Illustration	2.7%	23.4%	20.7%	36.0%	17.1%	0.0%
Country	**United Kingdom**	**United States**	**Canada**	**South Africa**	**Other**	**N/A**
Text	66.7%	9.4%	1.7%	5.1%	17.1%	0.0%
Illustration	57.7%	17.1%	6.3%	5.4%	11.7%	1.8%

Table 4: Participants' demographics information.

a 5-point scale and also rated their level of concern for three privacy issues. The survey randomly assigned them either a text description or a customized explanatory illustration to learn about Local DP. After reviewing the material, they rated their opinions again and saw their pre- and post-ratings. They were then asked to explain why their ratings changed or remained the same, as well as which design elements helped them understand Local DP and which elements they liked or disliked. Finally, participants retook the DP knowledge test.

Text Explanation Selection. The text explanations created by Xiong et al. in [9] were found to help individuals comprehend Local DP effectively. In contrast, text explanations from Cumming et al. [29] did not have the same impact. Thus, we adopted the *LDP Imp* description created by Xiong et al. as the reading material for our control group. The *LDP Imp* description primarily outlines the impact Local DP has on the user's data and the reasons why it ensures data privacy, which is also conveyed in our explanatory illustration. The text explanation's full content is in Appendix A.

Knowledge Test Design. The knowledge test used in our study was based on the one created by Xiong et al. [9] to assess participants' understanding of privacy risks and data availability. It evaluated their comprehension of privacy protections against attackers, organization employees, and third parties. The whole test questions content is in Appendix A. Unlike the open-book test used by Xiong et al., participants were not allowed to refer to the description/illustration of Local DP while taking the test.

7.2. Findings

The survey results have been divided into three sections. The first section showcases the participants' test results to determine the effectiveness of visual illustrations or textual explanations in improving understanding of Local DP (**RQ1**). The second section analyzes the changes and reasons for participants' willingness to share data. The final section collects participants' assessments of which design elements were effective in learning about Local DP.

Evidence of Knowledge Acquisition. Table 5 shows the knowledge test performance of the text group and the illustration group before and after learning the Local DP knowledge.

Test	Text			Illustration		
	pre	post	change	pre	post	change
T1 (Privacy Against Hackers)	0.214	0.752	0.538*	0.198	0.603	0.405*
T2 (Privacy Against Organizations)	0.111	0.188	0.077	0.108	0.622	0.514*
T3 (Privacy Against Third-party)	0.214	0.556	0.342*	0.261	0.441	0.180*
Total Score	0.539	1.496	0.957*	0.567	1.666	1.099*

Table 5: Comparison of knowledge test results between the text and illustration groups before and after reading the Local DP knowledge explanation. (* indicates that the score change is statistically significant.)

We performed a one-way ANOVA analysis to assess the impact of learning about Local DP on the test scores. The results showed that both the text description and explanatory illustration groups had statistically significant improvements in their knowledge test performance. The text group's average test score improved from 0.539 out of 5 to 1.496 ($F(1, 115) = 82.73$, $p < .001$) with a large effect size ($\eta^2 > 0.687$). Similarly, the illustration group's average score increased from 0.567 to 1.666 ($F(1, 109) = 78.49$, $p < .001$) with a large effect size ($\eta^2 > 0.708$). However, one-way ANOVA showed no significant difference between the two groups' score improvement ($F(1, 226) = 0.94$, $p = 0.333$).

An analysis of the test results for each question revealed an interesting difference between the two groups. While the text group's accuracy in determining if Local DP protects user data privacy from organizations did not change much after reading the text description, the illustration group's accuracy improved by 51.4% after reading the explanatory illustration. One-way ANOVA showed that the different communication designs had a statistically significant impact on the improvement of test performance ($F(1, 226) = 42.28$, $p < .001$) with a large effect size ($\eta^2 > 0.477$).

Evidence of Willingness Change. Our findings indicated that the text group's average willingness to share data increased, with an average score rising from 2.496 to 2.915 (Kruskal-Wallis $H = 7.14$, $p = 0.008$) and a large effect size ($r = 0.613$). For the illustration group, there was also an improvement in their average willingness to share data, but it was not statistically significant, going from 2.658 to 2.820 (Kruskal-Wallis $H = 0.70$, $p = 0.403$). Our comparison of the reasons given by both groups revealed that the illustration group was 10.1% more concerned about attacks by hackers that Local DP could not protect against.

Table 6 presents the themes that explain why participants were willing or unwilling to share data despite Local DP protection. The findings showed that 20.7% of the illustration group expressed concerns about the possibility of uploading real data, even if they could deny the authenticity of the data due to Local DP. However, 13.5% of the illustration group felt that sharing real data with a low probability of detection was acceptable, as it would be difficult for adversaries to determine the authenticity. In contrast, no responses from the text group mentioned the possibility of uploading real data.

Helpful Design Elements. Table 7 highlights the key design elements that aided participants in comprehending Local DP. Over 60% of the participants in the text group cited that the phrase "randomly modify data" helped their understanding of Local DP, whereas only 23% of the illustration group participants said the same. Nearly 47% of the illustration group found the analogy between Local DP and lottery helpful in comprehending the concept. They explained that they learned about Local DP randomly selecting real or altered data to upload based on specific probabilities and how this impacted their willingness to share data. This type of response was not

Theme	Design	Count	Representative Message
Feel safer under Local DP protection	Illustration	35	Noticed that once my data is protected hackers won't be able to have access to my data.
	Text	46	I'd feel slightly more reassured there is a privacy system in place, but not much since I don't understand it fully.
Distrust any privacy-enhancing technology	Illustration	15	Even with more technology to protect my data I am still not willing to share it.
	Text	15	Regardless of the security methods used to protect my data, there is always a risk that my personal data could be compromised.
Fear of other unknown attacks	Illustration	15	Hackers etc. are always innovative and will no doubt find a way to circumvent this at some point.
	Text	4	Still have vulnerable point of entries for hackers.
Distrust the DP technology for non-technical reasons	Illustration	9	I do not trust this new technology.
	Text	9	I do not trust the 'modified' privacy method.
Worry about denial of data authenticity	Illustration	23	There is still a small chance that my data will be accessed.
Fine with denial of data authenticity	Illustration	15	Being able to deny the data sent was my real data.

Table 6: Coding themes of the reasons to change or not change the willingness to share data under Local DP protection.

Theme	Design	Count	Representative Message
Wording of "randomly modify"	Illustration	26	The randomness of it makes it not perfect and I would prefer seeing something that just avoids your real data going out entirely.
	Text	71	Local DP will randomly modify my data before sending it to the organisation.
Lottery analogy	Illustration	48	Similar to the lottery and the box containing different balls as it shows that the chances of your data being shared is slim.
Visual coding	Illustration	26	The visual and colouring makes it easy to track the whole process.
Not understand	Text	10	I am confused about data modification and I guess this is why I feel I don't truly understand DP because that doesn't make sense.

Table 7: Key design elements that were helpful for participants to learn Local DP.

seen in the text group's feedback. Additionally, 8.5% of the text group participants stated that they still did not fully understand how Local DP protects data.

8. Discussion

In this section, we combine our findings to address **RQ2** and **RQ3**. We will first examine the unique illustration design elements that led to a higher proportion of the illustration group correctly answering the question about organizational data privacy protection (**RQ2**). Then, we will explore the reasons for the lack of increased willingness to share data despite awareness of Local DP protection (**RQ3**). Lastly, we will discuss the limitations of our study.

8.1. Effectiveness of Lottery Analogy Design

Our results showed that the explanatory illustration design was more effective in helping people understand data perturbation's process and result than a text-based explanation. This conclusion was drawn from the fact that individuals who read the illustration revised their incorrect answers about protecting user data privacy from organizations, while those who read the text did not. The difference in performance improvement was statistically significant. Additionally, many participants in the illustration group used their newfound understanding of the data perturbation process to explain their concerns and reasoning behind their willingness or reluctance to share data. However, we did not see this phenomenon with the text group, who offered vague reasons such as "a little reassured about having a privacy system" or "should be able to protect."

Previous work by Xiong et al. [9] demonstrated that it is challenging for people to comprehend the data perturbation process in DP through text alone. Our research shows that using a visual illustration design can make this process easier to understand. Our design's success can be attributed to its use of an analogy between the local DP data perturbation process and a lottery draw. This analogy is based on the approach used in K-12 math education [35] to teach probability. The systematic framework for teaching mathematical concepts with visual language and manipulatives in K-12 education [36] has been effective for students who have difficulty learning math through language [37, 38]. Our design leverages this framework to help explain the concept of probability to ordinary users.

Therefore, we recommend the following approach to tackle similar design challenges: First, identify the mathematical concepts necessary for explaining privacy-enhancing technologies. Second, seek out a visual teaching solution for those concepts from the K-12 math education toolkit. Finally, incorporate the teaching solution into the illustration design.

8.2. Influence of Risk Perception on Willingness to Share

Our survey findings indicated distinct outcomes in the data-sharing decisions of the two groups after learning about Local DP's privacy protection. The willingness to share data remained largely unchanged among participants in the illustration group, while it increased significantly among those in the text group. The illustration group showed a significant improvement in answering all quizzes correctly, while the text group struggled to correctly answer questions related to Local DP uploading data to organizations.

We also observed that the illustration group was more likely to consider the potential privacy risks associated with using a privacy-enhancing technology, as 34.2% of them considered whether it was acceptable for the technology to have a chance of uploading their authentic data. On the other hand, the text description did not prompt participants to think about potential privacy risks. This finding suggests that people are better equipped to assess privacy risks and make informed decisions when they have a clear understanding of how a privacy-enhancing technology operates.

Interestingly, our experiment revealed that providing a clear explanation of the technology is unnecessary to establish trust. Instead, participants were more inclined to trust a straightforward statement like "your privacy is protected even if the database is compromised." While using plain language has advantages in making information more understandable [39], it does not imply that visual illustrations should be overlooked. Our study demonstrated that users might not grow trusting as they gain more knowledge about the technology. Their reluctance mainly stemmed from newfound uncertainties regarding the technology's inability to provide 100% protection or unfounded fears and rejections. Previous work on nudges [40, 41] and emotional motivations [28] has demonstrated promise in addressing such problems and promoting the adoption of privacy-enhancing technologies. Hence, it is worthwhile to explore combining these strategies with visual illustration designs in future work.

8.3. Limitations

Our study had two key limitations. Firstly, the impact of our communication flow customization design was limited. This limitation could be due to either the general nature of the privacy attack examples or the design's inability to address people's new concerns that arose. Secondly, our explanatory illustration design only explains how DP protects numerical data and cannot yet be applied to other data types, such as geolocation data. Although Xiong et al. [12] recently developed an illustration that effectively explains DP protection for geolocation data, there is no straightforward way to combine the two designs. Future work can focus on creating an explanatory illustration design framework that can explain DP protection for various data types.

9. Conclusion and Future Work

With the increasing adoption of differential privacy technologies, it is crucial to understand how users perceive these technologies. Our research examined the impact of different explanation designs on people's understanding of privacy-enhancing technologies and data-sharing decisions. Our findings showed that illustration-based explanations are more effective in helping people understand these technologies than text-based explanations. Additionally, having a clear understanding of the technology does not necessarily lead to an increase in willingness to share data. Instead, it prompts individuals to reflect more on potential privacy risks associated with sharing their information.

Moving forward, there is an opportunity to enhance the design of our visual illustrations to cater to users who may still be confused or concerned about privacy. One potential approach is to combine the techniques of nudges [40, 41] and emotional motivations [28] with visual illustrations to address users' concerns and alleviate any fears or rejections they may have. Another approach could involve utilizing natural language processing technologies (e.g., InstructGPT model [42]) to facilitate a dialogue with users, gathering feedback on the DP explanations, and generating answers to their queries. Additionally, it would be interesting to explore how visual illustrations can be dynamically added to these explanations to improve users' comprehension. Finally, it would be useful to develop a framework that can automatically switch the type of explanation based on the type of data being shared, for example, using the lottery analogy to explain the addition of noise to numerical data and heat maps to explain the addition of noise to geolocation data.

Acknowledgments

The authors gratefully acknowledge support from the China Postdoctoral Science Foundation under grant number 2022M720889. The authors would like to thank the anonymous reviewers for their valuable comments and helpful suggestions.

References

[1] A. Narayanan, V. Shmatikov, Robust de-anonymization of large sparse datasets, in: IEEE Symposium on Security and Privacy, 2008, pp. 111–125.

[2] J. M. Abowd, The u.s. census bureau adopts differential privacy, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2867–2867.

[3] M. Christ, S. Radway, S. M. Bellovin, Differential privacy and swapping: Examining de-identification's impact on minority representation and privacy preservation in the us census, in: IEEE Symposium on Security and Privacy, IEEE Computer Society, 2022, pp. 1564–1564.

[4] Ú. Erlingsson, V. Pihur, A. Korolova, Rappor: Randomized aggregatable privacy-preserving ordinal response, in: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, 2014, pp. 1054–1067.

[5] A. D. P. Team, Learning with privacy at scale, `https://machinelearning.apple.com/research/learning-with-privacy-at-scale` (2017).

[6] P. Story, D. Smullen, Y. Yao, A. Acquisti, L. F. Cranor, N. Sadeh, F. Schaub, Awareness, adoption, and misconceptions of web privacy tools, Proceedings on Privacy Enhancing Technologies 2021 (3) (2021) 308–333.

[7] V. Ha, K. Inkpen, F. Al Shaar, L. Hdeib, An examination of user perception and misconception of internet cookies, in: Extended Abstracts on Human Factors in Computing Systems, 2006, pp. 833–838.

[8] J. Tang, H. Shoemaker, A. Lerner, E. Birrell, Defining privacy: How users interpret technical terms in privacy policies, Proc. Priv. Enhancing Technol. 2021 (3) (2021) 70–94.

[9] A. Xiong, T. Wang, N. Li, S. Jha, Towards effective differential privacy communication for users' data sharing decision and comprehension, in: IEEE Symposium on Security and Privacy, IEEE, 2020, pp. 392–410.

[10] D. Franzen, S. N. von Voigt, P. Sörries, F. Tschorsch, C. Müller-Birn, "am i private and if so, how many?"—using risk communication formats for making differential privacy understandable, arXiv preprint arXiv:2204.04061 (2022).

[11] F. Karegar, S. Fischer-Hübner, Vision: A noisy picture or a picker wheel to spin? exploring suitable metaphors for differentially private data analyses, in: European Symposium on Usable Security, 2021, pp. 29–35.

[12] A. Xiong, C. Wu, T. Wang, R. W. Proctor, J. Blocki, N. Li, S. Jha, Using illustrations to communicate differential privacy trust models: An investigation of users' comprehension, perception, and data sharing decision, arXiv preprint arXiv:2202.10014 (2022).

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

[14] E. Kang, E. Jackson, W. Schulte, An approach for effective design space exploration, in: Monterey Workshop, Springer, 2010, pp. 33–54.

[15] C. Dwork, A. Roth, et al., The algorithmic foundations of differential privacy, Foundations and Trends in Theoretical Computer Science 9 (3-4) (2014) 211–407.

[16] T. Wang, N. Li, S. Jha, Locally differentially private frequent itemset mining, in: S&P, IEEE Computer Society, 2018, pp. 127–143.

[17] B. Ding, J. Kulkarni, S. Yekhanin, Collecting telemetry data privately, Advances in Neural Information Processing Systems 30 (2017).

[18] Q. Ye, H. Hu, Local differential privacy: Tools, challenges, and opportunities, in: International Conference on Web Information Systems Engineering, Springer, 2020, pp. 13–23.

[19] E. Gabber, P. B. Gibbons, Y. Matias, A. Mayer, How to make personalized web browsing simple, secure, and anonymous, in: International Conference on Financial Cryptography, Springer, 1997, pp. 17–31.

[20] A. Abdou, A. Matrawy, P. C. Van Oorschot, Cpv: Delay-based location verification for the internet, IEEE Transactions on Dependable and Secure Computing 14 (2) (2015) 130–144.

[21] S. Agrawal, M. Chase, Fame: Fast attribute-based message encryption, in: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 665–682.

[22] J. Gluck, F. Schaub, A. Friedman, H. Habib, N. Sadeh, L. F. Cranor, Y. Agarwal, How short is too short? implications of length and framing on the effectiveness of privacy notices, in: Proceedings of the Twelfth Symposium on Usable Privacy and Security, 2016, pp. 321–340.

[23] N. Ebert, K. A. Ackermann, P. Heinrich, Does context in privacy communication really matter?—a survey on consumer concerns and preferences, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–11.

[24] E. M. Redmiles, A. R. Malone, M. L. Mazurek, I think they're trying to tell me something: Advice sources and selection for digital security, in: IEEE Symposium on Security and Privacy, IEEE, 2016, pp. 272–288.

[25] F. Schaub, R. Balebako, A. L. Durity, L. F. Cranor, A design space for effective privacy notices, in: Proceedings of the Eleventh Symposium on Usable Privacy and Security, 2015, pp. 1–17.

[26] A. Acquisti, L. Brandimarte, G. Loewenstein, Privacy and human behavior in the age of information, Science 347 (6221) (2015) 509–514.

[27] W. Bai, M. Namara, Y. Qian, P. G. Kelley, M. L. Mazurek, D. Kim, An inconvenient trust: User attitudes toward security and usability tradeoffs for {Key-Directory} encryption systems, in: Proceedings of the Twelfth Symposium on Usable Privacy and Security, 2016, pp. 113–130.

[28] M. Namara, D. Wilkinson, K. Caine, B. P. Knijnenburg, Emotional and practical considerations towards the adoption and abandonment of vpns as a privacy-enhancing technology, Proceedings on Privacy Enhancing Technologies (2020).

[29] R. Cummings, G. Kaptchuk, E. M. Redmiles, "i need a better description": An investigation into user expectations for differential privacy, in: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, 2021, pp. 3037–3052.

[30] F. Karegar, A. S. Alaqra, S. Fischer-Hübner, Exploring {User-Suitable} metaphors for differentially private data analyses, in: Proceedings of the Eighteenth Symposium on Usable Privacy and Security, 2022, pp. 175–193.

[31] M. A. Smart, D. Sood, K. Vaccaro, Understanding risks of privacy theater with differential privacy, Proceedings of the ACM on Human-Computer Interaction 6 (CSCW2) (2022) 1–24.

[32] M. Sadoski, A. Paivio, Imagery and Text: a Dual Coding Theory of Reading and Writing, Routledge, 2013.

[33] J. Saldaña, The Coding Manual for Qualitative Researchers, Sage, 2021.

[34] J. W. Creswell, Mixed-method research: Introduction and application, in: Handbook of educational policy, Elsevier, 1999, pp. 455–472.

[35] J. A. HANLEY, Lotteries and probability: Three case reports, Teaching Statistics 6 (3) (1984) 88–91.

[36] M. Boggan, S. Harper, A. Whitmire, Using manipulatives to teach elementary mathematics, Journal of Instructional Pedagogies 3 (2010).

[37] A. W. Hunt, K. L. Nipper, L. E. Nash, Virtual vs. concrete manipulatives in mathematics teacher education: Is one type more effective than the other?, Current Issues in Middle-Level Education 16 (2) (2011) 1–6.

[38] R. Satsangi, E. C. Bouck, T. Taber-Doughty, L. Bofferding, C. A. Roberts, Comparing the effectiveness of virtual and concrete manipulatives to teach algebra to secondary students with learning disabilities, Learning Disability Quarterly 39 (4) (2016) 240–253.

[39] G. R. Milne, M. J. Culnan, H. Greene, A longitudinal assessment of online privacy notice readability, Journal of Public Policy & Marketing 25 (2) (2006) 238–249.

[40] P. Story, D. Smullen, R. Chen, A. Acquisti, L. F. Cranor, N. Sadeh, F. Schaub, et al., Increasing adoption of tor browser using informational and planning nudges, Proceedings on Privacy Enhancing Technologies 2 (2022) 152–183.

[41] P. Story, D. Smullen, A. Acquisti, L. F. Cranor, N. Sadeh, F. Schaub, From intent to action: Nudging users towards secure mobile payments, in: Proceedings of the 16th Symposium on Usable Privacy and Security (SOUPS 2020), 2020, pp. 379–415.

[42] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al., Training language models to follow instructions with human feedback, arXiv preprint arXiv:2203.02155 (2022).

Appendix A. Survey Instruments

Appendix A.1. Knowledge Test

Before reading Local DP materials, the participant needs to take the knowledge test about Local DP. The questions in the test are as follows.

T1. Suppose that you have your health information collected by the organization, but your health information was protected by a privacy protection technique called differential privacy. If an attacker gets access to the database of the organization, will the attacker be able to see your real health information?

T2. The organization decided to deploy differential privacy to improve the privacy protection of its users. With the deployment of differential privacy, will the initial data received by the organization contain any noise?

T3. The organization decided to deploy differential privacy to improve the privacy protection of its users. For the third party companies with which the health organization shared data, will they be able to see the real answer that you submitted?

The options are ”Yes”, ”No”, ”Unsure”, or ”Prefer not to answer”.

Appendix A.2. Privacy Concerns

Imagine that during your next doctor’s visit, your primary care doctor informs you that they are part of a non-profit organization trying to push the boundaries of medical research. This non-profit is asking patients around the country to share their medical records, which will be used to help medical research on improving treatment options and patient care. Your doctor, with your permission, can facilitate the non-profit getting the information they need. The organization mainly uses your data for data aggregation, which is a process where raw data is gathered and expressed in a summary form for statistical analysis.

To further protect your data privacy, the organization would like to provide a customized privacy protection technology based on your privacy concerns. **The following three cases of data breach may happen either because the organization does not follow a valid standard data operating procedure or it is beyond the control of the organization.** Please rate the level of concern about these data breach cases. The level of concern ranges from 1 to 5 (1: not concerned at all, 2: slightly concerned, 3: somewhat concerned, 4: moderately concerned, and 5: extremely concerned).

C1. Hackers attack the database to access your real data.

C2. The government compulsorily acquires your real data.

C3. The organization directly uses or shares your real data.

Appendix A.3. Willingness to Share Data

Imagine that during your next doctor's visit, your primary care doctor informs you that they are part of a non-profit organization trying to push the boundaries of medical research. This non-profit is asking patients around the country to share their medical records, which will be used to help medical research on improving treatment options and patient care. Your doctor, with your permission, can facilitate the non-profit getting the information they need. The organization mainly uses your data for data aggregation, which is a process where raw data is gathered and expressed in a summary form for statistical analysis.

We would like you to rate your willingness to share your data. You can rate your willingness ranging from 1 to 5 (1: not willing at all, 2: slightly willing, 3: somewhat willing, 4: moderately willing, and 5: extremely willing).

Appendix A.4. Open-ended Questions

1. Your willingness to share data was rated XX before knowing that your data would be protected by the differential privacy technology then you re-rated your willingness level to be XX. What are the reasons that changed (or didn't change) your mind?
2. Your privacy concerns rating before was XX for hacker, XX for government and XX for organization. After learning about how your data would be protected by the differential privacy technology, your privacy concerns changed to XX for hacker XX for government and XX for organization. What are the reasons that changed (or didn't change) your mind? You may elaborate the reasons for each privacy concern item.

Appendix A.5. Helpful Design Elements

1. What are the 2 (or more) most helpful design elements that helped you understand the privacy protection technique and why?
2. What are the 2 (or more) most helpful design elements that alleviated your concerns about privacy breaches and why?
3. What are the design elements that you dislike and why?

Appendix A.6. Text Description of Local DP

To respect your personal information privacy and ensure better user experience, the data shared with the organization will be processed via the local differential privacy (LDP) technique. That is, the app will randomly modify your data on your local device before sending it to the organization. Since the organization stores only the modified version of your personal information, your privacy is protected even if the app server's database is compromised.