Replication: The Effect of Differential Privacy Communication on German Users' Comprehension and Data Sharing Attitudes

Patrick Kühtreiber
University of Göttingen

Viktoriya Pak
University of Göttingen

Delphine Reinhardt
University of Göttingen

Abstract

Differential privacy (DP) has become a standard for privacy-preserving data collection. However, there is little understanding of users' comprehension of this privacy technique, which could increase users' willingness to share personal data. Xiong et al.'s 2020 study tackles this problem by investigating the effect of differential privacy communication to laypeople, with an average of 466 participants per study primarily from USA and India. Since privacy decisions have been shown to depend on participants' culture in multiple past studies, we have replicated this study with German participants to compare the results with the original study and to gain further insights about differential privacy communication in a different cultural context. After having translated the original questionnaire into German, we conducted two studies with an average of 728 participants. While we could confirm that participants did not fully understand differential privacy and that a new method to communicate the effects of differential privacy is needed, participants in our study were more willing to share data than the participants from USA and India. This finding is surprising, as Germans have been shown to be more worried about their privacy than other cultures.

1 Introduction

The benefits of using personal data for machine learning are most prominent in healthcare applications [7, 9, 34]. Among ethical considerations, there are also privacy concerns [19] due to the fact that most applications require a lot of data to train the models. As data breaches appear to be ubiquitous [16], many people are reluctant to share their private information [20, 37]. One of the key points of the European *General Data Protection Regulation (GPDR)* is that data subjects (i.e. individuals whose personal data are collected) must consent to the data processing [8]. It is therefore of major interest to investigate steps that allow data subjects to consent easily if their personal data are protected.

Among methods to protect privacy in such a context, DP is a promising solution to this problem. DP was introduced by Cynthia Dwork in 2006 [15] and it has since influenced many different areas of research, such as federated learning [39], data mining [18], and location-based services [2]. In principle, DP sets a statistical bound on the privacy risk of individuals who share their data. It does that by introducing carefully calibrated noise into the data, which masks the contribution of each individual data subject to a certain degree but still maintains the usability of the collected data, albeit sacrificing accuracy. The underlying promise of DP is that nothing about an individual in a dataset should be learnable that could not have been learned if the individual was not in the dataset [14].

Furthermore, the original model of DP has been extended to a more privacy-preserving model, referred to as *Local Differential Privacy (LDP)* [25]. In this model, data perturbation happens on the user's device (instead of a central entity with the original DP). As a result, the raw data do not leave the device, thus providing more privacy. However, since the noise is locally applied, the utility cannot be optimized by taking into account other users' data. In the following, we will refer to both models as (L)DP, if no distinction is necessary. Already used in practice by Google [17], Apple [43], and Microsoft [13], amongst others, (L)DP promises to be a solution to many problems faced in collecting data. However, it is not very well known outside of the technical and research communities, especially not to laypeople.

Laypeople may be reluctant to share information, though, because they fear for their privacy [20, 37]. Helping them to understand how their privacy is protected may help them to make informed decisions about sharing their data. However, only few publications [5, 11, 48] tackle this challenge. Among

USENIX Symposium on Usable Privacy and Security (SOUPS) 2022.
August 7–9, 2022, Boston, MA, United States.

them, the studies conducted by Xiong et al., presented in [48], investigate the effects of DP communication on the users' comprehension and their willingness to share personal data. While the authors tested many different and creative ways to explain DP, the studies have been conducted with young and educated participants who were recruited via Amazon MTurk, which has been shown to include mostly users from USA and India [12]. Nevertheless, it has been shown in [28] that cultural/age differences may impact the results. Also, replication studies have been shown to enhance the understanding of a certain subject [35] and clarify potentially false assumptions drawn from previous research [21].

To investigate these potential differences and validate the results in a different context, our contributions are as follows: We have replicated the original studies with participants from a different cultural and demographic background, directly compared self-reported and actual understanding of differential privacy, and evaluated whether personal health app usage impacts the willingness to share personal data. Tab. 1 illustrates the differences in our study compared to [48].

	Original study	Our study
Country	USA/India	Germany
Age	80% < 45y	Representative of the
Education	60% bachelor's degree	German population
#Experiments	4	2
Avg. #participants	∼ 466	∼ 728

Table 1: Differences from the original study.

As a result, we conducted two studies to (1) test the willingness to share low- and high-sensitivity data with a health app and its respective server depending on different text-based descriptions of (L)DP and (2) to evaluate the trust in and comprehension of these techniques. Similar to the original study, we only evaluated one description of DP or LDP respectively in the first study, while we evaluated eleven different descriptions in the second study.
The obtained key results are as follows.

1. We can confirm that the participants' attitudes are similar in both groups DP and LDP. Unlike originally expected, participants in the LDP group did not share more data with the app server than participants in the DP group, even though it is safer to do so under LDP.

2. Participants who were presented with a description that emphasizes the implications of the LDP, i.e., that privacy is protected even if the company's data base is breached, participants, indicated the largest willingness to share personal data, as in the original study.

3. The communication of (L)DP has a greater effect in our study compared to the original study. Participants whose privacy was protected via (L)DP wanted to share significantly more personal data than those in the control group where no privacy protection was communicated.

4. Overall, we experience a smaller variance in the results of the different descriptions of (L)DP as compared to [48]. Moreover, we find that there exists a correlation between participants who used health apps in their private life and their willingness to share data and their trust in the app, the server, and (L)DP.

5. As in the original study, our participants' comprehension of (L)DP was not very high; thus more effective communication methods are needed.

The remainder of this paper is structured as follows: We summarize the theoretical and technical background of (L)DP in Sec. 2 and present related work, including the original study in Sec. 3. We present our methodology in Sec. 4 and our experiments in Sec. 5 and Sec. 6. We discuss our results in Sec. 7 and make conclusions in Sec. 8.

2 Backgrounds on differential privacy

The primary assumption of (L)DP is that users send their personal data to a data curator, e.g., a company's data base. A data analyst can then analyze the data. (L)DP guarantees the users' privacy to a certain extent while keeping the data usable for the data analyst. However, one key element of (L)DP is that data analysts never see raw or perturbed data but only receive answers to queries of the noisy dataset. The thread model only considers attacks on the data curator, but not on the user's device itself.

2.1 DP vs. LDP

The global or centralized model is the original form of DP. In this model, users' raw data is sent to a trusted curator. Only then is the perturbation of the data carried out (see Fig. 1). Perturbation of the data in the global model takes place via noise that is added, e.g., from the Laplacian or the Gaussian distribution [14].

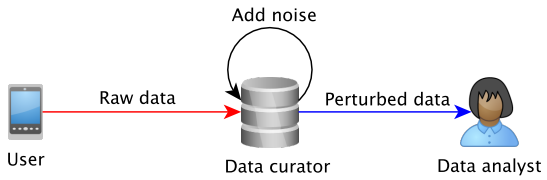


Figure 1: Differential Privacy

In the local model LDP, the data is perturbed on the device before it is being sent to the data curator. The privacy advantage in this case is that raw unperturbed data never leave the device (see Fig. 2). However, the accuracy of the data is lower, as the perturbation of data does not occur on data aggregates but on the data of single users. Perturbation is usually achieved via *randomized response (RR)* [46]. RR can be

best explained by imagining a scenario in which a participant has to answer a (sensitive) "Yes" or "No" question. However, before they answer, they first flip a coin. If it lands "heads" they answer truthfully, and if it lands "tails", the participant flips the coin again and answers "Yes", if it lands "heads" and "No", if it lands "tails". This way, there is a 25% chance of the answer being incorrect, thus providing plausible deniability to the participants and encouraging them to answer the questions truthfully (if the coin lands "heads"). Other than the previously described basic version of RR, one can also imagine biased coins or spinners representing the weights added to certain outcomes. This way, a data collector can emphasize privacy (by adding more weight to the randomized outcome) or accuracy (by increasing the weight of the true answer). Bullek et al. [5] conducted a study on biased spinners (see Sec. 3). The utility of LDP data is reduced by $O(\sqrt{N})$ compared to DP data, where N is the number of users [6]. In both cases, the data analyst receives only perturbed data.

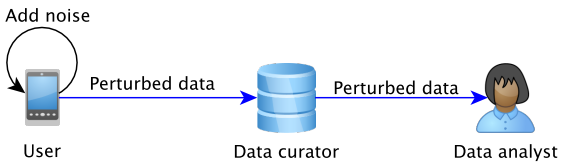


Figure 2: Local Differential Privacy

The most relevant fact for the data subject is that the privacy guarantee of LDP is higher than that of DP when considering only attacks on a company's application server, for example, and not directly on the user's device. This is because raw data never leaves the device and there is no centralized instance (like the trusted curator) you have to trust with your data.

3 Related work

In this section we first describe the impact of culture and privacy law on privacy attitudes, followed by relevant papers regarding usable (L)DP and the original study, which we replicate in this paper.

3.1 Cultural differences

There have been many studies investigating inter-cultural differences in regard to privacy. For example, studies have found that a country's culture impacts its privacy regulations [31] and its citizens' privacy regulation preferences [4]. Other studies focus on the difference in privacy attitudes in the context of digital government [10] or e-commerce adaption [32].

According to Hofstede's cultural comparisons [1], Germany is one of the countries that avoid uncertainty, especially compared to the US or India. Also, Germany can be seen as an individualistic country, although the US scores higher in this dimension. It has been shown that both dimensions, uncertainty avoidance and individualism, impact the risk-taking behavior of the country's citizens regarding personal data. Citizens of collectivist countries as well as those from countries with a high uncertainty avoidance place more emphasis on privacy [44]. For example, Germans are more conservative when sharing data on online social networks [28] and trust providers of activity trackers less [22] when compared to US-Americans. Further studies have found that the medical history is seen more sensitive in the US, while income level is a little more sensitive for German participants [30, 40]. Moreover, Germans tend to feel less in control about the processing of their personal data [33]. However, none of the existing studies comparing cultures has focused on (L)DP.

3.2 Differences in privacy law

Privacy and data protection rights are perceived differently in the US and the EU. Whereas in the EU data privacy is seen as an individual right, in the US the right to privacy is not directly granted by the constitution and is context-dependent [3]. The different European privacy laws were harmonized in 2018 within the GDPR, which grants extensive data privacy rights to all EU citizens and heavily fines companies that do not comply. Since 2020 the *California Consumer Privacy Act (CCPA)* has granted people in California more extensive privacy rights as well, but its scope regarding individual privacy rights is still limited compared to the GPDR [3].

Early research shows that the existence of privacy regulations such as the GDPR can reduce data subjects' privacy concerns [47]. However, more recent studies show that increased knowledge about these regulations does not yield the same result [36]. We can therefore assume that our sample — German citizens who are protected by the GDPR — might be more concerned about their privacy than the sample of the original study, which consisted mainly of US citizens.

3.3 Usable differential privacy

The first study concerning usable (L)DP was presented by Bullek et al. in 2017 [5]. This study focused on the participants' understanding of RR, which is used in LDP (see Sec. 2.1). The participants were presented with three spinners that all had a different bias towards the true answer (40%, 60%, and 80%). That means, that a participant has a 40/60/80 percent chance of having to answer truthfully and a 60/40/20 percent chance that the answer is randomized (equally between "Yes" and "No"). To make this concept more accessible to laypeople, the authors designed (animated) spinners that would land on a certain field that would tell the participant how to answer the sensitive questions asked in the questionnaire. The study provided some seemingly contradictory results. As expected, participants preferred the spinner that provided the most amount of privacy; however, the sec-

Low-sensitivity	High-sensitivity
reason to use the health app	date of birth
exercise experience	family medical record
exercise time	substance use
gender	surgery record
height	diagnostic record
weight	income level
vegetarianism	current medication

Table 2: Low- and high-sensitivity questions

ond most chosen one was the spinner that provided the least amount of privacy. Participants justified their choice of the least anonymous spinner by stating that it would otherwise feel like lying [5].
Another recent study in this area was conducted by Cummings et al. [11] and published while we were conducting the replication study presented in this paper. The goal was not only to evaluate the impact of DP communication on the willingness to share data but also how different DP explanations affect the users' expectations of DP. The authors synthesized 76 different DP descriptions into 6 short descriptions that all convey a certain theme (technique, trust, risk, etc.). The participants were presented with one of those descriptions and one of two relevant scenarios (disclosure of salary or medical records with DP). Being exposed to DP descriptions did raise the participants' privacy expectation; however, it did not increase their willingness to share data [11].

3.4 The original study by Xiong et al.

In the original study, Xiong et al. investigated effective communication of (L)DP and its impact on data-sharing decisions [48]. To this end, four experiments were conducted. These experiments consisted of online surveys, and their participants were recruited via Amazon MTurk.

3.4.1 Experiments 1 and 2

The participants in experiments 1 and 2 were presented with a scenario, in which they had to imagine downloading a health app that asks seven low-sensitivity and seven high-sensitivity questions (see Tab. 2). The participants did not actually have to provide these answers to the researchers, but instead had to answer how they would like their answers to be processed: 1.) not at all (opt out), 2.) only used by the app locally on the device (local only), or 3.) used by the app as well as the application server (both). To test the effect of (L)DP communication, participants in experiment 1 were randomly assigned to one of the four categories: DP, LDP, gain, and control. Participants in the DP and LDP groups were presented with a description of DP and LDP, respectively. The introduction to the questionnaire in the gain group was framed in a positive way (gain framing [45]), and the control group was presented with a neutral introduction. No descriptions of (L)DP or any other data protection technique were presented to neither the gain nor the control group.
After confirming the effects of the gain framing, the authors repeated the experiment with different descriptions of (L)DP in experiment 2 (which was split into two separate surveys). The findings of experiments 1 and 2 suggest that (L)DP communication has little effect overall; however, there was an increase in sharing high-sensitivity questions. Contrary to the actual privacy guarantee, DP ranked higher than LDP which suggests that LDP was not well understood. In experiment 2, the authors tested further descriptions of (L)DP, which only confirmed the findings of experiment 1. Participants found DP easier to understand. However, when the description of LDP emphasized the data perturbation process, participants were more willing to share data with the app locally.

3.4.2 Experiment 3

In experiment 3, the authors examined the understanding of eleven different descriptions of (L)DP and also investigated the reasoning behind the participants' sharing decisions via open questions. The findings indicate that terms like "random" and "noise" are hard to understand. Participants were willing to share more information if the implication of the presented technique was also mentioned. As reasons to share data, participants noted that they had no privacy concerns, wanted to improve the utility of the app, or that they simply trust the presented (L)DP technique. Participants who did not want to share their data wrote that they distrusted the techniques, the requested data was too sensitive, data breaches could still occur, or that they distrusted the application or tech companies in general.

3.4.3 Experiment 4

Finally, experiment 4 investigated whether the self-reported understanding rates were accurate by asking five comprehension questions. Findings revealed that participants did not fully understand the implication of (L)DP in most cases. Only one description that emphasized the implications of LDP generated a high correct response rate for the implication-question. As a result, we used the existing studies as a basis for our work. However, our participants have a different cultural and demographic background, we have changed the number of studies, and we analyze whether personal health app usage affects the outcome. This way we increase the generalizability of the findings in [48] and are also able to compare self-reported and actual understanding of (L)DP.

4 Methodology

We started the replication study by translating the English questionnaire in [48] into German. Two of the authors translated the questions (and answers) independently of each other

and then discussed and resolved the differences. For example, some expressions like "health app" have a literal German translation that we only used when we agreed that it is more common than the English term.

Following the translation, the questionnaires were created in LimeSurvey and the participants were recruited via an ISO 29362-certified panel provider. All participants were financially rewarded if they completed the study. We set age and gender quotas to ensure a representative sample of the German population [42]. Our university does not have an official IRB process, but we adhered to ethical standards set by the German Research Foundation. All questionnaires have been approved by the university's data protection officer.

4.1 Differences to the original study

We replicated the study conducted by Xiong et al. in order to compare the responses of different populations. However, we also made the following changes: (1) demographics as detailed in Sec. 4.1.1, (2) a reduced number of studies as detailed in Sec. 4.1.2, (3) the introduction of an additional question, and (4) correlation of the participants' self-reported and actual understanding of (L)DP. Note that we also performed additional statistical tests in Sec. 5 and Sec. 6.

4.1.1 Demographics

The participants in [48] were recruited via Amazon MTurk. Xiong et al. did not ask where their participants were from; however, we know from other research that the majority of MTurk users are from the USA (75%) and India (16%) [12]. In comparison we focused on German participants only. Another major difference is the age and education of the participants. The original study is heavily skewed towards college educated (60% bachelors degree) younger people (80% younger than 45). Instead, we used quotas in our questionnaire to recruit participants that are representative of the German population, as illustrated in Tab. 3. We also asked the participants an additional question to see whether they are currently using a health app.

	Categories	Exp 1 (518)	Exp 2 (937)
Gender	Male	50.95%	53.1%
	Female	46.8%	45.8%
	Other	0.15%	0%
	No answer	2.1%	1.1%
Age	18-24	15.4%	15.3%
	25-34	30.1%	21.8%
	35-44	27.1%	24.3%
	45-54	15.6%	28.2%
	55 or older	10.1%	9.8%
	No answer	1.7%	0.6%
Education	No high school	23.2%	33.7%
	High school	39%	34.3%
	Bachelor	14.7%	12.1%
	Master	18.1%	16.3%
	PhD	1.7%	2.2%
	No answer	3.3%	1.4%
IT background	Yes	15.4%	16.1%
	No	79.5%	82.1%
	No answer	5.1%	1.8%
Health app	Yes	47.9%	54.3%
	No	49.6%	45%
	No answer	2.5%	0.7%

Table 3: Demographics

4.1.2 Study design

As depicted in Fig. 3, Xiong et al. conducted four experiments (excluding pilot studies and the division of the second experiment into two sub-experiments). As detailed in Sec. 3.4, experiments 1 and 2 used the same questionnaire with different descriptions of (L)DP and tested these descriptions on four different groups. We used the best descriptions found by the authors and used them in our experiment A, thereby compressing experiments 1 and 2 of the original study. Another difference is that Xiong et al. had already confirmed the effect of framing the questions in a positive way (gain framing [45]), which is why we used three different groups: DP, LDP, and control, with control including the description of the gain framing of the original study. We examined the different descriptions of (L)DP in our experiment B, in which we not only asked for the participants' self-reported understanding of the presented descriptions but also checked their comprehension with knowledge questions. Both of these are taken from experiments 3 and 4 of the original study and were originally separated. As a result, we can directly correlate self-reported understanding and actual comprehension of (L)DP. See Fig. 3 for our study design compared to the original study.

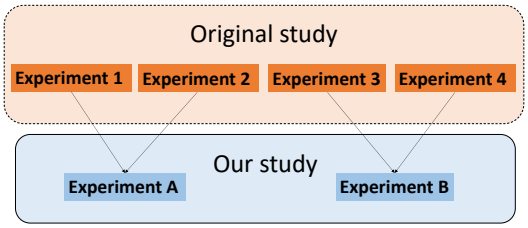


Figure 3: Study design

5 Experiment A

In this section we present the study design and the results of experiment A, before we discuss and compare the results with the ones from the original paper. The complete questionnaire for experiment A can be seen in Appendix A.

Group	Summary of description
DP	DP protects personal data via random noise added to aggregated data. Used by Harvard, US Census Bureau,...
LDP	LDP protects personal data via random noise added to every answer provided by the user. Used by Apple, Google.

Table 4: Summary of (L)DP descriptions for experiment A

5.1 Study design

With changes detailed in Sec. 4.1, we conducted our first experiment. After the introduction in which participants were informed about the goal of the research, they first answered demographic questions in order to ensure the targeted quotas in terms of gender and age. The participants were then divided into three groups: DP, LDP, and control. Next, the participants were presented with the scenario in which they had to imagine themselves. In the described scenario, they had just downloaded a health app, that needed some partially sensitive information from them. All three groups were presented with the same introduction, i.e., the gain framing [48]. The DP and LDP groups were then presented with their respective descriptions of differential privacy. In Tab. 4 you can see the high-level summary of the descriptions, and the complete descriptions for experiment A are available in Appendix C. Afterwards the participants had to answer a comprehension question. If the question was not answered correctly, the description was shown again.

In the next step, we asked our participants the same questions as in [48], i.e., the participants' willingness to share potential answers to questions of the downloaded health app with the app or the app server. As presented in Tab. 2, the questions are separated into seven low- and seven high-sensitivity questions in order to evaluate the difference in the participants attitudes towards their willingness to share low- and high-sensitivity information with the health app or the app server. The participants did not answer those questions but only chose how they would like their potential answers to be processed. They could choose not to share anything (opt out), to trust their data only to the app locally, or to share them with the app and the app server. The participants could also choose not to answer. In that case they were counted in the opt out category, as in [48].

5.2 Participants

Through our certified panel provider, a total of 990 participants were recruited This means that our three groups, DP, LDP, and control, comprised 330 participants each. We applied the same exclusion criteria to our participants as in [48]: (1) Completion time less than 120 seconds (57 DP, 46 LDP, 99 Control) and (2) wrong answers to the comprehension question (124 DP and 135 LDP). Consequently, 149 participants remained in the DP group, 138 participants in the LDP group, and 231 in the control group. The median completion time (before exclusions) was 199.17 seconds in the DP group, 201.55 seconds in the LDP group, and 148.27 seconds in the control group.

5.3 Results

In the following, we report all significant results of our experiment that can be directly compared to the original study and additional tests. Note that the complete results are available in Appendix E.

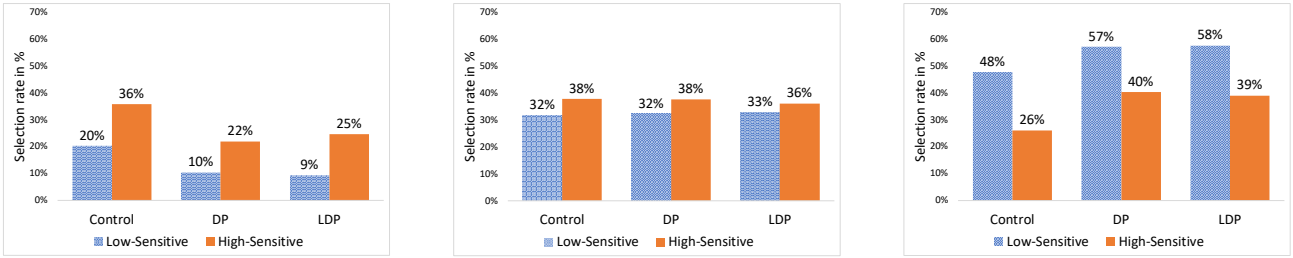
5.3.1 Replication tests

Similar to the original study, we first performed χ^2 tests on the three relevant decisions (opt out, local only, or both) for each question type (low-sensitivity, high-sensitivity) collapsed across participants.

Question sensitivity across all participants: We observed significant differences between low and high question sensitivity across participants of all groups. Participants chose to *opt out* more often when they were asked a high-sensitivity question (29%) than when they were asked a low-sensitivity question (15%), $\chi^2_{(1)} = 217.63$, $p < .001$. We observed a similar attitude in the decision *local only*, with 37% for the high-sensitivity questions and 32% for the low-sensitivity questions, $\chi^2_{(1)} = 21.48$, $p < .001$. Consequently, the decision to share with *both* was higher for the low-sensitivity questions (53%) than for the high-sensitivity questions (34%), $\chi^2_{(1)} = 280.5$, $p < .001$. This means that our participants chose to share low-sensitivity questions more often (locally and with the app server) than high-sensitivity questions.

Question sensitivity among groups: Differences among groups (control vs. DP vs. LDP) could only be observed for the decisions *opt out* and *both*. The decision rate to *opt out* was significantly larger in the control group (28%) than in the DP (16%) and LDP (17%) groups, $\chi^2_{(2)} = 139.21, p < .001$. In contrast, the decision rate to share with *both* was higher for the DP (49%) and LDP (48%) groups compared to the control group (37%) $\chi^2_{(2)} = 97.95, p < .001$). Post-hoc independent sample t-tests reveal that only the differences Control vs. DP and Control vs. LDP are significant ($p < .001$ for all four tests, Bonferroni corrected). These results indicate that (L)DP communication has the effect of increased willingness to share data. However, almost no difference between DP and LDP could be observed.

Two-way interaction of sensitivity $\times$ condition: Finally, we replicated the 2×3 cross-table *question sensitivity (low, high)* $\times$ *group (control, DP, LDP)* to perform χ^2 tests on this matrix (see Fig. 4) . Again, only the decisions to *opt out* ($\chi^2_{(2)} = 8.08, p = .018$) and to share with *both* ($\chi^2_{(2)} = 9.94, p = .007$) are significant. Pairwise tests reveal that only Control vs. DP and Control vs. LDP show significant differences. For the low-sensitivity questions, only the decision to *opt out* is statistically significant for the pairs Control vs.



(a) Opt out (b) Local only (c) Both

(d) Opt out (original study [48]) (e) Local only (original study [48]) (f) Both (original study [48])

Figure 4: Selection rates across the three different conditions and both question sensitivities for experiment A (a-c) in comparison with the selection rates of the original study's experiment 1 (d-f). As we used the gain framing in our control group, we compare the rates of our control group with the gain condition of the original study.

DP ($p = .002$) and Control vs. LDP ($p = .001$, Bonferroni corrected). For the high-sensitivity questions, there are significant differences in the two decision rates *opt out* (Control vs. DP, $p < .001$, Control vs. LDP, $p = .001$, Bonferroni corrected) and *both* (Control vs. DP, $p < .001$, Control vs. LDP, $p = .004$, Bonferroni corrected). This further confirms that (L)DP communication had a positive effect on data sharing and that participants in both (L)DP groups show little difference in their willingness to share.

5.3.2 Additional tests

In addition to the tests carried out by Xiong et al., we also tested if we could observe differences in the participants' trust in the app, the app server, and (L)DP and their willingness to share based on their demographics. To this end, we performed Kruskal-Wallis tests and we only report the significant results. The complete statistics are available in Tab. 9 and 10 in Appendix E.

Trust: Across all three groups, only participants who were already using a health app show a significant difference in the trust in the app (H(1) = 40.028, $p < .001$), the server (H(1) = 27.362, $p < .001$), and (L)DP (H(1)=26.31, $p < .001$). For example, 33% of participants who already used a health app agreed at least somewhat with the statement that they trusted that (L)DP was secure, in contrast to only 14% of those that did not use a health app. On the other hand, 26% of those who did not use a health app distrusted the app (somewhat disagree or lower) with their private information, whereas only 7% of health app users said the same. There was no difference in trusting the app or the server among the three conditions as well as no difference in trust in (L)DP between the groups DP and LDP.

Willingness to share: There are differences in gender when participants report their willingness to share. Female participants share more with *local only* (low-sensitivity: $\chi^2_{(28)} = 20.50, p = .005$; high-sensitivity: $\chi^2_{(28)} = 25.76, p = .001$), while male participants share more with *both* (low: $\chi^2_{(28)} = 16.33, p = .022$; high: $\chi^2_{(28)} = 18.85, p = .009$). The usage of health apps stands out again, as participants who used health apps decided more often to *opt out* (low: $\chi^2_{(7)} = 15.11, p = .035$; high: $\chi^2_{(7)} = 19.23, p = .007$) and to share with *both* (low: $\chi^2_{(7)} = 15.33, p = .032$; high: $\chi^2_{(7)} = 32.59, p < .001$). Further significant results are correlations between age and the decision to *opt out* ($\chi^2_{(28)} = 56.74, p = .001$) and to share with *local only* ($\chi^2_{(28)} = 56.25, p = .001$) for the high-sensitivity questions and to share the low-sensitivity questions with *both* ($\chi^2_{(28)} = 49.84, p = .007$). Also, participants who reported an IT background were significantly more willing to share high-sensitivity questions with *both*. ($\chi^2_{(7)} = 18.38, p = .010$)

5.4 Comparison and discussion

As in the original study, there was hardly any difference in the participants' willingness to share information between the DP and the LDP group. It could be expected that people share more with *both* under the LDP condition and share more *local only* with the DP condition. However, both conditions led participants to share more with *both* and to *opt out* less in very similar rates. We could confirm that the question sensitivity is significant across all three groups.
The major difference between this experiment A and the original study's experiments 1+2 is that in our case the communication of (L)DP had a significant effect on the participants' willingness to share, especially when looking at high-sensitivity questions. However, there was hardly any difference in the *local only* decision across the three groups, which suggests an "all or nothing" mindset of our participants.
Participants showed more trust in the app, the app-server, and (L)DP when they were already using health apps, which is in line with findings in [4], and more willingness to share if they had an IT background.

6 Experiment B

Here, we present the study design and the results of experiment B before we again discuss and compare the results with the ones from the original paper. The complete questionnaire for experiment B can be seen in Appendix B.

6.1 Study design

For our second experiment, we combined the last two experiments of the original study into one LimeSurvey questionnaire (see Fig. 3). By doing so, we could directly compare the self-reported understanding of (L)DP with the comprehension questions, while they were separated in the original study. This also allowed us to test all 11 descriptions of (L)DP, which also provides additional results compared to the original study's experiment 4.We first asked for the participants' demographics and then presented one of the 11 (L)DP descriptions provided by [48]. A high-level summary of these descriptions can be seen in Tab. 5, while the complete descriptions are available in Appendix D. After the participants' introduction to (L)DP, questions regarding trust and comprehension were asked. In the following, we present a short description of the questions, while the full questionnaire is available in Appendix B.

(Q1) Do you want to share personal data with the app server given the presented data protection technique?

(Q2) Why? / Why not? (open question depending on the answer to the previous question)

(Q3) The description of (L)DP was understandable. (7-point Likert scale)

(Q4) Please highlight the words you did not understand (based on a score < 4 on the previous question, participants could highlight words by clicking on them)

(Q5) Comprehension questions

- C1. Can an attacker see your data if they get access to the data base?
- C2. Can employees see your data?
- C3. Can third-party companies see your personal data?
- C4. The usability of the data is now ... when the presented data protection technique is in place (better/worse/the same)
- C5. Do the data stay useful for third-party companies?

The comprehension questions in Q5 were presented in random order. Participants also had the choice not to answer or to select that they were unsure.

Group	Summary of description
LDP Flow	Answers are changed before they are sent to the company. Focus on the *flow* of data.
DP Flow	Answers are sent to the company's data base; others only receive changed answers to queries.
US Census	DP introduces controlled noise into the data, personal information is protected.
Google	LDP guarantees users' privacy as with random coin tosses.
Apple	DP transforms the data before they leave the device; true data cannot be reproduced.
Uber	DP allows statistical analyses without revealing information about individuals.
Microsoft	DP allows privacy-preserving data analysis by introducing inaccuracies into the analyzes.
DP Imp	DP uses only a modified version of your data. Personal information is not protected if the data base is compromised. Focus on DP's *implication* on the data.
LDP Imp. w/o Local	DP changes your data on the app randomly before they are sent to the server. Privacy is protected if the data base is compromised. No mention of the word "local".
LDP Imp	LDP changes your data on the app randomly before they are sent to the server. Privacy is protected if the data base is compromised.
LDP Comp	LDP introduces random noise to raw data before they are sent to the server. Used by Google, Apple. Includes *company* names.

Table 5: Summary of (L)DP descriptions for experiment B

6.2 Participants

As in experiment A, we used a between-subjects factorial design for our questionnaires. Participants were divided into 11 groups. In each of these groups, a different German description of (L)DP was presented to the participants. We excluded

203 participants who did not want to answer the question of whether they want to share data (Q1), 67 participants who gave nonsensical responses to the open question why they did or did not want to share data (Q2), and 31 with a completion time of less than 60 seconds.

6.3 Results

Here, we present our results of experiment B, first starting with the replication tests and followed by our additional tests.

6.3.1 Replication tests

Willingness to share: Across all 11 groups, 53% wanted to share sensitive information (Q1) with the application server. The LDP Imp group had the largest sharing rate with 60%, and DP Flow had the lowest sharing rate of 47%, see Fig. 5.

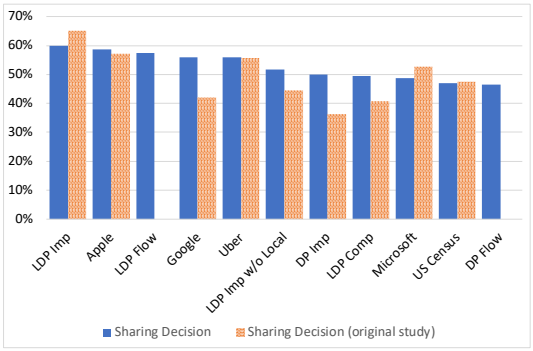


Figure 5: Sharing decision rates compared to the original study's experiment 3 [48]

Comprehension: Across all groups, only 13% indicated an easy-to-comprehend rating of less than 4 (on a 7-point Likert scale), which means that our participants were confident in their understanding of (L)DP. Participants in the LDP Imp w/o Local group showed the highest self-reported comprehension (M=5.3, SD=1.3), followed by Apple (M=5.3, SD=1.5), DP Imp (M=5.3, SD=1.3), and Uber (M=5.2, SD=1.3). Participants in the DP Flow group report the lowest understanding (M=4.8, SD=1.4), see Fig. 6.

Participants who indicated that the description of (L)DP was not understandable (score of less than 4 in Q3) could highlight the words that were less understandable (Q4). Across all groups, the most selected words were "differential" (22), "privacy" (21), "poise" (18), and "introduces" in combination with "controlled" (9). The correct response rates of the comprehension questions (Q5) are very low throughout all groups (see Tab. 6). Participants of all groups were able to answer correctly more often than 50% on average only for the question *C3 3rd party* when they were asked about the utility

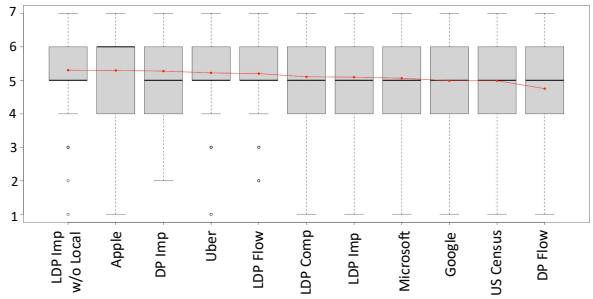


Figure 6: Self-reported easy-to-comprehend rating on a 7-point Likert scale sorted decreasing by mean (red dots).

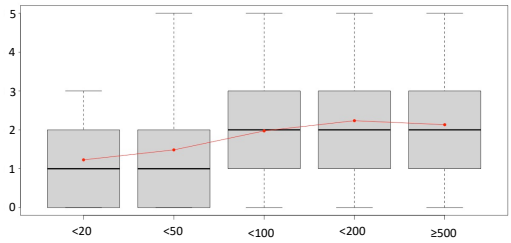


Figure 7: Time spent looking at the (L)DP descriptions in seconds vs. average number of correct responses on the comprehension questions in Q5. The red dots indicate the respective means.

of the perturbed data for third parties. Participants in the DP Imp group scored higher on the question *C1 Attacker* than the participants in other DP groups. However, this was to be expected, as the possibility of an attacker gaining access to the unperturbed data is mentioned within the DP Imp description. The only significant difference between DP vs. LDP is the correct response rate for *C1 Attacker* (30% vs. 49%, $\chi^2_{(1)} = 32.07, p < .001$, see Tab. 7).

We also tested whether participants who looked longer at the descriptions performed better at the comprehension questions. We computed Spearman's rank correlation between the time participants were spending on the description of (L)DP and their cumulative score on the comprehension questions (Q5) and found a positive correlation ($r(935) = .237, p < .001$). Due to technical reasons, the figure we compare regarding the time spent reading the description also includes the participants' answer to the questions whether they would like to share data (Q1) and their reasoning (Q2). The visualization of the differences in the average correct response rate based on the time spent looking at the (L)DP description can be seen in Fig. 7. There, we can see that the correct response rate peaks for participants who spent between 100 and 200 seconds reading the description and there is no improvement

when participants took longer than 200 seconds. The median time looking at the description was 57 seconds across all participants.

Sharing behavior: We used inductive coding to analyze the valid answers to the open question why participants decided (not) to share data (Q2) based on the established codes of the original study [48]. Two authors coded the answers independently and discussed the differences afterwards. If a participant's answer fell into two code categories, both were counted.

Why do participants want to share data?

Trust in DP and LDP techniques. 46% of responses fell into this category. Answers include "seems secure and recommended by experts", "statistical analysis without identification" (Uber), "randomized data gives a sense of security" (LDP Imp w/o Local) but also wrong assumptions such as "seems to be encrypted" (Apple).

Utility considerations. This category encompasses 31% of valid responses. Examples are "more data equals better recommendations" (Apple), "probably important for using the app" (DP Flow), and "brings advantages and seems secure" (LDP Imp).

Little privacy concern for asked or any information, learned helpless, and no fear of loss. This category holds 30% of responses. It is noteworthy that most answers in this category fall into the category *no fear of loss*: "nothing to hide" (LDP Imp), "most information is online anyway" (DP Imp), and "the requested data is not that important" (Microsoft).

Why do participants not want to share data?

Too sensitive to share. The majority of responses (51%) fall in this category. Participants wrote: "personal data should stay personal" (Microsoft), "data not relevant for health app" (Microsoft), and "no advantage for me" (US Census). Another common theme in this category are participants who are skeptical about sharing their income level.

Distrust differential privacy techniques. 31% of answers revealed little trust in general or more explicitly in (L)DP: "the term 'noise' is not explained well enough" (US Census), and "does not sound trustworthy" (Uber)

Risks of data leak, breach, or hack. Similar to the previous category, 12.99% of answers indicated that data breaches are always possible, no matter what the security promises: "no data is secure" (Uber), and "even the best software has holes in it" (LDP Imp w/o Local).

Distrust the app or tech companies. 18% of responses explicitly stated distrust of apps or tech companies in general: "as it is a private company I distrust these promises of data security" (LDP Imp w/o Local), and "my data is none of the app's business" (LDP Imp)

Some participants, however, have opposite opinions. For example, we have two participants who each stated that the mention of Google in the description (LDP Comp) influenced their decision whether they wanted to share their data (Q1). One participant did not wish to share, stating that it "does not seem secure especially since Google is involved", while another participant decided to share because "it is used by Google and Apple and therefore must be secure".

6.3.2 Additional tests

Again, we did some different additional tests to investigate potential differences in the participants' demographics.

Comprehension: We performed Kruskal-Wallis tests that revealed significant differences in IT background. Participants who indicated that they had an IT background found the description of (L)DP significantly easier to understand ($H(1) = 7.92, p = .005$) and answered correctly significantly more often to *C5 Utility 3rd party* ($H(1) = 4.652, p = .031$).

Willingness to share: Participants using apps to monitor their health were more willing to share information (Q1) than others ($H(1) = 37.47, p < .001$). Other demographics do not significantly impact the results.

Self-reported understanding vs. comprehension: We computed Spearman's rank correlation to investigate the relationship between self-reported understanding of the description (Q2) and the scores on the comprehension questions (Q5) (see Tab. 7). There was a significant positive correlation for *C1 Attacker* ($r(935) = .110, p = .001$), *C2 Employee* ($r(935) = .120, p < .001$), *C3 3rd party* ($r(935) = .168, p < .001$), and *C5 Utility 3rd party* ($r(935) = .152, p < .001$). Participants in LDP groups were on average significantly better at answering *C1 Attacker* correctly than the participants in the DP groups ($H(1) = 32.04, p < .001$).

6.4 Comparison and discussion

Compared to the original study, we obtained a similar sharing rate for the sensitive information (see Fig. 5). Across all conditions, 53% wanted to share sensitive information, opposed to 47.8% in the original study [48]. Also, we can report the largest sharing rate of 60% in the LDP Imp group, just as in the original study where the sharing rate of this condition was 65%. The major difference in this area is that no sharing rate is below 46% in our case, whereas there were some groups in the original study that had a sharing rate below that. An interesting similarity lies in the overall difficult-to-comprehend rate (A score less than 4 for Q3) of 13.4% compared to 13.3% in [48]. There is a difference in the lowest difficult-to-comprehend rate of a group: 0% in the original study's DP Imp group and 10% for our Apple group. However, our highest difficult-to-comprehend rate (DP Flow, 17%) is much lower than the one of the original study (DP w/o Names 30%). Overall, the differences in the difficult-to-comprehend rating and the sharing decision among groups are not as large as they were in the original study.

The participants' comments regarding the reasoning behind their decision to share or not to share data are very similar to the original study's. Two participants noted that they would

	DP						LDP				
	Apple	DP Flow	DP Imp	Microsoft	Uber	US Census	Google	LDP Comp	LDP Flow	LDP Imp	LDP Imp w/o Local
C1 Attacker	20%	24%	49%	29%	24%	27%	37%	45%	54%	55%	47%
C2 Employee	59%	34%	30%	33%	28%	29%	38%	32%	42%	48%	38%
C3 3rd party	68%	53%	48%	45%	49%	49%	47%	47%	62%	65%	53%
C4 Usability	6%	23%	7%	21%	6%	11%	11%	7%	18%	14%	9%
C5 Useful 3rd party	41%	49%	42%	48%	47%	43%	46%	47%	44%	35%	35%

Table 6: Correct response rates

		C1 Attacker	C2 Employee	C3 3rd party	C4 Usability	C5 Useful 3rd party
easy-to-comprehend	r	**.11**	**.12**	**.16**	-.03	**.15**
	p	**<.001**	**<.001**	**<.001**	.369	**<.001**
DP vs. LDP	χ^2	**32.07**	1.24	.30	.14	1.66
	p	**<.001**	.265	.583	.705	.197

Table 7: Correlations of easy-to-comprehend and difference DP/LDP regarding the comprehension questions in experiment B

like to see the data before it leaves the device in order to understand the data perturbation. There was no mention of this in the original study. An alarmingly large portion of answers fall into the privacy fatigue [26] category, with assumptions that their personal data is "never secure" and "out there anyway". Also, the comments regarding the participants' unwillingness to share their income level is in line with previous research about German data sharing preferences [23].

As in experiment A (see Sec. 5.4), personal health app usage has a significant impact on our participants' answers. This time, it shows a significant difference in the decision to share. Unsurprisingly, participants with an IT background showed a significantly higher score on the self-reported comprehension of the description of (L)DP. The self-reported understanding of the (L)DP description correlated positively with almost all comprehension questions; however, the correlation coefficients are below 0.2 which suggests a weak association and thereby a limited effect size [38]. The only significant difference between participants who were assigned to DP and those that were assigned to LDP descriptions lies in the comprehension question *C1 Attacker*, where most participants in the DP groups falsely believed that an attacker does not have access to the real answers if the company's data base is breached. As this is one of the key differences between DP and LDP, it shows once again that the difference is not clearly communicated and understood. One exception to this is the participants in the DP Imp group, where this scenario of a data base breach is explicitly mentioned. Still, even in this group most participants answered the question wrong. However, we found that participants who spent more time reading the (L)DP descriptions performed better on the comprehension questions.

7 Discussion

While our study partially confirms the findings of the original study by Xiong et al. [48], we also provide additional insights about (L)DP communication in a different culture (Germany), different demographics, and the impact of personal health app usage. Overall, participants who were told that their data would be protected by (L)DP decided to share more high-sensitivity data than those in the control group, which indicates that (L)DP communication had a positive effect on their data sharing attitudes. Similar to the original study, the participants' responses did not significantly differ between the LDP and the DP groups. This suggests that at least LDP was not completely understood. This also confirms the previously mentioned findings from Cummings et al. [11] that users misunderstand various descriptions of DP (see Sec. 3). Although self-reported understanding of the (L)DP descriptions was relatively high, the subsequent comprehension questions reveal that participants overestimated their understanding. Although participants with higher self-reported comprehension answered correctly more often on most comprehension questions, only few of them provided exclusively correct answers. As participants who spent more time reading the descriptions provided more correct answers, we can speculate that reading the description thoroughly improves the comprehension of (L)DP. However, it is also likely that the descriptions were not worded in a clear way. Due to the fact that users generally prefer not to read privacy statements [41], it is reasonable to assume that they do not want to read lengthy (L)DP descriptions either. As a result, alternative solutions based on more visual (L)DP communication like those proposed for privacy policies [27] should be investigated in the future. We observe the same pattern in the open answers about the participants' willingness to share their data as in [48]. However, some participants noted that they would need an example of "how noise changes the data", "what a hacker would have access to", or "of what use the small inaccuracies are". These statements indicate that users do not want only a vague privacy guarantee, which is probably too technical for laypeople to understand fully. They would rather see the actual perturbation of their data or at least a clearer

and more understandable presentation of (L)DP. Moreover, we could observe a pattern that the personal usage of health apps increases trust and the willingness to share data.

Besides the expected differences in attitudes due to cultural and regulatory differences, summarized in Sec. 3, it is also important to take the timeframe of the respective studies into account. Xiong et al. performed their study before March 2020, i.e., a time before worldwide lockdowns forced people and companies into digitalization. As our study was conducted during the summer of 2021, it is possible that our sample was more familiar with and presumably more trusting of digital technologies and less concerned about associated privacy risks.

8 Conclusion

We have replicated a study on the effect of DP communication on the willingness to share data and on the understanding of and trust in the privacy-preserving technique. Despite our different sample comprising German participants representative of the population, our results are similar to the original study in that participants' answers were not significantly different between LDP and DP models. However, the effect of DP communication could clearly be observed since the participants were significantly willing to share more data when (L)DP was applied. As a result, they trust the technology to protect their privacy. The big caveat is that even though self-reported understanding was high, follow-up comprehension questions revealed that participants did not fully understand the concept of (L)DP. Arguably, visual or otherwise more understandable differential privacy communication would help users' comprehension [24, 29].

References

[1] Hofstede Insights. Online: https://www.hofstede-insights.com/country-comparison/germany,india,the-usa/ (accessed in 02/2022).

[2] Miguel E Andrés, Nicolás E Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Geo-Indistinguishability: Differential Privacy for Location-Based Systems. In *Proceedings of the 20th ACM SIGSAC Conference on Computer & communications security (CCS)*, 2013.

[3] Catherine Barrett. Are The EU GDPR And The California CCPA Becoming the de facto Global Standards for Data Privacy and Protection? *Scitech Lawyer*, 2019.

[4] Steven Bellman, Eric J Johnson, Stephen J Kobrin, and Gerald L Lohse. International Differences in Information Privacy Concerns: A Global Survey of Consumers. *The Information Society*, 2004.

[5] Brooke Bullek, Stephanie Garboski, Darakhshan J Mir, and Evan M Peck. Towards Understanding Differential Privacy: When do People Trust Randomized Response Technique? In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2017.

[6] TH Hubert Chan, Elaine Shi, and Dawn Song. Optimal Lower Bound for Differentially Private Multi-Party Aggregation. In *Proceedings of the 20th European Symposium on Algorithms (ESA)*, 2012.

[7] Min Chen, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang. Disease Prediction by Machine Learning over Big Data from Healthcare Communities. *IEEE Access*, 2017.

[8] European Commission. General Data Protection Regulation (GDPR), 2016.

[9] Panos Constantinides and David A Fitzmaurice. Artificial Intelligence in Cardiology: Applications, Benefits and Challenges. *Br J Cardiol*, 2018.

[10] Rowena Cullen. Culture, Identity and Information Privacy in the Age of Digital Government. *Online Information Review*, 2009.

[11] Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles. "I need a better description": An Investigation Into User Expectations For Differential Privacy. In *Proceedings of the 28th ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2021.

[12] Djellel Eddine Difallah, Elena Filatova, and Panagiotis G. Ipeirotis. Demographics and Dynamics of Mechanical Turk Workers. *Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM)*, 2018.

[13] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (ICONIP)*, 2017.

[14] Cynthia Dwork. Differential privacy. In *Proceedings of the 33rd International Colloquium on Automata, Languages, and Programming (ICALP)*, 2006.

[15] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating Noise to Sensitivity in Private Data Analysis. In *Proceedings of the 3rd Theory of Cryptography Conference (TCC)*, 2006.

[16] Benjamin Edwards, Steven Hofmeyr, and Stephanie Forrest. Hype and Heavy Tails: A Closer Look at Data Breaches. *Journal of Cybersecurity*, 2016.

[17] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *Proceedings of the 21st ACM SIGSAC Conference on computer and communications security (CCS)*, 2014.

[18] Arik Friedman and Assaf Schuster. Data Mining with Differential Privacy. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2010.

[19] Kenneth Goodman, Diana Zandi, Andreas Reis, and Effy Vayena. Balancing Risks and Benefits of Artificial Intelligence in the Health Sector. *Bulletin of the World Health Organization*, 2020.

[20] Nadine Guhr, Oliver Werth, Philip Blacha, and Michael Breitner. Privacy Concerns in the Smart Home Context. *SN Applied Sciences*, 2020.

[21] Anne-Wil Harzing. Why Replication Studies are Essential: Learning from Failure and Success. *Cross Cultural & Strategic Management (CCSM)*, 2016.

[22] Aylin Ilhan and Maria Henkel. 10,000 Steps a Day for Health? User-based Evaluation of Wearable Activity Trackers. In *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS)*, 2018.

[23] Maria Karampela, Sofia Ouhbi, and Minna Isomursu. Exploring Users' Willingness to Share Their Health and Personal Data Under the Prism of the New GDPR: Implications in Healthcare. In *Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019.

[24] Farzaneh Karegar and Simone 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*, 2021.

[25] Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What Can We Learn Privately? *SIAM Journal on Computing (SICOMP)*, 2011.

[26] M. J. Keith, C. Maynes, P. B. Lowry, and J. Babb. Privacy Fatigue: The Effect of Privacy Control Complexity on Consumer Electronic Information Disclosure. In *Proceedings of the 35th International Conference on Information Systems, (ICIS)*, 2014.

[27] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. A "Nutrition Label" for Privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS)*, 2009.

[28] Hanna Krasnova, Natasha F Veltri, and Oliver Günther. Self-Disclosure and Privacy Calculus on Social Networking Sites: The Role of Culture. *Business & Information Systems Engineering (BISE)*, 2012.

[29] Patrick Kühtreiber and Delphine Reinhardt. Usable Differential Privacy for the Internet-of-Things. In *Proceedings of the 19th IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, 2021.

[30] Ereni Markos, George R Milne, and James W Peltier. Information Sensitivity and Willingness to Provide Continua: A Comparative Privacy Study of the United States and Brazil. *Journal of Public Policy & Marketing (JPP&M)*, 2017.

[31] Sandra J Milberg, H Jeff Smith, and Sandra J Burke. Information Privacy: Corporate Management and National Regulation. *Organization science*, 2000.

[32] Zareef A Mohammed and Gurvirender P Tejay. Examining Privacy Concerns and Ecommerce Adoption in Developing Countries: The Impact of Culture in Shaping Individuals' Perceptions Towards Technology. *Computers & Security*, 2017.

[33] Patrick Murmann, Matthias Beckerle, Simone Fischer-Hübner, and Delphine Reinhardt. Reconciling the What, When and How of Privacy Notifications in Fitness Tracking Scenarios. *Pervasive and Mobile Computing (PMC)*, 2021.

[34] Kee Yuan Ngiam and Wei Khor. Big Data and Machine Learning Algorithms for Health-Care Delivery. *The Lancet Oncology*, 2019.

[35] Caroline L Park. What is the Value of Replicating Other Studies? *Research Evaluation*, 2004.

[36] Christine Prince, Nessrine Omrani, Adnane Maalaoui, Marina Dabic, and Sascha Kraus. Are We Living in Surveillance Societies and Is Privacy an Illusion? An Empirical Study on Privacy Literacy and Privacy Concerns. *IEEE Transactions on Engineering Management*, 2021.

[37] Ismini Psychoula, Deepika Singh, Liming Chen, Feng Chen, Andreas Holzinger, and Huansheng Ning. Users' Privacy Concerns in IoT Based Applications. In *Proceedings of the 4th IEEE International Conference on Internet of People (IoP)*, 2018.

[38] Louis M Rea and Richard A Parker. *Designing and Conducting Survey Research: A Comprehensive Guide*. 2014.

[39] Nuria Rodríguez-Barroso, Goran Stipcich, Daniel Jiménez-López, José Antonio Ruiz-Millán, Eugenio Martínez-Cámara, Gerardo González-Seco, M Victoria Luzón, Miguel Angel Veganzones, and Francisco Herrera. Federated Learning and Differential Privacy: Software tools analysis, the Sherpa. ai FL framework and methodological guidelines for preserving data privacy. *Information Fusion*, 2020.

[40] Eva-Maria Schomakers, Chantal Lidynia, Dirk Müllmann, and Martina Ziefle. Internet Users' Perceptions of Information Sensitivity–Insights from Germany. *International Journal of Information Management (IJIM)*, 2019.

[41] Tomáš Sigmund. Attention Paid to Privacy Policy Statements. *Information*, 2021.

[42] Statistisches Bundesamt (Destatis). 12111-0004: Bevölkerung (Zensus): Deutschland, Stichtag, Geschlecht, Altersgruppen, 2021.

[43] Differential Privacy Team. Learning with Privacy at Scale. Online: https://machinelearning.apple.com/research/learning-with-privacy-at-scale (accessed in 12/2021).

[44] Sabine Trepte, Leonard Reinecke, Nicole B Ellison, Oliver Quiring, Mike Z Yao, and Marc Ziegele. A Cross-Cultural Perspective on the Privacy Calculus. *Social Media+ Society*, 2017.

[45] Amos Tversky and Daniel Kahneman. Rational Choice and the Framing of Decisions. In *Multiple Criteria Decision Making and Risk Analysis Using Microcomputers*. 1989.

[46] Stanley L Warner. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association (JASA)*, 1965.

[47] Jochen Wirtz, May O Lwin, and Jerome D Williams. Causes and Consequences of Consumer Online Privacy Concern. *International Journal of service industry management*, 2007.

[48] Aiping Xiong, Tianhao Wang, Ninghui Li, and Somesh Jha. Towards Effective Differential Privacy Communication for Users' Data Sharing Decision and Comprehension. In *Proceedings of the 41st IEEE Symposium on Security and Privacy (S&P)*, 2020.

APPENDIX

A Questionnaire for Experiment A

The questionnaire is taken from Xiong et al.'s original study [48] and has been translated to German. We provide the back-translated English version of our questionnaire which is almost verbatim to the original study. The PDF version of the German questionnaire is available online[1].

Introduction

In the digital age, everyone faces the question whether to share personal data in exchange for goods, services, or other advantages. The goal of this study is to understand what kinds of information you wish to share with a health app and how these data should be used.

Demographics

The demographics were checked first in order to fulfil the quotas of the questionnaire. Participants had the possibility to answer "No answer" to all questions.

What is your age group?
Please indicate your gender.
What is your highest school-leaving qualification?
Do you have an IT background?
Do you use apps or devices to monitor your health data?

Precondition

Please assume the following for this questionnaire:
1. You have just downloaded the health app Orange Health and you start using it immediately
2. To ensure suitable advice and recommendations regarding your health, the app asks for certain information, for example, your age and gender in regard to daily calorie intake.
3. At the same time, the app server requests permission to access and collect the information in order to provide you with a better user experience. For example, the information you share will be used to train machine learning algorithms that will subsequently will be used to provide more exact recommendations for all users.

Differential privacy communication

Here, the participants in the DP and LDP groups were shown the descriptions for DP and LDP respectively (see Sec. C). Afterwards, the following comprehension question was presented.

Please indicate which of the following descriptions of (local) differential privacy is correct:

[1] https://owncloud.gwdg.de/index.php/s/kDAUTawPDsJxAwp

- *A data protection technique that adds random noise to the collected data of user groups (e.g. average age) in order to protect the user's privacy just as if the user had not taken part in the data collection.*
- *A data protection technique, which adds random noise to every user response in order to protect the user's privacy just as if the user would not take part in the data collection.*
- *DP/LDP has not been used yet in any organization or company.*
- *I prefer not to answer.*

Participants were shown the respective description again if they answered incorrectly.

Questions of the Orange Health app

Participants first were presented with an explanation of how to answer the questions. Again, this is a direct translation and adaption of the original explanation from Xiong et al. [48].

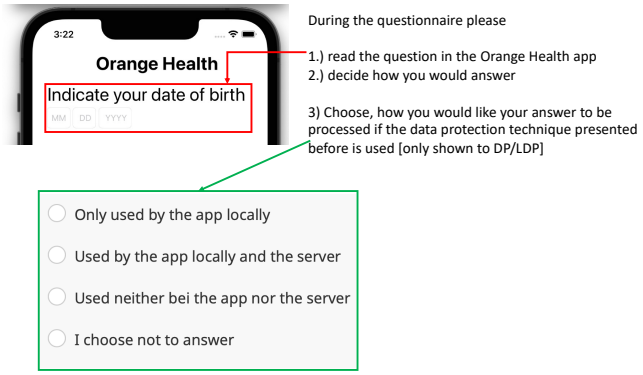


Figure 8: Back-translated explanation for the participants. Point 3 has only been provided to the groups DP and LDP.

Participants were provided with 14 screenshots of the questions in Tab. 2 similar to the one in Fig. 8 in random order and could choose the following answers.

- *Only used by the app locally*
- *Used by the app locally and the server*
- *Neither used by the app nor the server*
- *I prefer not to answer*

Trust questions

Participants could answer the following questions on a 7-point Likert scale ranging from "Strongly disagree" to "Strongly agree".

1. *I trust the Orange Health app to protect my personal information*
2. *I trust the app server to protect my personal information*
3. *I trust (local) differential privacy to protect my personal information*

The third question was only asked to participants in the DP and LDP groups.

B Questionnaire for experiment B

The PDF version of the German questionnaire is also available online[2].

Introduction

The goal of this study is to evaluate your willingness to share personal information when a data protection technique is used. The goal is also to understand why you made this respective decision. Furthermore, we want to evaluate your comprehension of this data protection technique.

Demographics

We asked the same demographic questions as in the first questionnaire A

Precondition

We presented the same precondition as in the first questionnaire A

Differential privacy

To respect your personal information and to guarantee a better user experience, the data that are shared with the Orange Health app are collected using a data protection technique. This data protection technique is presented in the following. Please read the description carefully.

The participants were randomly assigned to one of the eleven descriptions in D.

Trust in (L)DP

Under the condition that the above described data protection technique is in use: Would you share your personal data (e.g. date of birth, family medical record, income level, substance use, medical record, previous surgeries, current medication) with the app server?

- *Yes / No / No answer*

[2] https://owncloud.gwdg.de/index.php/s/s5hQeVmLNyy2kve

If the participant answered yes:
Please explain briefly why you would like to share your personal data if the described data protection technique is in use?

- Open question

If the participant answered no:
Please explain briefly why you would not like to share your personal data if the described data protection technique is in use?

- Open question

Self-reported understanding of (L)DP

Participants could answer the following questions on a 7-point Likert scale ranging from "Strongly disagree" to "Strongly agree".
Please indicate your agreement with the following statements: The previous description of the data protection technique was easy to understand.

If participants provided a score of 3 ("mildly disagree") or less they were presented with the description again to highlight words they did not understand.
You have indicated that the description of the data protection technique was not easy to understand. Please indicate the words you find hard to understand by clicking on them to highlight them.

Comprehension questions

C1 *Suppose you have answered truthfully to the questions in the Orange Health app and your answers have been collected with the presented data protection technique. If an attacker gets access to the data base of the Orange Health company, will he then be able to see your true answers?*

C2 *Suppose you have answered truthfully to the questions in the Orange Health app and your answers have been collected with the presented data protection technique. Are employees within the Orange Health company able to see your true responses?*

C3 *Suppose you have answered truthfully to the questions in the Orange Health app and your answers have been collected with the presented data protection technique. Are third parties with whom the Orange Health company shares data able to see your true answers?*

C4 *With the changes imposed through the data protection technique, the accuracy of the aggregated data the Orange Health company receives is . . . compared to the actual results without the data protection technique.*

C5 *Suppose you have shared data such as your family medical record with the health app. Do the results, which have been collected using the data protection technique to protect your privacy, stay useful for third party companies with whom the health app company shares data?*

Participants could answer *Yes / No / Unsure / No answer* for all answers except C4, which had the options *better / worse / unchanged / unsure*.

C Descriptions of (L)DP for experiment A

Again, here and in Appendix D we provide the back-translated English versions of our German (L)DP descriptions which are almost the same as the ones provided by Xiong et al. [48].

DP

Data shared with the app will be processed using differential privacy (DP) to protect your personal data and to ensure the best user experience. DP protects the users' privacy by adding random noise to the aggregated data, such as average age, so that the probability of deducing an individual person's information is low. DP is used in academia as well as in the corporate world, including Harvard University, the US Census Bureau and corporations such as LinkedIn and Uber.

LDP

Data shared with the health app will be collected using local differential privacy (LDP) to protect your personal information and to ensure the best user experience. LDP protects the users' privacy by adding random noise to every answer provided by a user. As a result, the probability of deducing a user characteristic is roughly as high as if the user had not taken part in data collection. LDP is used by companies such as Apple and Google.

D Descriptions of (L)DP for experiment B

LDP Flow

When local differential privacy (LDP) is used, the app changes the answers before they are sent from the user's device to the company. The company sees and stores only the changed version of each user's information and is unsure of the users' true answers. If changed answers from a large number of users are analyzed, however, the company can still gather useful results in aggregated form about the user population, although the accuracy is reduced compared to unchanged data.

DP Flow

When differential privacy (DP) is used the app sends the user's answers to the company. These answers are stored in

the company's data base. If the company wants to use these data either internally or with third parties, the company sends queries to the data base, uses DP techniques to change the results of the queries and uses only these changed results. The changed results only provide limited information concerning a specific user. If, however, the answers of a large number of users are analyzed, the company can still obtain useful results in aggregated form about the whole user population, even if the accuracy is lower compared to unchanged data.

US Census

Differential privacy has been developed by researchers at Microsoft and is used by many leading technology companies. There are many variants of differential privacy. The one used here introduces controlled noise into the data, so that the accuracy remains at higher levels. This method to protect privacy has been developed to maintain the data's usability and also to completely protect the personal information of each affected person.

Google

Building upon the concept of randomized response, local differential privacy (LDP) makes it possible to generate statistics about user behavior while guaranteeing the users' privacy. LDP builds upon this concept by allowing the app to send reports that are factually indistinguishable from random coin tosses and do not contain any unique user names. By aggregating reports, common statistics that are the same for many users can be derived.

Apple

Differential privacy transforms the information that is shared with the company before it leaves the device, so that the company can never reproduce the true data. The basic idea of differential privacy is to introduce statistical noise that hides the users' personal data before they are sent to the company. When a lot of people send the same kinds of data the introduced noise will cancel out on average and the company is able to gather useful information thanks to the huge amount of data.

Uber

Differential privacy is a formal definition of privacy and is accepted on a broad scale by industry experts because it provides robust privacy protection. In short, differential privacy allows general statistical analyses without revealing information about an individual within the data. That is why differential privacy provides an additional safety barrier against recognition attacks as well as attacks with auxiliary data.

Microsoft

Differential privacy is a technique that enables researchers and analysts to obtain useful analyses of data bases containing personal information. At the same time, it provides a strong protection for individual privacy. This seemingly contradictory result is reached by inserting relatively moderate inaccuracies into the analyses. These inaccuracies are large enough to protect the privacy but small enough so that the analyses remain useful for researchers and analysts.

LDP Imp. w/o Local

Data that is shared with the app will be processed with the help of the differential privacy (DP) technique to respect your personal information and to ensure the best user experience. The app will change the data on your app randomly before they are sent to the app-server. As the app-server now only stores the changed version of your personal information, your privacy is protected even if the data base of the app-server will be compromised.

LDP Imp.

Data that is shared with the app is processed with the help of the local differential privacy (LDP) technique to respect your personal information and to ensure the best user experience. The app changes the data on your app randomly before they are sent to the app server. As the app server now only stores the changed version of your personal information, your privacy is protected even if the data base of the app server is compromised.

DP Imp.

Data that is shared with the app is processed with the help of the differential privacy (DP) technique to respect your personal information and to ensure the best user experience. The health app company stores your data but only uses the modified total statistics, so that your personal information cannot be learned. Your personal information can be leaked, however, if the data base of the company is compromised.

LDP Comp

Data that is shared with the app is processed with the help of the local differential privacy (LDP) technique to respect your personal information and to ensure the best user experience. LDP protects your privacy by introducing random noise to the raw data BEFORE they are sent to the company (the raw data never leaves your device). LDP is used by companies such as Google and Apple.

E Statistics

	Local Only		Both		Opt out	
	χ^2	p	χ^2	p	χ^2	p
Question Sensitivity	**21.48**	**<.001**	**280.5**	**<.001**	**217.63**	**<.001**
Condition	.110	.947	**97.95**	**<.001**	**139.21**	**<.001**
Con vs. DP	N/A		**<.001**		**<.001**	
Con vs. LDP			**<.001**		**<.001**	
DP vs. LDP			.917		.729	
QS * Condition	.65	.722	**9.94**	**.007**	**8.08**	**.018**
Low vs. High QS						
Control	N/A		**<.001**		**<.001**	
DP			**<.001**		**<.001**	
LDP			**<.001**		**<.001**	
low-sensitivity						
Con vs. DP	N/A		.033		**.002**	
Con vs. LDP			.028		**.001**	
DP vs. LDP			.912		.718	
high-sensitivity						
Con vs. DP	N/A		**<.001**		**<.001**	
Con vs. LDP			**.001**		**.004**	
DP vs. LDP			.790		.465	

Table 8: Statistics for experiment A

		Trust in		
		App	Server	(L)DP
Age	$H_{(4)}$	3.297	4.705	2.007
	p	.509	.319	.734
Gender	$H_{(1)}$	.845	1.14	.818
	p	.358	.286	.366
Education	$H_{(4)}$	4.131	4.628	5.912
	p	.389	.328	.206
IT BG	$H_{(1)}$	1.41	7.43	.848
	p	.842	.115	.357
Health App	$H_{(1)}$	**40.028**	**27.362**	**26.31**
	p	**<.001**	**<.001**	**<.001**

Table 9: Kruskal-Wallis tests on correlations between demographics and trust in experiment A

		Opt out		Local only		Both	
		Low	High	Low	High	Low	High
Age	$\chi^2_{(28)}$	38.66	**56.74**	36.4	**56.25**	**49.84**	37.77
	p	.087	**.001**	.133	**.001**	**.007**	.103
Gender	$\chi^2_{(7)}$	5.86	10.41	**20.50**	**25.76**	**16.33**	**18.85**
	p	.556	.167	**.005**	**.001**	**.022**	**.009**
Education	$\chi^2_{(28)}$	29.32	38.9	35.39	24.39	32.38	33.36
	p	.396	.083	.159	.661	.259	223
IT BG	$\chi^2_{(7)}$	16.38	8.03	8.89	7.64	13.6	**18.36**
	p	**.022**	.330	.261	.365	.059	**.010**
Health App	$\chi^2_{(7)}$	**15.11**	**19.23**	6.794	4.55	**15.33**	**32.59**
	p	**.035**	**.007**	.451	.715	**.032**	**<.001**

Table 10: Correlation between demographics and willingness to share for experiment A

		Share	easy-to-comprehend
Age	$H_{(4)}$	6.488	.643
	p	.166	.958
Gender	$H_{(1)}$	1.871	2.852
	p	.171	.091
Education	$H_{(4)}$	1.017	4.833
	p	.907	.305
IT BG	$H_{(1)}$	.526	**7.918**
	p	.468	**.005**
Health app	$H_{(1)}$	**37.465**	1.937
	p	**<.001**	.164

Table 11: Kruskal Wallis tests for impact on demographics on willingness to share and self-reported easy-to-comprehend rate for experiment B

		C1	C2	C3	C4	C5
Age	$H_{(4)}$	.701	2.282	5.667	9.239	6.824
	p	.951	.684	.225	.055	.145
Gender	$H_{(1)}$	.27	.139	2.38	.028	3.294
	p	.603	.710	.123	.867	.070
Education	$H_{(4)}$	6.578	6.89	5.928	6.978	6.819
	p	.160	.142	.205	.137	.146
IT BG	$H_{(1)}$	.347	.001	.722	1.705	**4.652**
	p	.556	.981	.396	.192	**.031**
Health app	$H_{(1)}$	2.895	.176	.005	2.14	.389
	p	.089	.675	.944	.143	.533

Table 12: Kruskal Wallis tests for impact on demographics on the comprehension questions for experiment B