Am I Private and If So, how Many? Communicating Privacy Guarantees of Differential Privacy with Risk Communication Formats

Daniel Franzen
daniel.franzen@fu-berlin.de
Freie Universität Berlin
Germany

Saskia Nuñez von Voigt
saskia.nunezvonvoigt@tu-berlin.de
Technische Universität Berlin
Germany

Peter Sörries
peter.soerries@fu-berlin.de
Freie Universität Berlin
Germany

Florian Tschorsch
florian.tschorsch@tu-berlin.de
Technische Universität Berlin
Germany

Claudia Müller-Birn
clmb@inf.fu-berlin.de
Freie Universität Berlin
Germany

ABSTRACT
Every day, we have to decide multiple times, whether and how much personal data we allow to be collected. This decision is not trivial, since there are many legitimate and important purposes for data collection, for example, the analysis of mobility data to improve urban traffic and transportation. However, often the collected data can reveal sensitive information about individuals. Recently visited locations can, for example, reveal information about political or religious views or even about an individual’s health. Privacy-preserving technologies, such as differential privacy (DP), can be employed to protect the privacy of individuals and, furthermore, provide mathematically sound guarantees on the maximum privacy risk. However, they can only support informed privacy decisions, if individuals understand the provided privacy guarantees. This article proposes a novel approach for communicating privacy guarantees to support individuals in their privacy decisions when sharing data. For this, we adopt risk communication formats from the medical domain in conjunction with a model for privacy guarantees of DP to create quantitative privacy risk notifications.

We conducted a crowd-sourced study with 343 participants to evaluate how well our notifications conveyed the privacy risk information and how confident participants were about their own understanding of the privacy risk.

Our findings suggest that these new notifications can communicate the objective information similarly well to currently used qualitative notifications, but left individuals less confident in their understanding. We also discovered that several of our notifications and the currently used qualitative notification disadvantage individuals with low numeracy: these individuals appear overconfident compared to their actual understanding of the associated privacy risks and are, therefore, less likely to seek the needed additional information before an informed decision.

1 INTRODUCTION
We generate a large amount of mobility data daily, for example, when using online map services for navigation or when purchasing public transport tickets. Due to the ubiquitousness of smartphones, collecting location and mobility data has become a quasi-standard also for many other applications, and prediction algorithms use the data collected for various purposes (e.g., [6, 29, 49]). Mobility data is also essential for the development of urban areas [60, 78], for example, to identify where public transport can be improved, bicycle lanes can be added or how new pedestrian zones might influence a neighborhood. These tasks are vital also to meet current environmental challenges [48, 78]. To obtain the necessary data for such legitimate tasks we have to rely on either data donations by individuals [33] or on existing data collections by third parties. However, laypeople are often unaware [7, 9, 54], but location data can reveal sensitive information [37]. They can be used to identify locations of interest (e.g., home address, religious or political organizations), reveal daily routines (e.g., doing exercises), show
A central point in the definition of DP is the parameter \( \epsilon \) which acts as a privacy budget. It limits the amount of information obtained about any individual. Based on the value chosen for \( \epsilon \) a mathematically sound privacy guarantee can be computed, which describes an upper bound for the privacy risk of any individual [22]. Since the privacy protection offered by DP depends heavily on \( \epsilon \), the chosen value is an essential piece of information for individuals to decide about sharing data using DP.

Several techniques have been developed (e.g., PATE for machine learning [50]) over the last years that realize the idea of DP by adding carefully constrained noise to the results of computations [4]. Companies, such as Apple and Google, or governmental agencies, such as the US Census Bureau, already employ DP (cf. [22]). However, so far, individuals have not been made aware of the use of DP by explaining some aspects of its functionality. Information about the chosen privacy budget \( \epsilon \) and the resulting quantitative guarantees are not communicated [19]. Thus, in the current state, individuals cannot evaluate the privacy protection offered by DP and, therefore, cannot make an informed decision on sharing their data. Instead we need methods to communicate the quantitative privacy guarantees to individuals. However, as privacy risk, aside from DP, is hard to quantify, these methods have not been researched very well so far.

In our research, we take the first step to close this gap. We propose using quantitative privacy risk notifications to convey the DP privacy guarantee, i.e., the remaining privacy risk of sensitive information being exposed when sharing data. As opposed to existing research that explains the functionality of DP (e.g., [76]), we adapt empirically validated research from the medical domain, where quantitative risk communication is an essential part of education and decision-support processes. In particular, we identify recommended medical risk communication formats, such as percentages (52%) or simple frequencies (26 out of 50), which can also be used to communicate privacy risk. In a crowd-sourced study, we compare the resulting quantitative privacy risk notifications to conventional qualitative DP notifications and investigate the effect of these notifications on individuals’ understanding of privacy risks.

Utilizing our research, we want to enable people to become aware of the level of protection offered by DP. We assume that only then individuals can make an informed decision by weighing up the purpose of the data collection with the remaining privacy risk before sharing their personal data. With this research we want to make the following contributions:

- We propose the usage of quantitative privacy risk notifications to communicate the privacy guarantees of DP in an understandable way.
- We derive suitable, empirically validated risk communication formats from existing medical research and combine them with privacy risk values derived from a mathematical model of DP [43] into quantitative privacy risk notifications. These notifications represent the real-world privacy guarantees by DP accurately.
- We evaluate the understandability of quantitative privacy risk notifications incorporating different risk communication formats in a crowd-sourced user study and find that they perform equally well to currently used qualitative notifications.
- Based on our results, we discuss opportunities for future research on supporting informed decisions with quantitative privacy risk notifications.

The remainder of this paper is structured as follows: We provide a brief overview of the current state of DP in Section 2 and argue that a new communication concept for DP is necessary. We substantiate this by discussing recent research into the communication of DP and survey suitable formats from medical risk communication. Section 3 details how we design novel privacy risk notifications utilizing selected risk communication formats and in Section 4, we describe the design of the study to evaluate their understandability. The results are presented in Section 5 and discussed in Section 6. We conclude in Section 7 with limitations and future work.

## 2 PRIVACY RISK - CHALLENGES AND OPPORTUNITIES

Our research is situated in the following context: An individual (e.g., the user of an app) has been using a mobility service (e.g., a public transport app) for some time. This mobility service collects personal data (e.g., mobility data, such as routes traveled or choice of transport mode). One day, the mobility service provider requests the individual’s consent of sharing the collected data with a third party (e.g., public administration). For the individual, sharing the data involves a certain privacy risk, the risk that sensitive information (e.g., home address) can be learned, for example, by combining data from different sources [61]. For making an informed decision, the individual needs to understand this privacy risk. Since the data are being anonymized by using DP, the mobility service provider can compute a privacy guarantee. This guarantee provides an upper bound on the probability, that sensitive data can be learned about the individual. With this information, the mobility service provider shows a privacy risk notification, a compact user interface containing easy to understand information on the privacy of the individual. Within this notification, the mobility service provider can present the privacy guarantees in different risk communication formats, for example, “Your privacy risk is 52%” or “Your data might be leaked in 26 out of 50 cases”.

The described context highlights that well-communicated privacy guarantees provided by DP can allow individuals to make informed privacy-preserving decisions. However, a challenge in this context is, how we need to design such privacy risk notifications. The communicated privacy risk should be derived from the defined privacy guarantee provided by DP. Thus, we provide in Section 2.1 a short introduction to DP and explain how privacy guarantees are provided. Furthermore, we summarize research on existing approaches on communicating DP to individuals (cf. Section 2.2), which so far have focused on qualitative descriptions only. We, therefore, extend our focus into the medical domain, where
quantitative risk communication has been studied over the last twenty years (cf. Section 2.3).

In the following, we use the term individual (e.g., for the user of the mobility service) to emphasize that the collected data contain personal and potentially sensitive information about people. We assume that these individuals are laypeople who have a limited knowledge of privacy measures and technologies used, and might have different levels of experience with statistics and risk representations. Accordingly, privacy risks must be communicated to these individuals with particular care and circumspection. Additionally, we consider two groups of stakeholders, namely, data owners and data consumers. Data owners are companies or service providers that collect individuals’ data (e.g., the mobility service provider). Data consumers are public institutions, companies or other third parties using this data for statistical analysis (e.g., the public administration).

2.1 Differential Privacy

Differential privacy is a mathematical property that aims to protect the privacy of individuals against the data consumer when querying information from a data set stored in a statistical database [24].

The underlying principle of DP is to limit the impact of a single individual on the analysis outcome. More specifically, the presence or absence of an individual’s data must not lead to a significantly different outcome of the analysis. Formally, assume two data sets \( D_1 \) and \( D_2 \) differing in exactly one entry. A function \( f \) provides \( \varepsilon \)-DP if for all such \( D_1 \) and \( D_2 \), all outcomes \( S \subseteq \text{Range}(f) \) satisfy

\[
P[f(D_1) \in S] \leq e^\varepsilon \cdot P[f(D_2) \in S].
\]

The core component of DP is the privacy budget \( \varepsilon \) that bounds how far an outcome of a data set is allowed to deviate when adding or removing one single individual.

We illustrate this property in Figure 1 and explain it in more detail next. We consider two data sets \( D_1 \) and \( D_2 \) that differ only in whether Alice has contributed her data (the red circle) or not. Using DP to protect Alice, it is required that the resulting outcome of an analysis should be approximately the same independently of Alice’s contribution. To achieve this, the DP mechanism adds carefully tuned random noise to the outcome, which results in a randomized outcome that is close to the actual value without DP [74]. In Figure 1, the distributions (red: with Alice, blue: without Alice) represent the probability of a certain outcome of the statistics. The factor \( e^\varepsilon \) bounds the ratio between these probability distributions, i.e., how close the randomized outcomes of \( D_1 \) and \( D_2 \) have to be and, thus, determines the maximum influence of a single individual’s data point on the outcome. Since the probabilities of \( D_1 \) and \( D_2 \) are similar (only differing by the ratio that is set by \( \varepsilon \)), the data consumer cannot be certain whether \( D_1 \) or \( D_2 \) was used to generate the final outcome and, therefore, whether Alice is included in the data set or not.

The trade-off between privacy protection and accuracy of the result can be illustrated by the two edge cases: for \( \varepsilon = 0 \), the definition requires the outcome distribution to be identical for both data sets \( P[f(D_1)] = P[f(D_2)] \). Therefore, Alice’s data does not influence the outcome. DP requires this property not just for Alice but for any two data sets that differ by one entry. By iteratively applying the same argument for other individuals, any results obtained could also be obtained from the empty data set. Although this leads to perfect privacy, the outcome of \( f \) is completely useless for any statistical purpose [74]. However, even with the complete randomization due to \( \varepsilon = 0 \), a data consumer can still randomly guess whether Alice contributed towards the original result or not, and, thus, a trivial re-identification risk remains.

At the other extreme, i.e., if \( \varepsilon \) is very large, DP is already achieved with very little noise. The outcome of DP is very close to the actual value, almost as if no DP was used. In this case, DP provides a high accuracy but low privacy.

In order to balance between the accuracy of the outcome and privacy protection, we need to choose a reasonable small \( \varepsilon > 0 \) that results in an outcome of \( f \) which is “almost” independent of each single individual, at the same time, still accurately representing the body of the data as a whole. With such a choice for \( \varepsilon \), DP essentially “hides” the data of each individual and its influence on the outcome.

When we carefully choose a value for \( \varepsilon \), DP can be a valuable tool for protecting individuals’ privacy. By contrast, with a less careful choice of \( \varepsilon \), the effectiveness of the privacy protection offered by DP can be reduced or even completely suppressed. For this reason, the value used for \( \varepsilon \) and the privacy guarantee resulting from it is essential information for evaluating the privacy protection offered by DP. Thus, providing this information to individuals is essential for making privacy decisions. In the next section, we give an overview of existing approaches to communicate DP.

2.2 Privacy Risk Communication

Privacy decisions have become ubiquitous, not only due to legal requirements (e.g., General Data Protection Regulation). Their intention is to empower individuals during privacy decisions by requiring that data owners “shall implement appropriate technical and organizational measures” as part of “data protection by design and by default” [62]. However, research has shown, for example, that simply removing user identifiers is insufficient for anonymizing data adequately [47] and especially concerning mobility data, 4 randomly known data-points are enough to identify 95% of mobility traces uniquely [20].

DP and other PPTs are good candidates to match the technical requirements of data protection and they are already used in practice [1, 21, 26]. Especially DP provides a unique opportunity of a mathematically proven guarantee in form of an upper bound to the privacy risk, of sensitive data being exposed. When using DP this guarantee is an essential piece of information for the sharing decision process. Therefore, easy to understand quantitative privacy risk notifications are needed.

The design and impact of privacy notices in general is a widely researched field of work (e.g., [16, 27, 53]). While previous research has produced many recommendation on how to communicate qualitative privacy information and even qualitative privacy risk [51], an effective communication of the quantitative privacy risk information of DP is still missing.
Previous attempts to explain DP in particular have led to mixed feedback. The US Census Bureau, for example, experienced that people did not understand the functioning of DP\(^1\) or even the necessity for using DP\(^2\). This problem has inspired first academic investigations into the best strategy for communicating DP to laypeople, which we summarize in the following.

Even earlier, Bullek et al.\(^{[15]}\) conducted a study to explain DP using the metaphor of randomized response technique [70]. Instead of modifying the data after collection, the randomized response technique instructs participants to randomize their answers directly according to a Wheel of Fortune-like visualization: the participants either answer truthfully or answer “Yes” or “No” as indicated on the randomized visualization regardless of their true answer. The distribution of “answer truthfully”, “answer Yes”, and “answer No” on the wheel can be adjusted to tune the privacy protection in accordance with the DP parameter \(\varepsilon\). Based on this procedure, participants were able to correctly rank different DP options according to their privacy guarantee. These results suggest that providing individuals with an intuition on privacy risks is possible and beneficial for the individual’s decision-making. However, this positive result has two caveats. Firstly, Bullek et al. only measured the ability to compare multiple DP settings. We focus our research on the more common situation, where individuals have to decide whether a single privacy guarantee is strong enough to share personal data. Secondly, as noted by Bullek et al., participants perceived the procedure as lying when prompted to simply answer “Yes” or “No.” This influenced individuals in the study towards choosing a less private option against their privacy preference to avoid lying. By focusing on the privacy guarantees instead of the mechanism that achieves the guarantee, this effect should be avoided.

Xiong et al.\(^{[77]}\) compared different real-world notifications about DP. They examined, for example, whether certain descriptions of DP are easier to understand or whether a positive framing of DP, i.e., its privacy guarantee, leads to different results compared to a negative framing of DP, i.e., the privacy risk. The researchers recognized that several unintentional differences in descriptions (e.g., the mention of a known institution as data collector) had a stronger influence on individual’s decision-making than the influence of the differences in description. The addition of quantitative risk descriptions with their objective nature might have the potential of a less volatile effect.

Cummings et al.\(^{[19]}\) investigate the effect of the theme (e.g., techniques, trust or risk) of various real-world notifications about DP on individuals’ willingness to share data. They did not find any significant effects on the decision-making, postulating that none of the descriptions provided a meaningful enough mental model of DP. Nonetheless, they identified the descriptions that centered around privacy risk as the most promising, because this description conveyed the most accurate privacy expectations to individuals. As the privacy guarantees of DP describe a bound on the privacy risk, this result also sound promising for our proposed communication.

In summary, the existing research on how to explain DP to laypeople supports the claim that communicating DP is challenging. Even developers and technically skilled people face the challenge of understanding DP in all its details \(^{[25, 31]}\).

Therefore, we take a different approach and consider what transparency means in the context of DP. We argue, consistent with the results by Cummings et al.\(^{[19]}\), transparency with respect to DP should focus on the privacy risk and specifically on the offered guarantees as described by the DP definition. In particular, our approach is to translate the guarantee into a quantitative risk communication format derived from the actual value chosen for the privacy budget \(\varepsilon\). As the choice of \(\varepsilon\) heavily influences the resulting protection, only with this transparency DP can be a useful tool in supporting individuals in privacy decision situation.

However, communicating quantitative privacy risk has been explored little in privacy research. We rely, therefore, on risk communication in the medical domain which is a well-developed field of research. In the following, we provide an overview of existing medical risk communication formats.

2.3 Medical Risk Communication

The goal of medical risk communication is to enable an individual (here a patient), to make an informed decision about a particular treatment. There are a number of meta-studies that offer recommendations on how to communicate medical risks effectively which

---

1. https://nypost.com/2021/08/26/elsie-eiler-surprised-when-one-person-nebraskatown-doubles-in-census, accessed 2021-10-06.
2. https://www.nytimes.com/2018/12/05/upshot/to-reduce-privacy-risks-the-census-plans-to-report-less-accurate-data.html, accessed 2021-10-06.
we used for reviewing existing medical risk communication formats [10, 12, 28, 66, 68]. Based on our analysis, we decided to focus on communicating aleatory uncertainty, i.e., uncertainty about future outcomes, instead of epistemic uncertainty, i.e., uncertainty due to the lack of information. Furthermore, we do not consider 2-variate risk formats (e.g., natural frequencies) because they do not support our goal of explaining DP. From the meta-studies, we derived six basic risk communication formats which are often used in medical risk communications: percentages, simple frequencies, fractions, numbers needed to treat, 1-in-x and odds. Some of these formats are generally discouraged, since they can be misunderstood easily (1-in-x, numbers needed to treat), they are inflexible in use or inaccurate (fractions), or not well researched (odds).

The two remaining base formats percentages and simple frequencies, are both recommended in different settings. In addition, we identified several format variations that can be added to these basic formats: comparison to known risks, comparison to peers, comparison to status quo and outcome framing. In the following we briefly introduce the two suitable base formats and the variations.

Percentages are a widely used and well-known risk communication format (e.g., "23.8 \% of patients experience dizziness") [8, 12, 66]. Travena et al. [66] recommend this format especially when comparing multiple risk-related options. McDowell et al. [42] only caution about presenting risks lower than 1 \% in this format. For all other values, rounding to the nearest whole number is generally recommended.

Simple frequencies [12, 28, 42, 66] presents the number (numerator) of cases affected by the event compared to a meaningful baseline (denominator) (e.g., "13 out of 350 patients experience dizziness"). This format is commonly accepted and understood by laypeople. However, multiple options exist for the numerator and the denominator to represent the same risk and especially the choice of a meaningful denominator might affect the efficiency. There are multiple complementary recommendations on how to make an appropriate choice: Several of the meta-studies strongly recommend using consistent denominators to minimize the mental load on the individuals when comparing different risks in this format. Multiples or even powers of 10 as denominator have been shown to be beneficial [14]. Small denominators are recommended because they are better understood and easier to remember. Concerning the numerator, there is strong evidence that 1 as a numerator skews the risk perception of individuals [55] and, therefore, higher values are recommended.

In addition the meta-studies discuss the following variations, which can be added to the basic formats. Firstly, a valuable information is the comparison to the risk of a similar event [12, 28] (e.g., "The risk of side effects due to radiography is 15 \%, only slightly higher than the 13 \% risk due to other common radiation sources"). Secondly, the risk of an event can also be compared to the risk of the same event for peers [12] (e.g., "The risk of suffering a heart attack is 15 \%, which is higher than the 13 \% of the average US American"). Finally, the risk of an outcome with an action can also be compared against the risk of the outcome without the action, i.e., the status quo [28, 66]. This is often used to explain the reduction of risk due to a treatment (e.g., "Treatment A leads to a survival chance of 60 \% compared to a survival chance of 40 \% without the treatment").

Another common variation is the outcome framing. To avoid bias due to presenting the positive or the negative effect ("5 \% of participants have side effects" or "95 \% do not have side effects") several meta-studies (e.g., [57, 58]) recommend presenting the probabilities of both events together. "5 \% have side effects, 95 \% do not."

Overall, results of medical risk communication research show that the success of an approach depends on a multitude of factors including situational factors (e.g., decision between two options versus the decision for or against a single option, the range of the presented risk) and personal factors (e.g., experience with statistics). When transferring risk communication formats to the domain of privacy risks, we, therefore, need to re-evaluate the formats in the new context.

3 COMMUNICATING PRIVACY GUARANTEES

In Section 2.3 we identified risk communication formats from the medical domain. These formats are well researched in the medical domain and, therefore, good candidates to present privacy risks in quantitative privacy risk notifications, which communicating the privacy guarantees of DP.

As a first step into such quantitative privacy risk notifications this study investigates the following two research questions:

RQ1 How do established quantitative risk communication formats from the medical domain perform in terms of subjective and objective understanding of privacy risk, compared to existing qualitative privacy risk notifications?

RQ2 How is the effectiveness of privacy risk communication influenced by personal attributes of individuals, such as gender, statistical numeracy and privacy aptitude?

In Section 3.1, we discuss suitable risk communication formats from the medical domain, in Section 3.2 we explain how the guarantees provided by DP can be presented in risk communication formats and, in Section 3.3, we describe how the formats were incorporated into privacy risk notifications.

3.1 Selecting Risk Communication Formats

The decision situation in the medical domain is similar to the privacy domain: patients have to decide whether to undergo a treatment which changes the probability of an adverse outcome. The change in probability of the adverse outcome is specified in a quantitative risk format but the result of the treatment remains uncertain. Therefore, risk communication formats used in the medical domain can be used to represent privacy risks.

As discussed, two formats are recommended as best practice formats in different medical contexts, the percentage format and simple frequency format. In the following, we discuss the adaptation of the formats into the privacy domain and also consider the recommended variations.

The percentage formats are well-known and, therefore suitable for presenting privacy guarantees. An example of presenting a privacy guarantee in this format is: “When using DP, your risk of..."
being identified in the dataset is at most 52 %.” As recommended, we avoid decimal point by rounding to the nearest whole number. Since DP calculates worst-case upper bounds on the privacy risk of all individuals, the resulting privacy guarantees are never below 1 %, circumventing the known issues of the percentage format in this range.

The format simple frequencies is also suitable to represent privacy risk. An example that applies this format in the privacy domain is “When using DP, at most, 26 out of 50 participants are identified in the dataset.” Weighing the recommendations, we use multiples of 10 over powers of 10 in our representations to avoid unnecessary high denominators or inaccuracy due to rounding. Furthermore, we prioritize avoiding 1 as numerator over small denominators due to the strong evidence against its understandability.

In contrast to the medical domain, time frames are not relevant in DP privacy guarantees as all information is considered to be known from the date of publication. The individual is either identifiable immediately or remains anonymous.

Concerning the variations identified in Section 2.3 we consider outcome framing and comparison to status quo, since they relate well with DP. In both base formats, percentages and simple frequencies, we usually present the probability of the negative outcome, i.e., the risk of being identified when sharing the personal data. The variation comparison to status quo adds the ground-probability of being identified before sharing data. This variation is naturally related to DP, since DP compares by definition the probabilities in the two described situations. In a second variation, outcome framing, we additionally present the counter-probability, i.e., the probability that the re-identification does not occur. This variation is especially well suited, in the context of DP, since the mathematical definition of DP is formulated in terms of a privacy guarantee rather than a risk.

The other two comparison variations cannot be translated easily to the privacy domain. The concept of risk of an average person, as used in the variation comparison to peers, is difficult to translate into the privacy domain, as according to Bhatia et al. [11] privacy risk is highly subjective and can only be measured by unacceptably severe surveillance. Furthermore, for the variation comparison to similar event, the outcome of a possible privacy leak cannot be easily compared to any other well-known risky event because data for such a comparison are not available.

In summary, we selected two base formats: the percentages format and the simple frequencies format. We can use each base-format as-is or with either of the two variations comparison to status quo or outcome framing. Thus, we arrive at six (2x3) possible quantitative privacy risk communication formats.

Next we discuss possible approaches for translating a privacy budget $\varepsilon$ into one of the identified risk communication formats.

### 3.2 Representing DP guarantees

While a lot of work has addressed the mathematical mechanisms of DP, there is less research on how to interpret $\varepsilon$ as a common risk [31, 46, 67].

One possible approach for such a translation is provided by the model of Lee and Clifton [38]. They rephrase $\varepsilon$ as a probability of identifying any particular individual as having contributed to the result of the analysis depending on two additional factors: the number of individuals in the data set and the maximal impact one individual could theoretically have on the outcome of the query. Mehner et al. [43] simplified this model by considering the worst-case bound for the additional parameters. Thus, their notion of global privacy risk $P$ provides a translation of $\varepsilon$ into the desired percentage or frequencies format:

$$P = \frac{1}{1 + e^{-\varepsilon}}.$$

The advantage of the model by Mehner et al., in contrast to Lee and Clifton [38] and other models, is that the obtained values are sound privacy guarantees, independent of any parameters, which are unknown at the time of data collection. Therefore, this model is a good fit for the goal of this study.

### 3.3 Privacy risk notifications

Using the DP model, we can calculate the DP guarantee as a worst-case privacy risk and represent the resulting values using the six identified risk communication formats discussed in Section 3.1. These six formats are used to specify six novel quantitative privacy risk notifications, which are then evaluated. In addition, as notification BaseLine we use the best performing qualitative privacy risk notification from Xiong et al. (cf. notification “DP Imp.” in [77]). Thus, overall, we consider the seven privacy risk notifications in Figure 2.

Each privacy risk notification contains the same general DP information (extracted from the BaseLine notification), information about the privacy risk (in qualitative form or in one of the two quantitative base formats) and potentially additional information dependent on the variation. The wording used to incorporate the base formats was based on similar medical risk notifications compiled by Bansback et al. [10] and developed in multiple iterations with experts and laypeople on DP. We aimed for a compromise between mathematical correctness of the notification and understandability for general users.
The choice of $\epsilon$ is critical for the level of protection offered by DP. Following the recommendation by Dwork [24], we used $\epsilon = 0.1$ for strong privacy guarantees, matching the wording of the BaseLine notification. With this value and the model by Mehner et al. [43], we calculated the privacy guarantees $P = 0.525$, which results in the notifications “With differential privacy, at most 26 out of 50 | 52% of statistics will reveal whether you visited a location.”

These values also determine the counter-probability of not being identified, as used in the variation outcome framing. In the variation comparison to status quo we use the worst-case value $P_{\text{guess}} = 0.5$ as outlined by Mehner et al. This value describes the probability of correctly guessing the presence of an individual in the data set without information from the data set, which corresponds to the status quo, before sharing one’s data. We present this probability as a percentage “50%” and a frequency “25 out of 50”.

It is worth noting that all calculated risks present realistic values according to the chosen privacy budget.

4 EXPERIMENTAL STUDY

Based on the privacy risk notifications, we designed and conducted an experimental study on Amazon Mechanical Turk (MTurk) in order to evaluate the suitability of our approach and understand the differences in effectiveness between the selected formats. We decided for a between-subject study to avoid potential learning effects.

4.1 Study Design

Our study setup was divided into four main phases (cf. Figure 3): Participants (1) were introduced to a scenario with a focus on mobility data; (2) were shown one of seven conditions at random, i.e., a privacy risk notification, and seeing the notification (2a) were asked to self-rate their understanding of the notification; (2b) were asked to answer objective true/false questions about the privacy risks of DP; (3) completed standard tests about their privacy aptitude and their statistical numeracy, and, finally, (4) were asked to indicate their self-identified gender. We describe the design consideration for each phase in the remainder of this section.

We defined a fictional ridesharing scenario to capture the complexity of privacy risks. Ridesharing services show increasing popularity and, therefore, have a high likelihood of being relatable by the participants. The scenario describes that, after an app update, user data from the ridesharing app will now be shared with a local authority to improve the urban infrastructure. For simplicity, the scenario does not include a privacy decision. Our research goal is concerned with establishing the requirements for an informed decision, rather than facilitating the decision itself. The scenario text was refined in a thinking-out-loud session with English native speakers. We deliberately chose the scenario centered around location data for two reasons: firstly, collecting location data is omnipresent, and we assume highly relatable to study participants. Secondly, we hope that our research can benefit environmentally-aware urban mobility development by sharing location data in a privacy-preserving manner.

In the context of this scenario, the participants were exposed to one of the privacy risk notifications, presented as a typical pop-up design on a stylized smartphone screen. We then evaluated the participants’ subjective and objective understanding of the privacy guarantees: the subjective understanding measures how well the participants think they understood the notification, whereas objective understanding measures how well DP was actually understood based on the notifications. For the subjective understanding we asked participants to rate (S1) their perceived general understanding of the notification, (S2) their perceived ability to assess individual privacy risks based on the notification and (S3) whether they felt the descriptions provided all necessary information each on a 7-point Likert scale. Concerning the objective understanding, we asked participants to judge true/false statements about DP. We compared the privacy risk of DP with the risk (O1) when sharing the unprotected data and (O2) when sharing no data at all. In addition, we asked about two basic principles of DP: (O3) the assumption, that a basic privacy risk exists without sharing more data and (O4) that the distribution of the privacy risk cannot be manipulated by the data consumer, to reveal one specific target location. Where possible, we randomized the correct answer (true/false) of these questions by varying the qualifier (less/more, higher/lower).

In RQ2 we consider a number of context factors, which have shown an influence on risk communication in other contexts. An important influence for the effectiveness of risk communication is the ability of the individual to handle statistics and probability information in general. This attribute is commonly known as (statistic) numeracy. In our study we measure the numeracy of our participants with the Berlin Numeracy Test [17] in its adaptive format. Research also shows an influence of gender, especially on the perception of positive or negative framing of risk communication [32]. We, therefore, ask participants to self-report their gender following recommendations by Spiel et al. [59]. Finally, privacy aptitude is an obvious potential influence. We adapted the Internet Users’ Information Privacy Concerns test (IUIPC) [39] to evaluate the privacy aptitude of participants. This test divides the privacy aptitude into the three dimensions “control,”16 “awareness”7 and “collection”8 each with three to four questions. Since the participants did not have control over the data in our scenario, we measured only for “awareness” and “collection,” which are both directly relevant to our scenario.

We included comprehension check questions after the scenario and attention check questions as part of the objective understanding, and the privacy aptitude test to exclude inattentive participants. At the beginning of the study, we informed the participants that MTurk IDs and IP addresses would be stored for quality control purposes, but would be deleted once quality was verified. We also explained the collection of gender information to the participants and its purpose in the study.

---

1 How important is control of personal data and autonomy?
2 How important is the knowledge concerning how personal data is collected, processed and used?
3 How concerned is the participant with data collection?
4.2 Participant recruitment and study execution

Although MTurk has known disadvantages concerning nuanced representation of users [5], its advantage of reaching a wide audience in this early investigation into the suitability of the risk communication formats outweighs these concerns.

We restricted participants to MTurk workers with at least 95 \% approval rating and at least 1,000 approved assignments. As customary, we restricted participants to US-residents, to increase the likelihood of an adequate English proficiency. We determined a working time of about 12 minutes, with careful reading of all material, in a small pre-test run with a few selected participants (recruited from our research group as well as project partners). Based on a conservative working time of 15 minutes, each participant was paid $3.00 ($12.00/hour). This is significantly higher than the local minimum wage (~$10.01/hour at the time of recruitment) and about equal to the average hourly wage on MTurk (~$12.05/hour). However, we allowed up to 40 minutes per participant to avoid automatic timeouts of participants who were slower for any reason. The MTurk assignments were distributed throughout the day between reasonable US working hours.

To address potential ethical issues we followed common recommendations [5, 40, 45], and further informed participants before the survey of our data policies: we only used identifying data (MTurk IDs and IPs) for payment approval and removed them before the statistical analysis.

5 STATISTICAL ANALYSIS

In this section, we present the procedure and the results of our statistical analyses\(^\text{10}\). We start with the general preprocessing and insights into the collected data set, then report on the analysis regarding RQ1 and RQ2. Throughout the analysis, we use a significance-level of $\sigma=0.05$ and adjust the results of the post-hoc t-tests with Bonferroni-corrections\(^\text{11}\). We are confident, that we eliminated most unreasonable outlier with these test and, therefore, decided against further elimination of outliers based on values. The cleaned data set contains 343 submissions, which are approximately equally distributed between the seven groups (cf. Table 1)\(^\text{14}\). Average working time was around 8:40 minutes.

The three answers about subjective understanding were combined into its arithmetic mean. The resulting values lie between 1 and 7, with high values denoting that the participant felt confident in their understanding. For the objective understanding we marked each correct answer with 1 point. The summed score for all four answers constitutes the objective understanding value between 0-4.

Out of the 343 participants, 117 as female (F) and 223 reported themselves to be male (M). With 65.0 \% male participants the ratio is slightly higher than similar studies [41]. We combined the three participants reporting being non-binary or self-described into a third category (X).

According to the BNT, we categorized submissions into four numeracy groups\(^\text{15}\). The results were highly skewed towards low numeracy (lowestN: 175, lowN: 47, highN: 39, highestN: 82). The test is designed to result in four equal groups with university-affiliated participants. However, the lower rate of MTurk workers with a university degree (around 49 \% [45, 69]) plausibly explains the difference to the BNT evaluation study [17].

We used the results from the privacy aptitude questionnaire, to compute separate arithmetic means for “awareness” (three questions) and “collection” (four questions). We then calculated the arithmetic mean of these two values to get an aggregated value for the privacy aptitude\(^\text{16}\) on a scale between 1 and 7. Additionally, we sorted submissions into privacy aptitude groups (cf. Section 5.3.1). By visual inspection of the relationship privacy aptitude to subjective understanding (Figure 4), we identified rising values in subjective understanding up to a privacy aptitude value of 5.3, a plateau up to 6.2, and falling subjective understanding afterwards. According to those three different areas we categorized the privacy aptitude into three groups lowP, mediumP, and highP.

In the final data set each entry contains the following values: condition group (1-7), subjective understanding [1-7], objective understanding [0-4], Numeracy {lowestN, lowN, highN, highestN}, privacy aptitude group {lowP, mediumP, highP} as well as the accurate privacy aptitude value [1-7] and gender {F,M,X}.

---

\(^\text{10}\)The cleaned data set and the analysis script is available here: https://osf.io/2dkty/?view_only=8577de0996bd488fc8dc17b04cb799

\(^\text{11}\)The Berlin Numeracy Test (BNT)\(^\text{[17]}\) provides a flow diagram to classify participants into one of four levels

\(^\text{14}\)For example, if a participant answered (3,3,7) on awareness and (3,6,2,5) on collection, we calculated awareness $=(3+3+7)/3=5$, collection $=(3+6+2+5)/4=4$, and finally privacyaptitude $=(5+4)/2=4.5$.

\(^\text{15}\)The resulting rejection rate of 24\% is not uncommon.
Concerning objective understanding (cf. Table 1), we identify the best performing notifications to be FreqStatQuo (1.98) and PercPosNeg (1.88), both performing better than the BaseLine notification (1.79). However, the difference is not shown to be significant by the standard ANOVA test \((p = 0.3053)\). Consequently, we cannot claim a significantly better performance of the two mentioned quantitative risk communication formats.

### 5.3 The Effects of Individuals’ Characteristics

#### 5.3.1 Statistical Model and Requirements.

In Section 4 we hypothesized that numeracy, privacy aptitude or gender might have an influence on the effectiveness of the privacy risk notifications, i.e., on the subjective and objective understanding.

To evaluate our hypothesis, we constructed a linear regression model each for subjective and objective understanding. We include seven independent predicates for each model: (1) condition group, (2) numeracy, (3) privacy aptitude group, (4) gender, and three interaction terms (5) group: numeracy, (6) group: privacy aptitude group and (7) group: gender. The dependent variable is the objective or respectively subjective understanding. Since the relationship between privacy aptitude and objective/subjective understanding is not linear, we use the privacy aptitude groups in the linear regression model instead. All other requirements were checked without issues.

As base categories for the linear model, we used the BaseLine condition, the highest numeracy level highest\(N\), the medium privacy aptitude medium\(P\), and the gender male 19.

#### 5.3.2 Results.

In the linear regression model for the subjective understanding, we observe significant effects for PercStatQuo \((p = 0.0255)\) and for the interaction between the condition PercStatQuo and the numeracy low\(N\) \((p = 0.0248)\).

Figure 5a shows the subjective understanding depending on condition group and numeracy. The diagram reveals that the PercStatQuo notification (the right most light blue group) has one of the lowest results in the sub-groups lowest\(N\), high\(N\) and high\(N\)\(P\) (labeled "lowest\(N\).PercStatQuo", "high\(N\).PercStatQuo", "high\(N\).PercStatQuo")\(P\)), but the highest medium performance in the subgroup low\(N\) (labeled "low\(N\).PercStatQuo").

Furthermore, the four numeracy groups of the PercStatQuo notification show a distinct pattern; the subjective understanding is moderate in the lowest numeracy level lowest\(N\), then it continues with a higher value in the low numeracy level low\(N\), followed by a low value for the high numeracy level high\(N\), and finally again with a moderate value in the highest numeracy level highest\(N\). This pattern can also be seen for the BaseLine, PercStatQuo and PercPure notifications.

---

17The corresponding non-parametric Kruskal-Wallis test is not significant, either \((p = 0.2483)\).

18These base categories define the group to which differences are calculated. We aim to discover differences caused by the risk communication formats. Thus, we choose the BaseLine condition for comparison. Since we expected higher differences in lower levels of numeracy, we selected the highest level of numeracy to discover differences to the lower levels. For privacy aptitude we did not have an intuition and, therefore, chose the medium level to discover effects of the levels low and high simultaneously. The choice for gender base category is irrelevant with only two considered values. We kept the value male for gender as defaulted to by the linear regression model.

19The table of all regression coefficients is provided as part of the supplementary material.
the influence of numeracy on the effectiveness of the formats and finally we summarize our thoughts on the influence of the other attributes.

6.1 Suitability of Risk Communication Format

We enhanced existing privacy risk notifications from existing usable security and privacy research (cf. [19, 77]) with quantitative risk communication formats from medical research (cf. Section 2). The resulting six quantitative privacy risk notifications provide in a transparent way realistic quantitative guarantees, i.e., mathematically rigorous worst-case estimates (cf. Section 3.2), about the likelihood that certain sensitive data are revealed to unauthorized parties.

Our study results provide first encouragement that the inclusion of quantitative risk information does not significantly harm the objective understanding and, thus, the informed decision-making of the participants. Instead, the quantitative notifications perform similar to and, in some cases, slightly better (yet not significantly) than existing best practice by Xiong et al. [77]. This is an important result, because it encourages further research into the use of quantitative privacy risk notifications to transparently inform individuals about possible privacy risks when sharing their data.

Our results also support the insights by Bullek et al. [15], who highlight the usefulness of transparency when using DP mechanisms for understandability. Such understandability (cf [2, 3]) allows for two potential modes for DP. Trusted authorities could recommend suitable values for \( \varepsilon \) (global DP) and transparently communicate the resulting guarantees. Eventually, understanding of the guarantees would enable users to individually select [35] suitable values for \( \varepsilon \) according to their privacy needs (local DP).

Concerning differences between the different risk communication formats, in our study we could not identify significant effects in general. Therefore, we cannot recommend any risk communication format over others. Having selected only the best practice formats from the medical domain, the difference between the formats might be too small to be detected. Additionally, in order to maintain comparability with the qualitative risk communication, we did not measure the understanding of the level of privacy protection, which should be communicated by the quantitative risk values. Future research will have to evaluate whether the additional quantitative privacy risk information contributes to a higher overall understanding of the involved privacy risk and which of the six quantitative risk communication formats perform best in this regard.

6.2 The Influence on Subjective and Objective Understanding

The results regarding the overall effect of the notifications on subjective understanding show that the BaseLine notification outperforms all quantitative notifications according to the mean value, and even significantly outperforms the condition FreqStatQuo.

However, the results regarding objective understanding show a different behavior: no significant difference could be found, but two of the quantitative notifications, i.e., FreqStatQuo and PercPosNeg, outperform the BaseLine notification according to the mean value.

---

21The table of all regression coefficients is provided as part of the supplementary material.
Consequently, we assume that all conditions had indeed similar effects on objective understanding.

However, the results regarding objective understanding were generally surprisingly low for all notifications: with four true/false questions the expected value of random guesses is two correct answers\(^{22}\). The overall arithmetic mean on objective understanding in our sample set was 1.74, with only the \textit{FreqPosNeg} notification achieving a result (1.98) close to this expected value (cf. Table 1). This suggests that the level of information provided even in the currently used notifications is insufficient for understanding the properties of DP. We discuss this aspect in more detail in future research (cf. Section 7.2).

The two results on objective and subjective understanding seem to be contrary to each other. The quantitative notifications perform similarly to the \textit{BaseLine} notification on objective understanding, but less well on subjective understanding. This difference is especially pronounced regarding the notification \textit{FreqStatQuo}, which performs significantly lower than \textit{BaseLine} on subjective understanding, yet, achieves the highest overall value for objective understanding. One possible explanation is that the \textit{BaseLine} condition creates a feeling of confidence, while the quantitative notification cause a more cautious reaction. However, this confidence caused by the \textit{BaseLine} notification (5.6 mean out of 7) is not backed up by the objective understanding measured (1.74 mean out of 4), as discussed above. Thus, we consider the high subjective understanding of the \textit{BaseLine} notification as overconfidence.

Similar to research by Siroti et al. \cite{55}, future research should find privacy risk notifications, for which the subjective understanding corresponds best with the objective understanding. Such a format would support informed consent, as it enables individuals to know when to seek more information, as opposed to \textit{dark patterns} (e.g., \cite{56}), which try to maximize this overconfidence. The framework presented by Bhatia et al. \cite{11} could be a promising starting point for measuring the \textit{correct} perception of risk provided by the different privacy risk notifications.

### 6.3 The Effect of Numeracy

We discovered an effect of numeracy on subjective understanding significant for notification \textit{PercStatQuo}, and less pronounced for notifications \textit{FreqStatQuo}, \textit{FreqPure} and \textit{BaseLine}. The shape of the pattern (cf. Figure \ref{fig:pattern}) resembles the Dunning-Kruger effect \cite{36}; low-skilled people also lack the meta-skill of accurately assessing their skill and tend to overestimate their skill level. Higher-skilled people, in contrast, can better assess their shortcomings and tend to underestimate their skill level. The same shape can be seen in our data.

According to Agrawal et al. \cite{4} we already have a society, a technocracy, where the power to manage their privacy is reserved predominantly for people with experience in technology. The effect in our data only contributes further to this imbalance: individuals with the lowest and low numeracy are more confident of having understood the notification without grasping all the potentially severe consequences, leading to them sharing data more often and, therefore, being more vulnerable. As this effect is also found to a

\(^{22}\)Randomly guessing 4 questions with a probability of 50% each would result in an expected success rate of 2 answers

degree in the conventional \textit{BaseLine} notification this indicates a need for action.

With this knowledge, we see two different routes for counteracting the technocracy: for some of the notifications we tested the effect was less prominent. Using these would give individuals a more accurate sense of their own understanding and enable them to seek more information before an important decision. Alternatively, knowing this effect, privacy notifications could be tailored to the numeracy of individuals. Especially before more complex and important sharing decisions, the level of numeracy of an individual can be captured, for example, by a short questionnaire. Individuals with low numeracy could then be cautioned while affirming individuals with higher numeracy, counteracting the confidence effect. Personalization of risk communication has already been recommended in other contexts \cite{64} and so called nudges have been shown to successfully influence individuals’ privacy decisions. However, a main concern with these approaches is that influence of a decision is inherently biased. Our result provides an objective measure to evaluate the influence, i.e., to counteract the existing effect, which disadvantages individuals with low numeracy.

### 6.4 The Effect of Privacy Aptitude and Gender

Concerning Privacy Aptitude, we found that participants with low privacy aptitude in the group \textit{FreqPure} performed less well in objective understanding. Since this is an isolated finding we cannot speculate about the reason for this performance. If this effect can be confirmed in future studies, the use of the frequencies format in privacy notification without additional information could not be recommended.

We could not find any other significant effects of Privacy Aptitude or Gender. However, the IUIPC test used in this study only records a self-reported privacy aptitude. Research about the accuracy of self-reported privacy aptitude has produced mixed results \cite{34, 71}. Using enhanced methods to record a more accurate privacy aptitude in future studies could yield additional significant results.

## 7 LIMITATIONS AND FUTURE WORK ON RISK COMMUNICATION FORMATS

This study provides novel insights into using quantitative privacy risk notifications to inform laypeople about privacy risks when sharing their data. However, our study has a number of limitations; at the same time, our results are hypothesis-generating and open up a number of promising research directions that may help to shape the way in which we enable people to share their data in a private manner. We next discuss both limitations and future research.

### 7.1 Situating Privacy Risk Communication in the Real-World Context

MTurk enabled us to quickly recruit a large sample within the budgetary requirements. However, we are aware of existing caveats caused by such a study design, especially regarding ethical and quality assurance concerns, and concerns regarding the representativeness of its users.
To mitigate these concerns as much as possible, we considered existing academic (e.g., [52, 65]) and practical recommendations for using MTurk. We informed, for example, all participants of any personal data collected before they accepted the task, provided means of aborting and deleting the collected data throughout the task, and handled submissions anonymously where possible.

Secondly, using crowd workers as our participants, we have to consider MTurk’s incentive structure: MTurk awards fixed payments per task. This encourages spending as little time as possible on reading the study texts and thinking through all possible repercussions of a decision (cf. [23]). We employed comprehension and attention checks to minimize the effect of this disadvantage and to ensure results comparable to conventional participation recruitment procedures (cf. [5]).

Finally, we discussed the appropriateness of MTurk users for our research question. Recent surveys (e.g., [63]) revealed that MTurk Workers are reasonably representative of the general population in the US (population validity). Thus we believe our results explore the design space of risk communication formats when employing DP in a representative way. However, transferring these insights into other cultural settings (e.g., the European context) requires additional validation.

Furthermore, we acknowledge the participants’ awareness of the study situation: Literature has shown, that privacy behavior is highly situated and responses concerned with hypothetical personal data might differ from actual personal data. In response we will follow up this initial exploration study with participatory design workshops (e.g., [13, 73]) with citizens in the context of urban mobility. We will work closely with an institute for transport research that collects mobility data via smartphones to confirm and deepen the results in a realistic context. Such participatory design workshops will allow us to extend the variables considered, for example, how these descriptions relate to users’ concrete sharing decisions, similar to work by Wu et al. [75].

7.2 Exploring Alternative Risk Communication Formats

Our findings show that quantitative privacy risk notifications perform comparably to a description that explains DP without quantitative information (BaseLine condition derived from Xiong et al. [77]) in objective understanding. However, as noted before, the overall results in objective understanding were very low, which indicates general difficulties understanding the properties of DP. We defined our quantitative privacy risk notifications with similar content and approximately the same length as the BaseLine notification to ensure comparability. In future work, alternative privacy risk notifications (in terms of their content, length or their visual representation) might help to increase both the objective and subjective understanding of individuals and even avoid the possibly existing overconfidence effect. There has already been promising research done on visual aids in the field of medical risk communication and, concerning privacy, Xiong et al. [76] investigated the adaptation of a “Fact Box” 24 to explain DP. However, the results of their study are not yet available. In our study we deliberately focused on quantitative questions to explore the general feasibility of quantitative notifications. However, to evaluate the requirements for alternative notifications in future studies, the results from open-text questions will be especially valuable.

We use our mathematically rigorous worst-case estimate regarding the likelihood that certain sensitive data are revealed to unauthorized parties. We understand that the condition chosen as BaseLine also rises another kind of risk i.e., the risk of data theft, which is not reflected in our quantitative notifications. Additionally, we limited our privacy risk notifications to one value for \( \varepsilon \) only, as this is the default situation in current privacy decision settings. Future work could offer multiple values for \( \varepsilon \) within the privacy decision setting. Bullek et al. [15] have already provided promising results for such a setting. They showed that explanations of DP can convey the relationship between the privacy protections of several DP options, however, their explanations have not been evaluated towards the absolute privacy risks of each option. Explanations that can achieve both goals, i.e., conveying the relation between different options and the absolute privacy protection of each option, would enable laypeople to select the most appropriate \( \varepsilon \) value according to their personal privacy preferences in a specific situation. Such a scenario could benefit from additional results in the medical domain, where the risks of different drugs often have to be compared to each other. We discovered in our literature review (cf. Section 2.3) that certain risk communication formats are preferable in single risk communication, but others are recommended for comparing multiple risks. We, therefore, expect that the results of this study are not easily transferable to a setting when choosing between multiple \( \varepsilon \) values. Future work is necessary for evaluating the benefits of such approaches.

Finally, our findings suggest that especially numeracy influence the effectiveness of privacy risk notifications in certain situations. A possible direction for future research might be to design tailored privacy risk notifications according to the individual’s characteristics and, thus, equally support informed decision throughout the whole population. Such research can build on existing research in the area of explanations (e.g., [44]). However, the need for better explanations has already been highlighted within an interview study by Agrawal et al. [4], in which the participants suggested that explanations should be even part of a standardization process and included within software libraries. Such a standardization process is not restricted to the technical domain but might impact policy making as well, which we discuss next.

7.3 Towards Contextual Anonymity

Our research is driven by the vision to enable individuals to make informed decisions when sharing their data. An informed layperson should understand the consequences of sharing personal data. Thus, we used results from medical risk communications together with a mathematically rigorous model of DP for the translation from the parameter \( \varepsilon \) into one numeric privacy guarantee (cf. Section 3.2).

The major advantage of this approach is that the privacy notification we displayed shows accurate, universally true bounds on the privacy risk rather than vague or averaged risk values, which disadvantage marginal groups. However, by applying the interpretative privacy
risk model of Mehner et al. \cite{43}, we rely on a worst-case estimate to derive a risk value. In more specific situations, where factors such as sample size or sensitivity of a query are known, other models might provide less conservative bounds, which might not sound as alarming as the worst-case estimates used here. Future work should, in addition, consider such other models, for example, Naldi and D’Acquisto \cite{46} or Hsu et al. \cite{31}.

Furthermore, we hope that our research might provide the basis for making DP understandable and applicable to a broader range of people for making the abstract disadvantage of sharing the data, i.e., the privacy risk, tangible. Current research focuses on showing which data is being exchanged for what purposes, for example, based on icons (e.g., \cite{18}), or visualizations (e.g., \cite{30}). This research assumes that people are sharing their data, but what they are sharing is the risk of someone unauthorized accessing their personal information. We hope we can inform discussions in the policy domain by translating the abstract, intangible concept of “data sharing” into something more concrete. We envision the use of DP as a tool for contextual anonymity that can be realized in two directions. Firstly, instead of communicating the purpose of data sharing, such as required by the General Data Protection Regulation, protection measures should be defined for specific areas of applications (e.g., detailed location tracing versus on level of public transport stops). When applying DP, such protection measures define minimum requirements for the privacy budget \( \varepsilon \) (cf. Section 2.1). Secondly, based on such legally defined protection measures, laypeople can decide based on a privacy risk notification whether, in a concrete situation, the privacy protection proposed is sufficient for them and if not, lower the value of \( \varepsilon \) to increase noise and, consequently, to increase their protection.

8 CONCLUSION

Novel privacy-preserving technologies, like DP, protect the privacy of individuals, when used correctly. To support the sharing decisions, however, the privacy guarantees provided by these technologies need to be communicated to individuals in an understandable way (e.g., \cite{4}).

Our research takes the first step in this direction by suggesting the use of risk communication formats informed by results from the medical domain. We combined these risk communication formats with a mathematically rigorous DP model, which translates the less intuitive privacy budget \( \varepsilon \) into an understandable value for the privacy guarantee. With this combination, we achieve accurate quantitative privacy risk notifications, which not only inform about the use of DP, but also communicate the level of protection for the individual’s privacy.

We evaluated the understandability of our proposed notifications in a crowd-sourced study against a well-performing conventional DP notification and found that our quantitative privacy risk notifications perform similarly in objective understanding. That makes the risk communication formats a promising research area. In subjective understanding, however, the quantitative privacy risk notifications lag behind the conventional qualitative notification. Future research will have to show whether additional risk communication tools, such as visual aids, can compensate this lack of confidence and increase the overall understandability. Furthermore, we discovered that especially numeric can influence the subjective understanding, including with the conventional DP communication.

In future studies we want to benefit people in two ways: With a better understanding about supporting the sharing decision, using understandable and even tailored privacy risk notifications, we hope to enable individuals to accurately weigh the privacy risk against the benefit of sharing data. We hope that this encourages individuals, who are currently discouraged by the difficult to understand consequences, to share their data in a privacy-preserving way for legitimate causes. Furthermore, we hope we can inform policy makers for a better “data security by design”. With improved communication on the privacy guarantees, we envision official minimal recommendations for the privacy budget \( \varepsilon \) for different specific data collections contexts.

9 ACKNOWLEDGEMENTS

We thank the reviewers for their insightful comments and the statistical consulting team (fu:stat) at Freie Universität Berlin for their help with the statistical analysis. This work is supported by the Federal Ministry of Education and Research, 01UV2090B ("Transdisciplinary exploration of privacy-aware donation of movement data for sustainable urban mobility, sub-project: human-centered data security (freeMove)").

REFERENCES

\[1\] John M. Abowd. 2018. The U.S. Census Bureau Adopts Differential Privacy. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD ’18). ACM, London, United Kingdom, 2867. https://doi.org/10.1145/3219819.3226070

\[2\] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. Science 347, 6221 (Jan. 2015), 509–514. https://doi.org/10.1126/science.aaa1465

\[3\] Idris Adjjerid, Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2013. Slegtethical of privacy: framing, disclosures, and the limits of transparency. In Proceedings of the Ninth Symposium on Usable Privacy and Security - SOUPS ’13. ACM Press, Newcastle, United Kingdom, 1. https://doi.org/10.1145/2501604.2501613

\[4\] Nitim Agrawal, Reuben Binnis, Max Van Kleek, Kim Laine, and Nigel Shadbolt. 2021. Exploring Design and Governance Challenges in the Development of Privacy-Preserving Computation. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. ACM, Yokohama Japan, 1–13. https://doi.org/10.1145/3411764.3445677

\[5\] Herman Agunis, Isabel Villamor, and Ravi S. Ramani. 2021. MTurk Research: Review and Recommendations. Journal of Management 47, 4 (April 2021), 823–837. https://doi.org/10.1177/0149206320969787

\[6\] Najwa Alghandhi, Nora Alrajabeh, and Shizoo Al-Megren. 2019. Crowd Behavior Analysis using Snap Map: A Preliminary Study on the Grand Holy Mosque in Mecca. In Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing (CSCW ’19). ACM, New York, NY, USA, 137–141. https://doi.org/10.1145/3311957.3359473

\[7\] Hazim Almuhimedi, Florian Schaum, Norman Sadeh, Idris Adjjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. 2015. Your Location has been Shared 5,398 Times! A Field Study on Mobile App Privacy Nudging. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI ’15). ACM, New York, NY, USA, 787–796. https://doi.org/10.1145/2702123.2702210

\[8\] Jessica Ancker. 2020. The COVID-19 Pandemic and the Power of Numbers. Numeracy 13, 2 (May 2020), Article 2. https://doi.org/10.5038/1936-4660.13.2.1558

\[9\] Rebecca Balebako, Jaeyeon Jung, Wei Lu, Lorrie Faith Cranor, and Carolyn Nguyen. 2013. “Little brothers watching you”: raising awareness of data leaks on smartphones. In Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUPS ’13). ACM, New York, NY, USA, 12:1–12:11. https://doi.org/10.1145/2501640.2501636

\[10\] Nick Bansback, Madeleine Bell, Luke Spooner, Alysa Pompeo, Paul K. J. Han, and Mark Harrison. 2017. Communicating Uncertainty in Benefits and Harms: A Review of Patient Decision Support Interventions. The Patient 10, 3 (June 2017), 311–319. https://doi.org/10.1007/s40271-016-0210-z
