Communicating the Privacy-Utility Trade-off: Supporting Informed Data Donation with Privacy Decision Interfaces for Differential Privacy

DANIEL FRANZEN*, Freie Universität Berlin, Germany
CLAUDIA MÜLLER-BIRN*, Freie Universität Berlin, Germany
ODETTE WEGWARTH, Charité - Universitätsmedizin Berlin, Germany and Max Planck Institute for Human Development, Germany

Data collections, such as those from citizen science projects, can provide valuable scientific insights or help the public to make decisions based on real demand. At the same time, the collected data might cause privacy risks for their volunteers, for example, by revealing sensitive information. Similar but less apparent trade-offs exist for data collected while using social media or other internet-based services. One approach to addressing these privacy risks might be to anonymize the data, for example, by using Differential Privacy (DP). DP allows for tuning and, consequently, communicating the trade-off between the data contributors' privacy and the resulting data utility for insights. However, there is little research that explores how to communicate the existing trade-off to users. We contribute to closing this research gap by designing interactive elements and visualizations that specifically support people's understanding of this privacy-utility trade-off. We evaluated our user interfaces in a user study (N=378). Our results show that a combination of graphical risk visualization and interactive risk exploration best supports the informed decision, i.e., the privacy decision is consistent with users' privacy concerns. Additionally, we found that personal attributes, such as numeracy, and the need for cognition, significantly influence the decision behavior and the privacy usability of privacy decision interfaces. In our recommendations, we encourage data collectors, such as citizen science project coordinators, to communicate existing privacy risks to their volunteers since such communication does not impact donation rates. From a design perspective, we emphasize the complexity of the decision situation and the resulting need to design with usability for all population groups in mind. We hope that our study will inspire further research from the human-computer interaction community that will unlock the full potential of DP for a broad audience and ultimately contribute to a societal understanding of acceptable privacy losses in specific data contexts.

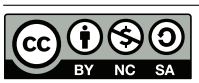
CCS Concepts: • **Human-centered computing** → *Empirical studies in interaction design*; • **Security and privacy** → **Human and societal aspects of security and privacy**; **Usability in security and privacy**.

Additional Key Words and Phrases: risk communication, differential privacy, informed choice, citizen science

ACM Reference Format:
Daniel Franzen, Claudia Müller-Birn, and Odette Wegwarth. 2024. Communicating the Privacy-Utility Trade-off: Supporting Informed Data Donation with Privacy Decision Interfaces for Differential Privacy. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 32 (April 2024), 56 pages. https://doi.org/10.1145/3637309

*Both authors contributed equally to this research.

Authors' addresses: Daniel Franzen, daniel.franzen@fu-berlin.de, Freie Universität Berlin, Germany; Claudia Müller-Birn, clmb@inf.fu-berlin.de, Freie Universität Berlin, Germany; Odette Wegwarth, odette.wegwarth@charite.de, Charité - Universitätsmedizin Berlin, Germany and Max Planck Institute for Human Development, Germany.



ACM 2573-0142/2024/4-ART32
https://doi.org/10.1145/3637309

1 INTRODUCTION

In citizen science, members of the public participate in scientific research in various ways, for example, through intellectual effort, by sharing their knowledge, or by using tools to collect data (see, e.g., [61]). Collecting data in the context of citizen science projects has recently gained renewed appeal, for example, in the context of biodiversity [70], mobility [58], or healthcare [21], due to the ease of use of sensors in smartphones or smartwatches. Other organizations already collect similar data as a byproduct of their primary service. This data could also enable society to make more demand-driven decisions. However, while such data collections may create valuable scientific insights or help the public to make decisions based on actual demand, it may also cause several privacy risks, such as revealing sensitive information (e.g., [55, 67, 75]). One such example of valuable data are location data. They are essential for implementing and verifying demand-driven improvements to urban infrastructure. However, they can also relate to an individual's home address or to visits to a sensitive location. Research has shown that two known spatio-temporal data points are sufficient to uniquely identify 50 % of the contributors in a typical dataset [19], and knowing someone's home address, even at the level of zip codes, is sufficient to uniquely identify 87 % of the U.S. population if gender and date of birth can be linked to the data [71]. This privacy risk highlights the need for consent before collected data can be used for any public purposes. However, we argue that data collectors and data consumers have an obligation to act responsibly in protecting volunteers' data (see also [7]), even when consent is explicitly requested. Although existing studies show that people are generally suspicious of the collection of personal data via mobile devices [8], so far, their concerns seem to play a minor role in data collection. However, in order to maintain the trust in such data collections and, furthermore, increase the willingness of volunteers to donate data for data-driven societal improvements, privacy protection should be given the same importance as other features of the respective services and projects. Building on this, Bowser et al. [6] emphasize the need for more flexible data donation practices that allow volunteers to change their data donation preferences depending on their respective context. The authors recommend that volunteers be given sufficient time to understand the data collection and privacy practices, a process that must be supported by appropriate interfaces that enable volunteers to make their donation decision based on the context of the collection.

Our research has been inspired by this recommendation since we were challenged by a similar situation. We are conducting research on data donation in the context of mobility. One of our primary concerns is how we can support volunteers to make informed decisions about sharing their data by communicating the privacy risk associated with their donation in an understandable way. There are several common approaches to reducing privacy risk, such as reducing spatial resolution, generating synthetic data, or introducing privacy zones. All of these approaches have an inherent trade-off: they increase privacy protection at the expense of data accuracy and, therefore, data utility. At the same time, these methods are not able to quantify this trade-off, which means that it is not possible to communicate the privacy risk associated with the donation to potential contributors. Research has shown (see [7]) that contributors need to understand the trade-off between privacy and utility in order to decide whether to donate their data. In order to make an informed decision, contributors need to weigh the benefits to a "greater good" against the potential privacy risk according to their privacy needs.

A potential solution to this situation is Differential Privacy (DP) [24]. The general idea of DP is to modify the data probabilistically, for example, by adding statistical noise [24]; the more the data is modified, the lower the utility of the data, but the higher the protection of the contributors. Unlike the methods mentioned above, DP, by definition, quantifies the privacy-utility trade-off using its parameter ε, which specifies how much the data must be modified. Using this parameter,

DP can be tuned sensitively to the context of the data (e.g., prioritizing data utility in high-impact data collections for the greater good or prioritizing more privacy protection for more sensitive data types) and to the specific privacy needs of the target group of contributors (e.g., increased privacy protection for vulnerable groups). According to Dwork et al. [25], "[DP] will allow society to reap the benefits of big data while protecting individual and collective interests in privacy." However, in addition to the many technical challenges in implementing DP, we have had to recognize that there is still little knowledge in the field of human-computer interaction on how to effectively facilitate the communication of the privacy-utility trade-off when donating data. Thus, our research aims to answer the question: *How can we enable laypeople to incorporate privacy risk information into data donation decisions for the benefit of an informed decision.*

To address this question, we designed privacy decision interface elements that aim to support laypeople, such as volunteers of citizen science projects, in understanding the conditions under which they are sharing their data when using DP. Building on existing design recommendations for privacy notification and choice interfaces (e.g., [27, 64, 69]), we developed user interface elements that allow potential data donors to explore the risk values visually and interactively during a donation decision. Since access to users in actual data donations is valuable, we decided to first evaluate these novel interface elements in an experimental study to ensure that our designs for representing the trade-off are meaningful before approaching actual volunteers. With our research, we make the following contributions:

- We propose different design concepts (textual, visual, interactive) for communicating privacy risk in the context of DP and demonstrate these concepts in a privacy decision interface.
- We provide empirical evidence gathered from an experimental study that shows that privacy risk communication for DP incorporating risk visualization and interactive risk exploration support informed privacy decision-making.
- From these insights, we derive recommendations on how to use privacy risk communication for DP in the context of data donations, for example, in citizen science projects and beyond.

We provide a review of literature on DP and other related areas in Section 2, which we use to propose our overall research methodology in Section 3. Based on that, we describe the study setup in Section 4 and the results in Section 5. We discuss our results in Section 6 and derive a number of recommendations in Section 7. We discuss the limitations of our study design (see Section 8) and, finally, provide an outlook for future research in Section 9.

2 RELATED WORK

In the following, after introducing the issues and requirements of privacy risk communication using the example of citizen science projects, we briefly introduce DP and its functionality. We highlight existing research that aims to communicate privacy protection and, specifically, DP to support contributors' privacy decisions. We extend this perspective by reviewing existing design recommendations for privacy notification and choice user interfaces from privacy and security research.

2.1 Privacy in Citizen Science Projects

Citizen science allows interested members of the public to voluntarily participate in research through various means (e.g., [61]). Existing projects range from sharing computing resources to annotating images. Increasingly, citizen science projects are being used to collect data, for example, on biodiversity, mobility, or health (see, e.g., [21, 58, 70]). Research has shown that volunteers have very different motivations for participating in citizen science projects, such as the desire to contribute to "real science," personal interest or personal relevance of a topic, or the satisfaction of

social interactions [46]. Although citizen science projects can provide a valuable service to science and communities, they raise some ethical concerns about privacy. Volunteers often collect data using their smartphones or smartwatches. This can lead to the disclosure of highly sensitive information, such as home addresses, either by explicitly sharing locations or via embedded metadata in shared images. Research has shown that the handling of personal data in citizen science projects is often not transparent. For example, survey results showed that only 10 out of 118 projects (8 %) asked their participants to sign an explicit consent form [63]. However, transparency in citizen science projects is essential as lack thereof decreases willingness to contribute [50]. This is crucial for the projects as the value of a dataset depends on its size.

Although the fact that the protection of volunteers' privacy is important is emphasized in citizen science literature (see, e.g., [42, 57]), what should be protected and how it should be protected in the context of citizen science is generally a less researched area [7]. One exception is the study by Bowser et al. [6], who conducted focus groups with volunteers and semi-structured interviews with citizen science coordinators. Their findings highlight that volunteers (in this case, experienced citizen scientists) understand data sharing as a prerequisite for achieving a "greater good" [6]. In addition, the norms and values of citizen science projects foster a culture that emphasizes the openness of data sharing rather than a discussion of potential privacy concerns. This puts contributors at risk often without an available way to raise concerns.

Existing research in the context of citizen science has been a strong motivation for our research, as it shows that considering the privacy of volunteers' data is important but still an open issue. While we use citizen science as a concrete motivation to improve the support during privacy decisions, similar decisions are, albeit less explicit, also at the heart of many network-enabled services, including various social networks. Therefore, in the next section, we introduce a privacy-preserving technology, i.e., DP, which allows for anonymizing contributors' data and, thus, works towards ensuring their privacy. We furthermore discuss how the privacy protection of DP can be conveyed to laypeople.

2.2 Communicating Differential Privacy

DP is a mathematically defined property of a data analysis method that aims to protect the privacy of each data contributor against a data consumer, in the example the citizen science projects, while preserving as much statistical accuracy as possible [24]. The analysis of a dataset is defined to be differentially private if there is a limit on how much the result of an analysis might change when removing or adding one more data contributor. The core component of DP is the *privacy budget* ε that specifies the maximal influence one contributor could have on the result. Different mechanisms can achieve DP (e.g., [24, 36, 76]) by altering the data in different ways. To illustrate one DP mechanism, we refer to the mechanism proposed by Dwork [24], which adds carefully tuned statistical noise to the data. However, the research presented here focuses on communicating the privacy risk guarantees achieved by DP to laypeople, regardless of the mechanisms used.

The value of ε determines how much the data needs to be altered, for example, how much noise needs to be added. Two edge cases can illustrate the influence of ε on the trade-off between privacy and utility: DP with $\varepsilon = 0$ makes the result of the analysis completely independent of the data donation of any single contributor, therefore, independent of the whole dataset. The outcome is perfectly private but useless for any statistical purpose [79]. At the other extreme, i.e., for very large ε, the DP property is achieved with very little noise, resulting in an exact analysis but meaninglessly low privacy guarantees. Values for ε between these extreme cases allow to tune in as much privacy as necessary while preserving as much utility as possible.

With this flexibility, DP can be fine-tuned to exactly fit the specific privacy needs of the data and collection context of a project. As an additional advantage, DP guarantees the chosen privacy

protection independent of the used anonymization method. While the DP mechanisms might be too complex to be explained to laybpeople, especially in the brevity of a privacy decision situation, the resulting guarantees can be expressed as a privacy risk [51] and, therefore, have the potential to be understood by laypeople without technical knowledge of the underlying mechanism or even the mathematical definition of DP. In this context, we understand the privacy risk as the probability that an adversary, Eve, guesses correctly whether a volunteer, Alice, visited a location. Even without Alice donating her data, Eve could make an educated guess as to whether or not Alice has visited the location, for example, based on the known age distribution at a certain location, or simply by tossing a coin to randomly guess "Yes" or "No". In the latter case, Alice's privacy risk would be 50 %, i.e., 50 % of the coin tosses would present Eve with the correct guess that Alice has visited the location ("Yes") or not ("No"). With the publication of a new data donation, Alice's privacy risk increases since, with the information from the new publication, Eve could take a more educated guess. The protection offered by DP can be thought of as a bound on the maximum increase of the privacy risk. For example, with $\varepsilon = 0.1$ and a privacy risk of 50 % before the donation (i.e., equivalent to tossing a coin), DP guarantees that Alice's new data donation increases her privacy risk by at most 3 % to 53 %.

We compare the privacy risk provided by DP with different anonymization methods, i.e., generalization [72]. Here the individual's data is "hidden" in an equivalency group. That means the sensitive information cannot be linked with the individual and the privacy risk is suggested to be 0 %, until the data can be correlated with a separate and unexpected dataset, at which point the privacy risk becomes 100 %. While the data published using DP by design cannot be de-anonymized, there is still a separate kind of risk involved posed by a leak of the original (i.e., un-anonymized) data being accessed by attackers. As DP is an anonymization strategy, it can not give any guarantees on this kind of risk. Appropriate countermeasures against these need to be enacted, like encryption of data in transit and at rest as well as a digitally and physically fortified IT infrastructure. These countermeasure also decrease the risk of leak of the personal information. However, in contrast to the privacy protection provided by DP, these risks are not quantifiable independently of future events (e.g., the development of stronger computing resources) or human actions (e.g., accidental or malicious exposure of confidential databases). Since this kind of risk has a different quality and is out of reach for the application of DP, in the following we focus on the privacy risk of the indented publication of the DP protected data.

When using DP correctly, once the appropriate value for the parameter ε is chosen, the provided privacy protection is guaranteed mathematically. The attack model, on which DP is based, assumes the attacker already knows all information on all participants except for Alice's information. This strong attacker model makes DP unique in the sense that the protection of DP is not vulnerable to currently unknown attack vectors, such as aggregation with future data collections, but, instead, can be expressed as the maximal privacy risk of one particular data donation. This property presents the otherwise unique opportunity to express the privacy risk of the specific data donation.

However, depending on the choice for ε, the offered privacy protection ranges from very strong to meaningless. Data contributors are not able to judge the level of protection unless the privacy risk resulting from the selected value for ε is communicated in an understandable way.

As DP is already being used in some practical applications (see Microsoft, Apple, Google [22], and the U.S. Census Bureau [40]), there are examples of how DP is presented to potential contributors. Currently, the chosen value of the privacy protection is not communicated in the notification or consent dialogues at all. The value for the parameter ε is often mentioned in the corresponding privacy or data protection policy documents. However, the ε-value is practically meaningless without an intricate knowledge of the complicated DP definition. Consequently, even the few users

who take the time to read the policy documents (see [2]), would not be able to properly judge their privacy risk with the provided information.

Nevertheless, some studies investigate the existing privacy decision interfaces for DP and what effect this communication has on contributors. Xiong et al. [80] compare textual descriptions in the form of privacy notifications inspired by real-world examples. They particularly explore how different aspects of DP, such as the used mechanisms of DP or the sensitivity of the information protected by DP, influence the users' comprehension and decision-making. They find that some aspects, such as trust towards the collecting organisation, have a more substantial influence than the aspects of DP they investigated. Kühtreiber et al. [45] confirm in a replication study the results found by Xiong et al. [80] showing that participants did not fully understand DP. The authors highlight that a new method is needed to communicate the effects of DP. In another study, Xiong et al. [81] investigate the influence of symbolic graphics and animations used in DP notifications. The results show, besides other things, that animations and static illustrations perform equally well, and participants prefer stronger privacy protection in most cases but are more willing to share data for the public good. However, the results also show that users' comprehension of DP is poor. Cummings et al. [18] classify 76 real-world textual descriptions used in DP notifications into six themes, such as "technique", "trust", and "risk". They then compile representative examples for each theme and investigate their effect on the individuals' willingness to share data. They find that participants' privacy expectations were raised by the notification but do not find any significant effects on decision-making. Nonetheless, they identify the descriptions centered around privacy risk as the most promising because they convey the most accurate privacy expectations to individuals. Building on these findings, with our own research we investigate whether the privacy expectation can be further supported for the benefit of an informed decision by not only focusing on the privacy risk but also communicating the specific numerical level of the privacy risk.

All real-world DP communications collected by Cummings et al. and Xiong et al. are focused on informing the contributors about the presence of privacy protection, for example, stating the privacy risk for the contributors as "almost no risk". However, no communication found by either study informs the contributors about the privacy-utility trade-off or chosen ε value. In contrast, in our study, we investigate how the communication of the specific value for the privacy-utility trade-off can support an informed decision.

To date, there are only two studies [11, 28] that incorporate the actual quantitative value for the privacy-utility trade-off. Bullek et al. [11] investigate the potential for communication and understandability of the randomized response technique [76][1]. They find that their chosen wheel-of-fortune metaphor enables laypeople to correctly rank the privacy protection of different settings for DP. The study shows that the actual level of protection can be assessed by individuals properly and support them in their privacy decision-making. However, they also find that contributors do not like "lying" when the randomized response technique instructs them to give a predetermined answer. This highlights the importance of the utility for the contributors but also shows that this specific metaphor influences volunteers' informed choices in an unintended way. Unlike Bullek et al. [11], who investigate the choice between multiple protection levels, current privacy decision interfaces usually only offer a binary choice to contributors: they can either donate data with a given and fixed privacy protection or decline the donation. With our study, we focus on this more common binary decision situation.

[1]Randomized response technique is a mechanism to achieve DP different to the Laplacian noise mechanism proposed by Dwork[24]. With the randomized response technique, DP is achieved by instructing contributors either to provide their real information or to answer with a given option, i.e., they lie, with specific probabilities.

The other study investigating the communication of quantitative privacy-utility trade-off information is by Franzen et al. [28], who extend the privacy decision interfaces compiled by Xiong et al. with privacy risk information. They adapt risk communication formats from the medical domain and presented the less intuitive privacy budget ε as an understandable value for the privacy guarantee. In their study, they evaluate laypeople's objective and subjective understanding of different quantitative privacy risk communication formats. Their results suggest that the qualitative and quantitative formats perform similarly in objective understanding, but quantitative notifications need to provide additional assistance to also support subjective understanding. While their results indicate the feasibility of quantitative privacy notifications, they also emphasize that the quantitative information needs to be incorporated into appropriate user interfaces.

In summary, there is not much research on communicating the quantitative privacy risks provided by DP to laypeople for the benefit of informed decision-making. However, the qualitative discussion of privacy risk in privacy notifications has shown promise [18] and the existing research on quantitative communication shows that laypeople are able to rank different privacy risk values [11]. However, text-only presentations are not able to communicate the privacy risk well enough [28]. We propose that with additional interface elements, the privacy risk information can be presented in an understandable way, thereby facilitating informed consent. In the next section, we review existing insights on how to design privacy choice and notification interfaces.

2.3 Design Recommendation for Privacy Decision Interfaces

Privacy notifications inform people about existing or potential data collections and the use and sharing practices regarding their personal data. Privacy choices (e.g., cookie banners, app permission interfaces) additionally provide people with control over certain aspects of such data sharing practices, for example, what data they want to share. Because the two terms, privacy choice and privacy notification, are often used interchangeably in literature, we introduce the term *privacy decision interface*, to emphasize our focus on the decision-making process. Finally, there are privacy policy documents which accompany a privacy decision interface by providing more comprehensive information about the data collection for the more interested or more concerned users.

From the privacy perspective, the goal of privacy risk communication is not to promote more or less data sharing, but instead is to support data contributors to take an informed decision. Bekker [5] defines an informed decision as a "choice [...] made [...] using relevant information about the advantages and disadvantages of all the possible courses of action, in accord with the individual's beliefs". Applied to the context of DP, the "relevant information" is the privacy risk provided by DP and the "personal beliefs" are represented by the privacy concerns of the volunteers.[2] Consequently, given an informed decision, users with fewer general privacy concerns should be more inclined to donate their data than those with more privacy concerns. According to this definition, the goal is to present the privacy risk provided by DP in a way that allows people to choose according to their privacy concerns. Therefore, a suitable privacy decision interface should result in a correlation between privacy concerns and privacy decision.

One way of understanding decision-making is the so-called *privacy calculus* [17]. It describes a privacy decision essentially as weighing benefits against risks. However, other research (e.g., [38]) showed that the people can, in fact, be influenced by more than just facts, for example, by the way in which the information is presented. Especially the point in time and the context are of

[2]We are aware of the "privacy paradox" [3] that describes the discrepancy between people's privacy concern, i.e., caring about online privacy, and their actual behavior, i.e., sharing personal data without much thought. The effect of the privacy paradox is, however, generally debated with studies both finding and denying its existence and even studies with reversed results [15]. We, therefore, think that our mode of study is appropriate for the stage of our research. Nevertheless, we emphasize the need to reevaluate our results in real-world settings (see Section 8)

utmost importance for people's privacy decisions [68, 78]. Before addressing the question of how such an informed decision can be supported by means of privacy decision interface elements, we summarize insights from research on designing privacy notifications and choices.

Current privacy decision interfaces present static information, use toggles as binary indicators [35] for opt-ins or permissions, or at most visualize privacy with categorical privacy indicators [69]. There exists a lot of research on cookie consent interfaces, which confront people with a very similar decision about sharing their movements on the world-wide-web. These interfaces mostly contain textual information about the purpose and the data consumer of the data collection, usually with toggles to allow a subset of data uses or buttons to enable/disable all cookie tracking [53]. However, even when avoiding the common dark patterns that are used to influence decisions towards consent [32], none of the currently used interfaces are suitable for representing the flexibility and expressiveness of DP's privacy protection.

Schaub et al. [64] carried out a literature review regarding privacy notifications in the context of cookie consents. They identify various design dimensions which influence the effectiveness of privacy notifications such as modality (e.g., text, icons, images) and control (e.g., blocking, non-blocking). Feng et al. [27] focus on privacy choices and extend these dimensions by the type of choice (e.g., binary, multiple) and the functionality (e.g., presentation, feedback). Both pieces of research highlight that each of these dimensions needs to be considered when designing effective privacy notifications. Acquisti et al. [1] extend this perspective with insights from behavioral science. They discuss different interventions that could be used to assist laypeople's privacy-preserving decision-making by ensuring their freedom of choice (so-called soft paternalistic interventions). They differentiate six dimensions: information, presentation, defaults, incentives, reversibility and timing. In the *information* dimension, the researchers describe that the design should provide a realistic perspective of risks, i.e., an interface should increase the user's awareness of existing privacy and security risks. This can be realized by educational elements that support such decision-making. An example of such elements are privacy icons (e.g., [16, 26, 35, 60]). Our research aims to support people in making an informed decision, thus, we build on insights from behavioral research in security and privacy research, but do not consider types of interventions in our research, which try to *nudge* people in one direction. Zimmermann and Renau [82] use visual and textual information in their privacy notifications and conclude that a combination of visual and textual information encourages more secure choices in some cases. These results reaffirm that incorporating visual information might also benefit contributors when taking a privacy decision with DP. Habib et al. [34] investigate cookie consent interfaces by identifying common design choices of such interfaces in the wild and evaluating a set of representative user interfaces against each other. In order to compare these user interfaces, they define attributes that concern the usability of consent interfaces. The insights from their comparison recommend to block the current work flow with the consent interface and to embed relevant cookie options directly in the main consent interface rather than offering a separate settings view.

In related research in medical decision-making, multiple studies find that personal attributes also contribute to differences in decision-making and should, therefore, be considered when designing decision interfaces. Since medical decision situations have many similarities to privacy decision situations, we also briefly consider these results. Studies in the medical field show that, for example, participants with low *Numeracy* need additional support in decision situations, especially when involving numbers [10, 56]. The personality trait *NeedForCognition* is also often related to influences in decisions [77][3]. Therefore, we expect that the performance of privacy decision interfaces might be influenced by people's *NeedForCognition* or *Numeracy*.

[3]We elaborate more on *Numeracy* and *NeedForCognition* in Section 3.2

In summary, current research provides a number of design recommendations for general privacy decision interfaces that we can also consider for DP. We generally differentiate textual, visual and interactive elements that can be used in unison [27, 35, 64, 69]. In the context of DP, existing privacy risk should be represented realistically [1] and, furthermore, account for human differences in decision-making [77, 82]. Thus, people should be provided with possibilities to explore the existing risk based on visual and textual presentations. However, a large number of studies and different recommendations on decision interfaces also show that the decision situation is complex and dependent on a variety of factors. Many different choices in the design can have undesired consequences. Therefore, the influence of a design on the data contributors needs to be carefully evaluated. These insights informed our concept of privacy risk communication when using DP, which we describe in the following section.

3 PRIVACY RISK COMMUNICATION FOR DIFFERENTIAL PRIVACY

The advantages of DP, namely the enabling of precise and context-sensitive tuning of privacy protection and the ability to communicate the remaining privacy risk without the need to understand the anonymization mechanism, provide a unique potential to support a truly informed decision. However, it is still unclear how the resulting privacy risk can be effectively communicated to contributors in such a way that they can incorporate this information into their decision making, as stated in our research question. In the following section, we formulate our hypotheses on how this can be achieved, introduce the measurements to evaluate these hypotheses and, finally, propose the design of the communication elements within privacy decision interfaces.

3.1 Hypotheses

In Section 2.2, we saw that privacy risk information is promising for supporting informed decisions but is not yet incorporated in real-world privacy decision communication. We also discussed that text-only interfaces are not suitable for communicating the risk information sufficiently. Therefore, we hypothesize that visual and interactive elements in the decision interface, when used correctly, enable people to better understand the privacy risk and, in turn, to make more informed decisions based on their individual privacy needs. Since privacy decisions are complicated in nature, we divide our research question into the following six hypotheses.

From Section 2.2 we know that the privacy risk information can benefit the informed decision of contributors, but these benefits depend on the presentation of the information in the interface. Therefore, with our first two hypotheses, we want to evaluate how well the chosen interface elements perform in enabling contributors to incorporate the privacy risk information into their privacy decision. We assume that one prerequisite for understanding the privacy risk is that the contributors perceive and retain the quantitative information presented in the decision interfaces. Therefore, our first hypothesis examines whether participants recall the quantitative information. However, we are aware that even though people might recall the risk, it does not necessarily mean that they take this information into account during their privacy decision. Therefore, we directly focus on the informed decision in our second hypothesis.

H-1: Including quantitative risk information into decision interfaces using suitable interface elements enables users to judge the risk associated with a donation decision better.

H-2: Including quantitative risk information into decision interfaces using suitable interface elements enables users to make an informed privacy decision, i.e., a decision according to their privacy concerns.

In contrast, adding quantitative risk information might have negative effects. First, previous research (see Section 2.2) has shown that adding quantitative information might lead to negative

effects in terms of user experience. Second, making potential contributors aware of potential risks might inadvertently decrease their willingness to donate. We investigate these two potential disadvantages with our second set of hypotheses:

H-3: Quantitative risk information, as part of privacy decision communication, decreases privacy usability.

H-4: Informing users about the quantitative risk of donating data decreases the donation rate.

Finally, we discussed in Section 2.3 that privacy decisions depend on many different contextual factors and in Section 2.1 that the contributors in citizen science projects have various backgrounds and motivations. Therefore, we want to evaluate the proposed privacy decision communication in relation to individual personal attributes, which have been linked to decision-making in different contexts. This leads to our last set of hypotheses.

H-3': Personal attributes (*NeedForCognition*, *Numeracy* and/or *GeneralDecisionMakingStyle*) influence the effect of the quantitative risk information on privacy usability.

H-4': Personal attributes (*NeedForCognition*, *Numeracy* and/or *GeneralDecisionMakingStyle*) influence the effect of the quantitative risk information on the donation rate.

Next, we will discuss how we measure the concepts relevant to the hypotheses.

3.2 Measurements

Within our hypotheses, we refer to three different aspects of the decision situation: *Privacy decision-making concepts* measure the decision and its influences, *privacy usability concepts* measure how participants perceived the interface, and *personal attributes* measure characteristics of the participants which potentially influence their decisions.[4] All measures are summarized in Table 1, and we discuss them in detail next.

3.2.1 Privacy Decision-Making. According to the definition in Section 2.2, we measure informed consent as a statistical correlation between privacy concerns and the donation decision. This way of measuring informed decisions, in particular, also eliminates any general effect an interface design might have on the willingness to donate, as such effects would act on all participants seeing an interface equally, independent of their level of privacy concern. The decisions of our study participants are recorded as a binary variable (*DecisionOutcome*), for or against data donation. The *Internet Users' Information Privacy Concerns* test (IUIPC) [49] is used to measure the participants' privacy concerns and is recorded in the measure *PrivacyConcern*. If the correlation between these two measurements, *PrivacyConcern* and *DecisionOutcome*, is statistically significant ($\alpha <= 5\%$), with higher privacy concerns correlating with lower willingness to donate, we say the respective interface supports informed decision-making.

Additionally, for H-1 we measured two more concepts related to the decision: *DecisionRecall* and *RiskEstimation*. For *DecisionRecall* we ask the participants after the privacy decision whether they decided for or against donation. We postulate this measure as an indicator of an informed choice since we assume informed choices to be more memorable. Finally, for *RiskEstimation*, we ask participants to recall the privacy risk (between 0 % and 100 %) related to the donation shown in the interface to measure the memorability and understandability of the risk presentation. Since not all of the considered privacy decision interfaces provide the quantitative risk information equally, this measure has to be adapted in the different conditions in two ways: First, the baseline condition does not display the quantitative risk information. In this condition, this measure serves as a risk

[4]Several of the measurements use 7-point Likert scales between "1 - strongly disagree" and "7 - strongly agree." We specifically chose to represent each answer option with its numerical representation (1-7) in addition to its textual representation ("strongly disagree" ... "strongly agree") to induce a sense of scale between the answer options.

Measure	Type (unit)	Description
Privacy Decision-Making		
DecisionOutcome	binary (0,1)	whether users share their data or not
PrivacyConcern	numeric (1-7)	how concerned users are about privacy
DecisionRecall	binary (true,false)	How well users remember their choice
RiskEstimation	numeric (%)	how users estimate their privacy risk when donating
Privacy Usability		
Ability	numeric (1-7)	users can accomplish a particular privacy goal
Awareness	numeric (1-7)	users are aware of privacy choices
Comprehension	numeric (1-7)	users are aware of privacy properties
Need	numeric (1-7)	user's privacy needs are addressed
Sentiment	numeric (1-7)	users are satisfied
PrivacyUsability (overall)	numeric (1-7)	combined from all five usability concepts
Personal Attributes		
GeneralDecisionMakingStyle	groups (Avoidant, Intuitive, Dependent, Rational, Spontaneous)	how people decide in a given situation
NeedForCognition	numeric (1-7)	whether users enjoy complicated decisions
Numeracy	groups (NumLowest, NumLow, NumHigh, NumHighest)	how proficient users are with numbers/statistics

Table 1. Overview of measures used in both experimental studies organized into *Privacy Decision-Making concepts, Privacy Usability concepts* and *Personal Attributes* showing besides the used measures the measurement type (with units) and providing a short description of its meaning.

estimation instead of a risk recall. Second, the privacy risk guaranteed by DP depends on the base risk. We ask participants to recall the privacy risk corresponding to a base risk of 20 % (more details on this decision in Section 3.3.2). This value is used as the initial base risk in the privacy decision interfaces but can be varied by the participants in some interfaces.

3.2.2 Privacy Usability. We use *privacy usability* (see Habib et al. [34] in Section 2.3) to measure the usability of our privacy decision interface elements. We adopt five of the aspects defined by

Habib et al. [34], which are suitable for our context: *Ability*, *Awareness*, *Comprehension*, *Need*, and *Sentiment* (see Table 1).[5] We adapted questions from [34] to our context where possible and added new questions with a similar spirit for the aspects where too few questions were adaptable. Our process of combining the question responses into measurements is explained in Section 4.2.1.

3.2.3 Personal Attributes Influencing Privacy Decisions. The *General Decision-Making Style* captures how individuals usually arrive at their decisions. We follow the test by Scott et al. [66], which considers the styles Avoidant, Dependent, Intuitive, Rational, and Spontaneous. According to the test, we assign each participant to one of these five decision-style groups recorded as *GeneralDecisionMakingStyle*.

Since DP is a complex mechanism, we assume that *Need for Cognition* (NFC) might influence the effects of our interfaces. We employ the abbreviated questionnaire NCS-6, developed by Lins de Holanda Coelho et al. [47], to measure the NFC.[6] We calculated the arithmetic mean of the six answers to obtain one *NeedForCognition* value (1-7) for each participant.

Numeracy was shown in previous studies (see, e.g., [28, 77]) to have a moderating effect on people's understanding of privacy risks. To capture *Numeracy*, we use the Berlin Numeracy Test (BNT), developed by Cokely et al. [14], in its adaptive form.

Fig. 1. Interface Design of the ■ Hybrid condition consisting of ① *qualitative information*, ② *textual information*, ③ *interactive sliders with lower risk (left) and higher risk (right) values*, ④ *icon array visualization with lower risk values (left) and higher risk values (right)*, and ⑤ *donation choice*.

[5]The remaining two aspects, *DecisionReversal* and *Nudging*, are outside the scope of this paper, as explained in Section A.1.4.
[6]The original questionnaire NCS-18 developed by Cacioppo et al. [13] is shown only to have a minor benefit in expressiveness while taking significantly more time and concentration. This is especially important in our study context, as we are employing multiple lengthy standard measurement questionnaires.

3.3 Supporting informed privacy decision with DP

As discussed in Section 2.2, there has not been much research on communicating quantitative privacy risk because the concept is unique to DP. Therefore, novel concepts to communicate the privacy risks are needed. Analogous to other privacy-related kinds of information, privacy risk communication would benefit from a layered approach: the gist of the privacy risk would be communicated within the privacy decision interface, in accordance with the generally brief attention span of users and screen space constraints. More details about the privacy risk can be provided in the privacy policy. Since we expect the communication incorporated into the privacy decision interfaces itself to have a stronger influence on the decision, based on the poor utilization of privacy policies by users ([2]), in this initial study we focus on communicating the privacy risk in privacy decision interfaces. However, we belief that the privacy risk communication could benefit from the additional attention and screen space budget of privacy policy by expanding on the proposed interface elements and providing additional elements to further support the risk understanding.

In this chapter, we show the results of a pre-study, which informed our interface elements. We present the final design and the rationale of the proposed privacy decision interface elements and, finally, discuss how we split them into five conditions for our study.

3.3.1 Pre-Study. As a first step, we conducted a pre-study in which we compared six different textual explanations of DP protection in a between-subject experiment (N=329). The aim of this pre-study was to evaluate whether any textual explanation is preferred but also to gain intuition on what the interfaces lack to meet the privacy needs of potential contributors. The general setup of this study is very similar to the main study described in Section 4. The interfaces, however, are kept simple with a short explanation of the employed privacy protection and a familiar choice between sharing data and not sharing data. To cover a wide range of possible but realistic explanations of DP, we considered the common explanation topics "What?", "How?", and "Why?" [59] and chose one fitting DP explanation for each question from the collection by Cummings et. al [18]. Furthermore, we added one explanation for each topic, which additionally incorporated the quantitative privacy risk information. A more detailed description of the pre-study and its results are provided in Appendix B.

The statistical analysis of the pre-study did not reveal any significant preferences for either of the text versions, neither the qualitative nor the quantitative texts. However, the responses to the free text questions[7] of the privacy usability portion of the study[8] informed the phrasing of the textual content of the interfaces for the main study, motivated the supporting elements, and helped us to refine the questions of the measurement instruments. The resulting privacy decision interface for the main study is shown in Figure 1. The elements of the interfaces will be discussed next.

3.3.2 Privacy Communication Elements. While our goal is to incorporate the novel privacy guarantees provided by DP in privacy communication, we do not aim to change the conditions of the decision situation. Ideally, with more advanced privacy mechanisms (e.g., local DP [37]), more experienced contributors would be given more options to customize the privacy protection individually, similar to the study by Bullek et al. [11]. However, the most common data donation decision is binary: contributors either choose to donate data with a fixed privacy protection level or not to donate data at all. Few more individualized privacy options exist, such as privacy zones, but they

[7]The full dataset is provided at https://osf.io/6u2qx/?view_only=a4c8289c02b140768bf57a25d92f980f.

[8]"What do you remember about the content of the notification you saw on the last screen. Please summarize in your own words?", "Do you feel there was information missing in the notification? If so, what information was missing?" and "Why did you decide to choose this option?".

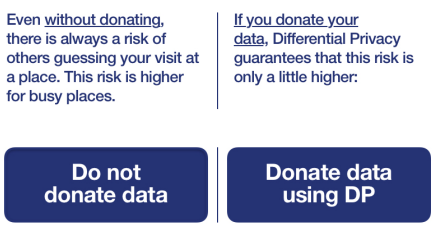
are typically offered after the basic privacy decision has been made, and their effectiveness has been debated [20]. In addition, there are decision situations that allow contributors to choose from a small set of options, such as different purposes (e.g., feature personalization, targeted advertising) or different publication modes (e.g., no publication, aggregated results, open data). However, we argue that these are orthogonal to the basic decision to donate and could, therefore, also benefit from the results on the basic binary decision. Therefore, in order to capture a common and familiar scenario in our study, we chose to examine a binary choice for / against a data donation using DP.

After completing the pre-study with the text-focused decision interfaces for DP, we identified the following requirements for communicating the DP privacy protection: (1) a short introduction to the use of DP, (2) the relation between two privacy risks, (3) supporting a mental model of the DP protection and (4) supporting intuition of risk in general. Refining the interfaces from the pre-test, we arrived at a privacy decision interface with five major parts for the main study as depicted in Figure 1. In the following, we explain each part of the interface in detail.

① *Introduction to DP:*

Within the privacy decision interfaces, our goal is not to communicate the definition nor the technical mechanism of DP. Due to DP's complexity this is not a reasonable task within the constraints of the decision interface. Instead, our goal is to communicate the limit on the privacy risk, which we postulate, can support the informed decision without understanding DP's technical background. However, in the pre-study, we noticed that a basic explanation is needed for contributors to accept DP as privacy protection and to understand the subsequent presentation of privacy risk. However, we did not see any significant difference between the descriptions addressing the questions "What?", "How?" and "Why?". Consequently, we included texts for both "How?" and "Why?" in two separate tabs in the new interface.[9] Either text is short enough to justify the origin of the privacy risk value, however, does not burden the potential contributors with technical details of DP. To avoid bias, we randomized which of the two texts is displayed upfront. Both descriptions are inspired by DP notifications compiled by Cummings et al. [18].

② *Relation between two privacy risks:*



The unique characteristic of the DP property is the relation between the two privacy risks with and without donation. The pre-test has shown that potential contributors need more support to

[9]The content addressing the question "What?" is already included in either of the contents and was therefore omitted as a separate tab.

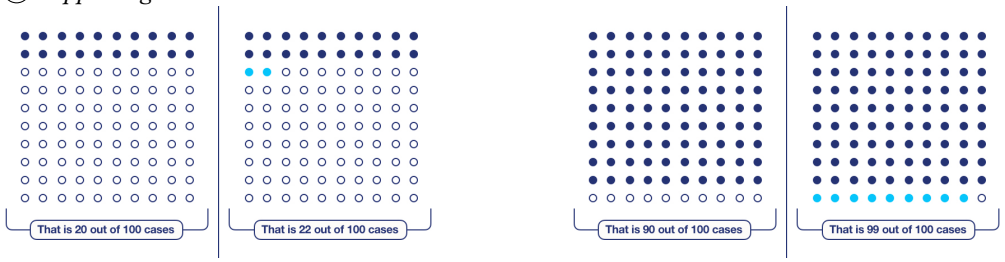
understand this unusual concept. For this reason, we divided the interface vertically; the left depicts the privacy risk without donation, and the right the increased privacy risk after donation. Each side concludes at the bottom with a button against / for donation (see Figure 1, ⑤). This way, the buttons are intuitively placed beneath the corresponding privacy risk.

③ *Supporting a mental model of DP with interactive exploration:*

For the pre-study, we used 50 % as the base risk. As mentioned in Section 2.2, this corresponds to randomly guessing whether a contributor visited a location or not. Using the definition of differential privacy, this results in a 53 % privacy risk when contributing; thus, there is a risk increase of 3 %. We encountered in the open text answers of the pre-study that rather than perceiving the 3 % increase as a strength of DP, participants perceived 53 % as a very high privacy risk. This resonates with a claim by Cummings et al. [18] that mental models of DP need to be better supported. A more accurate base risk than 50 % cannot be calculated without knowing too many details about each individual contributor. We solved the issue of the missing details for the base risk in two steps. As a first step, we provide a slider in the interface, on which the user can select the privacy risk before donation (between 1 and 100 %). When adjusting this slider, the privacy guarantee provided by DP is automatically calculated and displayed. To also allow for the reverse exploration - given a DP guarantee, how high is an acceptable base risk? - the interface displays the DP guarantee on an interactive slider, too. This results in the linked sliders (see Figure 1, ③). This interface allows users to explore the consequences of a data donation in a specific situation.

The interactive slider already solves the issue of choosing a reasonable base risk. However, this still assumes contributors can accurately asses their base risk. We added a metaphor to the sliders to support contributors in this matter. One factor for the base risk is the popularity of a visited location. If most people visit a location, an attacker could also guess Alice was there, too, with a higher probability. Therefore, the popularity of a location is an estimator for the base risk without any prior knowledge, i.e., the more popular a place, the higher the base risk. Using the metaphor of the popularity of a location as the base risk, we separated the sliders into the section between "not busy" and "as busy as it gets". This scale was derived from the "Popular times" interface element used in Google Maps. We assume these categories are known and intuitively understood by contributors. In contrast to Franzen et al. [28], we focus on representing risk values for the sliders in the percentage format, as Franzen et al. did not find any significant difference between percentages and frequencies.

④ *Supporting risk intuition with visualization:*



Furthermore, the pre-study showed that participants have difficulties comprehending the quantitative privacy risk values embedded in the textual explanations. Only a small fraction of the pre-study's participants, for example, were able to recall the risk values shown. Therefore, our privacy decision interface needs to provide a more intuitive understanding of the privacy risks. We

considered various graphical visualizations of risks (e.g., bar graphs, icons, or icon arrays), which have shown a positive effect on risk comprehension in the medical domain (e.g., [29]) and also when communicating privacy risk qualitatively (e.g., [34, 64, 69]).[10] Informed by this research, we decided to use *icon arrays* for explaining privacy risks when using DP. An icon array is a grid of singular icons (circles in our case) symbolizing a reference set. Based on the risk, the number of cases in the reference set affected by the negative consequences of the risk is shown in a different color. We chose 100 circles as our reference set to ensure consistent values between the icons and the percentages in the textual descriptions. The example above shows an icon array depicting on the left side a risk of 20 %. It displays 100 icons, 20 of which are marked differently to denote the number of cases affected.

We chose this visualization for DP, as it intuitively shows the number of expected privacy breaches in the reference set: A risk of 20 % in our DP scenario means that if a person guesses based on the information available whether an individual has visited each of the 100 places considered, this person will be correct for 20 places. We believe that this relationship between the 100 places considered and 20 successful guesses is presented very intuitively in the icon array visualization. Furthermore, DP always compares the chance of a privacy disclosure without donation against the chance of a privacy disclosure with donating the data. This difference can be easily presented in a different color in the icon array visualization, showing not only the absolute risk but also the risk increase due to the new data contribution (see Figure 1, ④, circles in lighter blue).

In order to avoid the rather pessimistically perceived risk of 53 %, we based the privacy risks in the main experiment on a base risk of 20 % instead of 50 % and picked a common value $\varepsilon = 0.1$ for the trade-off parameter. With this ε and the base risk of 20 % the privacy risk after donation is 22 %, according to the DP definition. Thus, the risk increases only a little (by 2 %) when sharing the data with DP. However, the base risk can be adjusted by the participants with the interactive slider to explore higher-risk scenarios. A corresponding example with a higher selected base risk (90 %) is shown on the right side of the exampe above.

3.3.3 Conditions. To evaluate the effect of the interaction ③ and visualization ④ on the decision-making of the users, we constructed five different versions of the interface (see Figure 2): while the interface ■ Hybrid contains both elements, the interactive sliders and the dynamically changing icon arrays, ■ Interactive only contains the interaction and ■ Static contains only the static visualization. As a comparison, we constructed two interfaces with purely textual interfaces (see Figure 2, bottom row): ■ BLQuant, the quantitative baseline, describes the trade-off in text form, and ■ BLQual, a qualitative baseline, contains a simple explanation 'DP protects privacy' only, as common in current notifications regarding DP in the wild. In summary, the five conditions only differ in how the privacy-utility trade-off is depicted: not at all (■ BLQual), text only (■ BLQuant), interactive sliders (■ Interactive), static icon array (■ Static), and interactive icon array (■ Hybrid).

4 STUDY: COMMUNICATING THE TRADE-OFF IN DP-BASED PRIVACY DECISION INTERFACES

This section provides an overview of the study process, the scenario used, and how we conducted the experiment. Our considerations for each part are described briefly in the following. We provide additional details on our study design in Appendix A, the used study materials in Appendix C, and our dataset and analysis script for download[11] in order to ensure the reproducibility.

[10] Privacy risk outside of DP is hard to measure; thus, there is no existing research on representing quantitative privacy risks.

[11] We provided all data as an anonymized repository on the OSF platform: https://osf.io/6u2qx/?view_only=a4c8289c02b140768bf57a25d92f980f.

Fig. 2. Main view of the five privacy decision interfaces

4.1 Study Design

4.1.1 Structure. We evaluated the effects of the designed interfaces in a between-subject online study. The participants interacted with one interface each (shown in Figure 2) in a typical privacy decision context, i.e., they had a choice between donating data protected by DP or not donating data.

The study consists of six parts (see Figure 3). The survey started (1) with an instruction screen informing the participants about the procedure of the study. After that, (2) we measured the privacy concerns of the participants, followed by (3) the scenario in which the privacy decision is to take

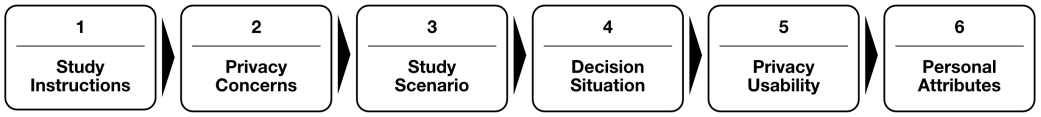


Fig. 3. Overall Study Design for both experimental studies consisting of six parts.

place. After ensuring the participants had understood the scenario sufficiently, we presented (4) the decision situation in the form of an interactive prototype simulating a smartphone app and recorded the decision measurements. We then (5) collected the privacy usability measurements by asking questions about the users' experience with the decision interface and, finally, (6) provided questionnaires to measure the personal attributes (*Numeracy*, *NeedForCognition*, and *GeneralDecisionMakingStyle*)[12] of the participants. The rationale for this study order is the following:

The way we measure informed consent, as a correlation between privacy concerns and privacy decision, considers privacy concerns as an independent variable in the planned statistical analysis. To avoid priming effects in the privacy concerns questionnaire (2), we, therefore, presented the IUIPC test at the start of the study before the scenario (3). An added benefit of this placement is that the participants do not have to answer all of the standardized questionnaires (IUIPC, BNT, NCS-6, General Decision-Making Style) in one block at the end of the test, where attention is assumed to be decreased. We are aware that this order might influence the participants towards higher privacy settings in the decision situation. However, the potential priming effect acts equally on all participants and should, therefore, not influence a potential correlation between privacy concerns and privacy decisions. In the reverse order, i.e., recording privacy concerns after the privacy decision, participants would potentially have answered the privacy concern questionnaire specific to the used scenario of the privacy decision rather than concerning their general level of privacy concern. We judge the probability higher that participants would then have recreated the sentiment of their privacy decision in the privacy concern measurements. Finally, our study is not designed to be deceptive and, therefore, does not rely on hiding the privacy focus.

The measurements of the privacy usability concepts need to be scheduled directly after the decision interface while the experience is still fresh in the minds of the participants, and confusion about what is considered to be part of the interface can be minimized. As the remaining questionnaires are considered covariates, we placed them at the end of the study procedure, as attention to them is the least critical to our hypotheses.

4.1.2 Scenario. We chose a mobility data donation scenario as the context for our study. Mobility data can reveal sensitive information about individuals [41], for example, home address, sexual orientation, information about an individual's health, or political inclinations. However, it also has a legitimate application in the improvement of city infrastructure or the validation of such efforts. Mobility data can be easily collected using smartphones. Mobility providers or location-based social networks have a legitimate primary purpose for collecting such data, for example, to ensure a specific functionality. We assume that the wider population knows about these aspects of mobility data based on presence in the news and the number of apps and services processing mobility data. Furthermore, the application of mobility data for improvements of urban infrastructure serves as a believable contribution to a "good cause", to which volunteers might be willing to contribute their data. To make the scenario imaginable, our study introduces a fictional car-sharing service, CityCar, which offers the option to donate already collected data. We purposefully decided against

[12]For all measurements, we randomized the order of the questions, where possible, to avoid biases.

a citizen science project description to position the scenario more closely to our study participants' more likely everyday reality. We iteratively improved the scenario text in a small pilot study (N=30), where we gathered feedback on how understandable the scenario and the content of the interface are and improved details based on feedback from the pre-study (see Section 3.3.1).

4.1.3 Online Experiment. We recruited the participants for our study on Amazon's Mechanical Turk crowdsourcing platform (MTurk)[13]. To ensure the validity of our study results and ethical treatment of our study participants, informed by research such as [23, 43, 73], we made the following considerations.

We restricted the participants to U.S.-based workers to assume a reasonable proficiency in English[14] and required 1,000 completed HITs[15] and a 95 % approval rate[16] to ensure the base quality of the submissions. Regarding compensating our study participants, we determined the working time of less than 15 minutes by using pre-testers who had not seen the survey before. We then calculated $3 as fair compensation[17]. We decided not to collect any sensitive information other than the answers to the measurement instruments. In particular, we did not collect any demographics since they were not needed to answer our research questions. Crowdsourcing studies have to be designed with particular care since the continuous attention of participants cannot be assumed. Therefore, we asked comprehension questions to ensure careful reading of the scenario text and included attention checks throughout the remainder of the survey. Based on these, we exclude submissions by inattentive participants from the dataset before analysis. All rejected submissions were invited for a retake survey to mitigate the downsides of a rejection for the workers.

Condition	Sample size (decision data / complete)	*RiskEstimation* mean (sd)	*Decision-Outcome* % donating	*Decision-Recall* % correct	Interaction with text tabs % interacted	*Privacy-Concern* [1-7] mean (sd)	*NeedFor-Cognition* [1-7] mean (sd)
BLQual	72/59	53% (29.2)	72.2%	86.1%	16.7%	5.79 (0.68)	4.74 (1.1)
BLQuant	82/71	48% (25.1)	67.1%	91.5%	13.4%	5.88 (0.61)	4.84 (1.13)
Interactive	77/63	50% (29.3)	71.4%	85.7%	18.2%	5.6 (0.77)	4.7 (1.25)
Static	76/61	53% (25.9)	63.2%	73.7%	9.2%	5.87 (0.66)	4.87 (1.14)
Hybrid	71/50	48% (25.6)	71.8%	87.3%	8.5%	5.79 (0.78)	4.43 (1.38)

Overview of the conditions showing (1) the number of recorded decisions / number of recorded privacy usability data per group, (2) the arithmetic mean and standard deviation of *RiskEstimation*, (3) the ratio of participants choosing donation in percent, (4) the ratio of participants remembering their decision, (5) the ratio of participants interacting with the qualitative information, (6-7) the arithmetic mean and standard deviation of the personal attributes.

Table 2. Overview of condition groups and personal attributes.

[13]The submissions were recorded on our local university servers using the software LimeSurvey (https://www.limesurvey.org/).

[14]Most of the MTurk workers are in either the U.S. or India, making the U.S. the only representative choice for surveys in the English language.

[15]Human Intelligence Tasks (HITs): task on MTurk

[16]Approval rate is a rating measure of MTurk, it measures how many submissions of a worker were approved and paid by the requester.

[17]Compensation is based on 12 $/h, which is considerably higher than the local minimum wage of 7.25 $/h and the average wage on MTurk of 10.61 $; see http://faircrowd.work/platform/amazon-mechanical-turk, accessed 27.12.2022.

4.2 Statistical Analysis

Based on a power analysis with a target power of 0.95, we determined a minimal sample size of 302 and collected a total of 581 submissions. We eliminated submissions with missing data (29)[18], by automated attention checks (149) and manual screening of open text answers (25). Due to technical issues[19], we had to exclude another 74 submissions from the analysis of the privacy usability concepts. The resulting dataset has 304 complete submissions (+74 submissions usable for the analysis regarding the privacy decision concepts only). These submissions were randomly assigned to the conditions, which resulted in a roughly equal distribution (see Table 2).

We chose an appropriate statistical test by the following procedure: effects on binary dependent variables are measured using χ^2 tests, and effects on non-binary dependent variables are measured using Kruskal-Wallis tests[20]. We then performed potential post-hoc Wilcox tests. Interaction effects are measured utilizing appropriate (linear or logistical) regression models. These regression models include seven explaining factors: (1) Group, (2) *Numeracy*, (3) *NeedForCognition*, (4) *GeneralDecisionMakingStyle*, as well as interaction terms (5) Group:*Numeracy*, (6) Group:*NeedForCognition*, and (7) Group:*GeneralDecisionMakingStyle*. We use linear regression for numeric target variables and the corresponding logistic regression for binary target variables. In the regression models, we use ▪ BLQuant as the ground condition, as this way, we can detect differences towards ▪ BLQual due to textual differences as well as differences due to the added interface elements in ▪ Interactive, ▪ Static, and ▪ Hybrid. For *Numeracy* we use NumLowest as the ground condition since we expect users with lower numeracy to have a higher disadvantage and, therefore, potentially more apparent differences. For *GeneralDecisionMakingStyle* we chose Avoidant as the first in alphabetical order, since we do not have an intuition on how the decision-making styles influence the different concepts.

All statistical tests are based on a 5 % significance level and we use Bonferroni corrections for all post-hoc tests.

4.2.1 Evaluation of multi-item measurements. We measured the personal attributes *NeedForCognition*, *Numeracy*, and *GeneralDecisionMakingStyle* using evaluated questionnaires. However, since we did not find a fitting evaluated measurement tool to measure privacy usability, we adopted the concepts by Habib et al. [34] to the context of our study. To strengthen the validity of these measurements, we performed a factor analysis on the results of the quantitative responses.

The result of the analysis suggests that four factors (instead of five concepts) are enough to describe the results ($p = 0.18$). Considering weights above 0.5 there is a clear 1-to-1 mapping of the question items into these four factors. Based on this mapping, we identified the theme connecting all questions of each factor as presented in Table 3. In our analysis, we calculated values for these four factors from the responses of each participant according to the weights resulting from the factor analysis[21]. All weights are provided in Section A.1.4 Afterward, we scaled the results of the factors back down to the range [1-7] of the original responses to preserve the comparability between these factors.

We proceeded similarly with the responses for the *PrivacyConcern* and got weights similar to the weights found by Malhotra et al. [49]. After applying the weights we again normalized the values to the familiar range between 1 and 7.

[18] Since the data loss was due to an oversight on our side, we approved and compensated these submissions.
[19] For 74 submissions one survey page with questions on privacy usability was not displayed correctly. These submissions were excluded from privacy usability analysis, but compensated like complete submissions
[20] Shapiro-Wilk tests could not confirm the normality of our Likert items. Therefore, we have decided on non-parametric instead of ANOVA tests, which are already controversial for Likert items.
[21] As customary, we ignore questions with weights < 0.1 in this calculation

Factor	Primary questions	Weight	Theme
PU_Decision	Comprehension1	0.726	How do participants feel about the presentation of the decision?
	Comprehension2	0.635	
	Sentiment2	0.555	
	Ability3	0.700	
	Awareness1	0.517	
PU_Security	Need1	0.815	How well does the interface meet the security / privacy needs of the participant?
	Need2	0.622	
	Comprehension3	0.655	
PU_Data	Ability1	0.839	Does the interface communicate the information about the shared data well enough?
	Ability2	0.640	
	Sentiment1	0.518	
PU_Options	Awareness2	0.843	How satisfied are participants with the available options and the information about them?
	Ability4	0.572	

Table 3. Privacy Usability factors identified in the factor analysis. We used all questions contributing to a factor with a weight of at least 0.5 as primary questions and derived the theme of the factor from the content of these questions.

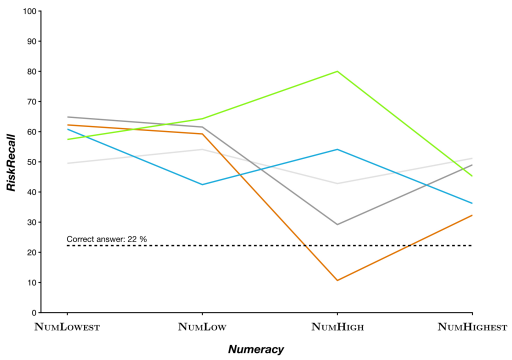
Hyp.	Condition	Moderated by	Effect on	Dir.	p-value	Effect size Cohan's d / (Regr. coeff.)	
H-1	■ Interactive	*Numeracy* NumHigh / NumHighest	*RiskEstimation*	+	0.02 / 0.03	1.66 / 0.84	large
H-2	■ Hybrid	-	Informed decision*	+	0.003	0.88	large
H-2	■ Static	-	*DecisionRecall*	-	0.03	0.38	small
H-3'	■ Interactive	*NeedForCognition*	*PU_Options*	~	0.04	(−0.39)	
H-3'	■ Hybrid	*Numeracy*	*PU_Data*	~	0.05	(−1.21)	
H-3'	-	*Numeracy*	*PU_Security*/ *PU_Options*	~	0.01 / 0.02	(1.49 / 1.34)	
H-4'	■ BLQual	*Numeracy* NumHighest	*DecisionOutcome*	-	0.01	(−2.49)	
H-4'	■ Interactive	*Numeracy* NumHighest	*DecisionOutcome*	-	0.02	(−2.40)	
H-4'	■ Hybrid	*Numeracy* NumHighest	*DecisionOutcome*	-	0.02	(−2.36)	

Overview of the significant effects found in the data, showing (1) the relevant hypothesis, (2) the significant condition, (3) the attribute moderating the effect, (4) the measurement affected, (5) the direction of the effect (+: positive, -: negative, ~: mixed), (6) the p-value, rounded to one significant digit and (7) Cohen's d as measure for effect size, or the respective regression coefficient in brackets
*: Measured as correlation *PrivacyConcern* → *DecisionOutcome*

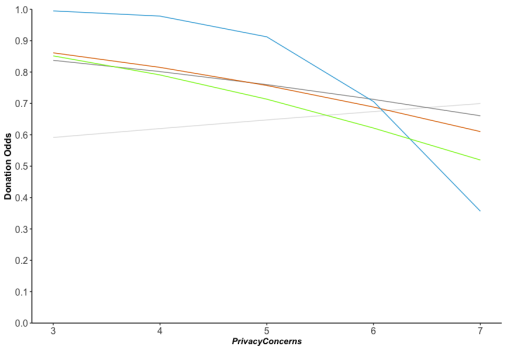
Table 4. Overview of significant effects.

5 RESULTS

In this section, we describe the statistical results of our study. All significant results are summarized in Table 4 and will be presented in detail in the following.



(a) Effect *Numeracy* on *RiskEstimation* per Condition. 22 % is the correct answer.



(b) Odds of donation in relation to *PrivacyConcern* according to the logistic regression model. Our definition of informed decision requires higher *PrivacyConcern* to reduce donation odds.

Fig. 4. Significant effects of personal attributes on *RiskEstimation* and *DecisionOutcome* for conditions ■ BLQual, ■ BLQuant, ■ Interactive, ■ Static, and ■ Hybrid. Even though we have discrete *Numeracy* groups, we deliberately chose line graphs to clarify the relation between the values of each group.

5.1 Hypothesis H-1: Risk Estimation

We did not get significant results ($p = 0.35$) regarding the influence of the interfaces on *RiskEstimation*, but we got several interaction effects in a linear regression: In the condition ■ Interactive users in the two *Numeracy* groups NumHigh ($p = 0.02$) and NumHighest ($p = 0.03$) can recall / estimate the correct privacy risk (22 %) significantly better (see Figure 4a). Cohen's d classifies these effects with 0.84 and 1.66 as *large*. Therefore, we can conclude that ■ Interactive succeeds in communicating the correct risk information to users who have at least some proficiency with statistics.

However, we also see that this effect is not present for ■ Hybrid, which also includes the same risk interaction as ■ Interactive. Here, the risk estimation is close to ■ Interactive only in NumHighest. We assume that the addition of the risk visualization in ■ Hybrid distracts from the interactive element. Also note the surprisingly high *RiskEstimation* value around 80 % for the group NumHigh in the condition ■ Static (Figure 4a), which could be the inverse risk of the given base risk 20 % or the correct answer 22 %. This is also an indicator that the risk visualization was confusing for some reason.

5.2 Hypothesis H-2: Informed Decision

We performed a logistic regression for the influence of *PrivacyConcern* on *DecisionOutcome* per condition to answer H-2. The results of these tests are shown in Table 5.

The decision in ■ Hybrid is indeed significantly influenced by *PrivacyConcern* ($p = 0.003$, Cohan's d: 0.88), and the mean *PrivacyConcern* values show that this influence is in favor of an informed decision: Users with lower *PrivacyConcern* are more likely to donate. As a comparison, the two textual conditions ■ BLQuant and ■ BLQual have the lowest difference between donors and non-donors in mean *PrivacyConcern* (see Table 2). This indicates that the text-only interfaces performed worst in enabling informed decision.

In order to confirm this influence, we also performed a logistic regression on the whole dataset with the explaining factors condition group, *PrivacyConcern*, and the interaction of *PrivacyConcern* with the condition group. The results of this test confirmed the influence of the condition ■ Hybrid

Condition	Mean *PrivacyConcern*			Logistic Regression *PrivacyConcern* → *DecisionOutcome*	Decision time
	No Donation	Donation	Difference	(p-Value)	mean (sd)
BLQual	5.8617796	5.7556765	0.1061031	0.54	16.50s (20.17110)
BLQuant	5.84995602	5.89770555	-0.04774953	0.76	19.99s (21.87422)
Interactive	5.7352862	5.5450807	0.1902055	0.32	21.60s (24.59207)
Static	5.9839865	5.8113672	0.1726193	0.26	24.20s (36.61257)
Hybrid	6.2539857	5.6061481	0.6478376	0.003 **	20.85s (24.47777)

Mean *PrivacyConcern* for decisions pro and contra donation in each condition, logistic regression p-values (*PrivacyConcern* → *DecisionOutcome*) and mean time taken for the decision
$^{**}p \leq 0.01$, $^{*}p \leq 0.05$, $^{+}p \leq 0.1$.

Table 5. Relation *PrivacyConcern* to *DecisionOutcome*.

($p = 0.01$) and the significance of the interaction term of Hybrid with *PrivacyConcern* ($p = 0.01$). The results of the model are visualized in Figure 4b as the odds of participants choosing to donate in relation to their *PrivacyConcern*. The odds of BLQual, Interactive, and Static decrease slightly with higher *PrivacyConcern*; however, only Hybrid has a significant trend with 99 % odds for participants with low *PrivacyConcern* [22], but down to 36 % for participants with high *PrivacyConcern*. This confirms hypothesis H-2 for the condition Hybrid.

An additional interesting finding is the time participants spend on the decision (see last column in Table 5). The mean decision times range from 16.50s to 24.20s. The conditions BLQual and BLQuant require less time than the augmented interfaces. This is to be expected as the augmented interfaces contain more interface elements, which need to be inspected. However, it is remarkable that condition Static required more time (24.2s) in the mean than the two interactive interfaces Interactive and Hybrid (21.6s and 20.85s), even though Hybrid contains more elements than Static. Additionally, Static shows a higher standard deviation (36.6) than the other interfaces (between 20.1 and 24.5). One reason might be that the interactive element of Hybrid and Interactive aids in understanding the visualization quicker for a broader audience.

We further analyzed the measurement *DecisionRecall*, which asked the participants which donation option they had chosen. The χ^2 test of the influence of condition on the *DecisionRecall* was significant ($p = 0.03$). A post hoc test of the residues of this test reveals that the only significant condition is Static ($p = 0.02$), which performed especially poorly regarding *DecisionRecall*, meaning that users in this condition group did not remember their choice as accurately as in other conditions. This is again expected in comparison to the two interactive conditions Interactive and Hybrid, as interaction might make the choice more memorable. However, it is unclear why Static performs worse than BLQual and BLQuant in this measure.

Overall the interface Hybrid is the clear preference for informed decision. The two baseline interfaces BLQuant and BLQual performed especially poorly with regards to enabling participants to incorporate their privacy concerns into the privacy decision, and interface Static took more time to consider without a benefit on the decision outcome or decision recall.

5.3 Hypothesis H-3: Privacy Usability

None of the privacy usability factors (*PU_Decision*, *PU_Security*, *PU_Data*, *PU_Options*) is influenced significantly by the conditions (see Table 7). However, the condition Static has a higher standard deviation in all four factors than the other conditions, similar to our results on decision time. This

[22] We do not have responses with *PrivacyConcern* < 3

Condition	*PU_Decision* mean (sd)	*PU_Security* mean (sd)	*PU_Data* mean (sd)	*PU_Options* mean (sd)
BLQual	5.41 (0.68)	3.54 (1.26)	5.27 (0.85)	4.07 (1.28)
BLQuant	5.65 (0.6)	3.48 (1.09)	5.46 (0.67)	4.07 (1.25)
Interactive	5.44 (0.75)	3.56 (1.1)	5.29 (0.75)	4.07 (1.23)
Static	5.42 (0.89)	3.66 (1.38)	5.28 (1.06)	4.12 (1.43)
Hybrid	5.49 (0.73)	3.5 (1.12)	5.27 (0.84)	4.07 (1.32)
Kruskal-Wallis (p-Value)	0.26	0.98	0.46	1.00

Results of the privacy usability factors. Each group shows the arithmetic mean and the standard deviation of the separate factors. The last row shows the p-values of the Kruskal-Wallis test of the corresponding concept $^{**}p \leq 0.01$, $^{*}p \leq 0.05$, $^{+}p \leq 0.1$.

Table 6. Effect of conditions on privacy usability factors.

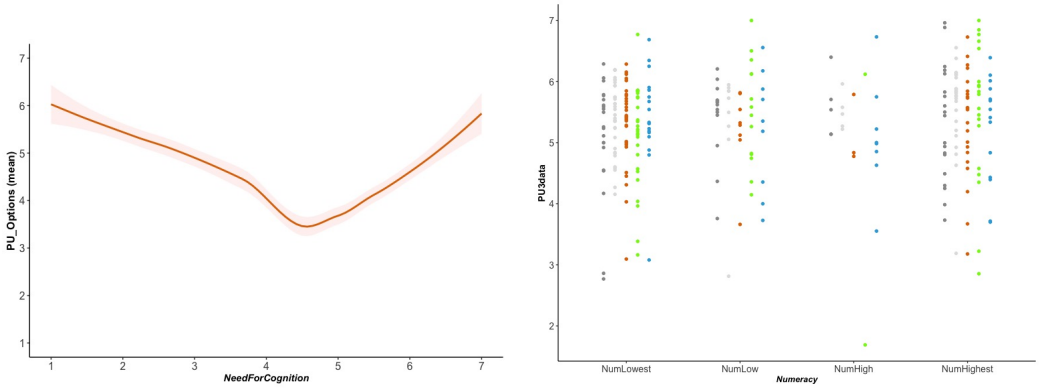
could again indicate that the usability of Static is more dependent on individual differences of the participants.

Additionally, we considered the element showing two textual explanations *Why* and *How*, which was available in all conditions. In our prototype, we also measured how many participants interacted with these elements. This measurement *Interaction* (see Table 2 in column 5) is generally very low and shows how short the attention span of users is. On the one hand, it is known that workers on MTurk try to complete the tasks by spending as little time and effort as possible [23]. Therefore, part of this result is due to the mode of our study. However, we argue that this behavior is not too different from real users regarding privacy decision interfaces: such decision interfaces usually interrupt the primary activity of the user, for example, booking a car in a car-sharing app. Consequently, actual users also try to deal with the interface expending as little time and effort as possible. The difference is, of course, that real users face real consequences from the decision, while workers on MTurk do not. Considering the differences in *Interaction* between the conditions, we see that the interactive interfaces Interactive and Hybrid have the lowest *Interaction* rate regarding the explanation type element. This could indicate that the attention of users is limited, and the introduction of additional interactive elements might impact the effectiveness of other elements. Another reason for the lower *Interaction* values in Interactive and Hybrid could be that users of these interfaces do not require textual explanations as the interactive risk element already satisfies the need for information sufficiently.

5.4 Hypothesis H-4: Donation Decision

The χ^2 test for the effect of the conditions on *DecisionOutcome* is not significant ($p = 0.70$). The donation rates of all conditions can be seen in Table 2. All conditions have similar donation rates with less than 10 points differences between the lowest condition Static (63 %) and the highest condition BLQual (72 %), with Interactive (71 %) and Hybrid (71 %) both very close to the qualitative baseline.

Therefore, it does not seem like the conditions influence *DecisionOutcome*.



(a) Influence of *NeedForCognition* on *PU_Options* in Condition ■ Interactive.

(b) Influence of *Numeracy* on *PU_Data*

Fig. 5. Significant effects of personal attributes on selected Privacy Usability factors in conditions ■ BLQual, ■ BLQuant, ■ Interactive, ■ Static, and ■ Hybrid.

5.5 Hypothesis H-3': Influence of Personal Attributes on Privacy Usability

The linear regressions for interaction effects of the personal attributes on the privacy usability factors yielded two significant results: *PU_Options* is significantly influenced by *NeedForCognition* in the condition ■ Interactive ($p = 0.04$) and *PU_Data* is significantly influenced by *Numeracy* in the condition ■ Hybrid.

Figure 5a shows that in condition ■ Interactive, the factor *PU_Options* is especially low for participants with medium *NeedForCognition*. We find this result surprising since interaction usually benefits mostly higher *NeedForCognition*. Regardless of speculations on the reason for this effect, we see that the usability rating is heavily impacted by the personal attributes. A similar but less pronounced shape of the data is also present in the other conditions. Therefore, we emphasize the need for testing and adopting decision interfaces with different personal attributes of the volunteers in mind.

To interpret the effect of *Numeracy* on *PU_Data* in the condition ■ Hybrid we plotted the data in Figure 5b. We see that in the affected *Numeracy* level NumHigh, Hybrid seems to perform slightly worse and with more variations than the other conditions. However, this shape of the ■ Hybrid condition is similar to the distribution of *PU_Data* in the other *Numeracy* levels. As such, we interpret this effect as a slightly better performance overall in terms of *PU_Data*, except for the interface ■ Hybrid. This could be caused by the various interface elements distracting from some information, which helps participants with NumHigh in the other condition to understand the kind of data being shared. However, the low outlier in condition ■ Static might suggest that with more data, ■ Static might be more similar to ■ Hybrid than to any other.

Finally, we see an effect of *Numeracy* on the factors *PU_Security* and *PU_Options*. Since this effect is independent of the conditions, this indicates that there is still work needed to make the interface equally usable for contributors with all levels of *Numeracy*.

We did not find any interaction effect which could explain the increased standard deviation of ■ Static on the privacy usability factors. This might be caused by an additional personality attribute, which we did not test for.

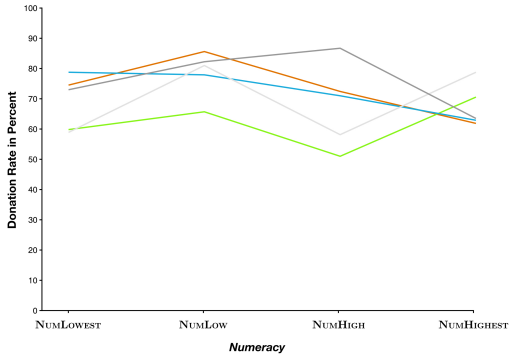


Fig. 6. Effect of *Numeracy* on *DecisionOutcome* for conditions ■ BLQual, ■ BLQuant, ■ Interactive, ■ Static, and ■ Hybrid

5.6 Hypothesis H-4': Influence of Personal Attributes on Donation Decision

We conducted a logistic regression analogously to H-3' to check for interaction effects on *DecisionOutcome*. We found four significant terms: NumHighest ($p = 0.02$) as well as the interaction terms between NumHighest and the conditions ■ BLQual ($p = 0.01$), ■ Interactive ($p = 0.02$) and ■ Hybrid ($p = 0.02$). We see in Figure 6, compared by *DecisionOutcome*, the conditions split into two groups. ■ BLQuant and ■ Static have a similar trend and ■ BLQual, ■ Interactive and ■ Hybrid have a different trend. The two conditions ■ BLQuant and ■ Static both introduce the quantitative information but are missing the interactive effort to enable the users to form a correct mental model. This result suggests that quantitative information can have an unintended effect on users with different numerical abilities if not combined with suitable opportunities to form a correct mental model.

6 DISCUSSION OF RESULTS

Our study provides a number of insights into how people's privacy decisions are influenced by the explanation of DP and the decision interface as a whole. Based on our results, we discuss our hypotheses derived from the research question: *How can we enable laypeople to incorporate privacy risk information into data donation decisions for the benefit of an informed decision.* In summary, we did find that the condition ■ Hybrid did benefit informed decision-making (H-2), while not affecting the overall donation rate (H-4). Additionally, we found several influences of personal attributes on the different measurements (H-3' and H-4'). The details of these results will be discussed in the following.

6.1 People's ability to judge the privacy risk of sharing data (H-1)

We hypothesized in H-1 that contributors could estimate the risk of sharing data better if the privacy risk information was included in the decision interface. While overall there is no significant effect of the interfaces on *RiskEstimation*, we saw in the interface ■ Interactive that contributors with high and higher *Numeracy* could recall the privacy risk significantly better. However, we also see that the improved *RiskEstimation* of the group ■ Interactive did not result in significantly improved informed decisions, nor did this effect occur in the other conditions with the interactive element. Thus, we conclude that interactive user interface elements to represent privacy risks can aid in perceiving and memorizing privacy risk values, but the effect is also dependent on other factors.

6.2 People's ability to make an informed decision based on their privacy concerns (H-2)

We assumed in H-2 that communicating the privacy risk enables potential contributors to make more informed decisions. We found in condition ■ Hybrid that users' decisions are significantly aligned with their level of privacy concerns, which is a strong indication that interactive visualizations can support people to make informed decisions. However, we also see that enabling an informed decision is complex: The interface ■ Hybrid has shown a slight negative privacy usability effect in comparison to the textual interfaces ■ BLQuant and ■ BLQual. This effect seems similar to Bucinca et al.'s [12] insight in the context of explainable AI. They showed that although their user interface enabled people to make better decisions, they forced people to think more; thus, study participants found these user interfaces more complex and less usable. Our privacy decision user interface seems to show a similar situation; even though our study participants made a more informed decision, they might have perceived the interface ■ Hybrid as less usable. Additionally, ■ Hybrid also did not improve *RiskEstimation* as the interface ■ Interactive did. Again, we assume that the complexity of the user interface causes this effect. Even though this is a promising result, it is concerning at the same time, especially for citizen science projects. Citizen science projects fundamentally rely on the participation of volunteers. Therefore, a complex user interface might hinder volunteers' participation, especially for newcomers.

6.3 People's characteristics influence how they experience the user interface (H-3, H-3')

Informed by existing research (see [28]), we assumed in H-3 and H-3' that our risk presentations might counteract the privacy usability of privacy decision interfaces. Indeed, we found some effects that support H-3', but overall our results regarding these hypotheses are inconsistent. Therefore, we suggest that future research should investigate how the presentation of privacy risks might impact privacy usability depending on the personal characteristics of people. Moreover, even though we found Habib et al.'s privacy usability concepts [34] valuable, the factor analysis of our adapted questions did not match with the intended usability concepts. To avoid these difficulties, we suggest instead using a validated user experience questionnaire (such as [65]) and adding missing privacy related questions.

6.4 People's willingness to donate data decreases when understanding the privacy risks (H-4, H-4')

We assumed in H-4 that our risk presentations can discourage users from donating data. Our results suggest that this hypothesis can be rejected. Donation rates were generally similar in all conditions. This is an important result since, as opposed to concerns raised by Rudnicka et al. [62], communicating existing privacy risks is not limiting participation.

However, our user interface influenced who donated their data (H-4'): While donation rates in conditions ■ BLQuant and ■ Static were high for users in the highest *Numeracy* group (NumHighest), we found that the donation rate was significantly lower for users in the groups NumLowest and NumHigh. This shows that personal attributes need to be considered when designing privacy decision interfaces to avoid unintended biases.

7 RECOMMENDATIONS FOR COMMUNICATING THE PRIVACY RISKS

Our findings have revealed several valuable insights from which we derive recommendations for providing DP-based privacy-preserving user interfaces for citizen science projects. Even though we focus on projects in the context of location data, our insights might also be valuable for other contexts, such as healthcare (see, e.g., [54]). While the first two sections focus on coordinators

of data collections, the third section addresses researchers and designers who work on privacy decision interfaces and the final section addresses opportunities for broader risk communications.

7.1 Data donation tasks should be enriched by privacy risk information

Based on our results, we suggest that privacy decision interfaces should include an interactive exploration and visualization of risk information. Even though further research is needed to differentiate the data donation behavior of people more precisely, i.e., how users' characteristics influence their user interface preferences, our results are promising to pursue further research in this regard:

Independently of how people carry out their decision-making[23], the perceived privacy risk is one of the best predictors of privacy concerns and also explains the behavioral intention, to some extent [30]. From an ethical point of view, information on the privacy risk is indispensable to ensure ongoing trust in data collections for data-driven improvements of society, such as citizen science projects (see [6]), and in our study, the fear of a more cautious decision due to an explicit statement of the involved risk was shown unwarranted.

Furthermore, we believe that DP can accommodate existing concerns since DP's parameter ε can enable far richer decision-making based on factoring in the trade-off between data utility and individual privacy. To recall, the lower the value for ε, the higher the privacy protection. However, that also means that more noise is added to the data, i.e., the accuracy and utility of the dataset are lowered. If volunteers favor data utility in a citizen science project, for example, they might agree with lower privacy protection to preserve more data utility. A variation of DP, such as the so-called local differential privacy [37], enables volunteers to choose individual values for ε, with which they are comfortable to donate their data [44]. Thus, we recommend incorporating various possible courses of action, i.e., different ε, in DP-based decision interfaces that go beyond the current binary decision between sharing data or refusing to share.

7.2 Considering and supporting privacy literacy in citizen science projects

In order to enable volunteers to judge the adequacy of the technology, the education of privacy literacy needs to be supported. Trepte [74] states that "online privacy literacy [is] a combination of factual or declarative ('knowing that') and procedural ('knowing how') knowledge about online privacy." The "knowing that" enables volunteers to be aware of existing privacy risks and the 'knowing how' to act in the respective data donation context. The sensitivity of location data, for example, might vary based on the context: bus stops might be less sensitive than the home address.

Compared to commercial data collection, such as social networks, citizen science projects have a special opportunity, as volunteers usually familiarize themselves more thoroughly with the project before donating data. Often citizen science projects will also have an onboarding process, which could also be used to improve on the privacy decision with the benefit of reinforced trust in the used privacy-preserving technologies and, in turn, in the project. Thus, we envision citizen science project coordinators implementing a kind of training as it already exists in citizen science games (see, e.g., [52]). These pieces of training can be, for example, part of the onboarding process of new volunteers and can also be staggered according to the type of participation. For example, initially, a high privacy level could be set automatically until a volunteer is familiar with the data collection tasks. After that, the trade-off could be communicated, and volunteers can decide whether to donate depending on the context.

[23]Three decision-making categories of people that affect their privacy choices are differentiated in research [4]: the rational calculation of the risk and benefits, the irrational risk-benefit calculation characterized by biased risk assessment, and a negligible or no risk assessment.

However, a narrow focus on individual privacy literacy might imply that individuals alone are responsible for protecting their privacy (e.g., [67]). Since DP provides a mathematically proven measure of privacy, it can also be used by the project coordination or even by regulatory institutions to define an acceptable privacy loss for specific data and include it as a standard in the project.

7.3 Accommodating the complexity in privacy-preserving decision-making

As expected, the privacy guarantees provided by DP were challenging to understand for our study participants and depended on numerous factors. However, our results suggest that interactive visualization approaches can support people in making decisions more aligned with their privacy concerns. We are convinced that DP's promises of anonymizing data (see [25]) can benefit society if we find suitable ways to communicate the privacy-utility trade-off properly. Such anonymizing approaches can ensure volunteers' privacy in citizen science projects, and insights might also be applicable in other data collecting contexts.

So far, only a few studies have dealt with the question of communicating DP (e.g., [11, 28, 80]). We provide the first study that employed interface elements specifically designed to communicate DP's privacy risk; however, DP is complex and different personality traits, which potentially impact the effectiveness of certain interface elements, further complicate the communication.

As Maeda states: "Accept the fact that some things can never be made simple" [48]. Enabling people to make informed decisions requires them to understand complex relationships. We believe that designing privacy decision interfaces, which communicate the relevant information in an understandable way and incorporating people's abilities would increase their acceptance. Instead of using soft paternalistic interventions (e.g., [1]), where designers decide on behalf of people in order to make the interaction effortless, we should focus on communicating such complex information more suitably. The challenges of communicating complex relationships exist also in the context of explainable AI (see, e.g., [12]). We hope our study inspires more research in contexts where we need to communicate complex technical relationships without endangering the usability of user interfaces.

Our experimental approach can only be a first step towards supporting informed decision-making since the complexity and individuality of the decision situation require approaches that are more closely centered on human abilities and real-world scenarios. Therefore, as the next step, we plan to use our insights to carry out co-creation workshops with volunteers in an ongoing citizen science project. This way, we can factor in their individual concerns and preferences directly and mitigate them in the user interface design.

7.4 Broader communication of privacy risks

In our experiment we have seen the communication of privacy risks can have a benefit on the informed decisions of potential data contributors. We have intentionally limited our notion of privacy risks on the kind of privacy risk of the intentional publication of data collections protected by DP. So far no other anonymization method has the same property as DP, the independently quantifiable privacy risk. However, apart from mentioning DP as the source of the privacy risk value, our interface elements are not specific to DP, instead they are designed with regards to general privacy risk. Therefore, if the research community discovers additional privacy enhancing technologies in the future, which have a similar property, we assume that our results are transferable. Since we could show the benefit of the communication of privacy risks on informed decision, we encourage the development of further such methods in order to fine-tune the privacy-utility trade-off to match the specific requirements of each collection context. Equally, we have not considered the risks inherent to security measurements. However, we believe that our results would also apply

to communicating accurate risk estimations for existing measures or new measures with inherent risk guarantees.

Finally, risk communication not only benefits the potential contributors, but could also help to inform coordinators of data collections to find the correct initial ε value for their data collection. If they understand the implications on individuals' privacy, they also can weight the privacy-utility trade-off in a more informed way. One step further, similar interaction and visualization elements could also be used for policy makers with political power. Supporting their understanding of the privacy risks of different DP settings could result in reasonable minimum requirements for the privacy protection when collecting different kinds of data.

8 LIMITATIONS OF OUR RESEARCH

In the following, we discuss the limitations of our study design. First, we used the MTurk crowdsourcing platform for recruiting our participants. MTurk sets different incentives than an actual decision situation would. We have employed standard procedures to mitigate as many of these limitations as possible (see description in Section 4.1.3). More importantly, simulating a data donation does not necessarily elicit the same decisions as in a real context[24]. The results from the study should be reconfirmed with contributors donating real data. Second, due to the population of MTurk and our requirement for proficiency in English, we were limited to U.S.-based participants. Privacy and trust are sociocultural concepts that differ greatly between different societies, as shown by a recent replication study of Xiong at al.'s DP user interfaces (see [45]) in the European context. Consequently, results from our study should be re-evaluated for different cultural contexts. Moreover, we employed specific visualizations and interactions in our privacy decision interfaces that were based on previous results in related research and the characteristics of DP. However, the design space of visualization and interaction elements is much broader, and different elements might support different users better. Additionally, there are various design dimensions [64], which we did not consider to keep our study design expressive.

Our study has some limitations primarily caused by our chosen experimental study design. However, existing research on DP-based privacy-preserving user interfaces is rare. We indeed discussed a more qualitative approach in the form of co-creation workshops. Yet, talking to and interacting with volunteers in citizen science projects is a valuable resource. Thus, we decided to first improve our understanding of existing design dimensions for DP-based privacy decision interfaces before approaching our volunteers. We hope that our results enable future studies to utilize the attention and time of volunteers in a focused manner.

9 CONCLUSIONS AND FUTURE RESEARCH

Data donated by volunteers in citizen science projects and the results obtained from similar data collections are valuable for research and demand-driven improvements in various contexts. At the same time, such data donations might expose sensitive information about volunteers. Consequently, a particular duty exists to ensure volunteers' privacy. However, ensuring people's privacy lowers the quality, i.e., utility, of the data. Thus, with our research, we want to enable the donors of the data to make informed decisions about their data donations based on individual privacy concerns. We investigate an approach that is especially suitable in this regard - Differential Privacy (DP). DP, with its parameter ε, allows tuning the trade-off between privacy and utility, but the effective communication of this trade-off and the resulting privacy risk in an understandable way is a desideratum. Based on previous research, we propose novel privacy decision interfaces which

[24]Warberg et al. [77], for example, argue that non-realistic scenarios might miss risk perception if participants are asked to make "authentic" decisions.

communicate this trade-off using graphical risk visualizations and interactive risk exploration. The promising results of our experimental study indicate that the combination of visualizations and interactive explorations of the privacy-utility trade-off of DP can enable informed decision-making. However, our results can only provide a first indication of how we can communicate the privacy risks of DP to support an informed decision. As a next step, we will transfer our results to an ongoing citizen science project on bike mobility and scrutinize our study results with actual volunteers. Within co-creation workshops, we will explore how the bike mobility community perceives our privacy decision interfaces and to what extent such interfaces influence volunteers' decisions to donate their data. Unlike our present experimental study, a real-world context allows people to experience and factor in the actual consequences of privacy risks into their decisions. We hope our study inspires additional research in this field. For example, researchers might study the interplay between the various interface elements in more detail to understand better how to support users best. Since we have considered only a small subset of the possible design space for privacy decision interfaces, we are curious about which elements and metaphors might further enhance the understanding of privacy risks. We also see a potential for synergy by joining forces with the explainable AI community that also wants to enable informed decision-making, even though the technology being explained is different. Finally, our results suggest that our study participants' personal characteristics influenced our interfaces' effectiveness. These insights are critical in order to avoid unintended discrimination against certain user groups and should, therefore, be investigated in more detail. Our results make a first step towards deepening our understanding of privacy risk communication. This understanding is essential to unfolding the potential of DP for a broad audience. Effective risk communication that supports informed decision-making has the potential to encourage more citizens to consider a data donation. Ultimately, this might inform policymakers and lead to a societal understanding of acceptable privacy losses in specific data contexts.

ACKNOWLEDGMENTS

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, under grant 01UV2090B "Transdisciplinary exploration of privacy-aware donation of movement data for sustainable urban mobility, sub-project: human-centered data security (freeMove)".

REFERENCES

[1] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. 2017. Nudges for Privacy and Security: Understanding and Assisting Users’ Choices Online. *Comput. Surveys* 50, 3 (Aug. 2017), 44:1–44:41. https://doi.org/10.1145/3054926

[2] A. Acquisti and J. Grossklags. 2005. Privacy and rationality in individual decision making. *IEEE Security Privacy* 3, 1 (Jan. 2005), 26–33. https://doi.org/10.1109/MSP.2005.22 Conference Name: IEEE Security Privacy.

[3] Susan B. Barnes. 2006. A privacy paradox: Social networking in the United States. *First Monday* (Sept. 2006). https://doi.org/10.5210/fm.v11i9.1394

[4] Susanne Barth and Menno D. T. de Jong. 2017. The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics* 34, 7 (Nov. 2017), 1038–1058. https://doi.org/10.1016/j.tele.2017.04.013

[5] Hilary Bekker, J. G. Thornton, C. M. Airey, JlBl Connelly, J. Hewison, M. B. Robinson, J. Lilleyman, M. MacIntosh, A. J. Maule, and S. Michie. 1999. Informed decision making: an annotated bibliography and systematic review. *Health Technol Assess* 3, 1 (1999), 1–156.

[6] Anne Bowser, Katie Shilton, Jenny Preece, and Elizabeth Warrick. 2017. Accounting for Privacy in Citizen Science: Ethical Research in a Context of Openness. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative*

Work and Social Computing (CSCW '17). Association for Computing Machinery, New York, NY, USA, 2124–2136. https://doi.org/10.1145/2998181.2998305

[7] Anne Bowser, Andrea Wiggins, Lea Shanley, Jennifer Preece, and Sandra Henderson. 2014. Sharing data while protecting privacy in citizen science. *Interactions* 21, 1 (Jan. 2014), 70–73. https://doi.org/10.1145/2540032

[8] danah boyd and Kate Crawford. 2012. Critical Questions for Big Data. *Information, Communication & Society* 15, 5 (June 2012), 662–679. https://doi.org/10.1080/1369118X.2012.678878 Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2012.678878.

[9] R Brauneis and Ellen P. Goodman. 2018. Algorithmic transparency for the smart city. *Yale Journal of Technology and Law* 20 (2018), 103. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/yjolt20Âğion=4

[10] Cameron Brick, Michelle McDowell, and Alexandra L. J. Freeman. 2020. Risk communication in tables versus text: a registered report randomized trial on 'fact boxes'. *Royal Society Open Science* 7, 3 (March 2020), 190876. https://doi.org/10.1098/rsos.190876 Publisher: Royal Society.

[11] Brooke Bullek, Stephanie Garboski, Darakhshan J. Mir, and Evan M. Peck. 2017. Towards Understanding Differential Privacy: When Do People Trust Randomized Response Technique? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3833–3837. https://doi.org/10.1145/3025453.3025698

[12] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 188:1–188:21. https://doi.org/10.1145/3449287

[13] J. T. Cacioppo, R. E. Petty, and C. F. Kao. 1984. The efficient assessment of need for cognition. *Journal of Personality Assessment* 48, 3 (June 1984), 306–307. https://doi.org/10.1207/s15327752jpa4803_13

[14] Edward Cokely, Mirta Galesic, Eric Schulz, Saima Ghazal, and Rocio Garcia-Retamero. 2012. Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making* 7 (Jan. 2012). https://doi.org/10.1037/t45862-000

[15] Jessica Colnago, Lorrie Cranor, and Alessandro Acquisti. 2023. Is There a Reverse Privacy Paradox? An Exploratory Analysis of Gaps Between Privacy Perspectives and Privacy-Seeking Behaviors. *Proceedings on Privacy Enhancing Technologies* (2023). https://petsymposium.org/popets/2023/popets-2023-0027.php

[16] Lorrie Faith Cranor. 2021. Informing California privacy regulations with evidence from research. *Commun. ACM* 64, 3 (Feb. 2021), 29–32. https://doi.org/10.1145/3447253

[17] Mary J. Culnan and Robert J. Bies. 2003. Consumer Privacy: Balancing Economic and Justice Considerations. *Journal of Social Issues* 59, 2 (2003), 323–342. https://doi.org/10.1111/1540-4560.00067 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1540-4560.00067.

[18] Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles. 2021. "I need a better description": An Investigation Into User Expectations For Differential Privacy. In *Proceeding of 28th ACM Conference on Computer and Communications Security (CCS)*. ACM, Seoul, South Korea, 16 pages. https://cs-people.bu.edu/kaptchuk/publications/ccs21_dp.pdf

[19] Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports* 3, 1 (March 2013), 1376. https://doi.org/10.1038/srep01376

[20] Karel Dhondt, Victor Le Pochat, Alexios Voulimeneas, Wouter Joosen, and Stijn Volckaert. 2022. A Run a Day Won't Keep the Hacker Away: Inference Attacks on Endpoint Privacy Zones in Fitness Tracking Social Networks. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22)*. Association for Computing Machinery, New York, NY, USA, 801–814. https://doi.org/10.1145/3548606.3560616

[21] Daniel Diethei, Jasmin Niess, Carolin Stellmacher, Evropi Stefanidi, and Johannes Schöning. 2021. Sharing Heartbeats: Motivations of Citizen Scientists in Times of Crises. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3411764.3445665

[22] Josep Domingo-Ferrer, David Sánchez, and Alberto Blanco-Justicia. 2021. The limits of differential privacy (and its misuse in data release and machine learning). *Commun. ACM* 64, 7 (July 2021), 33–35. https://doi.org/10.1145/3433638

[23] Julie S. Downs, Mandy B. Holbrook, Steve Sheng, and Lorrie Faith Cranor. 2010. Are your participants gaming the system? screening mechanical turk workers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 2399–2402. https://doi.org/10.1145/1753326.1753688

[24] Cynthia Dwork. 2006. Differential Privacy. In *Automata, Languages and Programming (Lecture Notes in Computer Science)*, Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener (Eds.). Springer, Berlin, Heidelberg, 1–12. https://doi.org/10.1007/11787006_1

[25] Cynthia Dwork, Nitin Kohli, and Deirdre Mulligan. 2019. Differential Privacy in Practice: Expose your Epsilons! *Journal of Privacy and Confidentiality* 9, 2 (Oct. 2019). https://doi.org/10.29012/jpc.689 Number: 2.

[26] Zohar Efroni, Jakob Metzger, Lena Mischau, and Marie Schirmbeck. 2019. Privacy Icons: A Risk-Based Approach to Visualisation of Data Processing. *European Data Protection Law Review* 5, 3 (2019), 352–366. https://doi.org/10.21552/

edpl/2019/3/9 Publisher: Lexxion Publisher.

[27] Yuanyuan Feng, Yaxing Yao, and Norman Sadeh. 2021. A Design Space for Privacy Choices: Towards Meaningful Privacy Control in the Internet of Things. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–16. https://doi.org/10.1145/3411764.3445148

[28] Daniel Franzen, Saskia Nuñez von Voigt, Peter Sörries, Florian Tschorsch, and Claudia Müller-Birn. 2022. *"Am I Private and If So, how Many?" – Using Risk Communication Formats for Making Differential Privacy Understandable.* Technical Report arXiv:2204.04061. arXiv. https://doi.org/10.48550/arXiv.2204.04061 arXiv:2204.04061 [cs] type: article.

[29] Rocio Garcia-Retamero and Edward T. Cokely. 2017. Designing Visual Aids That Promote Risk Literacy: A Systematic Review of Health Research and Evidence-Based Design Heuristics. *Human Factors* 59, 4 (June 2017), 582–627. https://doi.org/10.1177/0018720817690634 Publisher: SAGE Publications Inc.

[30] Nina Gerber, Paul Gerber, and Melanie Volkamer. 2018. Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & Security* 77 (Aug. 2018), 226–261. https://doi.org/10.1016/j.cose.2018.04.002

[31] Gerd Gigerenzer and Kai Kolpatzik. 2017. How new fact boxes are explaining medical risk to millions. *BMJ (Clinical research ed.)* 357 (May 2017), j2460. https://doi.org/10.1136/bmj.j2460

[32] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3411764.3445779

[33] David Gunning. 2019. DARPA's explainable artificial intelligence (XAI) program *(the 24th International Conference)*. SIGAI, ACM Special Interest Group on Artificial Intelligence, New York, New York, USA, ii. https://doi.org/10.1145/3301275.3308446 Journal Abbreviation: IUI '19 Publication Title: IUI '19.

[34] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. 2022. "Okay, whatever": An Evaluation of Cookie Consent Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–27. https://doi.org/10.1145/3491102.3501985

[35] Hana Habib, Yixin Zou, Yaxing Yao, Alessandro Acquisti, Lorrie Cranor, Joel Reidenberg, Norman Sadeh, and Florian Schaub. 2021. Toggles, Dollar Signs, and Triangles: How to (In)Effectively Convey Privacy Choices with Icons and Link Texts. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–25. https://doi.org/10.1145/3411764.3445387

[36] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. 2015. *Extremal Mechanisms for Local Differential Privacy.* Technical Report arXiv:1407.1338. arXiv. https://doi.org/10.48550/arXiv.1407.1338 arXiv:1407.1338 [cs, math] type: article.

[37] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2008. What Can We Learn Privately?. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*. 531–540. https://doi.org/10.1109/FOCS.2008.27 ISSN: 0272-5428.

[38] Mark J. Keith, Samuel C. Thompson, Joanne Hale, Paul Benjamin Lowry, and Chapman Greer. 2013. Information disclosure on mobile devices: Re-examining privacy calculus with actual user behavior. *International Journal of Human-Computer Studies* 71, 12 (2013), 1163–1173. https://doi.org/10.1016/j.ijhcs.2013.08.016

[39] Carmen Keller, Michael Siegrist, and Vivianne Visschers. 2009. Effect of risk ladder format on risk perception in high- and low-numerate individuals. *Risk Analysis: An Official Publication of the Society for Risk Analysis* 29, 9 (Sept. 2009), 1255–1264. https://doi.org/10.1111/j.1539-6924.2009.01261.x

[40] Christopher T. Kenny, Shiro Kuriwaki, Cory McCartan, Evan T. R. Rosenman, Tyler Simko, and Kosuke Imai. 2021. The use of differential privacy for census data and its impact on redistricting: The case of the 2020 U.S. Census. *Science Advances* 7, 41 (Oct. 2021), eabk3283. https://doi.org/10.1126/sciadv.abk3283 Publisher: American Association for the Advancement of Science.

[41] Carsten Keßler and Grant McKenzie. 2018. A geoprivacy manifesto. *Transactions in GIS* 22, 1 (2018), 3–19. https://doi.org/10.1111/tgis.12305 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tgis.12305.

[42] Sunyoung Kim, Jennifer Mankoff, and Eric Paulos. 2013. Sensr: evaluating a flexible framework for authoring mobile data-collection tools for citizen science. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 1453–1462. https://doi.org/10.1145/2441776.2441940

[43] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. Association for Computing Machinery, New York, NY, USA, 453–456. https://doi.org/10.1145/1357054.1357127

[44] Nitin Kohli and Paul Laskowski. 2018. Epsilon Voting: Mechanism Design for Parameter Selection in Differential Privacy. In *2018 IEEE Symposium on Privacy-Aware Computing (PAC)*. IEEE, Washington, DC, 19–30. https://doi.org/10.1109/PAC.2018.00009

[45] Patrick Kühtreiber, Viktoriya Pak, and Delphine Reinhardt. 2022. Replication: The Effect of Differential Privacy Communication on German Users' Comprehension and Data Sharing Attitudes. 117–134. https://www.usenix.org/conference/soups2022/presentation/kuhtreiber

[46] Anne Land-Zandstra, Gaia Agnello, and Yaşar Selman Gültekin. 2021. Participants in Citizen Science. In *The Science of Citizen Science*, Katrin Vohland, Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Roeland Samson, and Katherin Wagenknecht (Eds.). Springer International Publishing, Cham, 243–259. https://doi.org/10.1007/978-3-030-58278-4_13

[47] Gabriel Lins de Holanda Coelho, Paul H. P. Hanel, and Lukas J. Wolf. 2020. The Very Efficient Assessment of Need for Cognition: Developing a Six-Item Version. *Assessment* 27, 8 (Dec. 2020), 1870–1885. https://doi.org/10.1177/1073191118793208 Publisher: SAGE Publications Inc.

[48] John Maeda. 2006. *The laws of simplicity*. MIT Press, Cambridge, Mass. http://www.books24x7.com/marc.asp?bookid=18546 OCLC: 65205172.

[49] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research* 15, 4 (Dec. 2004), 336–355. https://doi.org/10.1287/isre.1040.0032 Publisher: INFORMS.

[50] Helia Marreiros, Mirco Tonin, Michael Vlassopoulos, and M. C. Schraefel. 2017. "Now that you mention it": A survey experiment on information, inattention and online privacy. *Journal of Economic Behavior & Organization* 140 (Aug. 2017), 1–17. https://doi.org/10.1016/j.jebo.2017.03.024

[51] Luise Mehner, Florian Tschorsch, and Saskia Nunez von Voigt. 2021. Towards Explaining Epsilon: A Worst-Case Study of Differential Privacy Risks. In *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*. 328–331. https://doi.org/10.1109/EuroSPW54576.2021.00041 ISSN: 2768-0657.

[52] Josh Aaron Miller, Britton Horn, Matthew Guthrie, Jonathan Romano, Guy Geva, Celia David, Amy Robinson Sterling, and Seth Cooper. 2021. How do Players and Developers of Citizen Science Games Conceptualize Skill Chains? *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 244:1–244:29. https://doi.org/10.1145/3474671

[53] Trung Tin Nguyen, Michael Backes, and Ben Stock. 2022. Freely Given Consent? Studying Consent Notice of Third-Party Tracking and Its Violations of GDPR in Android Apps. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22)*. Association for Computing Machinery, New York, NY, USA, 2369–2383. https://doi.org/10.1145/3548606.3560564

[54] Giovanna Nunes Vilaza, Raju Maharjan, David Coyle, and Jakob Bardram. 2020. Futures for Health Research Data Platforms From the Participants' Perspectives. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3419249.3420110

[55] Frank Pasquale. 2015. *The black box society: The secret algorithms that control money and information*. Harvard University Press.

[56] Ellen Peters, Judith Hibbard, Paul Slovic, and Nathan Dieckmann. 2007. Numeracy Skill And The Communication, Comprehension, And Use Of Risk-Benefit Information. *Health Affairs* 26, 3 (May 2007), 741–748. https://doi.org/10.1377/hlthaff.26.3.741 Publisher: Health Affairs.

[57] Jennifer Preece and Anne Bowser. 2014. What HCI can do for citizen science. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 1059–1060. https://doi.org/10.1145/2559206.2590805

[58] Aare Puussaar, Kyle Montague, Sean Peacock, Thomas Nappey, Robert Anderson, Jennine Jonczyk, Peter Wright, and Philip James. 2022. SenseMyStreet: Sensor Commissioning Toolkit for Communities. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 324:1–324:26. https://doi.org/10.1145/3555215

[59] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173677

[60] Arianna Rossi and Monica Palmirani. 2019. DaPIS: A data protection icon set to improve information transparency under the GDPR. *Knowledge of the Law in the Big Data Age* 252, 181-195 (2019), 5–5.

[61] Dana Rotman, Jenny Preece, Jen Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 217–226. https://doi.org/10.1145/2145204.2145238

[62] Anna Rudnicka, Anna L. Cox, and Sandy J. J. Gould. 2019. Why Do You Need This? Selective Disclosure of Data Among Citizen Scientists. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300622

[63] Sven Schade and Chrysi Tsinaraki. 2016. Survey report: data management in Citizen Science projects. https://doi.org/10.2788/539115 ISBN: 9789279583872 9789279639258 ISSN: 1831-9424.

[64] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2015. A Design Space for Effective Privacy Notices. 1–17. https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub

[65] Martin Schrepp, Jörg Thomaschewski, and Andreas Hinderks. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, Regular Issue (2017). https://www.ijimai.org/journal/bibcite/reference/2634

[66] Susanne G. Scott and Reginald A. Bruce. 1995. Decision-Making Style: The Development and Assessment of a New Measure. *Educational and Psychological Measurement* 55, 5 (1995), 818–831. https://doi.org/10.1177/0013164495055005017 Publisher: SAGE Publications Inc.

[67] Patrick Skeba and Eric P S Baumer. 2020. Informational Friction as a Lens for Studying Algorithmic Aspects of Privacy.. In *Proc. ACM Hum. Comput. Interact.*, Vol. 4. 1–22. https://doi.org/10.1145/3415172

[68] Daniel J. Solove. 2021. *The Myth of the Privacy Paradox*. SSRN Scholarly Paper ID 3536265. Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.3536265

[69] Alina Stöver, Nina Gerber, Sushma Kaushik, Max Mühlhäuser, and Karola Marky. 2021. Investigating Simple Privacy Indicators for Supporting Users when Installing New Mobile Apps. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 366. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3411763.3451791

[70] Brian L. Sullivan, Jocelyn L. Aycrigg, Jessie H. Barry, Rick E. Bonney, Nicholas Bruns, Caren B. Cooper, Theo Damoulas, André A. Dhondt, Tom Dietterich, Andrew Farnsworth, Daniel Fink, John W. Fitzpatrick, Thomas Fredericks, Jeff Gerbracht, Carla Gomes, Wesley M. Hochachka, Marshall J. Iliff, Carl Lagoze, Frank A. La Sorte, Matthew Merrifield, Will Morris, Tina B. Phillips, Mark Reynolds, Amanda D. Rodewald, Kenneth V. Rosenberg, Nancy M. Trautmann, Andrea Wiggins, David W. Winkler, Weng-Keen Wong, Christopher L. Wood, Jun Yu, and Steve Kelling. 2014. The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation* 169 (Jan. 2014), 31–40. https://doi.org/10.1016/j.biocon.2013.11.003

[71] Latanya Sweeney. 2000. Simple demographics often identify people uniquely. *Health (San Francisco)* 671, 2000 (2000), 1–34.

[72] Latanya Sweeney. 2002. k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, 05 (Oct. 2002), 557–570. https://doi.org/10.1142/S0218488502001648 Publisher: World Scientific Publishing Co..

[73] Kyle A. Thomas and Scott Clifford. 2017. Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior* 77 (Dec. 2017), 184–197. https://doi.org/10.1016/j.chb.2017.08.038

[74] Sabine Trepte, Doris Teutsch, P. K. Masur, C. Eicher, Mona Fischer, Alisa Hennhöfer, and Fabienne Lind. 2015. Do People Know About Privacy and Data Protection Strategies? Towards the "Online Privacy Literacy Scale" (OPLIS). In *Reforming European Data Protection Law*, Serge Gutwirth, Ronald Leenes, and Paul de Hert (Eds.). Springer Netherlands, Dordrecht, 333–365. https://doi.org/10.1007/978-94-017-9385-8_14

[75] Janice Y Tsai, Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. 2010. Location-Sharing Technologies: Privacy Risks and Controls. *Journal of Law and Policy for the Information Society* 6, 2 (2010), 1–26.

[76] Yue Wang, Xintao Wu, and Donghui Hu. 2016. Using Randomized Response for Differential Privacy Preserving Data Collection. In *EDBT/ICDT Workshops*. CEUR-WS.org, Bordeaux, France, 1–8.

[77] Logan Warberg, Alessandro Acquisti, and Douglas Sicker. 2019. Can Privacy Nudges be Tailored to Individuals' Decision Making and Personality Traits?. In *Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society (WPES'19)*. Association for Computing Machinery, New York, NY, USA, 175–197. https://doi.org/10.1145/3338498.3358656

[78] Odette Wegwarth and Gerd Gigerenzer. 2018. The Barrier to Informed Choice in Cancer Screening: Statistical Illiteracy in Physicians and Patients. *Recent Results in Cancer Research. Fortschritte Der Krebsforschung. Progres Dans Les Recherches Sur Le Cancer* 210 (2018), 207–221. https://doi.org/10.1007/978-3-319-64310-6_13

[79] Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David O'Brien, Thomas Steinke, and Salil Vadhan. 2018. Differential Privacy: A Primer for a Non-Technical Audience. *Vanderbilt Journal of Entertainment, Technology & Law* 21, 17 (2018), 209–275. https://doi.org/10.2139/ssrn.3338027

[80] Aiping Xiong, Tianhao Wang, Ninghui Li, and Somesh Jha. 2020. Towards Effective Differential Privacy Communication for Users' Data Sharing Decision and Comprehension. *arXiv:2003.13922 [cs]* (March 2020). http://arxiv.org/abs/2003.13922 arXiv: 2003.13922.

[81] Aiping Xiong, Chuhao Wu, Tianhao Wang, Robert W. Proctor, Jeremiah Blocki, Ninghui Li, and Somesh Jha. 2022. Using Illustrations to Communicate Differential Privacy Trust Models: An Investigation of Users' Comprehension, Perception, and Data Sharing Decision. *arXiv:2202.10014 [cs]* (Feb. 2022). http://arxiv.org/abs/2202.10014 arXiv: 2202.10014.

[82] Verena Zimmermann and Karen Renaud. 2021. The Nudge Puzzle: Matching Nudge Interventions to Cybersecurity Decisions. *ACM Transactions on Computer-Human Interaction* 28, 1 (Jan. 2021), 7:1–7:45. https://doi.org/10.1145/3429888

Appendix

A EXTENDED STUDY INFORMATION

A.1 Extended considerations

A.1.1 Online Experiment. As personally identifiable information, we only collected the worker IDs as well as the IP addresses of the participants. These are necessary for managing submissions and removing repeated submissions by the same worker. After completing these tasks, we deleted these personally identifiable attributes from our dataset and performed all statistical analyses anonymously. Furthermore, we informed MTurk-workers of this practice before the start of the survey. They were also given the option to delete all collected data at any time during the survey.

Eventhough we think the working time of the survey is less than 15 Minutes, we allowed a working time of 40 minutes to avoid automatic timeouts of participants who were slower for any reason. The calculated actual working time of 15 Minutes was appropriate, based on worker responses on turkerView[25], one of the most popular worker forums. To encourage workers to pay attention, we promised an additional bonus of 0.50$ for coherent answers. We collected the submissions for both studies (pre-study and main study) in August 2022 throughout the day within reasonable US working hours.

We emphasized the importance of attentive reading especially for the scenario text. After the scenario, we asked comprehension questions in the form of single-choice questions about the three most important pieces of information in the scenario (What kind of app? What data is shared? With whom is the data shared?). The incorrect answer options were chosen to be considerably different from the correct answers. Participants had two chances to answer the comprehension checks correctly; otherwise, we informed them immediately that we would reject their submission and that they should not spend any more time on the rest of the survey. In addition, to support the immersion into the scenario, we asked participants to imagine a possible location which could have been collected about them. We manually evaluated these open-text answers, flagging answers that did not discuss any concrete or abstract place.

Throughout the remainder of the survey, we included attention checks in the form of "Please pick <answer option> for this question." or framed like "My answer should be accepted/rejected.".

In order to fairly accept or reject workers, we counted the number of attention checks which were flagged as incorrect. Submissions with two or more flagged answers would be rejected directly. For submissions with only one flagged answer we inspected the remaining open-text answers manually and excluded submissions where the answers did not relate to the questions asked. In particular, misconceptions were not excluded as long as they provided a possible answer to the question.

In order to mitigate the downsides of a rejection on the acceptance rate of the workers, we invited all rejected submissions for a retake survey. In total, 14 (8 in the pre-study, 6 in the main study) workers resubmitted, of which 7+4 answered attentively. The rejection on their original submission was reverted. However, since these workers were pre-exposed to the survey, we did not consider these answers in the analysis.

[25] https://turkerview.com

A.1.2 Scenario considerations. Our study introduces a fictional car-sharing service, CityCar, to make the scenario tangible for study participants. CityCar provides cars distributed throughout the city, which can be rented for short-distance trips. We inform our participants about the features of the CityCar app and ask them to imagine that they have used the service frequently to go to multiple locations. Subsequently, the participants are informed about an update of the CityCar app, which includes an option for a data donation of already collected mobility data.

In order to ensure the quality of scenario texts, we set up a small pilot study with the actual texts, asking for feedback on how understandable and imaginable the scenario are and how we can improve. For feedback on the notification content, we let each participant perform the interaction in one random interface as they would in the real study. Additionally, we showed the remaining notifications in a static display and ask for open-text feedback about understandability or missing information. We collected 33 submissions for the pilot study. We check all submissions using the employed attention checks and manually inspect all open-text answers. Unfortunately, 13 submissions did not answer the open-text answers attentively[26]. The suggestions from the remaining 19 submissions were incorporated into the survey texts for the main experiments.

A.1.3 Measurements. Several of the chosen instruments measure on a 7-Point Likert scale. We used the following answer options throughout the whole study: "1: strongly disagree", "2: disagree", "3: More or less disagree", "4: undecided", "5: more or less agree", "6: agree", "7: strongly agree". We specifically chose to represent each answer option with its numerical representation (1-7) in addition to its textual representation ("strongly disagree" ... "strongly agree") to induce a sense of scale between the answer options.

To measure *PrivacyConcern* we used the test IUIPC [49]. It evaluates privacy concern on three dimensions, each with 3-4 questions on a 7-point Likert scale. The final value for *PrivacyConcern* was computed from the values *Collection*, *Control* and *Awareness*[27] using a factor analysis, similar to the original procedure to evaluate the IUIPC questionnaire by Malhotra et al. [49]. The obtained weights obtained from the factor analysis (*Control*: 0.793, *Collection*: 0.734 and *Awareness*; 0.775) are similar to the weights obtained by Malhotra et al.(*Control*: 0.75, *Collection*: 0.78 and *Awareness*; 0.91), although in our analysis *Awareness* factors less into the *PrivacyConcern* value.

A.1.4 Privacy Usability. To measure *PrivacyUsability* we adopted five of the aspects by Habib et al. [34] which are suitable for our context, namely: *Ability*, *Awareness*, *Comprehension Need*, and *Sentiment*. The remaining two aspects *Decision reversal* and *Nudging* are out of scope of this paper for the following reasons: The aspect *Decision reversal* does not apply to the context of data donation, as once the data has been analyzed using DP in a larger dataset, the decision cannot easily be reversed. Similar questions might apply to the continuous donation of data. However, we are considering this to be a different research question. In contrast, the aspect *Nudging* can be applied to the chosen decision situation. However, investigating nudge effects is a research area by itself and would confound our research question. So, instead, we design our interface carefully to avoid possible nudge effects and keep the design of the different interfaces as equal as possible to avoid effects in favor of any condition.

To derive questions for each aspect we first screened the questions used by Habib et al. [34] for relevance to data donation and modified them to fit our scenario. We adapted the items for each of the aspects *Ability* (three questions), *Comprehension* (two questions) and *Sentiment* (two questions).

[26]These submissions instead pasted unrelated random texts, some of which were found using a google search.

[27]The factor *Collection* captures the user's concern about how personal data may be possessed by others, *Control* represents how important a free choice about data collection is and *Awareness* indicates how much users want to be informed about data practices.

The remaining items on *Awareness*, *Ability* and *Need* are created similar to the questions by Habib et al. adopted to the privacy decision domain.

The final privacy usability factors *PU_Data*, *PU_Decision*, *PU_Options*, and *PU_Security* are calculated according to the following weights resulting from the factor analysis:

Question	*PU_Decision*	*PU_Security*	*PU_Data*	*PU_Options*
Comp1	**0.726**		0.144	0.171
Comp2	**0.635**		0.359	
Comp3		**0.655**	0.160	0.284
Sent1	0.452	0.105	**0.518**	
Sent2	**0.555**		0.146	0.140
Ability1	0.290		**0.839**	
Ability2	0.425		**0.640**	-0.106
Ability3	**0.700**	0.137	0.194	
Ability4	0.146	0.356		**0.572**
Aware1	**0.517**	-0.101	0.168	
Aware2	0.135	0.394		**0.843**
Need1		**0.815**		0.209
Need2	-0.222	**0.622**	0.143	0.112

A.1.5 Personal Attributes Influencing Privacy Decisions. Based on existing research, we chose to consider *General Decision-Making Style* (GDMS), *Need for Cognition* (NFC) and *Numeracy*. Where possible, we use evaluated instruments to measure the desired attributes, which are explained next.

GeneralDecisionMakingStyle is measured by using a questionnaire developed by Scott and Bruce [66]. It contains five scales, representing the different styles, namely Avoidant, Dependent, Intuitive, Rational, and Spontaneouswith 5 Likert items each. To compute the *GeneralDecisionMakingStyle* of each participant, we calculated the arithmetic means of the 5 questions of each sub-scale and then sort each participant into the decision style group (Avoidant, Intuitive, Dependent, Rational, Spontaneous), where they achieved the highest rating. If there were multiple sub-scales with an equally high result, we assigned a random style from the possible options.

The NCS-6 test used to measure *NeedForCognition* contains 6 Likert items. We calculated the arithmetic mean of the 6 answers to obtain one *NeedForCognition* value (1-7) for each participant.

The Berlin Numeracy Test (BNT) developed by Cokely et al. [14] is used to measure *Numeracy*. In the adaptive form of the BNT, participants need to answer general questions about probabilities. Depending on the answer to the first question one of two different second questions is shown and the answer to the second question determines, whether a third question is necessary. Based on the correctness of the answers BNT sorts participants into four groups NumLowest, NumLow, NumHigh, and NumHighest according to a flow chart provided in [14].

A.2 Additional Evaluations and Results

A.2.1 DecisionOutcome. Surprisingly, in condition ■ BLQuant, the average *PrivacyConcern* of donating participants is lower than the average *PrivacyConcern* of non-donating participants. This is especially surprising, since condition ■ BLQuant contains strictly more information than condition ■ BLQual and should, therefore, support informed decision better. Even if this result is not significant, it serves as an additional indication that informed decision is a highly volatile concept as it depends on a lot of different factors. Even providing more information or the same information in a different way can lead to an unpredictable difference in the informed decision.

Condition	*Comprehension* mean (sd)	*Ability* mean (sd)	*Awareness* mean (sd)	*Sentiment* mean (sd)	*Need* mean (sd)	*PrivacyUsability* mean (sd)
BLQual	4.87 (0.90)	5.17 (0.82)	4.88 (1.07)	5.56 (1.06)	3.14 (1.47)	4.70 (0.74)
BLQuant	5.01 (0.73)	5.30 (0.86)	5.08 (1.08)	5.84 (0.78)	2.92 (1.26)	4.82 (0.62)
Interactive	4.90 (0.93)	5.18 (0.78)	4.85 (1.17)	5.64 (0.86)	3.07 (1.30)	4.72 (0.70)
Static	4.95 (1.14)	5.16 (1.15)	4.90 (1.15)	5.65 (1.06)	3.16 (1.49)	4.75 (0.99)
Hybrid	4.88 (0.91)	5.24 (0.86)	4.86 (1.16)	5.59 (1.01)	2.97 (1.31)	4.71 (0.66)
Kruskal-Wallis (p-Value)	0.35	0.85	0.74	0.29	0.88	0.55

Results of the privacy usability concepts. Each group shows the arithmetic mean and the standard deviation of the separate concepts and the combined *PrivacyUsability* score. The last row shows the p-values of the Kruskal-Wallis test of the corresponding concept $^{**}p \leq 0.01$, $^{*}p \leq 0.05$, $^{+}p \leq 0.1$.

Table 7. Effect of conditions on privacy usability concepts.

A.2.2 Original Privacy Usability Concepts. Before performing the factor analysis, we already analyzed the effect of the interfaces on the original privacy usability concepts. The resulting values are shown in Table 7 and the resulting significant results of the regression model are shown in Table 8.

Hyp.	Condition	Moderated by	Effect on	Effect	p-value
H-3'	Static	*NeedForCognition*	*Comprehension*	~	0.05
H-3'	Interactive	*NeedForCognition*	*Comprehension*	~	0.04
H-3'	Interactive	*NeedForCognition*	*PrivacyUsability*	~	0.04

Overview of the significant effects found in the data, showing (1) the relevant hypothesis, (2) the significant condition, (3) the attribute moderating the effect, (4) the measurement effected, (5) the direction of the effect (+: positive, -: negative, ~: mixed), (6) the p-value, rounded to one significant digit and (7) Cohen's d as measure for effect size, where applicable
*: Measured as correlation *PrivacyConcern* → *DecisionOutcome*

Table 8. Overview of significant effects.

A.2.3 Effects independent of conditions. Additionally, we found significant results which are independent of the interfaces:

- The *Sentiment* is influenced by the *GeneralDecisionMakingStyle* Intuitive ($p = 0.009$). Intuitive users feel better about their privacy decisions in our test (see Figure 7a).
- *Awareness* is significantly influenced by the *GeneralDecisionMakingStyle* Rational ($p = 0.004$). Rational users feel most aware of the privacy choices in our test (see Figure 7a).
- *Awareness* is also significantly influenced by *NeedForCognition* ($p = 0.04$). Users with low or high *NeedForCognition* have higher *Awareness* than users with medium *NeedForCognition* (see Figure 7c).

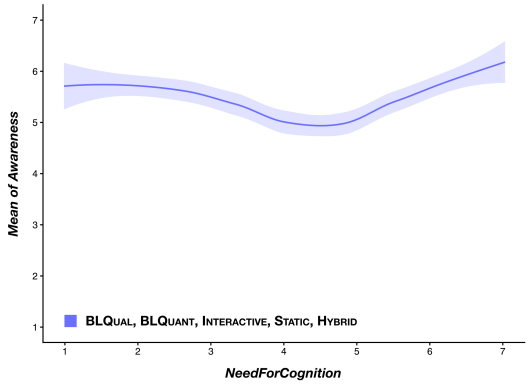
GeneralDecision-MakingStyle	*Sentiment* mean	*Awareness* mean	*Comprehension* mean
Avoidant	5.34	4.51	4.62
Dependent	5.62	4.20	4.61
Intuitive	**5.79**	4.96	4.84
Rational	5.61	**5.28**	**5.20**
Spontaneous	5.74	4.48	4.55

Values of significant effects are printed in bold.

(a) Effect of *GeneralDecisionMakingStyle* on *PrivacyUsability* concepts.

Numeracy	*Need* mean
NumLowest	2.59
NumLow	3.14
NumHigh	**3.63**
NumHighest	3.47

(b) Effect of *Numeracy* on *Need*



(c) Effect of *NeedForCognition* on *Awareness*.

Fig. 7. Effects independent of used interfaces in the conditions.

- *Need* is positively influenced by high *Numeracy* ($p = 0.02$), i.e., users with high *Numeracy* feel more satisfied (see Figure 7b). This could be an inherent property of DP, which, in itself, deals with probabilities (noise) and might be generally more accessible to users familiar with statistics.
- *Comprehension* is, in addition to the influences already discussed, also influenced by the *GeneralDecisionMakingStyle* Rational. ($p = 0.004$). The *Comprehension* values in this *GeneralDecisionMakingStyle* group are the highest (see Figure 7a), which suggests that rational deciders have understood the privacy choice. This makes sense, as they are expected to be able to extract the facts from the interfaces.

B PRE-STUDY ON TEXTUAL EXPLANANTION OF DP

B.1 Study Design

In our pre-study, we focus on how DP can be explained textually to laypeople and how these explanations inform their privacy decisions. Regulations such as the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the U.S. require that consent interfaces consider the principle of transparency. We make use of the concept of *meaningful transparency* from the area of eXplainable Artificial Intelligence (XAI) [33] which relates to "knowledge sufficient to approve or disapprove of the algorithm's performance" [9]. However, in the context of DP, such sufficient knowledge can focus on two different *transparency*

subjects: On the one hand, on different aspects of the mechanism of DP (e.g., [18]). Such decision interfaces are usually qualitative, i.e., they omit the ε chosen and the resulting quantitative privacy risks, and on the other hand, on the values of ε and the resulting quantitative guarantee on the privacy risk (e.g., [28]). Depending on ε, the DP guarantee can range from highly protected to almost no protection at all. This quantitative information is necessary for users to fully understand the consequences of a data donation. Therefore, it is desirable (or even a prerequisite) to include quantitative information in the decision interface in order to support an informed decision.

Either *transparency subject* (quantitative/qualitative) can be explored regarding different *explanation types*. Building on existing research [59], we differentiate three possible types[28]:

What? Reveal only that DP is applied or how high the privacy guarantee is.
How? Describe which steps are taken to achieve the DP protection or the DP guarantee.
Why? Provide justification for why DP protects or can guarantee the privacy without describing how it works.

Based on these considerations, we designed different user interfaces using the two design dimensions *transparency subject* and *type*, resulting in six conditions: ■ QualWhat, ■ QualHow, ■ QualWhy, ■ QuantWhat, ■ QuantHow, and ■ QuantWhy.

We selected explanations for the qualitative conditions from Cummings et al. [18], by comparing their representative descriptions with the central theme of each explanation type, and found the following descriptions to be a good fit: "Trust" (what), "Techniques" (how) and "Enables" (why).[29] The quantitative explanations were created resembling the corresponding qualitative explanations, but explain the quantitative risk instead of the mechanism of DP. These conditions, as shown in Table 9, were already included in the pilot study mentioned earlier (see Section 4.1.2).

All six conditions were then inserted into relatively simple static interfaces: The first screen shows basic information about the donation proposed (what data is shared, with whom and for what reason). The second screen (e.g., Figure 8 shows ■ QuantWhy) explains the privacy protection, including one of the six conditions and the privacy choice, either to donate the data using DP or decline donation.

All study materials used in this experiment are provided in Appendix C.1.

B.2 Statistical Analysis

Based on an a priori power analysis, we determined a sample size of 329 participants for a power of 95 %. We collected a total of 483 submissions. We rejected submissions for repeated participation (2), by the automatic attention checks (115) and by manually verifying the open text answers (36) (see Section 4.1.3 for further details on the rejection procedure). The resulting 330 submissions are shown in Table 10.

These regression models contain seven input variables: (1) Group, (2) Numeracy, (3) NFC, (4) GDMS, as well as interaction terms (5) Group:Numeracy, (6) Group:NFC, and (7) Group:GDMS. We use linear regression for numeric target variables and the corresponding logistic regression for binary target variables.

B.3 Results

All significant results of the pre-study are shown in Table 11.

[28] Radar et al. [59] presented these in the area of XAI. They differentiated four explanation types that can be considered by explanations: "what?", "how?", "why?" and "objective". The type "objective" describes how the system was developed. We disregard this explanation type in our study because we focus on the user's understanding rather than the correctness of the system implementation.
[29] The descriptors "Trust", "Techniques" and "Enables" are the names used by Cummings et al. [18].

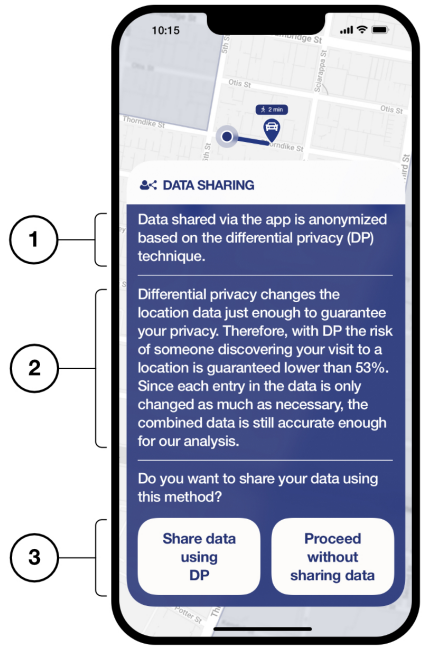
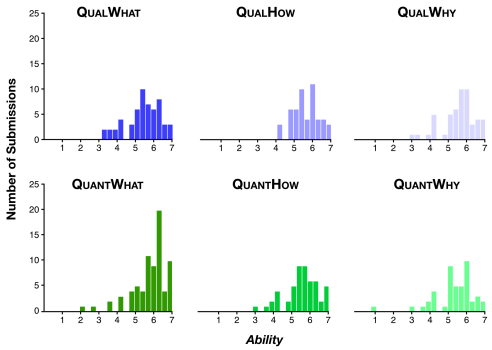


Fig. 8. ■ QuantWhy as one example of the structure and the explanations provided in a DP-based privacy decision interfaces consisting of ① *the privacy notification* showing a description of how privacy is ensured, ② *the privacy risk* depicting an explanation of the existing privacy risk using the *Why* explanation type, ③ *the privacy choice* enabling users to decide about sharing.



(a) Distribution of *Ability* by Group.

		(1)	(2)	(3)	(4)	(5)
■ QualWhat	(1)	-				
■ QualHow	(2)	1	-			
■ QualWhy	(3)	1	1	-		
■ QuantWhat	(4)	0.12	0.58	0.79	-	
■ QuantHow	(5)	1	1	1	0.48	-
■ QuantWhy	(6)	1	1	1	0.05 *	1

Pairwise Wilcox tests between Conditions and measure *Ability*. All p-values are Bonferroni-corrected. $^{**}p \leq 0.01$, $^{*}p \leq 0.05$, $^{+}p \leq 0.1$.

(b) Post-hoc Pairwise Wilcox, Condition → *Ability*.

Fig. 9. Influence of Condition on *Ability*.

Concerning H-3, we found little difference in the privacy usability concepts depending on the different conditions. The only significant effect ($p = 0.03$) shows a higher *Ability* in ■ QuantWhat (see the pairwise post hoc test and histograms in Figure 9). This indicates that the quantitative information can even benefit privacy usability. Thus, there is no objection to including quantitative information in privacy decision interfaces.

Condition	Focus	Text
■ QualWhat	DP Mechanism	Differential privacy injects statistical noise into collected data in a way that protects privacy without significantly changing conclusions. (Cummings, Techniques).
■ QualHow	DP Mechanism	Differential privacy allows analysts to learn useful information from large amounts of data without compromising an individual's privacy. (Cummings, Enables).
■ QualWhy	DP Mechanism	Differential privacy is a novel, mathematical technique to preserve privacy. (Cummings, Trust).
■ QuantWhat	DP Risk	With differential privacy the probability of someone discovering your visit at a location is guaranteed lower than 53 %.
■ QuantHow	DP Risk	Differential privacy adds/subtracts just enough randomness to the result of an analysis, such that no one can be certain, whether one more or one less person visited the location. Based on the amount of randomness added, we can calculate that the risk of someone discovering your visit to a location is guaranteed lower than 53 %.
■ QuantWhy	DP Risk	Differential privacy changes the location data just enough to guarantee your privacy. Therefore, with DP the risk of someone discovering your visit to a location is guaranteed lower than 53 %. Since each entry in the data is only changed as much as necessary, the combined data is still accurate enough for our analysis.

Conditions in the pre-study; the qualitative conditions are chosen from Cummings et al. [18].

Table 9. Conditions used in the first experiment study, defined by the design dimensions *transparency subject* (qualitative describing the DP mechanism and quantitative describing the privacy risk) and *explanation type* (what, how, why).

Considering the effect of quantitative information on the donation rate, we see no significant influence of the conditions on the donation decision (see Table 10). The quantitative conditions, i.e., ■ QuantWhat, ■ QuantHow, ■ QuantWhy, have a very similar donation rate to the qualitative conditions (within 1 %). Quantitative information, as used in our conditions, does not discourage users from donating.

However, the conditions regarding the explanation type "why", i.e., ■ QualWhy and ■ QuantWhy, and "how", i.e., ■ QualHow and ■ QuantHow, seem to encourage users in donating their data with DP more than "what": "what" 54 %, "why" 60 % and "how" 64 %.

The linear regression models (H-3') for *PrivacyUsability*, *Sentiment* or *Ability* did not return any significant results. We found a significant result for ■ QuantWhy (0.0255) and the interaction ■ QuantWhy with *NeedForCognition* ($p = 0.04$) on *Comprehension*. Figure 10 shows, as expected, that higher *NeedForCognition* results in higher *Comprehension*. This relationship shows that the *Comprehension* is indeed influenced by *NeedForCognition* in the condition ■ QuantWhy, confirming H-3' in part. The equivalent logistic regression model for donation rate (H-4') did not show any significant results.

Condition	Size	*RiskEstimation* mean (sd)	*Compreh.* mean (sd)	*Sentiment* mean (sd)	*Ability* mean (sd)	*Decision-Outcome* % donating	*Privacy-Concern* mean (sd)	*NeedFor-Cognition* mean (sd)
QualWhat	56	51 % (27.3)	6.0 (0.78)	5.8 (0.80)	5.4 (0.93)	53.6 %	5.7 (0.79)	4.6 (1.2)
QualHow	51	42 % (28.7)	6.0 (0.64)	5.8 (0.80)	5.6 (0.74)	60.8 %	5.9 (0.68)	4.8 (1.2)
QualWhy	51	44 % (29.8)	6.2 (0.74)	5.7 (0.98)	5.6 (0.9)	64.7 %	5.8 (0.75)	5.0 (1.1)
QuantWhat	74	50 % (25.3)	6.1 (0.75)	5.7 (0.98)	5.8 (1.00)	55.4 %	5.8 (0.70)	5.0 (1.1)
QuantHow	52	51 % (23.8)	6.0 (0.86)	5.8 (0.94)	5.5 (0.91)	67.3 %	5.8 (0.80)	4.8 (1.2)
QuantWhy	46	47 % (23.3)	6.0 (0.73)	5.6 (1.10)	5.3 (1.09)	54.3 %	5.7 (0.85)	5.0 (1.1)

Overview of the conditions showing (1) the size of each group/condition, (2) the arithmetic mean and standard deviation of *RiskEstimation*, (3-5) the arithmetic mean and standard deviation of the privacy usability concepts, (6) the ratio of participants choosing donation in percent and (7-8) the arithmetic mean and standard deviation of the personal attributes.

Table 10. Dataset overview

Condition	Moderated by	Effect on	Effect direction	p-value
QuantWhat	-	*Ability*	positive	0.03000
QuantWhy	*NeedForCognition*	*Comprehension*	positive	0.03540
QuantWhy	-	Informed decision (*PrivacyConcern* → *DecisionOutcome*)	positive	0.00736

Table 11. Significant Effects of the pre-study.

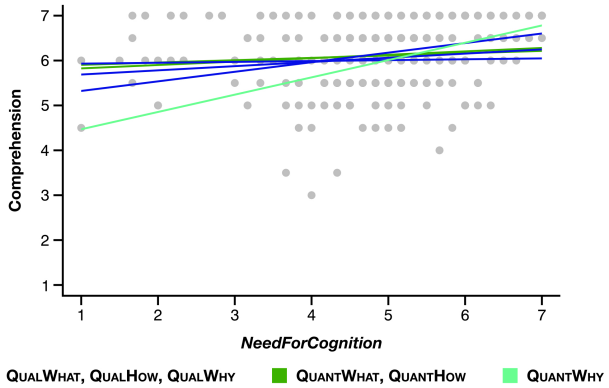


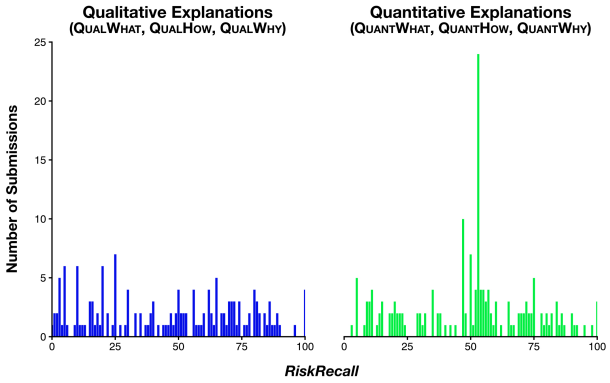
Fig. 10. Relationship between *NeedForCognition* and *Comprehension* in QuantWhy in comparison to the other qualitative and quantitative explanations.

In summary, we see some evidence that the quantitative conditions are more influenced by personal attributes of the users. The influence of *NeedForCognition* on *Comprehension* in the condition QuantWhy is significant.

The distribution of the estimation of privacy risk shows a visible accumulation of answers between 50 and 53 % in the quantitative conditions (see Figure 11b). The latter also show a lower standard deviation (24.2 vs. 28.7) and the reduced interquartile range (34 vs. 50) (see Figure 11a),

Condition	Percentiles 25	50	75	Stand. Dev.
qualitative	20	49	70	28.67337
quantitative	32	53	66	24.22278

(a) Spread of *RiskEstimation*.



(b) Distribution of *RiskEstimation* for qualitative and quantitative explanations.

Fig. 11. Risk Estimations of participants for qualitative vs. quantitative explanations.

Condition	Mean *PrivacyConcern* donating	not donating	Difference	Logistic Regression *PrivacyConcern* → *DecisionOutcome* (p-Value)
■ QualWhat	5.660185	5.80235	0.1422	0.498
■ QualHow	5.815412	5.995833	0.1804	0.350
■ QualWhy	5.805556	5.776235	-0.0293	0.893
■ QuantWhat	5.7771	5.8569	0.0798	0.625
■ QuantHow	5.6683	5.9788	0.3105	0.188
■ QuantWhy	5.374444	6.109788	0.7353	0.00736 **

Mean *PrivacyConcern* for donating and not donating participants in each condition and logistic regression p-values for informed decision
$^{**}p \leq 0.01$, $^{*}p \leq 0.05$, $^{+}p \leq 0.1$.

Table 12. Relation of *PrivacyConcern* to *DecisionOutcome*.

which indicate that users are more informed about the remaining privacy risk. A Kruskal-Wallis test, however, does not confirm this difference significant ($p = 0.36$).

In summary, some participants are able to recall the correct privacy risk value. However, the fraction able to do so is not significant.

The correlation in group ■ QuantWhy between *PrivacyConcern* and the *DecisionOutcome* is confirmed significant by a logistic regression ($p = 0.007$). Therefore, ■ QuantWhy supports an informed decision. None of the other conditions show this effect significantly (see Table 12).

B.3.1 Result on Privacy Usability . The Kruskal-Wallis tests for influence of the conditions are not significant on *Sentiment* ($p = 0.87$), *Comprehension* ($p = 0.61$) nor the combined privacy usability ($p = 0.37$). The arithmetic means per group (see Table 10) also do not show much variation. Therefore, we do not think that the quantitative information as presented in our interfaces negatively impacts the *Comprehension* or *Sentiment* of the users.

The Kruskal-Wallis test for *Ability* is significant ($p = 0.03$). A post-hoc pairwise Wilcox test returns a significant difference between ■ QuantWhat and ■ QuantWhy. However, judging by all p-values, ■ QuantWhat seems to be the differing condition.

In the arithmetic mean and the histogram of *Ability* shows a slightly higher value for the group ■ QuantWhat.

B.3.2 Result on Donation decision. A χ^2 test for the effect of condition groups on the *DecisionOutcome* results in no significant results ($p = 0.58$ comparing 6 groups, 0.34 comparing What vs. How vs. Why, $p = 0.97$ comparing quant. vs. qual.).

The willingness to donate is highest in group ■ QuantHow (67.3%) and lowest in ■ QualWhat (53.6%).

Based on the open-text answers about the reason for the choice, the overall high willingness to donate can partly be attributed to the good cause of the collection.

B.3.3 Result on Personal Attributes. Considering the overview of the *Sentiment* (see Table 10), we see that for each question (*What*, *How*, *Why*) the standard deviation of the quantitative condition is always higher than the corresponding qualitative condition. This indicates that the quantitative conditions have more variance in the effect they have on users, which might be due to differences in the personal attributes.

B.3.4 Result on RiskEstimation. The quantitative data on *RiskEstimation* shows some evidence in support of H-1: Between the three quantitative conditions 24 out of 172 participants answered correctly with 53%, 56 out of 172 participants fell in the range 45% up to 55%, roughly remembering the given risk. In contrast the most answered percentages in the qualitative conditions are 5% (6 participants), 10% (6 participants), 20%(6 participants) and 25% (7 participants).

B.3.5 Result on Informed decision. To test hypothesis H-2, we first inspected the mean privacy aptitude depending on condition and donation decision (see Table 12). The highest difference between donors and non-donors can be found in the conditions ■ QuantHow (0.31) and ■ QuantWhy (0.74). That means, seeing the condition ■ QuantWhy the donating participants have on average a 0.74 lower privacy aptitude. This correlation in group ■ QuantWhy is even confirmed by a logistic regression ($p = 0.007$).

However, in the condition ■ QualWhy this relation is reversed. This indicates that the text used here confused the participants so much, that they acted against their privacy aptitude.

B.3.6 Additional test: Ability single-choice. Even though the results from the question *Ability*, "Which of the following best describes the way you selected the answer" were skewed, we conducted a χ^2 test, which was not significant ($p = 0.82$).

B.3.7 Additional test: Decision Recall. The decision recall question, where participants were asked to select the option they had decided on in the interactive interface shows recall rates around 80% in all conditions. A χ^2-test is not significant ($p = 0.60$). We, therefore, conclude that none of the conditions influenced the decision recall in any way.

Aside from the significant findings, we also evaluated the open-text answers provided by our study participants.

The open-text questions for "comments on the interface" and for "reasons for their decision" showed users are well aware of the privacy issues, which could arise with such collections. One participant (in ■ QualWhat) put it clearly: "It's overwhelming to think about all the way we as consumers are losing our personal data in various apps and sites that are completely out of our control."

However, the general perception of our privacy decision interface was positive. Many participants commented on the understandability of the text or on the fairness of the choice, even specifically on the quantitative information, for example "It's good that the notification provide a clear understanding of what was intended with the data and the percentage of possibility that the data could be observed and realized by another person."(■ QuantWhat)

On the other hand, multiple participants in the quantitative conditions indicated that the presented risk of "less than 53%" is too high. For example, one participant remarked about ■ QuantWhat "i would agree to share the data as long as it was anonymized but it is not very comforting at 53% probability". This shows that people were able to incorporate the quantitative information into their decision-making, however, the risk information rather seems to discourage than to inform the users, which is not the intention.

Furthermore we asked participant in an open-text question what they remember from the interface. Most often, the answers mentioned that the interface was about data sharing / donation (197), and that location data was the subject (130). Many also remembered that the data was anonymized (61) or that DP was used (46). Some participants even mentioned the risk was provided (10). This is a further indicator that the quantitative information is perceived by the users and can, consequently, help support an informed decision.

Another question asked for "What is missing in the notification". Many answers mentioned "What DP is and how it works" (9 participants). This shows that at least the conditions answering the conditions *What* did not include enough information. Other common answers included legitimate questions: "Who would have access" (8), "How long will the data be stored" (9) and "How the data is used and stored" (12). We used these comments to improve the interfaces for experiment 2.

B.3.8 Additional tests: Ability *and* Awareness. In addition to these hypotheses, we also analyzed the remaining measurements: The results from the single-choice question regarding *Awareness* "How carefully did you consider the information?" and the question regarding *Ability* "Which of the following best describes the way you selected the answer?" are highly skewed but consistent across conditions (see Table 13). We realize that these questions might have an obvious socially desirable answer option, which might have let the workers to mistake this question for a badly written attention check. Therefore, do not think the results for this question are reliable. In the following experiment 2, we measured both of these concepts with Likert-items instead.

B.4 Discussion

B.4.1 Focusing on vulnerable users. We found some effects dependent on personal attributes. Since these characteristics are hard to measure for each privacy decision, we should instead primarily support vulnerable users, for example, with lower *Numeracy* or *NeedForCognition*. For representing risks in other contexts, such as therapy decisions, visualizations have shown to be beneficial(e.g., [10, 31, 39]). To evaluate these effects on risk visualizations, we include visualizations representing the privacy risk in the main experiment.

B.4.2 Interface variances. Despite our interesting finding in this first experiment, only a few results are statistically significant. We attribute this to the fact, that the quantitative information only accounted for a small part of the explanations. We already remarked that decision interfaces need to be adapted more towards the qualitative information in the hope of supporting informed decision better. In this pre-study, we limited ourselves to explanations, which are similar in style and length to traditional qualitative explanations of DP. With the further adaptations arising from the affordances of the quantitative information rather than mimicking the traditional decision interfaces we hope to see the influences of the quantitative information stronger in the results.

Condition	a) Randomly	b) Informed	c) Prior preference	d) Other
■ QualWhat	1	50	5	0
■ QualHow	0	46	5	0
■ QualWhy	0	44	6	1
■ QuantWhat	0	62	11	1
■ QuantHow	0	47	5	0
■ QuantWhy	0	40	6	0
Total	1	289	38	2

Answers to "Which of the following best describes the way you selected the answer?", answer option were: a) I picked randomly, b) I picked based on my preference in accordance with the presented information, c) I knew what I wanted to pick before the information was shown, d) Other.

(a) Distribution of the *Ability* single-choice questions.

Condition	a) Skipped it.	b) Skimmed it.	c) Read it.	d) Carefully.
■ QualWhat	0	2	17	37
■ QualHow	0	2	13	36
■ QualWhy	0	1	19	31
■ QuantWhat	0	1	18	55
■ QuantHow	0	1	14	37
■ QuantWhy	0	0	12	34
Total	0	7	93	230

Answers to "How carefully did you consider the the information?", answer options were: a) I skipped it and just pressed any button. b) I skimmed over it. c) I read it normally. d) I read it carefully, some parts multiple times.

(b) Distribution of the *Awareness* single-choice questions.

Table 13. Distribution of single choice measurements.

B.5 Insights from the pre-study on the study design of the main study

The basic design of the pre-study has been retained in the main experiment. In addition to minor adjustments to the scenario texts, based on the feedback we got in the pre-study, we considered two changes: we decided for a mix of interactive and static (visualization and textual) user interfaces and we extended the questions for the privacy usability.

Our used visualization of privacy risk is inspired by medical risk communications. We believe these decision scenarios are similar: In both cases the patient/user has a binary choice which has benefits (treatment of a condition or data-driven results and features) but might also have a negative effect (side effects or privacy loss). The risk of these negative effects can be calculated, but whether or not they happen is uncertain.

Besides the inclusion of the augmented elements, the interfaces for the main experiment were designed with the following considerations:

All interfaces in the main experiment, with the exception of ■ BLQual include the privacy risk without donation and the privacy risk with donation. Accordingly we split the design of the interface vertical. This way all information about the risk without donation could be displayed above the button to decline donation, while all information about the risk with donation could be displayed above the donation button. As this is consistent throughout the conditions, this does not influence the validity of our measurements.

Due to the inclusion of the linked sliders, we also needed to modify our question regarding *RiskEstimation*. In the pre-study we simply asked for an estimate how high the privacy risk would be when donating the data. This time we had to ask for a privacy risk based on the privacy risk without donation. To keep the answer manageable even for the non-interactive conditions, we chose the displayed risk of 20% here and asked what the privacy risk with donation would be in this case. This way, we are measuring the recall of the information rather than a transfer of learnings onto a new situation.

Since the probability of someone guessing a visit without donation is difficult to estimate by laypeople, we included the popularity of a place as an intuitive analogy. The more popular a place is in the general population, the higher is the probability that someone would guess that the user had also been there. To design the interface, where the user can select the popularity of a place, we took inspiration from the “Popular times” interface element in google maps, which displays the popularity of a place on a scale between “not busy” and “as busy as it gets”. We know that this is only one of many factors, which determine the probability of someone guessing a visit without donation. For example, gender and age of a user could also alter this probability. However, for simplicity towards the user we decided not to include any other factors.

As in the first study, we preface the decision interface with a first page containing all necessary information about the proposed donation.

C STUDY RESOURCES

C.1 Pre-study

C.1.1 Welcome Text. Welcome to the survey "Traffic data for city improvements".

Please read the following introduction carefully:

In this survey we want to evaluate how users feel about sharing their location data to improve the city. We hope that you can give us an inside into the users’ perspective.

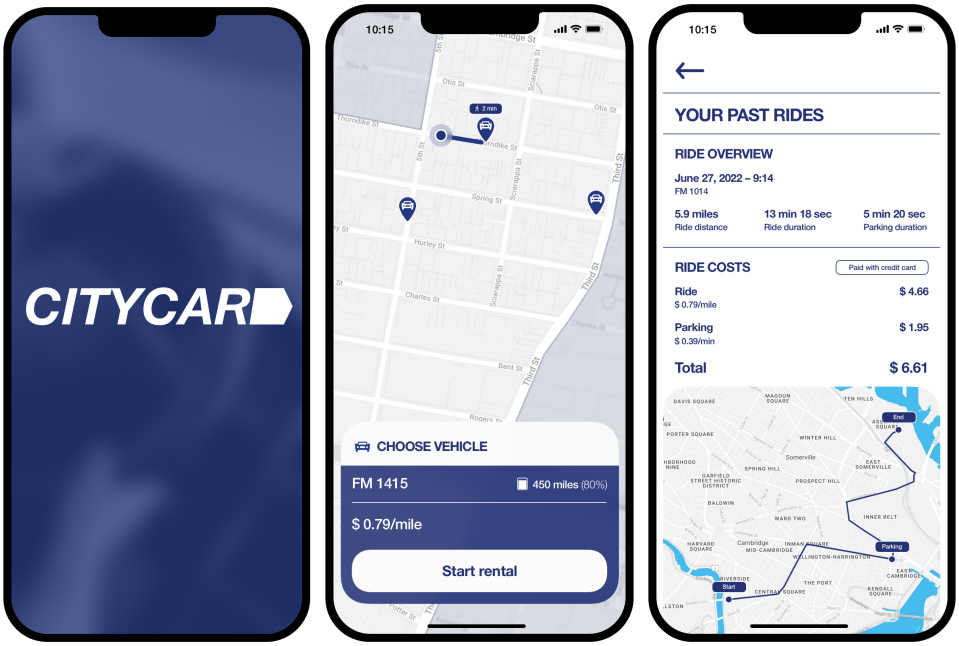
For our study it is important, that you read all texts in the following survey carefully and answer the questions honestly. This survey consists of 10 pages with 20 questions (some with a few subquestions) in total. We expect a working time of about 15 Minutes. In addition to the compensation for the HIT, we will be paying a bonus of $0.50 for coherent submissions. In this survey, we do not ask for sensitive information. We do, however, automatically collect your IP address and the MTurk-IDs for quality assurance. After checking the data and approving the HIT towards Amazon Mechanical Turk, the identifying information will be deleted from the dataset. We continuously store your responses during your participation. If you change your mind at any time and do not want to continue to participate in the study, you can delete the answers already stored by clicking the button "Exit and clear survey". Please do not use the browser navigation (e.g., the "back" button) during the survey.

Please note: We will be screening the answers for careful reading. If we see evidence of inattentive submissions or repeated participation, we will reduce the bonus or reject your submission, depending on the severity.

C.1.2 Scenario Introduction. Imagine the following scenario as motivation for the whole survey:

"CityCar" is a car-sharing service in your city. With it you can rent cars, which are distributed throughout the city, for short time use. You have used CityCar for some time and you have booked cars on multiple occasions to go to different locations. At the top, you can see some examples of the CityCar app screens. The app also stores data about your trips. You can, for example, review the exact route you drove on a given date and time. Last week, the CityCar app was updated.As one new change, CityCar would like to share statistics about the number of people driving through various locations in the city with the city’s department of transportation.Since this was not part of the terms and conditions before, the app now shows a notification including a choice on whether you want to participate in this data sharing agreement or not.You will see this notification on the next page.

The following screenshots were displayed alongside the scenario description to strengthen the immersion.



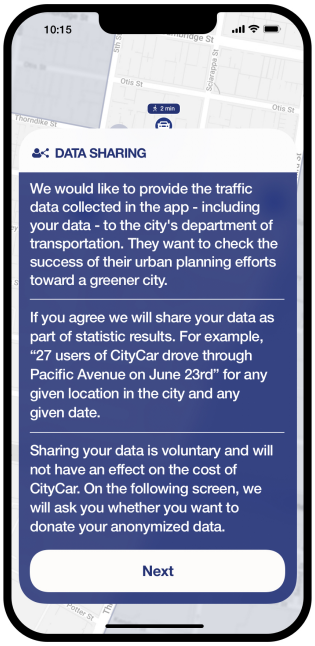
On the same page the following three comprehension questions were asked with randomized answer order:

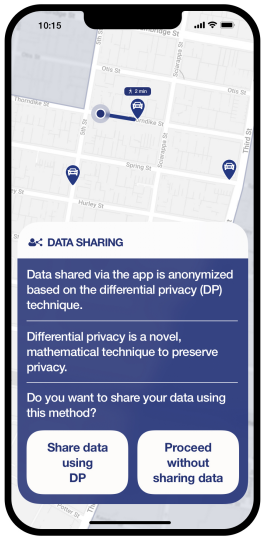
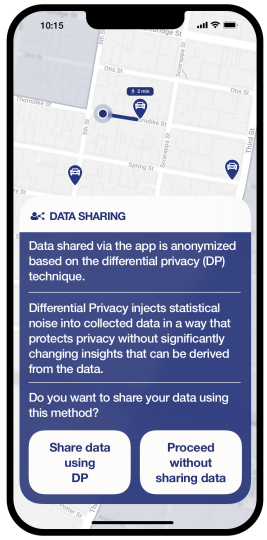
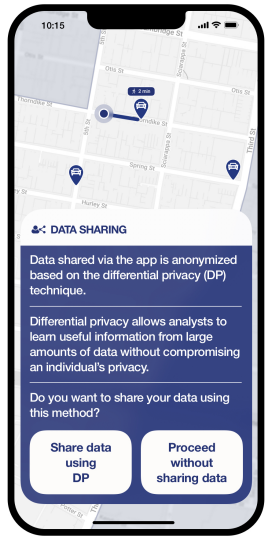
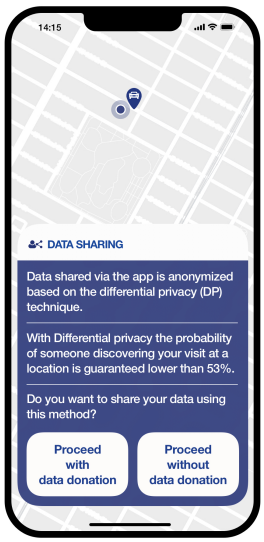
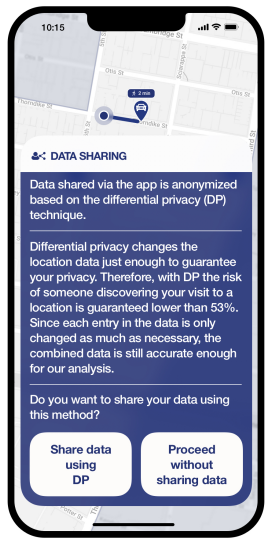
- What kind of app is the Scenario about? (**Car Sharing app**, Financial app, Camera app)
- What kind of data is mentioned in the scenario? (**Location data from trips I have taken**,My recorded sleeping schedule,Weather information)
- According to the scenario, who might get your data, if you agree to the notification? (**The city's department of transportation**, My employer, All people in your contact list)
- In the scenario you have used the app to go to different locations. Imagine and write down one location, where you could have gone using the app. (open text)

C.1.3 Conditions. The following interactive application shows the mentioned notification in the CityCar app.

Please read the text of the notification carefully and interact with the application to make a choice as you usually would with the available information. The notification consists of several screens.

You can only progress to the next page of the survey after you have made your choice in the application.


10:15
DATA SHARING
We would like to provide the traffic data collected in the app - including your data - to the city's department of transportation. They want to check the success of their urban planning efforts toward a greener city.
If you agree we will share your data as part of statistic results. For example, "27 users of CityCar drove through Pacific Avenue on June 23rd" for any given location in the city and any given date.
Sharing your data is voluntary and will not have an effect on the cost of CityCar. On the following screen, we will ask you whether you want to donate your anonymized data.
Next

	What	How	Why
Qual			
Quant			

If you have any thoughts about the notification, write them down here. (open text)
†

C.1.4 Privacy usability survey items. The order of questions / answer options were randomized where possible and suitable.

- (InterfaceRecall) What do you remember about the content of the notification you saw on the last screen. Please summarize in your own words? [open-text]
- (Needs-openText) Do you feel there was information missing in the notification? If so, what information was missing? [open text]
- (RiskRecall) Assume after reading the notification, you agreed to sharing your data.With the used method called "differential privacy", how high would you estimate the risk of someone discovering your visit of a location? [Range 1-100]

- (Comprehension1) I understand what the notification is about. [7-Likert]
- (Comprehension2) I find it easy to understand the notification. [7-Likert]
- (Sentiment1) I am satisfied with the information provided. [7-Likert]
- (Sentiment2) I am confident that I took the right choice. [7-Likert]
- (Ability1) I feel informed about the collected data. [7-Likert]
- (Ability2) I feel informed about the used method "differential privacy". [7-Likert]
- (Ability3) I feel capable of making the decision with the provided information. [7-Likert]
- (Awareness-singleChoice) How carefully did you consider the the information? [I skipped it and just pressed any button / I skimmed over it / I read it normally / I read it carefully, some parts multiple times]
- (ChoiceRecall) Which option did you choose? [Share data using DP / Share data without using DP / Proceed without sharing data / I do not remember]
- (Ability-openText) Why did you decide to choose this option? [open-text]
- (Ability-singleChoice) Which of the following best describes the way you selected the answer? [I picked based on my preference in accordance with the presented information / I knew what I wanted to pick before the information was shown / I picked randomly]

C.2 Main Study

C.2.1 Welcome Text. In this survey we want to evaluate how users feel about sharing their location data to improve the city. We hope that you can give us an inside into the users' perspective.

For our study it is important, that you read all texts in the following survey carefully and answer the questions honestly. This survey consists of 10 pages with 18 questions (some with a few subquestions) in total. We expect a working time of about 15 Minutes. In addition to the compensation for the HIT, we will be paying a bonus of $0.50 for coherent submissions. In this survey, we do not ask for sensitive information. We do, however, automatically collect your IP address and the MTurk-IDs for quality assurance. After checking the data and approving the HIT towards Amazon Mechanical Turk, the identifying information will be deleted from the dataset. We continuousely store your responses during your participation. If you change your mind at any time and do not want to continue to participate in the study, you can delete the answers already stored by clicking the button "Exit and clear survey". Please do not use the browser navigation (e.g., the "back" button) during the survey. At the end of the survey we will report the completion of the survey automatically back to Amazon Mechanical Turk. No manual code is necessary.

Please note: We will be screening the answers for careful reading. If we see evidence of inattentive submissions or repeated participation, we will reduce the bonus or reject your submission, depending on the severity.

We have had some reports that ad-blockers might interfere with our prototype. Please deactivate them before you proceed to make sure that you can complete the survey without issues.

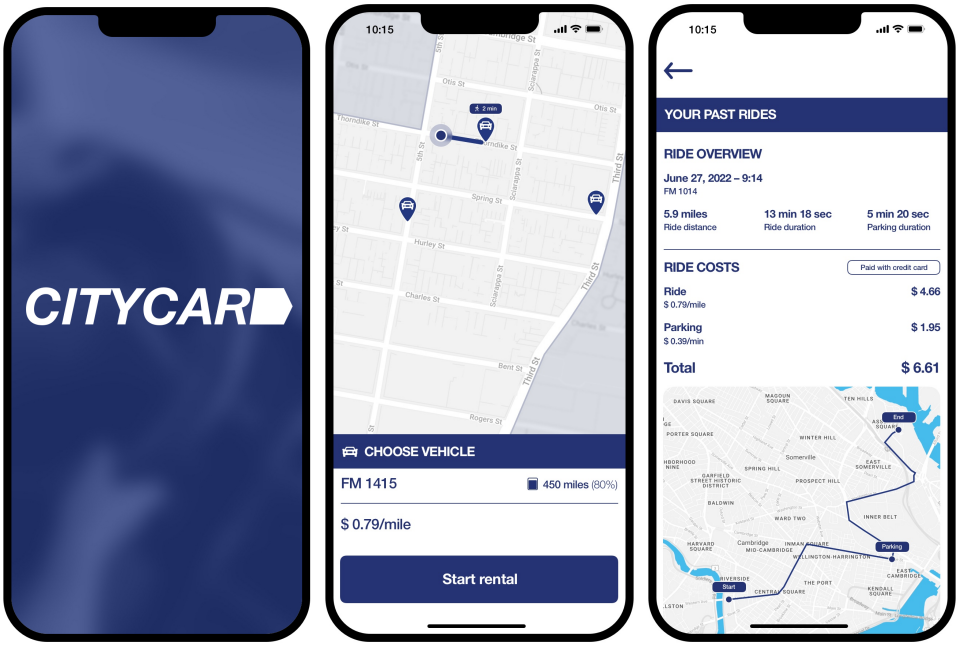
Click Next to begin

C.2.2 Scenario Introduction. "CityCar" is a car-sharing service in your city. With it you can rent cars, which are distributed throughout the city, for short time use. You have used CityCar for some time and you have booked cars on multiple occasions to go to different places. At the top, you can see some examples of the CityCar app screens. The app also stores data about your trips. You can, for example, review the exact route you drove on a given date and time.

Last week, the CityCar app was updated. As one new change, CityCar would like to share statistics about the number of people driving through various places in the city with the city's department of transportation. Since this was not part of the terms and conditions before, the app

now shows a notification including a choice on whether you want to participate in this data sharing agreement or not. You will see this notification on the next page.

The following screenshots were displayed alongside the scenario description to strengthen the immersion.



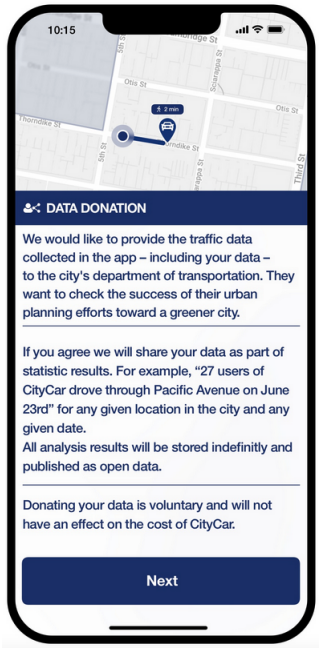
On the same page the following three comprehension questions were asked with randomized answer order:

- What kind of app is the Scenario about? (**Car Sharing app**, Financial app, Camera app)
- What kind of data is mentioned in the scenario? (**Location data from trips I have taken**,My recorded sleeping schedule,Weather information)
- According to the scenario, who might get your data, if you agree to the notification? (**The city's department of transportation**, My employer, All people in your contact list)
- In the scenario you have used the app to go to different places. Imagine and write down one place, where you could have gone using the app. This place does not need to be from the map above, but could be any place, where you could imagine yourself going to using CityCar. (open text)

C.2.3 Conditions. The following interactive application shows the mentioned notification in the CityCar app.

Please read the text of the notification carefully and interact with the application to make a choice as you usually would with the available information. The notification consists of several screens.

You can only progress to the next page of the survey after you have made your choice in the application.



If you have any thoughts about the notification, write them down here. (open text)

C.2.4 Privacy usability survey items . The order of questions / answer options were randomized where possible and suitable. Questions with a * have inverse rating.

- (RiskRecall) Imagine, one of the trips you did using CityCar was to a hospital. Since the hospital is not too busy, without sharing the data others would guess you went there with a probability of 20%. How high do you thing the probability is of others guessing you have been to that hospital, when you share your data using the proposed anonymization technique? [Range 1-100]
- (Comprehension1) I understand what the notification is about. [7-Likert]
- (Comprehension2) I find it easy to understand the notification. [7-Likert]
- (Comprehension3*) I still have many questions about privacy risks of the donation.
- (Sentiment1) I am satisfied with the information provided. [7-Likert]
- (Sentiment2) I am confident that I took the right choice. [7-Likert]
- (Ability1) I feel informed about the collected data. [7-Likert]
- (Ability2) I feel informed about the used method "differential privacy". [7-Likert]
- (Ability3) I feel capable of making the decision with the provided information. [7-Likert]
- (Ability4*) The option I wanted was not available. [7-Likert]
- (Awareness1) I know which options are available. [7-Likert]
- (Awareness2*) I am unsure what the available options mean. [7-Likert]
- (Need1*) I need more information to decide. [7-Likert]
- (Need2*) I need better security for my data. [7-Likert]
- (AttentionCheck1) My submission should be rejected. [7-Likert]
- (InterfaceRecall) What do you remember about the content of the notification you saw on the last screen. Please summarize in your own words? [open-text]
- (ChoiceRecall) Which option did you choose? [Share data using DP / Share data without using DP / Proceed without sharing data / I do not remember]

- (Ability-openText) Why did you decide to choose this option? [open-text]
- (AttentioCheck2) What would you like us to do with your submission on Amazon Mechanical Turk? [You can reject my submission, I did not pay attention., **I am taking the questions serious, you can accept my submission.**, I have not read some parts in the beginning. Let me do the whole survey again.]
- (Needs-openText) Do you feel there was information missing in the notification? If so, what information was missing? [open text]

Received January 2023; revised July 2023; accepted November 2023