Caring about Sharing: User Perceptions of Multiparty Data Sharing

Bailey Kacsmar
University of Waterloo

Kyle Tilbury
University of Waterloo

Miti Mazmudar
University of Waterloo

Florian Kerschbaum
University of Waterloo

Abstract

Data sharing between companies is typically regarded as one-size-fits-all in practice and in research. For instance, the main source of information available to users about how a company shares their data is privacy policies. Privacy policies use ambiguous terms such as 'third-parties' and 'partners' with regard to who data is shared with. In the real-world, data sharing has more nuance than is captured by these overarching terms. We investigate whether users perceive different data sharing scenarios differently through an online survey with scenarios that describe specific types of multiparty data sharing practices. We determine users' perceptions when explicitly presented with how their data is shared, who it is shared with, and why. We show that users have preferences and that variations in acceptability exist which depend on the nature of the data sharing collaboration. Users caring about sharing, necessitates more transparent sharing practices and regulations.

1 Introduction

Collaborations and contracts between companies increasingly involve the disclosure of data. Mastercard sold a stockpile of transaction data to Google to track whether Google ran ads that led to a sale at a physical store [9]. Data moving between companies is not limited to direct sales or targeted advertising. Data sharing can also occur through the purchase or merging of companies such as Google purchasing Fitbit [26]. Although Google's purchase of Fitbit includes a statement that the health and wellness data will not be used for Google advertising, it does not clarify how other data could be used and whether the health and wellness data can be used in ways not related to advertising. Through a legal request one user determined Tim Hortons' loyalty program app shared its users' precise location regularly with a third-party (Radar Labs Inc.) that identified users' home, work, travel destinations, as well as visits to a competitor. The third-party ultimately shared the users' precise locations with Tim Hortons' parent company, Restaurants Brand International [46].

There are even collaborations between technology and health companies that can and do occur. There are collaborations between Google and Ascension [69], Microsoft and Providence St. Joseph Health [12], and COVID-19 contact tracing tools [29]. Some of these collaborations only include the use of services, but others require sharing data in some form to perform computations, including machine learning. In addition to these forms of collaborations, the line dividing health and technology companies is blurring with the development of new services such as Amazon Care[1] and Telus Health[2]. Amazon's health care service specifies that patient information is exclusively used for supporting Care Medical, however, it is unclear how this could affect users' understanding and perceptions of health care data being used by technology companies.

We refer to companies that acquire or share data in these ways as collaborating for multiparty data sharing. Mechanisms to perform privacy-enhanced multiparty data sharing exist in the literature as secure computation, such as private set intersection [13, 62] and federated machine learning [47, 73]. While companies, such as Microsoft and Google, may choose to use privacy-enhanced computation in their collaborations, how to convey these practices fairly to users and indeed how users feel about enhanced computations is a question we address within this paper. Multiparty data sharing can be one-way, where only one of the companies in the exchange acquires data, two-way where the parties involved pool their collective data, or an exchange involving more than two-parties.

Although privacy policies should contain information for users about the data a company collects and how that company uses the data, such documents are hard to read and rarely read, making them inaccessible to users [44, 56]. Users who trust one company with their data may not understand that their data could be shared or purchased nor the corresponding privacy risks. However, it can be confusing for people reading privacy policies about sharing their data to understand what their data will be used for and make informed decisions based

[1] Available across the United States, https://amazon.care/
[2] Manages Canadian health care records, https://www.telus.com/en/health

on their perceptions of it.

Research Questions. We study users' perceptions of multiparty data sharing via an online survey. We analyze users' perceptions of various data sharing events (termed as scenarios), what potential controls users want, and identify avenues for improving regulations and engineering better systems to meet those needs. To this end we address the following research questions (*with salient results emphasized*):

RQ1: How does the overall acceptability vary across different types of multiparty collaborations? How do the types of companies involved further impact it?

> *The overall acceptability of multiparty data sharing is lower for collaborations that are not reciprocal. The inclusion of a health company in non-reciprocal collaborations is even less acceptable. (Section 4.2).*

RQ2: How does acceptability vary in multiparty data sharing for different user controls (consent, purpose, retention)?

> *Across user controls, preferences for consent vary the most between collaboration types, however, opt-in consent is, generally speaking, the most acceptable. (Section 4.1 and 4.2)*

2 Related Work

Privacy Perceptions. Users' perceptions of privacy have shown many changes over the years and so have their preferences [3, 17, 35]. Past work has often focused on data sharing for advertising purposes [15, 45, 77, 80], with the additions of privacy perceptions for IoT, mobile, and smart homes in more recent years [4, 24, 39, 51, 74, 78]. Regardless of whether the data is shared intentionally or unintentionally leaked via a data breach, user perceptions tend to perceive such treatments of their data negatively [25, 31, 43, 45, 64, 72].

Even when a users' data is only disclosed to a single company, different contexts influence what trade-offs users are willing to make at the expense of their privacy in terms of benefits, or how their data is being used [5, 7, 8, 19, 53, 76]. Further complicating matters are 'third parties' or 'partners' that data can be shared with. Users do not understand what these third parties are and how their data can be shared with these parties [64]. In cases where such terms are used in a privacy policy, it can remain ambiguous to users as to who their data can be shared with and thus prevent them from making an informed decision [22, 41].

In general, survey methodology research cautions that respondents may have difficulty predicting their behaviour or be inclined to report the perceived desirable response [63]. In the case of security research, recent work from Redmiles et al. [65, 66] shows that surveys can provide meaningful results for general constructs. We use a similar survey design to previous work on acceptability for IoT and data breaches [4, 32].

Thus far, research has primarily treated third-parties or partners in much the same manner as privacy policies do. Third parties are treated as monolithic black-box entities that can take many forms and treat data in different ways. Ebert et al. [20] include 'data sharing' among the legal principles of their study, but again it is left as a general concept. In this work, we build on past investigations into user perceptions of data sharing by specifically providing respondents with scenarios based on real-world examples of how their data could be shared with one or more other parties. We revisit whether policy and design decisions relating to these continually evolving multiparty data sharing scenarios can rely on past results, or whether different structures of data sharing result in different perceptions that need to be addressed.

User Controls and Accessibility. Though not strictly targeting the multiparty data sharing setting, methods to provide users with controls include toggles [28], permission settings [33, 40], and privacy nudges [2]. Despite this, such controls can still be hard for users to understand and use [1, 2, 6, 21, 24, 27, 67]. Difficulties associated with providing users with controls to set their own privacy preferences are not limited to the design of such controls. That is, users can be manipulated or tricked such that opting out of behavioural based advertising is limited [27, 38]. With this in mind, we specify explicitly details users may want to have user controls for in the survey. These aspects for potential controls include what purposes users find acceptable for their data, how they want to be informed (to get consent), and how long they will permit their data to be used in this way.

Park and Sandhu proposed *usage control* to generalize these controls and the idea that beyond privacy policies for all users there can be individual controls required for each user [60, 61]. Ebert et al. [20] referred to usage control variables such as storage and retention as legal preferences in their analysis. They do not focus on types of data sharing, but instead on the effect of the contexts of a fitness tracker versus a rewards card. Similar to Park and Sandhu's application to social media controls, in the case of multiparty data sharing, there are many potential parties that users may or may not want to share their data with and the type of data they are willing to share may vary for different companies [60].

Law and Policy. There are a number of regulations, both old and more recent, that apply to the privacy of users' data [16, 54, 58, 68]. However, they do not necessarily provide protections for all of the possible treatments of users' data [42, 57]. Even with the recent California Consumer Privacy Act (CCPA[3]), the right to opt-out of data sales does not

[3] https://oag.ca.gov/privacy/ccpa

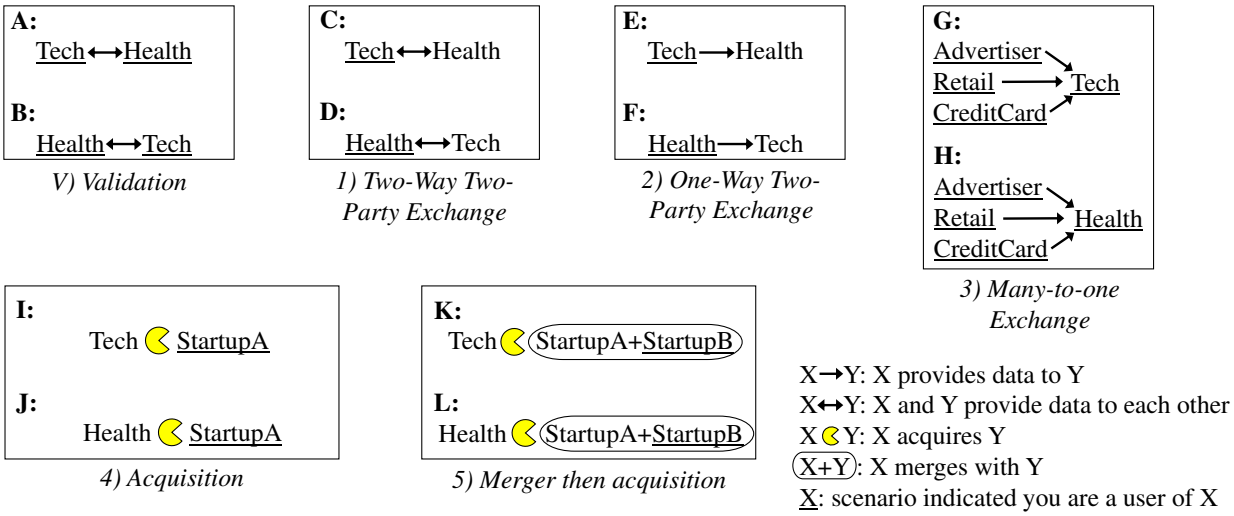

Figure 1: Overview of scenarios (A-L) presented in our survey and collaboration types (V, 1-5) that we investigate. For reference, Scenario C, "TechForYou is a large internet company that offers a search engine, email accounts and smartphone platforms to users. GoodHealth runs a chain of hospitals across the country and stores health data for millions of patients during its day-to-day operations. TechForYou and GoodHealth will share the customer data they hold with one another. You are a customer of TechForYou".

stop companies from manipulating users such that it is difficult or unappealing to opt-out [57]. Furthermore, it can only prevent companies from *selling* users' data, it does not prevent companies from sharing or exchanging data with other companies or affiliates. Multiparty data sharing needs to be better understood with respect to user preferences and perceptions to produce more specific regulations addressing all types of collaborations.

3 Methodology

We collected 1025 responses to our online survey through SurveyMonkey in March 2021. Each participant was compensated $3.04 for their response and spent, on average, four minutes to complete the survey. Our final participant set is $N = 916$ after excluding the 109 respondents that failed an attention checking question. Respondents could exit the survey at any time and could skip any question in the survey. Our study received ethics approval from our institution's office of research ethics (ORE). See survey at https://bkacsmar.github.io/files/SurveyUsenix2022.pdf.

3.1 Survey Design

Prior to the final survey, we ran a pilot study with $N = 26$ participants. We asked participants in the pilot study what they *would* agree to in a multiparty data sharing setting. The pilot had one scenario, between a technology company and a financial institution, to introduce the concept of multiparty data sharing. We used a free-form text response question to gather participants initial thoughts on this scenario and what could influence their perceptions. Our pilot study free-form responses report a desire for user controls that we incorporate into our final survey.

3.2 Survey Structure

Each survey provides one of twelve scenarios to respondents followed by a series of questions on user controls and privacy mechanisms. The twelve scenarios are categorized by the number of companies and which companies send and receive data (see Figure 1 for an overview of the scenarios). Each collaboration scenario is based upon real-world examples from Canada and the United States. For each question, excluding the free-form responses and correctness checks, respondents select a value from a five-point semantic differential [59] acceptability scale: "Completely Unacceptable", "Somewhat Unacceptable", "Neutral", "Somewhat Acceptable", and "Completely Acceptable" as in Apthorpe et al. [4]. Respondents rate acceptability given specified variables (shown as (a) through (k) in Table 1). For analysis, the values we assign to our scale are 1-5 where 1 is "Completely Unacceptable" and 5 is "Completely Acceptable".

3.3 Nature of Collaboration

The nature, or type, of the collaboration encodes the number of participating companies and how the data flows between those companies. Notably, we test the inclusion of a health company versus a technology company within the collaboration types. To check whether the ordering of the companies influences respondents, we include two identical scenarios, Scenarios A and B, where the only difference between them is the order in which the health and technology company are introduced. The following defines our five collaboration types with examples.

Two-way, Two-party Exchange (Type 1). In a 'two-way two-party exchange' there are two participating companies. During the exchange, the two companies send data to and receive data from one another. Four of our scenarios are a 'two-way two-party exchange' (Scenarios A-D). We use two of these four scenarios (C and D) in our collaboration type analysis, and we use the remaining two (A and B) for validation only. Examples of such a collaboration would be two companies that perform a computation, such as private set intersection dual execution, that uses extended methods to ensure both companies receive the result [48].

One-way, Two-party Exchange (Type 2). Perhaps the most conventional and well understood collaboration type is the 'one-way two-party exchange' (Scenarios E and F). In this case there are two companies where one acquires data from the other, perhaps in exchange for a monetary amount. Such collaborations could be two parties computing the intersection of data they hold where one party receive the resulting intersection [9]. Other examples of this collaboration type include insurance telematics (use-based insurance) [37] and computing joint cyber threats [10].

Many-to-One Exchange (Type 3). A company may acquire data related to their users from multiple other companies or data brokers. We include two scenarios of this form (Scenarios G and H) with a total of four participating companies. In these 'many-to-one' scenarios, three of the companies are providing data to one other company. This structure in practice, could of course take many forms depending on the number of participating companies and which companies provide or receive data. We chose this structure based on the real-world examples of companies acquiring data from a series of other 'partner' companies. For example, advertising networks may acquire data from any number of sources, including other apps, websites, and their competitors, depending on users' permission settings [23, 34].

Acquisition (Type 4). In our 'acquisition' scenarios, a single party purchases, or acquires, another (Scenarios I and J). Examples of acquisitions relating to data sharing include Google acquiring Fitbit [26], Microsoft acquiring LinkedIn [11], and WealthSimple acquiring SimpleTax [30]. The company SimpleTax promised to never sell its users' data, however, this did not account for when the company itself was sold. In such acquisitions the data held by a company may be included in its assets and upon purchase becomes available to the acquiring company depending on the applicability of regulations such as the FTC Act [4]. In the case of the purchase of SimpleTax, the explicit promise to never share its users' data was removed from its privacy policy going forward (only affecting data since the purchase) [30].

Merger then Acquisition (Type 5). Generally speaking, the difference between a merger and an acquisition can be thought of as two companies equally choosing to come together as one company in a merger versus one company taking ownership of another during an acquisition. In both cases, assets, which may include data, are consolidated in some manner. We include a scenario where two startups merge, forming a new company, which is then acquired by a third company (Scenarios K and L). In this case it is possible for an individual to have shared their data with one of the original start-ups, with no expectation that these two additional companies they have no connection with would come to possess it. Sometimes a merger with other acquired companies can be a part of an acquisition, and sometimes they are separate events; but they are both possible outcomes for smaller companies [75].

3.4 User Controls

Usage control enforcement mechanisms are components that can be written into designs or regulations which give users the ability to specifically set what they agree to. We use eleven usage control variables (listed in Table 1 as (b) through (k)) within our survey. The variables are selected from responses to our pilot study and real-world examples. We investigate how purpose of use, data retention, and the method of acquiring consent or notifying users can impact the acceptability of multiparty data sharing scenarios.

Purpose. There are three purposes of data sharing in our survey. These purposes are 'generating advertising revenue', 'providing users with a monetary reward' (e.g., free service, reduced rate [37], or gift-card), and 'improving services' [21, 71]. Note that while we included a variety of examples within the monetary return question, these examples may not have been viewed the same by all respondents. That is, respondents may have interpreted free service as an advertising funded service rather than an additional bonus service. Respondents that interpreted a free service in such a way may have been less inclined to consider the service as a monetary benefit in the same sense as a gift card or discount.

[4] https://www.ftc.gov/about-ftc/what-we-do/enforcement-authority

Data Retention. Users are known to have misconceptions about what happens when their data is deleted [50]. To prevent misconceptions, our data retention questions provide an explicit duration for each of the three retention questions. The duration values include keeping the data 'indefinitely', keeping the data for a 'specified duration' of time (e.g., three months, one year, etc.), or more ambiguously, keeping the data until the company (or companies) is 'finished using it'. We note here that the deliberate inclusion of the more ambiguous 'after they finished using it' does leave the potential for respondents to interpret it differently. Data may be used by companies in computations such as aggregate statistics, private set intersection, or to train machine learning models. Respondents may differ in whether they believe that continuing to use computations on data means that a company continues to use the data. We left interpretation of when the use ends open to the respondents.

Notification and Consent. We avoid directly asking participants whether they would consent, which would likely be influenced by perceived socially desirable behaviour [36]. Instead, we focus on notification strategies that inform users. Depending on local laws and regulations companies use a variety of methods to inform (or not inform) users how personal data can be used. We select a subset of those methods to evaluate any potential influence on the acceptability of multiparty data sharing.

In our survey we include four questions relating to informing users. First, 'concealed consent', where no formal notification is provided, and the respondents learns of the collaboration via the media. Second, there is 'assumed consent' where an email or app notification is sent which indicates to the user that by continuing to use the service, they are agreeing to the data sharing. Third, there is 'opt-out consent' that provides an option to specifically disallow the data to be shared. Fourth, 'opt-in consent', where the data is not shared by default and requires explicit permission.

3.5 Privacy Mechanisms

Our survey includes questions on how acceptability is influenced by privacy mechanisms. The five privacy mechanisms we included are local differential privacy (LDP), central differential privacy (CDP), data anonymization, data aggregation, and encryption [52, 79, 81]. Respondents each received one of the five privacy mechanisms and rated the acceptability of the data sharing scenario, if it were to include that privacy mechanism. To validate that respondents understood the mechanisms, our research team manually generated informal descriptions of the mechanisms, and the survey asks respondents to match their privacy mechanism to the most accurate description. This unfortunately suggested respondents had low comprehension of the privacy mechanisms provided to them. Thus, we exclude privacy mechanism related results.

3.6 Demographics

We report an overview of demographics rounded to the nearest percent. All survey respondents are located in the United States. Of the total $N = 916$ participants, when asked to specify their gender, 47% specified man, 50% specified woman, 1% specified non-binary, 2% preferred not to say, and less than one percent chose to self-describe. Respondents specified an age range with 17% of respondents selecting 18-24, 22% 25-34, 15% 35-44, 21% 45-54, 22% 55-64, and 3% prefer not to say. In terms of employment, 70% reported the industry of their current form of employment, 18% reported being unemployed, 5% as student, and 6% responded with prefer not to say. The industries reported by those that were employed were diverse with the most frequent industry being education at 10%. A slight majority of participants reported completing a degree at 59% (bachelor, graduate, or associate). The remainder of participants education can be broken down as 23% with some college but no degree, 14% completed high school, 3% less than high school, and 1% prefer not to say.

3.7 Limitations

We recognize that our scenarios are not all encompassing of multiparty data sharing. We have included varying companies, data types, and structures such that it may guide the focus of future work. The companies we selected for this study include a focus on health companies and health data. This focus may have influenced respondents in hard to predict ways based on respondents presumptions about how health data is regulated as well as their willingness to share such data. Further, we use a semantic differential acceptability scale, but acknowledge that such scales could still result in bias over the duration of the questions presented. Responses were gathered while the COVID-19 pandemic was ongoing [14]. We cannot know how this may have affected respondents' answers, but it may have contributed to the higher unemployment percentage.

We further note that our participants, from across the United States, are WEIRD (Western, educated, industrialized, rich and democratic) [70]. We do not presume to make global assertions from our study but instead show that even within this group there is a diverse set of expectations and preferences not currently supported by technology nor required by regulations. Our scenarios are based on examples located in North America, where our respondents live. This is critical as different regions, even within WEIRD participant pools, have different existing laws and expectations. For example, EU citizens already have different protections than non-EU citizens. Finally, we acknowledge the potential for bias towards perceived socially desirable behaviour [66]. We attempt to mitigate this bias by using the more neutral term 'acceptable'. We ensure there are no mentions of privacy until the end of the survey, and we give participants the opportunity to provide their own views in free-form text.

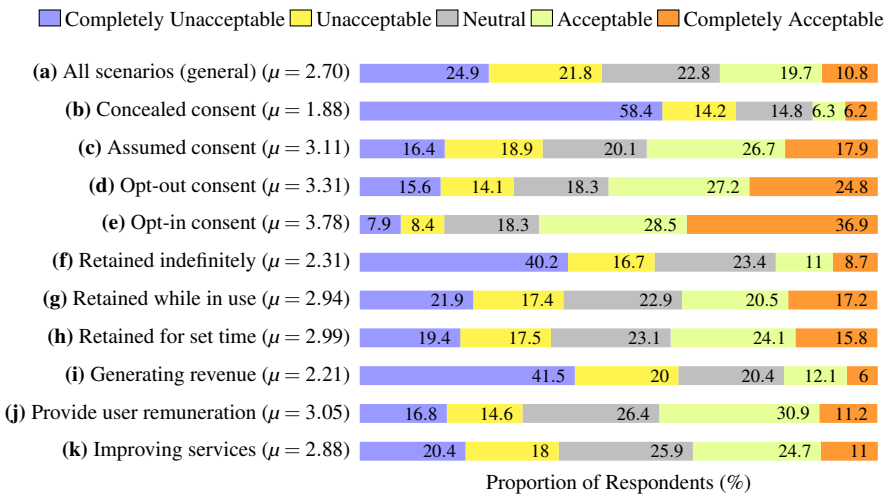

Figure 2: The acceptability distribution of multiparty data sharing across all scenarios for each variable. Acceptability is measured on a five-point semantic difference scale and each segment corresponds to the proportion of respondents who select that level of acceptability ($N = 916$).

4 Results

We first present respondents' overall perceptions of multiparty data sharing and related user controls. Second, we examine the differences in acceptability between and within each sharing type. This is followed by our analysis of demographic based variations in perceptions. Finally, we present an exploration of respondents' free-form responses. Recall, the labels for the variables and collaboration types are found in Table 1.

The results we present highlight our statistically significant findings. For interpretability, we report mean values for acceptability in this section. When we refer to statistically significant differences, we are not referring to these means, but include them as the statistical mean ranks are less interpretable. We use non-parametric statistical tests, which use mean ranks, as our data is not normally distributed. This decreases the risk of incorrectly saying a difference is significant. All statistical results use a significance level of 0.05. The details of our statistical analysis are in Appendix A.

Note that although we asked respondents questions with respect to how privacy mechanisms could impact acceptability, unfortunately respondents' comprehension of the privacy mechanisms definitions was low (based on our validation definitions) and so we exclude the acceptability results from this work. Please refer to Appendix B for details of respondent comprehension.

4.1 Overall Perceptions

We begin by determining a base understanding of how acceptable respondents find multiparty data sharing and our defined variables, regardless of the type of collaboration they received. The acceptability of the data sharing scenario in 'general' (a), is completely unacceptable or somewhat unacceptable to 45% of respondents. Without additional details about the collaboration, participants respond slightly more towards the unacceptable end of the scale, but almost 30% of respondents do find it to be at least somewhat acceptable. The distributions of how acceptable respondents found each variable are shown in Figure 2.

Within Informed Consent. All user control variables for consent, (b) through (e), have statistically significant differences in terms of acceptability. Overall, in terms of notification and consent, participants find data sharing more acceptable when they are explicitly informed or have more control over whether their data was used. 'Concealed consent', when they receive no formal notification, is overwhelmingly unacceptable to 73% of respondents ((b), $\mu = 1.88$). Unacceptability is substantially reduced when users are notified in any manner, regardless of control (e.g., even if opt-in or opt-out options are not available). 'Opt-out consent' ((d), $\mu = 3.31$), where users can toggle a setting to indicate they do not want their data shared, skews slightly more towards the acceptable end of the scale than the 'assumed consent' case ((c), $\mu = 3.11$). 'Opt-in consent' achieves the highest acceptability ((e), $\mu = 3.78$) within the consent/notification grouping with approximately 58% of respondents finding it at least somewhat acceptable.

Within Data Retention. We investigate respondents' perceptions with respect to data retention, (f) through (h), and find significant differences in their acceptability. Respondents find 'retaining data indefinitely' ((f), $\mu = 2.31$) to be less acceptable than retaining the data until the company is 'finished

using it' ((g), $\mu = 2.94$) and less acceptable than retaining the data for a 'specified time' limit ((h), $\mu = 2.99$). There is no significant difference in the distributions of how acceptable respondents find data between retention for a 'set period of time' and 'as long as the company uses it'. However, in practice there could be no real difference in how long the data is retained between indefinite retention and retaining the data as long as the company is using it. This result highlights the risk of influencing users consent based on phrasing; something not currently strictly defined across regulations on data sharing.

Within Purpose. In terms of purpose of use, (i) through (k), there are statistically significant differences in how acceptable respondents find each purpose. Respondents' overall perceptions are summarized as follows. It is least acceptable when the company (or companies) uses the data to generate revenue ((i), $\mu = 2.21$). Respondents find it somewhat more acceptable when there is an explicit tangible or perceived benefit to the user, such as a monetary reward ((j), $\mu = 3.05$) or improved service ((k), $\mu = 2.88$).

4.2 Nature of Collaboration

Recall the five types of collaboration defined in Section 3.1 and shown in Figure 1. First, we examine between group differences, that is, the differences in acceptability between different collaboration types. Second, we present within group differences, more specifically, the difference in acceptability between the scenarios that comprise a collaboration type.

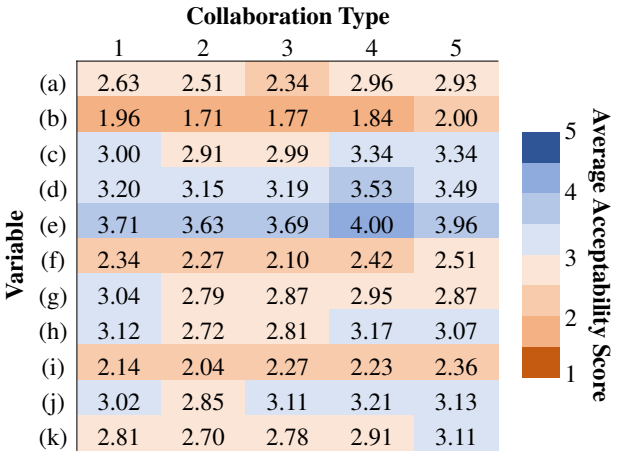

Figure 3: Average acceptability of variables for each collaboration type. The labels for collaboration types and variables correspond to those shown in Table 1.

Variable	Label
All scenarios (general)	(a)
Concealed consent	(b)
Assumed consent	(c)
Opt-out consent	(d)
Opt-in consent	(e)
Retained indefinitely	(f)
Retained while in use	(g)
Retained for set time	(h)
Generating revenue	(i)
Provide user remuneration	(j)
Improving services	(k)
Collaboration Type	**Label**
Validation	(V)
Two-way Two-Party Exchange	(1)
One-way Two-Party Exchange	(2)
Many-to-One Exchange	(3)
Acquisition	(4)
Merger then Acquisition	(5)

Table 1: Reference table for labels corresponding to usage controls and collaboration types.

4.2.1 Between Collaboration Types.

We compare our five types of multiparty data sharing to investigate whether some sharing types are more acceptable to respondents. The different average acceptability scores across types of collaborations for variables (a) to (k) are shown in Figure 3. To determine which types of collaboration are more or less acceptable we perform a subsequent pairwise analysis.

With respect to acceptability in 'general' (a), the different collaboration types, (1) through (5), are statistically significantly different. Both 'acquisition' ((4), $\mu = 2.96$) and 'merger then acquisition' ((5), $\mu = 2.93$) are more acceptable than a 'one-way two-party exchange' ((2), $\mu = 2.51$) and 'many-to-one exchange' ((3), $\mu = 2.34$). A possible attribution to the greater acceptability for mergers and mergers then acquisition rather than exchanges could be the indirectness by which data is acquired. Unlike in the specific exchange scenarios ('one-way two-party' and 'many-to-one') where data can be seen as a commodity, within the merger-acquisition scenarios nobody is explicitly seen as 'selling' users' data. Additionally, in the case of mergers and acquisitions, the company acquiring the data may be seen as the new shepherd of the data, continuing to provide the user with the services that led them to originally use the acquired companies' services.

User Controls Between Collaboration Types. We further compare between collaboration types for each of the user control mechanisms. We continue to observe statistically significant differences between mergers and acquisitions com-

pared to the other exchange types. Specifically, 'improving services' (k) is more acceptable for a 'merger then acquisition' ((5), $\mu = 3.11$) than a 'one-way two-party exchange' ((2), $\mu = 2.70$). 'Assumed consent' (c) is more acceptable for both merger collaboration types ('acquisition' ((4), $\mu = 3.34$) and 'merger then acquisition' ((5), $\mu = 3.34$)) than for a'one-way two-party exchange' ((2), $\mu = 2.91$). Finally, 'retained for a set time' (h) is more acceptable for an 'acquisition' ((4), $\mu = 3.17$) than a 'one-way two-party exchange' ((2), $\mu = 2.72$) The difference in acceptability between types for data retention and purpose could again be potentially attributed to the indirectness by which data is acquired in mergers and acquisitions.

There are no notable differences between collaboration types for 'concealed consent', 'generating revenue', 'retained indefinitely', and 'retained while in use'. This unchanging negative perception is likely because these attributes are considered more uniformly unacceptable. These results demonstrate another avenue where users would benefit from transparency in terms of the purpose and other contextual information, to make an informed decision of whether to consent, when companies are merged or acquired.

4.2.2 Within Collaboration Types.

Each collaboration type consists of two possible scenarios. We compare the scenarios within each collaboration type to one another to identify differences that exist depending on the sending and receiving companies as well as who the respondent is a user of. In our analysis we do not consider the order that the companies are introduced as a factor. This exclusion is based on our validation test for collaboration Type V; which found no statistically significant differences between the response distributions whether a health or technology company is introduced first, across variables (a) through (k).

We summarize the remainder of our results within collaboration types by their common themes. Overall, the within collaboration types analysis suggests that the inclusion of a health company negatively influences users' perceptions of the multiparty data sharing.

Collaboration over Commodification for Health Data. We find an interesting result within the 'one-way two-party exchange', an exchange type where the key distinction between scenarios is a tech company giving away user data (Scenario E) versus a health company giving away user data (Scenario F). We identified statistically significant differences across seven of the eleven measured variables. The four non-significantly different variables are 'concealed consent', 'assumed consent', 'opt-out consent', and 'retained indefinitely'. For the seven variables that do have significant differences, they are all more acceptable for Scenario E when compared to Scenario F. In Scenario E, respondents are framed as a user of a technology company which is providing its data to a health company. Whereas, in Scenario F respondents are framed as a user of a health company which is providing its data to a technology company. In both Scenario E and F, respondents are a user of the company giving away data.

This suggests the difference in acceptability could be attributed to the commoditization of health data being more objectionable than in the case of tech data. While respondents may be used to, or even have come to expect to have their data treated as a commodity by technology companies (Scenario E), the same may not be true for health companies. To further this idea, we look within 'two-way two-party exchanges' (Scenarios C and D), wherein the health company shares its data but also receives data in return. Respondents seem to interpret this reciprocity as providing some benefit to them, as opposed to being a 'sale'. When this reciprocity is absent in Scenario F, we see lower acceptability overall, possibly due to this commodification of health data which has an expectation to be the most protected data.

Health Companies Complicate Data Sharing. Health companies being involved negatively impact user perceptions of multiparty data sharing even when the health company is only receiving data. This is shown, first within 'many-to-one exchange', wherein a number of companies are sharing data with either a tech company (Scenario G) or a health company (Scenario H). We found a significant difference in acceptability of 'assumed consent'. Respondents who received the scenario where a technology company acquired the data (Scenario G, $\mu = 3.25$), found 'assumed consent' to be more acceptable than when a health company received the data (Scenario H, $\mu = 2.76$). This result implies that users were not as satisfied with simply being informed of data sharing, when it is shared with a health company, in contrast with a technology company. As both scenarios involve sharing financial data, we can hypothesise that users do not want their financial records to influence any future medical diagnoses. Users may be concerned for discrimination while receiving medical treatment or processing insurance, if a health company obtained their financial records.

The negative impacts of health company in data sharing is also shown within the 'acquisition' collaboration type. Scenarios within 'acquisition' involve a start-up that tracks user data on diet, fitness, and social habits being acquired by either a technology company (Scenario I) or a health company (Scenario J). Respondents found 'opt-in consent', the "strictest" consent option of the ones we tested, to be more acceptable when a technology company (Scenario I, $\mu = 4.24$), rather than a health company (Scenario J, $\mu = 3.77$) acquired a startup. We expect that respondents are more comfortable with their fitness habits influencing technology products, like in Scenario I, rather than having the potential to influence their medical treatment or insurance as in Scenario J.

As a final note on the inclusion of health companies and how they may influence respondents, we note that health

data has certain laws surrounding it that respondents may believe will protect them. Further, respondents concerns with data transferring to or from a health company may also be attributed to respondents being unsure as to the purpose. From our free-form responses we know that the purpose of use for the data was a frequent condition for acceptability.

4.3 Demographic Variations

We evaluate responses across all scenarios for differences based upon demographic groupings. For demographic differences due to gender, we compare men versus women as we did not have enough respondents representing other genders, leaving us with $N = 887$. We compare the two groups comprised of 432 men and 455 women across the variables (a through k). From our data we identify a significant difference in (b) 'concealed consent'. We can conclude that men ($\mu = 1.98$) found their consent not being explicitly granted, to be significantly more acceptable than women ($\mu = 1.76$) did.

To examine demographic variations due to age, we compare five age groups. From our data we identify a significant difference due to age group across all variables except for 'assumed consent', 'opt-out consent', and 'opt-in consent'. Across the variables 'concealed consent', 'retained indefinitely', 'generating revenue', and 'improving services', respondents aged 55-64 find each variable to be significantly less acceptable than their otherwise aged counterparts (18-24, 25-34, 45-54). Respondents aged 35-44 find the 'general scenario', 'retained for set time', and 'generating revenue' less acceptable than respondents aged 45-54. Additionally, those aged 35-44 find 'improving services' less acceptable than respondents aged 18-24. While older people, such as our respondents aged 55-64, could be generally expected to have more conservative views, we do not know why the middle age group, respondents aged 35-44, have a similar lower level of acceptability across variables. These results demonstrate that different demographics have different desires. User controls need to have sufficient individualization to support these differences.

4.4 Free-from General Perceptions

We analyze 789 non-empty free-form responses to the question 'In general, what are your thoughts on companies sharing data with other companies'. We exclude 62 responses that are either not interpretable or indicated a desire not to respond. For example, exclusions include: single or random character responses (e.g., 'a', 'alskj'), 'N/A', and 'no'.

Responses were coded in terms of their positive or negative perceptions of the practice of sharing data. Positive or negative responses can have a conditional component that indicates what improves or worsens their perceptions. The codes for the free-form responses perceptions were developed through discussion and definition by two members of the research team based on a sampling of the response set and the predefined 'positive' and 'negative' codes. A 'resigned' and a 'neutral' code were added after initial sampling to more accurately describe all responses. This methodology follows the process of Oates et al.'s [55] analysis and Miles et al.'s Qualitative Data Analysis: A Methods Sourcebook [49].

	Frequency by Collaboration Type					
Code	**All**	**1**	**2**	**3**	**4**	**5**
Neg.	305	46	45	54	49	48
Neg. Cond.	107	19	15	12	17	22
Neutral	66	12	7	10	15	16
Resigned	32	6	4	5	7	10
Pos. Cond.	165	26	25	26	33	37
Pos.	51	5	10	4	13	8

Table 2: Frequency given Nature of Collaboration. Columns correspond to: 1. One-way two-party exchange, 2. Two-way two-party exchange, 3. Many-to-one exchange, 4. Acquisition, and 5. Merger then acquisition.

The final codebook used to code the free-form responses is *neg* for unconditionally negative, *neg. Cond.* for overall negative response but permitted cases, *neutral* for neither positive nor negative, *resigned* for negative but accepted, *pos. Cond.* for overall positive but forbidden cases, and *pos.* for unconditionally positive.

Two members of the research team each independently applied the perceptions codebook to the response set. The coded responses were reviewed for agreement by the two team members. The process for handling a disagreement in coding was for both coders to check their responses. If the difference could be attributed to having mislabelled the code, a correction would be made. The coders would come back together and check the new agreement. If disagreement persisted, it went to a tie-breaker coder. We include an overview of common themes that were indicated as influencing perceptions or requirements in 'conditional' responses. The final code counts are summarized in Table 2.

4.4.1 Polarity of Base Perceptions

Of the total (789) responses coded for positive and negative perceptions, 32 required a third coder to break the disagreement. The original two independent coders agreed on the codes for 757 responses, or 96% of responses after checking for errors. One of the 32 responses shared with the third coder was coded differently by all three coders and the consensus was to remove it due to the ambiguity.

Unconditionally negative responses formed the largest group of responses and included a breadth of subjects relating to purpose, laws and regulations, distrust, and risks. Objections include users' data being used for generating revenue for the company or for marketing purposes.

P58310: I think companies after having acquired

data as an asset has one intention and it's making money through exploitation"

P78909: "These companies are reprehensible! I will not consent to my data being shared for marketing purposes"

Other negative responses report distaste for being coerced into agreeing to data sharing in order to access services. Respondents consider such requirements an uneven trade given the risks associated with a breach that exist whenever data is collected, saved, and shared.

P20322: "I'm not happy about it because if you do agree you can't choose who it will be shared with. If you don't agree, you can't use the service"

P53560: "I hate it. Cookies and data thieves. Opting out often renders the website unaccessible- so it's coercion/entrapment. Data breaches wouldn't really happen if data wasn't retained"

Other possible risks of such sharing, according to respondents, include malicious outsiders and malicious companies. Respondents express concern with targeted manipulation by a company, such as advertising, using the shared data. Concern with breaches or leaks also includes concern for data leaking out in ways users do not expect.

P69036: "While there are clear and logical reasons for utilizing and selling this data it does have potential for targeted manipulation"

P36717: "no. It makes me feel like my personal information is keep leaking out. i feel more vulnerable"

The responses coded as 'resigned' essentially express that respondents know such sharing occurs, do not necessarily like it, but accept it as reality. Respondents also express a need for law or regulations, a belief that such events are likely more or less frequent than they know, a feeling of futility, and the implied agreement to such things when using apps. One participant's response encompasses each of the above themes.

P07944: "It's a gray area: users make and agreement with companies for information use based upon the scope and identity/reputation of a company. What happens with an individual's information in the event of a the business/organization being sold. Legally speaking, the matter is an open and shut case. However, a user may not want to have the same information use agreement with the new company...and their rights to having a say in how their information is being used are clearly being violated by the new company which technically owns the rights to the information they have purchased since the company never negotiated terms with users and can use that information according to the company's desire and purposes. It's legal; but it sucks"

The neutral responses include two main types. First, some respondents directly say they are neutral or do not care about such sharing. Second, some respondents express some potential limitations on such sharing, but that they still did not have strong feelings about it either way.

P79659: "I don't have definite objections to companies sharing data with other companies"

P60109: "It depends on what it's used for and must have complete consent from an individual that isn't forced"

Few of the unconditionally positive responses say more than a one to three word answer. For example, 'good', 'epic', and 'sounds great' are common. The positive responses beyond sharing a generic response include some benefit to the individual or to the company. Benefits include personalizing advertising, ad opportunities, and new developments. While distaste for data being used for advertising was found in many negative responses, such as the earlier examples, this distaste was not universal.

P14505: "I think that it is acceptable because they need to use this data for advertising opportunities"

P98147: "Data sharing encourages more connection and collaboration between researchers, which can result in important new findings within the field. In a time of reduced monetary investment for science and research, data sharing is more efficient because it allows researchers to share resources"

4.4.2 Conditionals and User Control

In this section we focus on the responses coded as conditional. We highlight requirements users report as necessary for the scenarios to be acceptable. Specifically, we review 'positive conditional', 'negative conditional', and 'neutral' coded responses with respect to their conditionals. We include 'neutral' as our code definition of 'neither positive nor negative', does not prevent conditions from being specified in the response. Whether respondents viewed the scenario positively or negatively, they expressed similar themes.

Consent. The importance of consent and transparency is prominent in both positive and negative conditionals, with an emphasis on informed consent. Respondents express a need for easily accessible opt-out options and that consent (to data sharing) should not be a requirement for using a service.

P66884: “It’s inappropriate unless the user consents explicitly and should never be a requirement for use"

P10652: “I do not think it is acceptable unless they have the users permission. Or an option to cancel information sharing. If the user has a choice and is OK with it then I believe it’s fine"

P19193: “If they make people aware (in BIG print, not small, easy-to-miss print) then it’s fine"

When expressing the importance of users’ consent, some respondents highlight that data sharing should not be taken lightly. There are risks that can be associated with data being provided to other entities that cannot be properly evaluated without details as to where the data is going, what the data is, and why it is being shared. Receiving user consent requires full transparency with respect to each of those attributes.

P91741: “It should not be shared unless the individual gives authority to do so. It is private information that should not be shared on a whim"

P09262: “I don’t think companies should share customer’s personal information unless specific consent is received from the customer to where/what the information is shared to, as well as why"

Furthermore, consent can be withdrawn and cannot be assumed to be transferable between entities, even in the case of a company being purchased.

P41281: “Information collected, with the users permission, should never be shared with another company or assumed to be the property of said company if they merge with another company. This would be true regardless of whether the original company remains in the same business, or moves into a different service."

However, some respondents highlight that sufficient transparency can be advantageous to companies building goodwill after mergers or acquisitions.

P48036: “...The company can email its acquired users and them that they bought out the nicestartup and they want to use the data in order to improve their services and then list their services so people can decide for themselves. You’ll be surprised how many people will agree to continue, there’s no need to hide, lie, or manipulate anything. Just be honest! You’ll earn respect and loyalty as well"

Data Type and Processing. Respondents indicate preferences for the kind of data and how the data is processed.

P31222: “I do not like the idea of any personal, individual information being shared with other companies, either for free or for a price, but if a study is performed on that data and then the study results are shared I completely think that is okay"

The type of data that is acceptable or unacceptable is not universal. Respondents mention opposition to medical or health data generally, although there is some acknowledgment of possible exceptions. While personally identifiable information (PII) is generally expressed as inappropriate to share, what counts as PII is less universal. Some respondents consider buying habits to be fine while others highlight the private nature of such financial transactions [18].

P45732: “I don’t mind sharing information as long as it’s not financial"

P71169: “I have a problem with this when it’s sensitive personal information such as health information. I don’t have as much of a problem with this when it’s something less sensitive, such as my buying habits"

Purpose. The acceptability of different data sharing purposes, at least as far as the free-form responses are concerned, is highly individualized to what each respondent considers beneficial or detrimental. Some respondents find advertising acceptable while others do not. Sharing data to improve services or scientific investigations are spoken of positively while selling users’ data for monetary gain is aggressively opposed.

P24797: “It depends upon the purpose (my benefit or detriment), the data security to ensure the original personally identifiable data is secure or destroyed and the trust based on the history of how the company previously handled data"

Health. Health data is the most controversial type of data sharing, and a number of respondents express concern for whether legitimate sharing purposes exist. Many respondents that mention health data do so with intense negativity and concerns over the relevant ethics and legality of the exchange or purchase of health-related data.

P94865: “Repugnant, especially in light of for-profit health systems attempting to maximize profitability from patient interactions"

P72271: “There are stringent rules about sharing data under HIPAA in the US and this clearly violates it, along with potentially exposing PII"

P77878: “worried that data will be mined for insurance companies so they can eliminate or remove costly illnesses"

Even within the topic of health data, some respondents reflect upon the potential for acceptable data sharing settings. Privacy protections are key to improving the acceptability of health data sharing. Protections could include regulations, privacy mechanisms, and greater transparency.

P20986: “It depends. I think it can be beneficial under certain circumstances, but I would be hesitant having any healthcare data shared outside my practitioners. However, I recognize how it can improve goods/services, but there has to be a lot of protection in place anytime data is shared"

P44838: “I believe for health records it should be acceptable for continuance of care but not for advertising or making money"

5 Discussion

Disambiguate Third Parties. Privacy policies that give companies unrestricted ability to share data with ‘third-parties’ and ‘partners’ do not encapsulate the details that influence users’ preferences. Our results show users care about who data is being shared with, what is being shared, and the structure of the collaboration. In terms of ‘who’, health companies sharing data is less acceptable than technology companies. In terms of ‘what’, it is more acceptable to share fitness data with a technology than a health company. Structurally, reciprocity improves acceptability over one-way ‘sale’ type transactions. Transparency with respect to the nature of any collaboration is required to support the preferences our respondents expressed. Thus, regulations, such as CCPA, need to have detailed requirements for companies to clearly outline the properties we identify for data sharing.

Explicit over Implicit Consent. Implied consent is inferred based on a person’s actions or circumstances. When companies make consent conditional for the use of their service, the use of the service is taken as consent. In contrast to implied consent, explicit consent is unmistakably provided by the user, possibly in writing. It is specific, can be rescinded, and is non-transferable. Informed consent requires users to have an understanding of the implications and extent of what their agreement applies to when using an app or tool. Respondents in our study expressed a clear preference for explicit consent that requires them to opt-in over implied consent (e.g., ‘concealed consent’ or ‘assumed consent’). Respondents’ preference for, and emphasis on consent and transparency, held for both statistical analysis and free-form responses.

Reduce Ambiguities to Communicate Privacy. Although user controls affect the acceptability of collaborations, the effect does not always correspond to the impact on privacy in practice. For example, retaining data ‘while in use’ and ‘indefinitely’ may have no practical difference. Despite this, respondents found it more acceptable for companies to retain data ‘while they are using it’. Companies could abuse such misunderstandings by making something seem more private in practice than it actually is.

Similarly, each of the five privacy mechanisms we included have a different effect on privacy in practice. Respondents to our study had difficulties understanding our descriptions of the privacy mechanisms. Unless users can distinguish between accessible descriptions, they will not be making informed decisions. Therefore, when companies use privacy mechanisms, they should be compelled by law to ensure it is either easy to understand or that users are not required to understand the privacy mechanism used to successfully make an informed choice. Going forward, researchers and policy makers must focus on conveying the significance of different privacy implications and changing the information provided to users such that it is clear and concise and not perceived as minor details.

6 Conclusion

We presented the results of our survey on user perceptions of multiparty data sharing. Our results indicate that the type of data sharing collaboration affects acceptability as do the available user controls. Based on these results, we recommend that regulations for data sharing do not solely rely on past work that focused on only one company receiving data from another (whether for advertising or other purposes). We hope the recommendations we have made help other privacy researchers and regulators mitigate the inequity imposed on users by data commodification.

Acknowledgments

We gratefully acknowledge the support of NSERC for grants RGPIN-05849, CRDPJ-531191, IRC-537591, CRDPJ-534381, the NSERC Alexander Graham Bell Canada Graduate Scholarship program, and the Royal Bank of Canada for funding this research.

References

[1] Alessandro Acquisti. Privacy in Electronic Commerce and the Economics of Immediate Gratification. In *Proceedings of the 5th ACM Conference on Electronic Commerce*, pages 21–29, 2004.

[2] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian

Schaub, Manya Sleeper, et al. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Computing Surveys*, 50(3):1–41, 2017.

[3] Annie I Antón, Julia B Earp, and Jessica D Young. How Internet Users' Privacy Concerns Have Evolved Since 2002. *IEEE Security & Privacy*, 8(1):21–27, 2010.

[4] Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. Discovering Smart Home Internet of Things Privacy Norms Using Contextual Integrity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–23, 2018.

[5] Naveen Farag Awad and M. S. Krishnan. The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization. *MIS Quarterly*, 30(1):13–28, 2006.

[6] Vinayshekhar Bannihatti Kumar, Roger Iyengar, Namita Nisal, Yuanyuan Feng, Hana Habib, Peter Story, Sushain Cherivirala, Margaret Hagan, Lorrie Cranor, Shomir Wilson, et al. Finding a Choice in a Haystack: Automatic Extraction of Opt-Out Statements from privacy Policy Text. In *Proceedings of The Web Conference 2020*, pages 1943–1954, 2020.

[7] Susanne Barth, Menno DT de Jong, Marianne Junger, Pieter H Hartel, and Janina C Roppelt. Putting the Privacy Paradox to the Test: Online Privacy and Security Behaviors Among Users with Technical Knowledge, Privacy Awareness, and Financial Resources. *Telematics and Informatics*, 41:55–69, 2019.

[8] Alastair R Beresford, Dorothea Kübler, and Sören Preibusch. Unwillingness to Pay for Privacy: A Field Experiment. *Economics letters*, 117(1):25–27, 2012.

[9] Mark Bergen and Jennifer Surane. Google and Mastercard Cut a Secret Ad Deal to Track Retail Sales. Online, 2018. https://www.bloomberg.com/news/articles/2018-08-30/google-and-mastercard-cut-a-secret-ad-deal-to-track-retail-sales.

[10] Eric W. Burger, Michael D. Goodman, Panos Kampanakis, and Kevin A. Zhu. Taxonomy Model for Cyber Threat Intelligence Information Exchange Technologies. In *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security*, page 51–60, New York, NY, USA, 2014. ACM.

[11] Microsoft News Center. Microsoft to Acquire LinkedIn. Microsoft News Center, 2016. https://news.microsoft.com/2016/06/13/microsoft-to-acquire-linkedin/.

[12] Microsoft News Center. Microsoft and Providence St. Joseph Health Announce Strategic Alliance to Accelerate the Future of Care Delivery. Online, 2019. https://news.microsoft.com/2019/07/08/microsoft-and-providence-st-joseph-health-announce-\strategic-alliance-to-accelerate-the-future-of-care-delivery/.

[13] Hao Chen, Zhicong Huang, Kim Laine, and Peter Rindal. Labeled PSI from Fully Homomorphic Encryption with Malicious Security. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 1223–1237, New York, New York, USA, 2018. ACM, ACM.

[14] Evgenia Christoforou, Pinar Barlas, and Jahna Otterbacher. It's About Time: A View of Crowdsourced Data Before and During the Pandemic. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. ACM.

[15] Lorrie Faith Cranor. Can Users Control Online Behavioral Advertising Effectively? *IEEE Security & Privacy*, 10(2):93–96, 2012.

[16] Lorrie Faith Cranor. Necessary but Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice. *J. on Telecomm. & High Tech. L.*, 10:273, 2012.

[17] Lorrie Faith Cranor, Joseph Reagle, and Mark S Ackerman. Beyond Concern: Understanding Net Users' Attitudes About Online Privacy. *The Internet upheaval: raising questions, seeking answers in communications policy*, pages 47–70, 2000.

[18] Yves-Alexandre De Montjoye, Laura Radaelli, Vivek Kumar Singh, et al. Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata. *Science*, 347(6221):536–539, 2015.

[19] Tobias Dienlin and Sabine Trepte. Is the Privacy Paradox a Relic of the Past? An In-Depth Analysis of Privacy Attitudes and Privacy Behaviors. *European Journal of Social Psychology*, 45(3):285–297, 2015.

[20] Nico Ebert, Kurt Alexander Ackermann, and Peter Heinrich. Does Context in Privacy Communication Really Matter? — A Survey on Consumer Concerns and Preferences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–11, New York, NY, USA, 2020. ACM.

[21] José Estrada-Jiménez, Javier Parra-Arnau, Ana Rodríguez-Hoyos, and Jordi Forné. Online Advertising: Analysis of Privacy Threats and Protection Approaches. *Computer Communications*, 100:32–51, 2017.

[22] Benjamin Fabian, Tatiana Ermakova, and Tino Lentz. Large-Scale Readability Analysis of Privacy Policies. In *Proceedings of the International Conference on Web Intelligence*, pages 18–25, 2017.

[23] Federal Trade Commission. Android Flashlight App Developer Settles FTC Charges It Deceived Consumers. https://goo.gl/Zf18jI, 2013. Accessed 2019-08-09.

[24] Yuanyuan Feng, Yaxing Yao, and Norman Sadeh. 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*, pages 1–16, 2021.

[25] Casey Fiesler and Blake Hallinan. "We Are the Product" Public Reactions to Online Data Sharing and Privacy Controversies in the Media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[26] Chaim Gartenberg. Google Buys Fitbit for $2.1 Billion. The Verge. https://www.theverge.com/2019/11/1/20943318/google-fitbit-acquisition-fitness-tracker-announcement.

[27] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. "It's a Scavenger Hunt": Usability of Websites' Opt-Out and Data Deletion Choices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.

[28] Hana Habib, Yixin Zou, Yaxing Yao, Alessandro Acquisti, Lorrie Cranor, Joel Reidenberg, Norman Sadeh, and Florian Schaub. 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*, pages 1–25, 2021.

[29] Eszter Hargittai and Elissa M Redmiles. Will Americans be Willing to Install COVID-19 Tracking Apps? *Scientific American*, pages Epub–ahead, 2020.

[30] Anis Heydari. The Canadian Tech Company that Changed its Mind about Using Your Tax Return to Sell Stuff. CBC, 2020. https://www.cbc.ca/radio/costofliving/indigenous-nations-and-the-economy-plus-why-it-s-so-hard-to-fly-for-cheap-in-canada-1.5469919/the-canadian-tech-company-that-changed-its-mind-about-using-your-tax-return-to-sell-stuff-1.5471400.

[31] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. "My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security. In *Eleventh Symposium On Usable Privacy and Security 2015)*, pages 39–52, 2015.

[32] Sowmya Karunakaran, Kurt Thomas, Elie Bursztein, and Oxana Comanescu. Data Breaches: User Comprehension, Expectations, and Concerns with Handling Exposed Data. In *Fourteenth Symposium on Usable Privacy and Security 2018)*, pages 217–234, 2018.

[33] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as Part of the App Decision-Making Process. In *Proceedings of the CHI conference on human factors in computing systems*, pages 3393–3402, 2013.

[34] John Koetsier. Apple's Ad Network Gets 'Preferential Access To Users' Data' vs Facebook, Google, Others. Forbes, 2021. https://www.forbes.com/sites/johnkoetsier/2020/08/07/apple-ad-network-gets-special-privileges-that-facebook\-google-wont-on-ios14/.

[35] Spyros Kokolakis. Privacy Attitudes and Privacy Behaviour: A Review of Current Research on the Privacy Paradox Phenomenon. *Computers & security*, 64:122–134, 2017.

[36] Ivar Krumpal. Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review. *Quality & Quantity*, 47(4):2025–2047, 2013.

[37] Ed Leefeldt and Amy Danise Ed. The Witness Against You: Your Car. Forbes, 2021. https://www.forbes.com/advisor/car-insurance/telematics-data-privacy/.

[38] Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor. Why Johnny Can't Opt Out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 589–598, 2012.

[39] Jialiu Lin, Shahriyar Amini, Jason I Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy Through Crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 501–510, New York, New York, USA, 2012. ACM, ACM.

[40] Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I Hong. Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings. In *Tenth Symposium On Usable Privacy and Security 2014)*, pages 199–212, 2014.

[41] Thomas Linden, Rishabh Khandelwal, Hamza Harkous, and Kassem Fawaz. The Privacy Policy Landscape After the GDPR. *arXiv preprint arXiv:1809.08396*, 2018.

[42] Maureen Mahoney. California Consumer Privacy Act: Are Consumers' Digital Rights Protected. Technical report, Technical Report. Consumer Reports., 2020. https://advocacy.consumerreports.org/press_release/consumer-reports-study-finds-significant-obstacles-to-exercising-california-privacy-rights/.

[43] Peter Mayer, Yixin Zou, Florian Schaub, and Adam J Aviv. "Now I'm a Bit Angry:" Individuals' Awareness, Perception, and Responses to Data Breaches that Affected Them. In *The Thirtieth USENIX Security Symposium 2021*, 2021.

[44] Aleecia M McDonald and Lorrie Faith Cranor. The Cost of Reading Privacy Policies. *ISJLP*, 4:543, 2008.

[45] Aleecia M McDonald and Lorrie Faith Cranor. Americans' Attitudes About Internet Behavioral Advertising Practices. In *Proceedings of the Ninth Annual ACM Workshop on Privacy in the Electronic Society*, pages 63–72, 2010.

[46] James McLeod. Double-Double Tracking: How Tim Hortons Knows Where You Sleep, Work and Vacation. Financial Post, 2020. https://financialpost.com/technology/tim-hortons-app-tracking-customers-intimate-data.

[47] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *the 20th International Conference on Artificial Intelligence and Statistics*, pages 1273–1282, Fort Lauderdale, FL, USA, 2017. PMLR.

[48] Catherine Meadows. A more efficient cryptographic matchmaking protocol for use in the absence of a continuously available third party. In *1986 IEEE Symposium on Security and Privacy*, pages 134–134. IEEE, 1986.

[49] Matthew B Miles, A Michael Huberman, and Johnny Saldaña. *Qualitative Data Analysis: A Methods Sourcebook*. Sage publications, 2018.

[50] Ambar Murillo, Andreas Kramm, Sebastian Schnorf, and Alexander De Luca. "If I Press Delete, It's Gone"-User Understanding of Online Data Deletion and Expiration. In *Fourteenth Symposium on Usable Privacy and Security 2018)*, pages 329–339, 2018.

[51] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. Privacy Expectations and Preferences in an IoT World. In *Thirteenth Symposium on Usable Privacy and Security 2017*, pages 399–412, 2017.

[52] Arvind Narayanan, Joanna Huey, and Edward W Felten. A Precautionary Approach to Big Data Privacy. In *Data Protection on the Move*, pages 357–385. Springer, 2016.

[53] Helen Nissenbaum. Contextual integrity up and down the data food chain. *Theoretical Inquiries in Law*, 20(1):221–256, 2019.

[54] Thomas B Norton. The Non-Contractual Nature of Privacy Policies and a New Critique of the Notice and Choice Privacy Protection Model. *Fordham Intell. Prop. Media & Ent. LJ*, 27:181, 2016.

[55] Maggie Oates, Yama Ahmadullah, Abigail Marsh, Chelse Swoopes, Shikun Zhang, Rebecca Balebako, and Lorrie Faith Cranor. Turtles, Locks, and Bathrooms: Understanding Mental Models of Privacy Through Illustration. *Proceedings on Privacy Enhancing Technologies*, 2018(4):5–32, 2018.

[56] Jonathan A Obar and Anne Oeldorf-Hirsch. The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services. *Information, Communication & Society*, pages 1–20, 2018.

[57] Sean O'Connor, Ryan Nurwono, and Eleanor Birrell. (Un) clear and (In) conspicuous: The Right to Opt-Out of Sale Under CCPA. *arXiv preprint arXiv:2009.07884*, 2020.

[58] Office of the Privacy Commissioner of Canada. PIPEDA in brief. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda_brief/, 2019. Accessed 2019-06-18.

[59] Charles Egerton Osgood, George J Suci, and Percy H Tannenbaum. *The Measurement of Meaning*. University of Illinois press, 1957.

[60] Jaehong Park and Ravi Sandhu. A Position Paper: A Usage Control (UCON) Model for Social Networks Privacy. 2000.

[61] Jaehong Park and Ravi Sandhu. The UCON_{ABC} Usage Control Model. *ACM Trans. Inf. Syst. Secur.*, 7(1):128–174, February 2004.

[62] Benny Pinkas, Thomas Schneider, Christian Weinert, and Udi Wieder. Efficient Circuit-Based PSI via Cuckoo Hashing. In Jesper Buus Nielsen and Vincent Rijmen, editors, *Advances in Cryptology – EUROCRYPT 2018*, pages 125–157, Cham, 2018. Springer International Publishing.

[63] Stanley Presser, Mick P Couper, Judith T Lessler, Elizabeth Martin, Jean Martin, Jennifer M Rothgeb, and Eleanor Singer. Methods for Testing and Evaluating Survey Questions. *Public opinion quarterly*, 68(1):109–130, 2004.

[64] Emilee Rader. Awareness of Behavioral Tracking and Information Privacy Concern in Facebook and Google. In *Tenth Symposium On Usable Privacy and Security 2014*, pages 51–67, 2014.

[65] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 1326–1343, 2019.

[66] Elissa M Redmiles, Ziyun Zhu, Sean Kross, Dhruv Kuchhal, Tudor Dumitras, and Michelle L Mazurek. Asking for a Friend: Evaluating Response Biases in Security User Studies. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 1238–1255, New York, New York, USA, 2018. ACM, ACM.

[67] Joel R Reidenberg, N Cameron Russell, Alexander J Callen, Sophia Qasir, and Thomas B Norton. Privacy Harms and the Effectiveness of the Notice and Choice Framework. *ISJLP*, 11:485, 2015.

[68] John A Rothchild. Against Notice and Choice: The Manifest Failure of the Proceduralist Paradigm to Protect Privacy Online (Or Anywhere Else). *Clev. St. L. Rev.*, 66:559, 2017.

[69] Christophe Olivier Schneble, Bernice Simone Elger, and David Martin Shaw. Google's Project Nightingale Highlights the Necessity of Data Science Ethics Review. *EMBO molecular medicine*, 12(3):e12053, 2020.

[70] Jonathan Schulz, Duman Bahrami-Rad, Jonathan Beauchamp, and Joseph Henrich. The Origins of WEIRD Psychology. *Available at SSRN 3201031*, 2018.

[71] Tariq Shaukat. Our partnership with Ascension. Google Cloud, 2019. https://cloud.google.com/blog/topics/inside-google-cloud/our-partnership-with-ascension.

[72] Irina Shklovski, Scott D Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. Leakiness and Creepiness in App Space: Perceptions of Privacy and Mobile App Use. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2347–2356, New York, New York, USA, 2014. ACM, ACM.

[73] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet S Talwalkar. Federated Multi-Task Learning. In *Advances in Neural Information Processing Systems 30*, pages 4424–4434. Curran Associates, Inc., 2017.

[74] Madiha Tabassum, Tomasz Kosinski, and Heather Richter Lipford. "I Don't Own the Data": End User Perceptions of Smart Home Device Data Practices and Risks. In *Fifteenth Symposium on Usable Privacy and Security 2019*, 2019.

[75] StartUp HERE Toronto. Waterloo-Based Bonfire Acquired for $140 Million CAD in 'Govtech' Merger. StartUp HERE Toronto, 2018. https://startupheretoronto.com/partners/betakit/waterloo-based-bonfire-acquired-for-140-million-cad\-in-govtech-merger/.

[76] Janice Y Tsai, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti. The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Information Systems Research*, 22(2):254–268, 2011.

[77] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. In *Eighth Symposium on Usable Privacy and Security 2012*, pages 1–15, 2012.

[78] Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David Wagner, and Konstantin Beznosov. The Feasibility of Dynamically Granted Permissions: Aligning Mobile Privacy with User Preferences. In *2017 IEEE Symposium on Security and Privacy*, pages 1077–1093. IEEE, 2017.

[79] Aiping Xiong, Tianhao Wang, Ninghui Li, and Somesh Jha. Towards Effective Differential Privacy Communication for Users' Data Sharing Decision and Comprehension. *arXiv preprint arXiv:2003.13922*, 2020.

[80] Yaxing Yao, Davide Lo Re, and Yang Wang. Folk Models of Online Behavioral Advertising. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1957–1969, 2017.

[81] Xun Yi, Russell Paulet, and Elisa Bertino. Homomorphic encryption. In *Homomorphic Encryption and Applications*, pages 27–46. Springer, 2014.

A Statistics

This appendix details the process and results of our statistical analysis. All statistical results presented use a significance level of 0.05. We use non-parametric statistical tests as our data is not normally distributed. This leaves the potential for incorrectly finding a difference insignificant. However, it decreases the risk of incorrectly saying a difference is significant. Additionally, when presenting the results of multiple comparison procedures, we report the p-value adjusted using the Bonferroni correction to account for the increased chance of false positive results due to multiple comparisons.

A.1 Within Informed Consent, Data Retention, and Purpose

The statistics shown in this section correspond to the results presented in Section 4.1.

We perform a Friedman's two-way analysis of variance by ranks for each of the distributions of acceptability: within informed consent groups, within data retention groups, and within purpose groups. For all variables within groups $N = 916$. Results show that the distributions of acceptability is not the same for: consent groups (b), (c), (d), (e) with test stat $= 899.29$ $p < 0.001$, retention groups (f), (g), (h) with test stat $= 255.08$ $p < 0.001$, and purpose groups (i), (j), (k) with test stat $= 435.79$ $p < 0.001$.

We perform Dunn's multiple comparison procedure to identify which variables within a group differ and in what direction, for example within data retention, how do variables (f), (g), and (h) differ (see Table 3). The difference in mean rank (e.g., the mean rank of Var 1 subtract the mean rank of Var 2) shows the direction of the difference in acceptability of the pair. All pairs of variables have significantly different distributions of acceptability except for the (g), (h) variable pair from within data retention.

Var 1, Var 2	Difference in Mean Rank	Std. Test Statistic	*p*
	Informed Consent		
(a), (b)	-0.85	-14.098	<0.001
(a), (c)	-1.07	-17.663	<0.001
(a), (d)	-1.44	-23.798	<0.001
(b), (c)	-0.22	-3.565	0.002
(b), (d)	-0.59	-9.700	<0.001
(c), (d)	-0.37	-6.135	<0.001
	Data Retention		
(f), (g)	-0.46	-9.778	<0.001
(f), (h)	-0.51	-10.864	<0.001
(g), (h)	-0.05	-1.086	0.832
	Purpose		
(i), (j)	-0.71	-15.186	<0.001
(i), (k)	-0.55	11.764	<0.001
(k), (j)	-0.16	-3.423	0.002

Table 3: Dunn's multiple comparison test results for the distribution of acceptability compared pairwise between variables within informed consent, data retention, and purpose groups. All p-values are adjusted for multiple comparisons (6 comparisons for the consent group, 3 for each of the data retention and purpose groups).

A.2 Collaboration Types

This section corresponds to Section 4.2 results.

Between collaboration types. We perform a Kruskal-Wallis test on the distribution of acceptability of each collaboration type (1-5) for each variable ((a) through (k)) and report those with significant differences in Table 4. We perform a post-hoc analysis for variables that have significant differences from the Kruskal-Wallis test to identify which collaboration types have pairwise differences. We use Dunn's multiple comparison procedure and show the results in Table 5. Only the collaboration type pairs that have significantly different distributions of acceptability are reported. The difference in mean rank (e.g., the mean rank of Type X subtract the mean rank of Type Y) shows the direction of the difference in acceptability collaboration types.

Between collaboration types, the acceptability distribution of...	**Test Statistic**	***p***
...(a) is the same	26.724	<0.001
...(c) is the same	15.113	0.004
...(d) is the same	10.340	0.035
...(e) is the same	12.058	0.017
...(h) is the same	13.261	0.010
...(k) is the same	10.337	0.035

Table 4: Kruskal-Wallis test results for the distribution of acceptability of variables between sharing types {1 ($N = 140$), 2 ($N = 150$), 3 ($N = 134$), 4 ($N = 162$), 5 ($N = 170$)} for which the acceptability of the variable differs significantly between data sharing types.

Within sharing types. Each sharing type (1-5) is comprised of two scenarios, so within each type we perform a Mann-Whitney U test for each variable ((a) through (k)). For 'two-way two-party exchange' (type 1), we fail to identify any significant differences in the distribution of acceptability for its constituent scenarios C ($N = 73$) and D ($N = 67$). In 'one-way two-party exchange' (type 2), we identify significant differences between scenarios E and F in seven variables which can be seen in Table 6. For 'many-to-one exchange' (type 3), we identify one significant difference between scenario G ($N = 64$) and H ($N = 70$) for 'assumed consent' (variable (c), $p = 0.035$, std. test statistic$= -2.107$, mean rank difference$= 13.84$). For 'acquisition' (type 4), we identify a significant difference for 'opt-in consent' (variable (e)) between scenarios I ($N = 79$) and J ($N = 83$) ($p = 0.004$, std. test statistic$= -2.915$, mean rank difference$= 20.24$). For 'merger then acquisition' (type 5), we fail to identity any significant differences in acceptability of variables for scenario K ($N = 74$) compared with L ($N = 96$).

Collaboration Type X, Type Y	Difference in Mean Rank	Std. Test Statistic	p
(a) All scenarios (general)			
2, 4	-75.46	-3.124	0.018
2, 5	-69.42	-2.907	0.037
3, 4	-104.31	4.190	<0.001
3, 5	-98.27	3.990	0.001
(c) Assumed consent			
2, 4	-68.28	-2.825	0.047
2, 5	-68.23	-2.855	0.043
(d) Opt-out consent No pairwise differences due to Bonferroni correction.			
(e) Opt-in consent No pairwise differences due to Bonferroni correction.			
(h) Retained for set time			
2, 4	-71.96	-2.973	0.030
(k) Improving services			
2, 5	-70.38	-2.948	0.032

Table 5: Dunn's multiple comparison test results for the distribution of acceptability compared pairwise between collaboration types. All p values are adjusted for multiple comparisons (10 comparisons per variable).

Within One-Way Two-Party Exchange (E, F), the acceptability distribution of...	Difference in Mean Rank	Std. Test Statistic	p
...(a) is the same	16.04	-2.322	0.020
...(e) is the same	17.47	-2.550	0.011
...(g) is the same	16.11	-2.315	0.021
...(h) is the same	15.19	-2.188	0.029
...(i) is the same	17.22	-2.603	0.009
...(j) is the same	22.22	-3.202	0.001
...(k) is the same	15.24	-2.196	0.028

Table 6: Mann-Whitney U test results for the One-Way Two-Party Exchange (collaboration type 2) scenarios {E ($N = 81$), F ($N = 69$)}.

A.3 Demographics

The statistics in this section correspond to the results in Section 4.3. We show the statistical results for demographic variations, first, due to gender and, second, due to age.

Gender acceptability variations. For gender, we performed a Mann-Whitney U Test with two groups comprised of 432 men and 455 women compared for each of the variables (a through k). We found a significant result for 'concealed consent' (variable (b)). We can conclude that men found their consent not being explicitly granted, to be significantly more acceptable than women did ($p = 0.008$, std. test statistic$= -2.647$). The difference in mean rank between men and women for 'concealed consent' was 40.45.

Age acceptability variations. To examine how age group influences acceptability for each of the variables, (a) through (k), we performed a Kruskal-Wallis test comparing the five age groups {18-24 ($N = 154$), 25-34 ($N = 201$), 35-44 ($N = 140$), 45-54 ($N = 197$), 55-64 ($N = 201$)}. We find that between age groups, the acceptability distribution of (a) $p = 0.006$, (b) $p < 0.001$, (f) $p < 0.001$, (g) $p = 0.019$, (h) $p = 0.012$, (i) $p < 0.001$, (j) $p = 0.018$, and (k) $p < 0.001$ is not the same.

B Privacy Mechanism Comprehension

Respondents predominantly fail the comprehension check as to whether they understand their privacy mechanism. Only 37% of total respondents correctly identified the corresponding "layperson" description of the privacy mechanism they received. Data aggregation was the most correctly identified with 64% correctness. Respondents had the most difficulty comprehending LDP and CDP. As LDP and CDP are essentially modifications to aggregation when described less formally, it is not surprising that they were frequently thought to correspond to the aggregation description. Privacy mechanism comprehension results are shown in Figure 4.

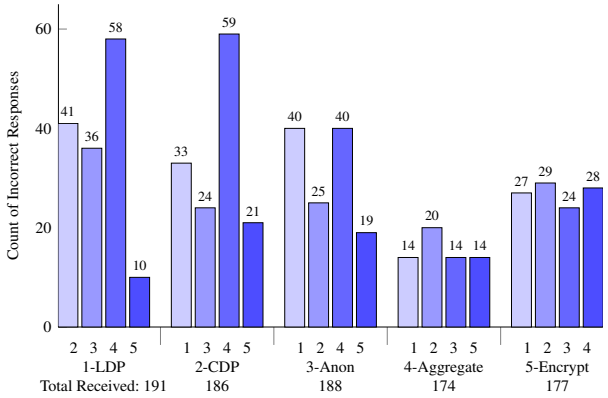

Figure 4: The counts of privacy mechanism received versus incorrectly guessed. Respondents receive the definition of a privacy mechanism and attempt to identify the layperson description that corresponds to that same privacy mechanism. For example, of the 191 respondents that received LDP (privacy mechanism 1), 41 incorrectly guessed they received CDP (privacy mechanism 2).