The Fair Value of Data Under Heterogeneous Privacy Constraints in Federated Learning

Justin S. Kang *justin_kang@berkeley.edu*
UC Berkeley

Ramtin Pedarsani *ramtin@ece.ucsb.edu*
UC Santa Barbara

Kannan Ramchandran *kannanr@berkeley.edu*
UC Berkeley

Reviewed on OpenReview: https://openreview.net/forum?id=ynG5Ak7n7Q

Abstract

Modern data aggregation often involves a platform collecting data from a network of users with various privacy options. Platforms must solve the problem of how to allocate incentives to users to convince them to share their data. This paper puts forth an idea for a *fair* amount to compensate users for their data at a given privacy level based on an axiomatic definition of fairness, along the lines of the celebrated Shapley value. To the best of our knowledge, these are the first fairness concepts for data that explicitly consider privacy constraints. We also formulate a heterogeneous federated learning problem for the platform with privacy level options for users. By studying this problem, we investigate the amount of compensation users receive under fair allocations with different privacy levels, amounts of data, and degrees of heterogeneity. We also discuss what happens when the platform is forced to design fair incentives. Under certain conditions we find that when privacy sensitivity is low, the platform will set incentives to ensure that it collects all the data with the lowest privacy options. When the privacy sensitivity is above a given threshold, the platform will provide no incentives to users. Between these two extremes, the platform will set the incentives so some fraction of the users chooses the higher privacy option and the others chooses the lower privacy option.

1 Introduction

From media to healthcare to transportation, the vast amount of data generated by people every day has solved difficult problems across many domains. Nearly all machine learning algorithms, including those based on deep learning rely heavily on data and many of the largest companies to ever exist center their business around this precious resource of data. This includes directly selling access to data to others for profit, selling targeted advertisements based on data, or by exploiting data through data-driven engineering, to better develop and market products. Simultaneously, as users become more privacy conscious, online platforms are increasingly providing *privacy level* options for users. Platforms may provide incentives to users to influence their privacy decisions. This manuscript investigates how platforms can fairly compensate users for their data contribution at a given privacy level. Consider a platform offering geo-location services with three privacy level options:

i) Users send no data to the platform — all data processing is local and private.

ii) An intermediate option with federated learning (FL) for privacy. Data remains with the users, but the platform can ask for gradients with respect to a particular loss function.

iii) A non-private option, where all user data is stored and owned by the platform.

If users choose option (i), the platform does not stand to gain from using that data in other tasks. If the user chooses (ii), the platform is better off, but still has limited access to the data via FL and may not be able to fully leverage its potential. Therefore, the platform wants to incentivize users to choose option (iii). This may be done by providing services, discounts or money to users that choose this option. Effectively, by choosing an option, users are informally selling (or not selling) their data to platforms. Due to the lack of a formal exchange, it can be difficult to understand if this sale of user data is *fair*. Are platforms making the cost of choosing private options like (i) or (ii) too high? Is the value of data much higher than the platform is paying?

A major shortcoming of the current understanding of data value is that it often fails to explicitly consider a critical factor in an individual's decision to share data—privacy. This work puts forth two rigorous notions of the fair value of data in Section 3 that explicitly include privacy and make use of the axiomatic framework of the *Shapley value* from game theory (Shapley, 1952).

Compelled by the importance of data in our modern economy and a growing social concern about privacy, this paper presents frameworks for quantifying the fair value of private data. Specifically, we consider a setting where users are willing to provide their data to a platform in exchange for some sort of payment and under some privacy guarantees depending on their level of privacy requirements. The platform is responsible for running the private learning algorithm on the gathered data and making the fair payments with the objective of maximizing its utility including statistical accuracy and total amount of payments. Our goal is to understand fair mechanisms for this procedure as depicted in Fig. 1.

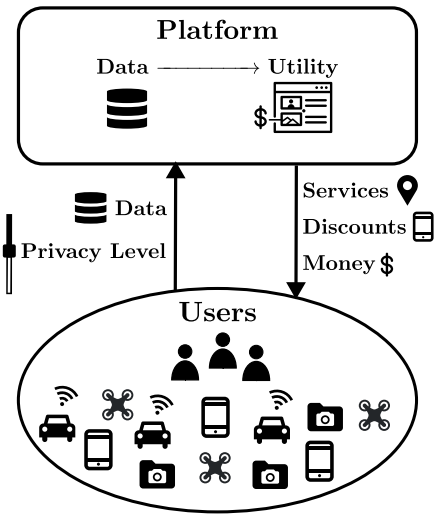


Figure 1: Depiction of interactions between platform and users. Users generate data with phones, cameras, vehicles, and drones. This data goes to the platform but requires some level of privacy. The platform uses this data to generate utility, often by using the data for learning tasks. In return, the platform may provide the users with payments in the form of access to services, discounts on products, or monetary compensation.

1.1 Related Work

Economics With the widespread use of the internet, interactions involving those that have data and those that want it have become an important area of study (Balazinska et al., 2011), and a practical necessity (Spiekermann et al., 2015b). Among these interactions, the economics of data from privacy conscious users has received significant attention in Acquisti et al. (2016) and Wieringa et al. (2021). The economic and social implications of privacy and data markets are considered in Spiekermann et al. (2015a). In Acemoglu et al. (2019) the impact of data externalities is investigated. The leakage of data leading to the suppression of its market value is considered.

Privacy Currently, popular forms of privacy include federated learning (Kairouz et al., 2021) and differential privacy (DP) (Dwork, 2008; Bun & Steinke, 2016) either independently or in conjunction with one another. Our work uses a flexible framework that allows for a rage of different privacy models to be considered.

Optimal Data Acquisition One line of literature studies *data acquisition*, where platforms attempt to collect data from privacy conscious users. Ghosh & Ligett (2013) consider the case of uniform privacy guarantees (homogeneous DP), where users have unique minimum privacy constraints, focusing on characterizing equilibria. Ghosh & Roth (2011) allows for heterogeneous DP guarantees with the goal to design a dominant strategy truthful mechanism to acquire data and estimate the sum of users' binary data. In Fallah et al. (2022) the authors consider an optimal data acquisition problem in the context of private mean estimation in two different local and central heterogeneous DP settings. It is assumed that players care about both the estimation error of the common estimator generated by the platform and any payments made to them by the platform in their decision making. By assuming linear privacy sensitivity represented by scalars and drawn from a distribution,

they devise a mechanism for computing the near-Bayes optimal privacy levels to provide to the players. Cummings et al. (2023) focuses on the central setting, under both the linear privacy sensitivity and the privacy constraints model, offering insights into the optimal solution. Hu & Gong (2020) goes beyond linear estimation to consider FL, where each user has a unique privacy sensitivity function parameterized by a scalar variable. Users choose their privacy level, and the platform pays them via a proportional scheme. For linear privacy sensitivity functions, an efficient way to compute the Nash equilibrium is derived. Roth & Schoenebeck (2012); Chen et al. (2018); Chen & Zheng (2019) also follow Ghosh & Roth (2011) and design randomized mechanisms that use user data with a probability that depends on their reported privacy sensitivity value.

Fairness In Jia et al. (2019), Ghorbani & Zou (2019) and Ghorbani et al. (2020) a framework for determining the fair value of data is proposed. These works extend the foundational principles of the Shapley value (Shapley, 1952), which was originally proposed as a concept for utility division in coalitional games to the setting of data. Our work takes this idea further and explicitly includes privacy in the definition of the fair value of data, ultimately allowing us to consider private data acquisition in the context of fairness constraints. Finally, we note that we consider the concept of fairness in data valuation, not algorithmic fairness, which relates to the systematic failure of machine learning systems to account for data imbalances.

1.2 Main Contributions

- We present an axiomatic notion of fairness that is inclusive of the platforms and the users in Theorem 1. The utility to be awarded to each user and the platform is uniquely determined, providing a useful benchmark for comparison.
- In the realistic scenario that fairness is considered between users, Theorem 2 defines a notion of fairness based on axioms, but only places restriction on relative amounts distributed to the players. This creates an opportunity for the platform to optimize utility under fairness constraints.
- Section 4 contains an example inspired by online platform advertisement to heterogeneous users. We use our framework to fairly allocate payments, noticing how those payments differ among different types of users, and how payments change as the degree of heterogeneity increases or decreases. We numerically investigate the mechanism design problem under this example and see how heterogeneity impacts the optimal behavior of the platform.
- Finally, Section 5 explores the platform mechanism design problem. In Theorem 3 we establish that there are three distinct regimes in which the platform's optimal behavior differs depending on the common privacy sensitivity of the users. While existing literature has investigated how a platform should design incentives for users to optimize its utility, this is the first work to consider fairness constraints on the platform. When privacy sensitivity is low, the platform will set incentives to ensure that it collects all the data with the lowest privacy options. When the privacy sensitivity is above a given threshold, the platform will provide no incentives to users. Between these two extremes, the platform will set the incentives so some fraction of the users chooses the higher privacy option and the remaining chooses the lower privacy option.

Notation Lowercase boldface $\mathbf{x}$ and uppercase boldface $\mathbf{X}$ symbols denote vectors and matrices respectively. $\mathbf{X} \odot \mathbf{Y}$ represents the element-wise product of $\mathbf{X}$ and $\mathbf{Y}$. We use $\mathbb{R}_{\geq 0}$ for non-negative reals. Finally, $\mathbf{x} \geq \mathbf{y}$ means that $x_i \geq y_i \ \forall i$. For a reference list of all symbols and their meaning, see Appendix A.

2 PROBLEM SETTING

2.1 Privacy Levels and Utility Functions

Definition 1. A *heterogeneous privacy framework* on the space of random function $A : \mathcal{X}^N \to \mathcal{Y}$ is:

1. A set of *privacy levels* $\mathcal{E} \subseteq \mathbb{R}_{\geq 0} \cup \{\infty\}$, representing the amount of privacy of each user. We use ρ to represent an element of $\mathcal{E}$ in the general case and ϵ when the privacy levels are referring to DP parameters (defined below).

2. A constraint set $\mathcal{A}(\boldsymbol{\rho}) \subseteq \{A : \mathcal{X}^N \to \mathcal{Y}\}$, representing the set of random functions that respect the privacy levels $\rho_i \in \mathcal{E}$ for all $i \in [N]$. If a function $A \in \mathcal{A}(\boldsymbol{\rho})$ then we call it a $\boldsymbol{\rho}$-private algorithm.

We maintain this general notion of privacy framework because different notions of privacy can be useful in different situations. For example, the lack of rigor associated with notions such as FL, may make it unsuitable for high security applications, but it may be very useful in protecting users against data breaches on servers, by keeping their data local. One popular choice with rigorous guarantees is DP:

Definition 2. Pure heterogeneous $\boldsymbol{\epsilon}$-DP, is a heterogeneous privacy framework with $\mathcal{E} = \mathbb{R}_{\geq 0} \cup \{\infty\}$ and the constraint set $\mathcal{A}(\boldsymbol{\epsilon}) = \{A : \Pr(A(\mathbf{x}) \in S) \leq e^{\epsilon_i} \Pr(A(\mathbf{x}') \in S)\}$ for all measurable sets S.

Henceforth we will use the symbol $\boldsymbol{\epsilon}$ to represent privacy level when we are specifically referring to DP as our privacy framework, but if we are referring to a general privacy level, we will use $\boldsymbol{\rho}$. Fig. 2, depicts another heterogeneous privacy framework. $\rho_i = 0$ means the user will keep their data fully private, $\rho_i = 1$ is an intermediate privacy option where user data is securely aggregated with other users before it is sent to the platform, which obfuscates it from the platform. Finally, if $\rho_i = 2$, the users send a sufficient statistic for their data to the platform. The platform applies an $\boldsymbol{\rho}$-private algorithm $A_{\boldsymbol{\rho}} : \mathcal{X}^N \mapsto \mathcal{Y}$ to process the data, providing privacy level ρ_i to data x_i. The output of the algorithm $y = A_{\boldsymbol{\rho}}(\mathbf{x})$ is used by the platform to derive utility U, which depends on the privacy level $\boldsymbol{\rho}$.

For example, if the platform is estimating the mean of a population, the utility could depend on the mean square error of the private estimator.

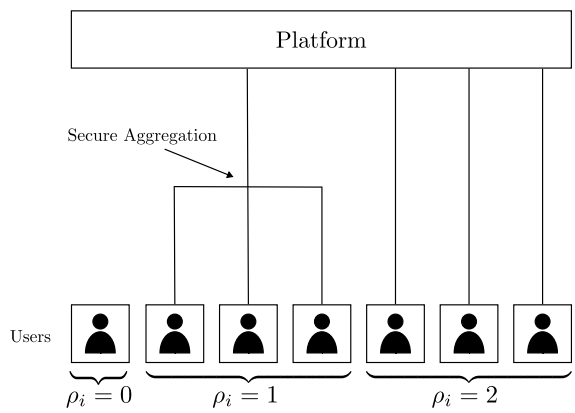


Figure 2: Users choose between three levels of privacy. If $\rho_i = 0$, users send no data to the platform. If $\rho_i = 1$, a user's model is securely combined with other users who also choose $\rho_i = 1$, and the platform receives only the combined model. If $\rho_i = 2$, users send their relevant information directly to the platform.

Differences from prior work This formulation differs from the literature of optimal data acquisition (i.e., Fallah et al. (2022)), where *privacy sensitivity* is reported by users, and the platform chooses privacy levels ρ_i based on this sensitivity. Privacy sensitivity is the cost that a user experiences by choosing a particular privacy level. Their formulation allows for a relatively straightforward application of notions like incentive compatibility and individual rationality from mechanism design theory. In this work, we instead emphasize that users choose a privacy level, rather than report a somewhat nebulously defined privacy sensitivity. Despite this difference, the notions of fairness described in the following section can be applied more broadly.

2.2 The Data Acquisition Problem

The platform generates transferable and divisible utility $U(\boldsymbol{\rho})$ from the user data. In exchange, distributes a portion of the utility $t_i(\rho_i; \boldsymbol{\rho}_{-i})$ to user i, where $\boldsymbol{\rho}_{-i}$ denotes the vector of privacy levels $\boldsymbol{\rho}$ with the ith coordinate deleted. These incentives motivate users to lower their privacy level, but each user will also have some sensitivity to their data being shared, modelled by a *sensitivity function* $c_i : \mathcal{E} \to [0, \infty)$, $c_i(0) = 0$. The behavior of users can be modelled with the help of a utility function:

$$u_i(\boldsymbol{\rho}) = t_i(\rho_i, \boldsymbol{\rho}_{-i}) - c_i(\rho_i). \tag{1}$$

The payment to user i will tend to increase with a lower privacy level, as the platform can better exploit the data, but their sensitivity c_i will increase with ρ_i, creating a trade-off. By specifying $t_i(\rho_i; \boldsymbol{\rho}_{-i})$, the platform effectively creates a game among the users. This situation is depicted in Fig. 3. Each user's action is the level of privacy that they request for the data they share. Users (players) select their privacy level ρ_i

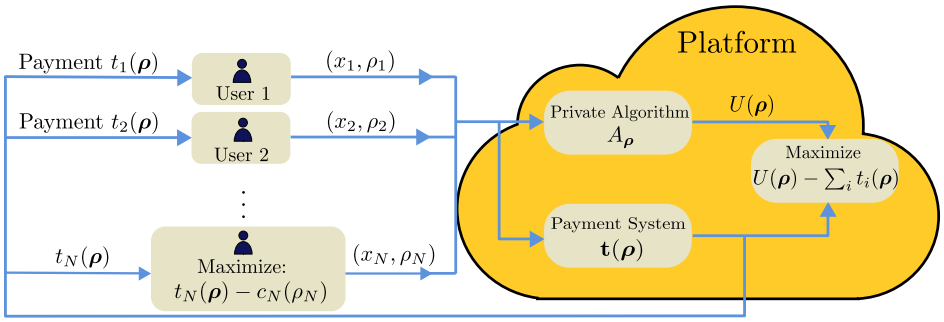


Figure 3: Users send their data x_i and a privacy level ρ_i to the central platform in exchange for payments $t_i(\rho_i; \boldsymbol{\rho}_{-i})$. The central platform extracts utility from the data at a given privacy level and optimizes incentives to maximize the difference between the utility and the sum of payments $U(\boldsymbol{\rho}) - \sum_i t_i(\rho)$.

by considering their utility function u_i and the potential actions of the other players. The platform's goal is to design the payments $t_i(\rho_i; \boldsymbol{\rho}_{-i})$ that maximize its net utility $U(\boldsymbol{\rho}) - \mathbb{1}^T \mathbf{t}(\boldsymbol{\rho})$. One way to formulate this problem is to consider maximizing this difference at equilibrium points:

$$\begin{aligned} \underset{\mathbf{t}(\cdot), \boldsymbol{\rho}}{\text{maximize}} \quad & U(\boldsymbol{\rho}) - \mathbb{1}^T \mathbf{t}(\boldsymbol{\rho}) \\ \text{subject to} \quad & \boldsymbol{\rho} \in \text{NE}(\mathbf{t}). \end{aligned} \tag{2}$$

$\text{NE}(\mathbf{t})$ denotes the set of NE strategies induced by the payment function $\mathbf{t}$, which is the vector with payment function t_i at index i. Recall that the NE is a stable state such that no user gains by unilaterally changing their strategy. Depending on the circumstances, we may also want to consider equilibrium points in mixed strategies (distributions) over the privacy space. This could make sense if we expect users to continually interact with the platform, making a different choice each time, such that users ultimately converge to their long-run average payoff. In such cases, we can formulate the problem for the platform as:

$$\begin{aligned} \underset{\mathbf{t}(\cdot), \mathcal{P}}{\text{maximize}} \quad & U(\mathcal{P}) - \mathbb{1}^T \mathbf{t}(\mathcal{P}) \\ \text{subject to} \quad & \mathcal{P} \in \text{NE}(\mathbf{t}). \end{aligned} \tag{3}$$

where we have used the shorthand $f(\mathcal{P}) = \mathbb{E}_{\boldsymbol{\rho} \sim \mathcal{P}}[f(\boldsymbol{\rho})]$ and $\mathcal{P}$ represents a distribution over the privacy space $\mathcal{E}$. Note that in both equation 3 and equation 2 restrictions must be placed on $\mathbf{t}$, otherwise it can be made arbitrarily negative. *Individual rationality* is a common condition in mechanism design that says that a user can be made no worse off by participation. In Section 5, we consider a fairness constraint.

2.3 Model Limitations

Known sensitivity functions To solve equation 3, the platform requires the privacy sensitivity c_i of each user, and our solution in Section 5 depends on this information. This can be justified when platforms interact with businesses. For example, an AI heath platform may interact with insurance companies and hospitals and can invest significant resources into studying each of its partners. Another example is advertisement platforms and sellers. Another justification is that the privacy sensitivity c_i is learned by the platforms over time, and we are operating in a regime where the estimates of c_i have converged. An interesting future direction could be investigating this learning problem.

Data-correlated sensitivity In Section 5 we treat the sensitivity function c_i as fixed and known, but a practical concern is that c_i may depends on the data x_i. Say x_i is biological data pertaining to a disease. Those users with the diseases may have higher c_i. Without taking this into account, the collected data will be biased. If our utility function is greatly increased by those users who do have the disease though, they may receive far more payment, compensating for this correlation. We leave a investigation of data-correlated sensitivity and fairness to future work.

Known transferable and divisible utility Solving equation 3 also requires knowledge of the utility function. In some cases, the platform may dictate the utility entirely on its own, perhaps to value a diverse set of users. In other cases, like in the estimation setting of Example 3.2, it may represent a more concrete metric, like a risk function that is easily computed. In some cases, however, the utility function may not be easily computed. For example, it may depend on the revenue of a company's product, or the downstream performance of a deep network. We also note that $t_i(\rho_i; \boldsymbol{\rho}_{-i})$ may not represent a monetary transfer. Individuals are often compensated for data via discounts or access to services. A shortcoming of our model is that we assume a divisible and transferable utility, which may fail to capture these nuances of compensation.

Informed and Strategic Users We also assume that users can compute and play their equilibrium strategy, which is a standard assumption in game theory. Practically this also means that the platform must be transparent about the incentives, fully publishing this information to the users.

3 Axiomatic Fairness with Privacy

What is a fair way to distribute incentives? One approach is to view the users and platforms as a coalition jointly generating utility. Following an axiomatic approach to fairness, the celebrated Shapley value (Shapley, 1952) describes how to fairly divide utility among a coalition. In this section, we take a similar approach to defining fairness. This coalitional perspective is not a complete characterization of the complex dynamics between users and platforms, but we argue that it is still a useful one. One of the benefits of this concept of fairness is that it deals with intrinsic value (i.e., how much of the utility comes from the data). This is in contrast to the *market value* that users are willing to sell for (potentially depressed). This information is particularly useful to economists, regulators, and investors, who are interested in characterizing the value of data as capital for the purposes of analysis, taxation, and investment respectively.

3.1 Platform as a Coalition Member

We define a coalition of users and a platform as a collection of s users, with $0 \leq s \leq N$ and up to 1 platform. Let $z \in \{0, 1\}$ represent the action of the platform. Let $z = 1$ when the platform chooses to join the coalition, and $z = 0$ otherwise. Let $U(\boldsymbol{\rho})$ be as defined in Section 2. We augment the utility to take into account that the utility is zero if the platform does not participate, and define $\boldsymbol{\rho}_S$ as follows:

$$U(z, \boldsymbol{\rho}) := \begin{cases} U(\boldsymbol{\rho}) & z = 1 \\ 0 & z = 0 \end{cases}, \quad [\boldsymbol{\rho}_S]_i := \begin{cases} \rho_i & i \in S \\ 0 & \text{else} \end{cases}. \tag{4}$$

Let $\phi_p(z, \boldsymbol{\rho})$ and $\phi_i(z, \boldsymbol{\rho})$, $i \in [N]$ represent the "fair" amount of utility awarded to the platform and each user i respectively, given z and $\boldsymbol{\rho}$, otherwise described as the "value" of a user. Note that these values depend implicitly on both the private algorithm $A_{\boldsymbol{\rho}}$ and the utility function U, but for brevity, we avoid writing this dependence explicitly. The result of Hart & Mas-Colell (1989) show that these values are unique and well defined if they satisfy the following three axioms:

A.i) **Two equally contributing users should be paid equally**. For any $i, j \in [N] : U(z, \boldsymbol{\rho}_{S \cup \{i\}}) = U(z, \boldsymbol{\rho}_{S \cup \{j\}}) \; \forall S \subset [N] \backslash \{i, j\} \implies \phi_i(z, \boldsymbol{\rho}) = \phi_j(z, \boldsymbol{\rho})$.

In addition, for any user $i \in [N]$, $U(1, \boldsymbol{\rho}_{S \cup \{i\}}) - U(1, \boldsymbol{\rho}_S) = 0 \;\; \forall S \subset [N] \backslash \{i\} \implies \phi_i(z, \boldsymbol{\rho}) = 0$.

A.ii) **The sum of all payments is the total utility**. The sum of values is the total utility $U(z, \boldsymbol{\rho}) = \phi_p(z, \boldsymbol{\rho}) + \sum_i \phi_i(z, \boldsymbol{\rho})$.

A.iii) **If two utility functions are combined, the payment for the combined task should be the sum of the individual tasks**. Let $\phi_p(z, \boldsymbol{\rho})$ and $\phi_i(z, \boldsymbol{\rho})$ be the value of the platform and users respectively for the utility function U, under the $\boldsymbol{\rho}$-private $A_{\boldsymbol{\rho}}$. Let V be a separate utility function, also based on the output of $A_{\boldsymbol{\rho}}$, and let $\phi'_p(z, \boldsymbol{\rho})$ and $\phi'_i(z, \boldsymbol{\rho})$ be the utility of the platform and individuals with respect to V. Then under the utility $U + V$, the value of user i is $\phi_i(z, \boldsymbol{\rho}) + \phi'_i(z, \boldsymbol{\rho})$ and the value of the platform is $\phi_p(z, \boldsymbol{\rho}) + \phi'_p(z, \boldsymbol{\rho})$.

Theorem 1. *Let $\phi_p(z, \boldsymbol{\epsilon})$ and $\phi_i(z, \boldsymbol{\epsilon})$ satisfying axioms* (A.i-iii) *represent the portion of total utility awarded to the platform and each user i from utility $U(z, \boldsymbol{\epsilon})$. Then they are unique and take the form:*

$$\phi_p(z, \boldsymbol{\rho}) = \frac{1}{N+1} \sum_{S \subseteq [N]} \frac{1}{\binom{N}{|S|}} U(z, \boldsymbol{\rho}_S), \tag{5}$$

$$\phi_i(z, \boldsymbol{\rho}) = \frac{1}{N+1} \sum_{S \subseteq [N] \setminus \{i\}} \frac{1}{\binom{N}{|S|+1}} \left(U(z, \boldsymbol{\rho}_{S \cup \{i\}}) - U(z, \boldsymbol{\rho}_S) \right). \tag{6}$$

Theorem 1 is proved in Appendix B.2, and resembles the classic Shapley value result (Shapley, 1952).

3.2 Fairness Among Users

Though we can view the interactions between the platform and the users as a coalition, due to the asymmetry that exists between the platform and the users, it also makes sense to discuss fairness among the users alone. In this case, we can consider an analogous set of axioms that involve only the users.

B.i) **Two equally contributing users should be paid equally**. For any $i, j \in [N] : U(\boldsymbol{\rho}_{S \cup \{i\}}) = U(\boldsymbol{\rho}_{S \cup \{j\}}) \ \forall S \subset [N] \setminus \{i, j\} \implies \phi_i(\boldsymbol{\rho}) = \phi_j(\boldsymbol{\rho})$.

In addition, for any user $i \in [N]$, $U(\boldsymbol{\rho}_{S \cup \{i\}}) - U(\boldsymbol{\rho}_S) = 0 \ \forall S \subset [N] \setminus \{i\} \implies \phi_i(\boldsymbol{\rho}) = 0$.

B.ii) **The sum of all payments is an $\boldsymbol{\alpha}(\boldsymbol{\epsilon})$ fraction of the total utility**. The sum of values is the total utility $\alpha(\boldsymbol{\rho}) U(\boldsymbol{\rho}) = \sum_i \phi_i(\boldsymbol{\rho})$. Where if $U(\boldsymbol{\rho}) = U(\tilde{\boldsymbol{\rho}})$ then $\alpha(\boldsymbol{\rho}) = \alpha(\tilde{\boldsymbol{\rho}})$ and $0 \leq \alpha(\boldsymbol{\rho}) \leq 1$.

B.iii) **If two utility functions are combined, the payment for the combined task should be the sum of the individual tasks**. Let $\phi_i(\boldsymbol{\rho})$ be the value of users for the utility function U, under the $\boldsymbol{\epsilon}$-private algorithm $A_{\boldsymbol{\rho}}$. Let V be a separate utility function, also based on the output of the algorithm $A_{\boldsymbol{\epsilon}}$, and let $\phi'_i(\boldsymbol{\rho})$ be the utility of the users with respect to V. Then under the utility $U + V$, the value of user i is $\phi_i(\boldsymbol{\rho}) + \phi'_i(\boldsymbol{\rho})$.

A notable difference between these axioms and (A.i-iii) is that the efficiency condition is replaced with pseudo-efficiency. Under this condition, the platform may determine the sum of payments awarded to the players, but this sum should in general depend only on the utility itself, and not on how that utility is achieved.

Theorem 2. *Let $\phi_i(\boldsymbol{\rho})$ satisfying axioms* (B.i-iii) *represent the portion of total utility awarded to each user i from utility $U(\boldsymbol{\rho})$. Then for $\alpha(\boldsymbol{\rho})$ that satisfies axiom* (B.ii) *ϕ_i takes the form:*

$$\phi_i(\boldsymbol{\rho}) = \frac{\alpha(\boldsymbol{\rho})}{N} \sum_{S \subseteq [N] \setminus \{i\}} \frac{1}{\binom{N-1}{|S|}} \left(U(\boldsymbol{\rho}_{S \cup \{i\}}) - U(\boldsymbol{\rho}_S) \right). \tag{7}$$

The proof of Theorem 2 can be found in Appendix B.2. This result is similar to the classic Shapley value (Shapley, 1952), but differs in its novel asymmetric treatment of the platform.

Computational Complexity At first glance it may seem that both notions of fairness have exponential computational complexity of $N |\mathcal{E}|^N$. This is only true for a worst-case exact computation. In order for these notions to be useful in any meaningful way, we must be able to compute them. Thankfully, in practice, U typically has a structure that makes the problem more tractable. In Ghorbani & Zou (2019), Jia et al. (2019), Wang & Jia (2023) and Lundberg & Lee (2017) special structures are used to compute the types of Shapley value sums we are considering with significantly reduced complexity, particularly in cases where the U is related to the accuracy of a deep network. This is critical because we want to compute fair values for large number of users. For example, our platform could be a medical data network with hundreds of hospitals as our users, or a smartphone company with millions of users, and we need to be able to scale computation to accurately compute these fair values.

Example: Differentially Private Estimation In this example, we use DP as our heterogeneous privacy framework. Let X_i represent independent and identically distributed data of user i respectively, with $\Pr(X_i = 1/2) = p$ and $\Pr(X_i = -1/2) = 1 - p$, with $p \sim \text{Unif}(0, 1)$. The platform's goal is to construct an $\boldsymbol{\epsilon}$-DP estimator for $\mu := \mathbb{E}[X_i] = p - 1/2$ that minimizes Bayes risk. There is no general procedure for finding the Bayes optimal $\boldsymbol{\epsilon}$-DP estimator, so restrict our attention to $\boldsymbol{\epsilon}$-DP linear-Laplace estimators of the form:

$$A(\mathbf{X}) = \mathbf{w}(\boldsymbol{\epsilon})^T \mathbf{X} + Z, \tag{8}$$

where $Z \sim \text{Laplace}(1/\eta(\boldsymbol{\epsilon}))$. In Fallah et al. (2022) the authors argue that unbiased linear estimators are nearly optimal in a minimax sense for bounded random variables. We assume a squared error loss $L(a, \mu) = (a - \mu)^2$ and let $\mathcal{A}_{\text{lin}}(\boldsymbol{\epsilon})$ be the set of $\boldsymbol{\epsilon}$-DP estimators satisfying equation 8. Then, we define:

$$A_{\boldsymbol{\epsilon}} = \underset{A \in \mathcal{A}_{\text{lin}}(\boldsymbol{\epsilon})}{\arg\min}\ \mathbb{E}[L(A(\mathbf{X}), \mu)] \qquad r(\boldsymbol{\epsilon}) = \mathbb{E}[L(A_{\boldsymbol{\epsilon}}(\mathbf{X}), \mu)]. \tag{9}$$

In words, $A_{\boldsymbol{\epsilon}}$ is an $\boldsymbol{\epsilon}$-DP estimator of the form equation 8, where $\mathbf{w}(\boldsymbol{\epsilon})$ and $\eta(\boldsymbol{\epsilon})$ are chosen to minimize the Bayes risk of the estimator, and $r(\boldsymbol{\epsilon})$ is the risk achieved by $A_{\boldsymbol{\epsilon}}$. Since the platform's goal is to accurately estimate the mean of the data, it is natural for the utility $U(\boldsymbol{\epsilon})$ to depend on $\boldsymbol{\epsilon}$ through the risk function $r(\boldsymbol{\epsilon})$. Note that if U is monotone decreasing in $r(\boldsymbol{\epsilon})$, then U is monotone increasing in $\boldsymbol{\epsilon}$. Let us now consider the case of $N = 2$ users, choosing from an action space of $\mathcal{E} = \{0, \epsilon'\}$, for some $\epsilon' > 0$. Furthermore, take U to be an affine function of $r(\boldsymbol{\epsilon})$: $U(\boldsymbol{\epsilon}) = c_1 r(\boldsymbol{\epsilon}) + c_2$. For concreteness, take $U(\mathbf{0}) = 0$ and $\sup_{\boldsymbol{\epsilon} \in \mathbb{R}} U(\boldsymbol{\epsilon}) = 1$. Note that this ensures that U is monotone increasing in $\boldsymbol{\epsilon}$, and is uniquely defined. Considering the example of a binary privacy space $\mathcal{E} = \{0, \infty\}$ ($\epsilon' = \infty$), the utility can be written in matrix form as:

$$\mathbf{U} = \begin{bmatrix} U([0,0]) & U([0,\epsilon']) \\ U([\epsilon',0]) & U([\epsilon',\epsilon']) \end{bmatrix} = \begin{bmatrix} 0 & 2/3 \\ 2/3 & 1 \end{bmatrix}. \tag{10}$$

Derivations are available in Appendix B.1. Note from equation 5 and equation 6, it is clear that $\phi_p(0, \boldsymbol{\epsilon}) = \phi_i(0, \boldsymbol{\epsilon}) = 0$. Let $\boldsymbol{\Phi}_p$ and $\boldsymbol{\Phi}_i^{(1)}$ represent the functions $\phi_p(1, \boldsymbol{\epsilon})$ and $\phi_i(1, \boldsymbol{\epsilon})$ in matrix form akin to $\mathbf{U}$. Then using equation 5 and equation 6, we find that the fair allocations of the utility are given by:

$$\boldsymbol{\Phi}_p = \begin{bmatrix} 0 & 1/3 \\ 1/3 & 5/9 \end{bmatrix}, \ \boldsymbol{\Phi}_1^{(1)} = \begin{bmatrix} 0 & 1/3 \\ 0 & 2/9 \end{bmatrix}, \ \boldsymbol{\Phi}_2^{(1)} = \begin{bmatrix} 0 & 0 \\ 1/3 & 2/9 \end{bmatrix}. \tag{11}$$

Consider the utility function defined in equation 10, for the $N = 2$ user mean estimation problem with $\mathcal{E} = \{0, \infty\}$. By Theorem 2 the fair allocation satisfying (B.i-iii) must be of the form:

$$\boldsymbol{\Phi}_1^{(2)} = \mathbf{A} \odot \begin{bmatrix} 0 & 2/3 \\ 0 & 1/2 \end{bmatrix}, \quad \boldsymbol{\Phi}_2^{(2)} = \mathbf{A} \odot \begin{bmatrix} 0 & 0 \\ 2/3 & 1/2 \end{bmatrix}, \ \mathbf{A} = \mathbf{A}^T, \ \ 0 \le [\mathbf{A}]_{ij} \le 1. \tag{12}$$

4 Fair Incentives in Federated Learning

FL is a distributed learning process used when data is either too large or too sensitive to be directly transferred in full to the platform. Instead of combining all the data together and learning at the platform, each user performs some part of the learning locally and the results are aggregated at the platform, providing some level of privacy. Donahue & Kleinberg (2021) consider a setting where heterogeneous users voluntarily opt-in to federation. A natural question to ask is: how much less valuable to the platform is a user that chooses to federate with others as compared to one that provides full access to their data? Furthermore, how should the platform allocate incentives to get users to federate? This section addresses these questions.

Each user $i \in [N]$ has a unique mean and variance $(\theta_i, \sigma_i^2) \sim \Theta$, where Θ is some global joint distribution. Let θ_i represent some information about the user critical for advertising. We wish to learn θ_i as accurately as possible to maximize our profits, by serving the best advertisements possible to each user. User i draws n_i samples i.i.d. from its local distribution $\mathcal{D}_i(\theta_i, \sigma_i^2)$, that is, some distribution with mean θ_i and variance

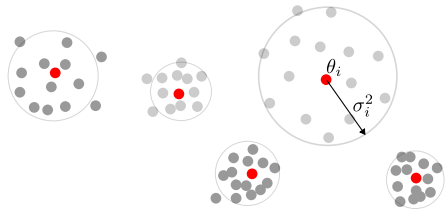


Figure 4: Each user $i \in [N]$ has mean and variance $(\theta_i, \sigma_i^2) \sim \Theta$, where Θ is a global joint distribution. Let $s^2 = \mathrm{Var}(\theta_i)$ and $r^2 = \mathbb{E}[\sigma_i^2]$ for all i. In this case s^2 is large relative to r^2, and the data is very heterogeneous.

σ_i^2. Let $s^2 = \mathrm{Var}(\theta_i)$ represent the variance between users and $r^2 = \mathbb{E}[\sigma_i^2]$ represent the variance within a user's data. When $s^2 \gg r^2/n_i$ the data is very heterogeneous, and it is generally not helpful to include much information from the other users when estimating θ_i, however, if $s^2 \ll r^2/n_i$, the situation is reversed, and information from the other users will be very useful. The goal of the platform is to construct estimators $\hat{\theta}_i^p$ while respecting the privacy level vector $\boldsymbol{\rho}$:

$$\mathrm{EMSE}_i(\boldsymbol{\rho}) := \mathbb{E}\left[\left(\hat{\theta}_i^p(\boldsymbol{\rho}) - \theta_i\right)^2\right]. \tag{13}$$

Fig. 2 summarizes our FL formulation. Users can choose from a 3-level privacy space $\mathcal{E} = \{0, 1, 2\}$. In this case the privacy space is not related to DP, but instead encodes how users choose to share their data with the platform. Let N_j be the number of users that choose privacy level j. The *heterogeneous privacy framework* is given in Table 1. Note that the error in estimating θ_i depends not just on the privacy level of the ith

Level	Description	Platform gets
$\rho_i = 2$	Provide local estimator directly to the platform.	$\hat{\theta}_i$
$\rho_i = 1$	Provide securely aggregated model with other users of same privacy level.	$\hat{\theta}^f = \frac{1}{N_1}\sum_{i:\rho_i=1}\hat{\theta}_i$
$\rho_i = 0$	Provide no data to the platform.	Nothing

Table 1: Privacy Level Description

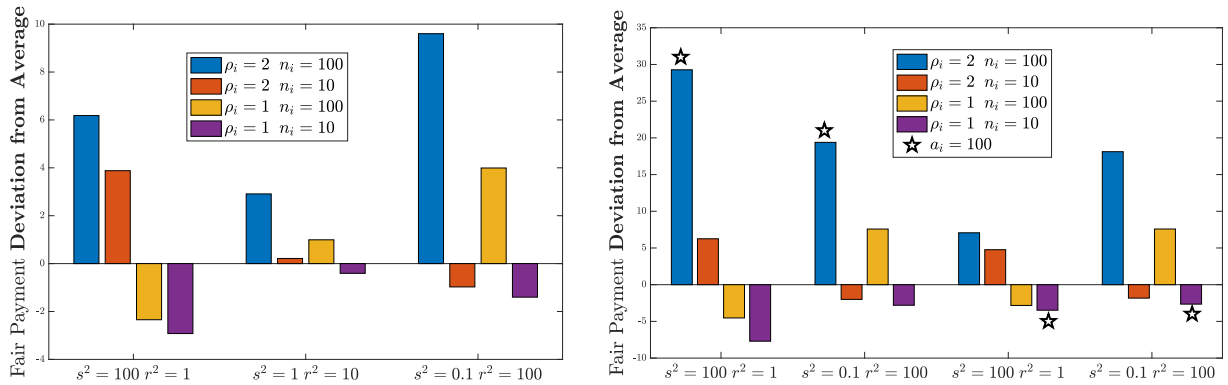
user ρ_i, but on the entire privacy vector. Let the users be ordered such that ρ_i is a non-increasing sequence. Then for each i the platform constructs estimators of the form:

$$\hat{\theta}_i^p = w_{i0}\hat{\theta}^f + \sum_{j=1}^{N_2} w_{ij}\hat{\theta}_j, \tag{14}$$

where, $\sum_j w_{ij} = 1$ for all i. In Proposition 5, found in Appendix B.3, we calculate the optimal choice of w_{ij} which depends on $\boldsymbol{\rho}$. From these estimators, the platform generates utility $U(\boldsymbol{\rho})$. The optimal w_{i0} and w_{ij} in equation 14 are well defined in a Bayesian sense if $\rho_i > 0$ for some i, but this does not make sense when $\boldsymbol{\rho} = \mathbf{0}$. We can get around this by defining $\mathrm{EMSE}_i(\mathbf{0}) := r^2 + 2s^2$. For the purposes of our discussion, we assume a logarithmic form utility function. This logarithmic form is common in utility theories dating back at least to Kelly (1956). In the following section, we make a *diminishing returns* assumption to derive our theoretical result, which the logarithmic utility satisfies. The exact utility function we consider is:

$$U(\boldsymbol{\rho}) := \sum_{i=1}^{n} a_i \log\left(\frac{(r^2 + 2s^2)}{\mathrm{EMSE}_i(\boldsymbol{\rho})}\right). \tag{15}$$

a_i **represents the relative importance of each user**. This is important to model because some users may spend more than others, and are thus more important to the platform i.e., the platform may care about computing their θ_i more accurately than the other users. The argument of the log is increasing as the EMSE decreases. The log means there is diminishing returns as each $\hat{\theta}_i$ becomes more accurate.



(a) (b)

Figure 5: **(a)** Plot of *difference from the average* utility per user $U(\boldsymbol{\rho})/N$ for each of the four different types of users, for three different regimes of $s^2 = \mathrm{Var}(\theta_i)$ and $r^2 = \mathbb{E}[\sigma_i^2]$, with heterogeneity decreasing from left to right. In left (most heterogeneous) plot users who choose $\rho_i = 2$ are more valuable compared to those that choose $\rho_1 = 1$. In the center there is an intermediate regime, where all users are paid closer to the average, with users with more data being favored slightly. In the rightmost graph, with little heterogeneity users with more data are paid more, and privacy level has a lesser impact on the payments.
(b) In each case there is one user i with $a_i = 100$ (indicated with a star), while all other users $j \neq i$ have $a_j = 1$ (a_i represents the relative importance of the user in the utility function). In the two leftmost set of bars, we see that the user with $\rho_i = 2$ and $n_i = 100$ receives by far the most payment, when heterogeneity is high, but this becomes less dramatic as heterogeneity decreases. This shows that when users are very heterogeneous, if a_i is large for only user i, most of the benefit in terms of additional payments should go to user i. Likewise, comparing the second from the left and the rightmost plots we see little difference, showing that the opposite is true in the homogeneous case: any user can benefit from any other user having a large a_i.

4.1 Fair Payments Under Optional Federation

In this section, we focus on our definition of fairness in Theorem 2 and analyze the fair values $\phi_i(\boldsymbol{\rho})$ that are induced when using that notion of fairness. Let there be $N = 10$ users. $N_1 = 5$ of these users opt for federating ($\rho_i = 1$), $N_2 = 4$ directly provide their data to the platform ($\rho_i = 2$), and finally, $N_0 = 1$ user chooses to not participate ($\rho_i = 0$). In this subsection (and Fig. 5), without loss of generality, we assume $\alpha(\boldsymbol{\rho}) = 1$, and the results of this section can be scaled accordingly. The choices of ρ_i for each user depends on their individual privacy sensitivity functions c_i, but we defer that discussion to the next subsection.

Different Amounts of Data Fig 5a plots the difference from an equal distribution of utility, i.e., how much each user's utility differs from $U(\boldsymbol{\rho})/N$. We assume $a_i = 1$ for all users. In the bars furthest to the left, where $s^2 = 100$ and $r^2 = 1$, we are in a very heterogeneous environment. Intuitively, this means that a user j will have data that may not be helpful for estimating θ_i for $j \neq i$, thus those users that choose $\rho_i = 2$ are paid the most, since at the very least, the information they provide can be used to target their own θ_i. Likewise, users that federate obfuscate where their data is coming from, making their data less valuable (since their own θ_i cannot be targeted), so users with $\rho_i = 1$ are paid less than an even allocation. On the right side, we have a regime where $s^2 = 0.1$ and $r^2 = 100$, meaning users are similar and user data more exchangeable. Now users with larger n_i are paid above the average utility per user, while those with lower n_i are paid less. Users with $\rho_i = 2$ still receive more than those with $\rho_i = 1$ when n_i is fixed, and this difference is significant when $n_i = 100$. In the center we have an intermediate regime of heterogeneity, where $s^2 = 1$ and $r^2 = 10$. Differences in payments appear less pronounced, interpolating between the two extremes.

More Valuable Users Fig 5b is like Fig 5a, except now in each set of graphs, exactly one user has $a_i = 100$, meaning that estimating θ_i for user i is 100 times more important than the others. Looking at the two leftmost sets of bars in Fig 5b we see that when user i with $\rho_i = 2$ and $n_i = 100$ is the most important

one, when s^2 is large compared to r^2, it is user i who receives most of the benefit in terms of its payment but when s^2 is smaller, other users also benefit. This can be intuitively explained as follows: if users are very heterogeneous, other users $j \neq i$ do not have data that is helpful for determining θ_i, thus they do not benefit when user i has a larger a_i. Likewise, when s^2 is small compared to r^2 not just user i benefits, but also all those users that contribute more data, as those users with $\rho_i = 1$ and $n_i = 100$ are also paid over the average utility per user. Another key point is the similarity between the second and fourth set of graphs. This tells an interesting story: when users are not very heterogeneous, regardless of which user is has $a_i = 100$, it is those users with large n_i that will benefit.

4.2 Platform and User Game - Mechanism Design

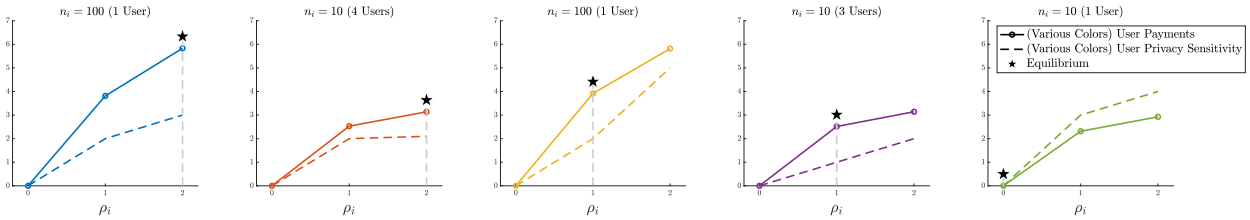


Figure 6: Plot of payments and user sensitivities c_i for 5 different types of users with different sensitivity functions. Some user types are repeated, (indicated in the title), such that there is a total of $N = 10$ Users. Users choose the privacy level ρ_i to maximize the difference between their payment and privacy sensitivity function. A star in each plot marks each user's top choice. The payment functions $\phi_i(\boldsymbol{\rho})$ changes based on the privacy levels of all the users. At the given configuration, which matches the left-most plot in Fig 5b ($s^2 = 100, r^2 = 1$), no user benefits by changing their choice of ρ_i, thus the configuration is a Nash Equilibrium.

Nash Equilibrium Under Fair Payments We have just discussed how the fair values $\phi_i(\boldsymbol{\rho})$ change depending on the model parameters, and the choices of privacy level ρ_i. Now we discuss *how* the users come to decide their ρ_i. We again focus on the notion of privacy in Theorem 2, but we now restrict $\alpha(\boldsymbol{\rho}) = \alpha \in [0, 1]$. To avoid overly complicating the model, we take $a_i = 1$ for all users. User i chooses their privacy level ρ_i based on both the payments that they receive from the platform $\alpha\phi_i(\rho_i; \boldsymbol{\rho}_{-i})$ as well as their own unique privacy sensitivity function $c_i(\rho_i)$. In this context, we can view each user as optimizing their own local utility function:

$$u_i(\rho_i, \boldsymbol{\rho}_{-i}) := \alpha\phi(\rho_i; \boldsymbol{\rho}_{-i}) - c_i(\rho_i). \tag{16}$$

User i individually optimizes this function to determine their *best response* (denoted BR_i) to the choices of the other users and the platform:

$$\mathrm{BR}_i(\boldsymbol{\rho}_{-i}, \alpha) := \arg\max_{\rho_i} u_i(\rho_i, \boldsymbol{\rho}_{-i}) \tag{17}$$

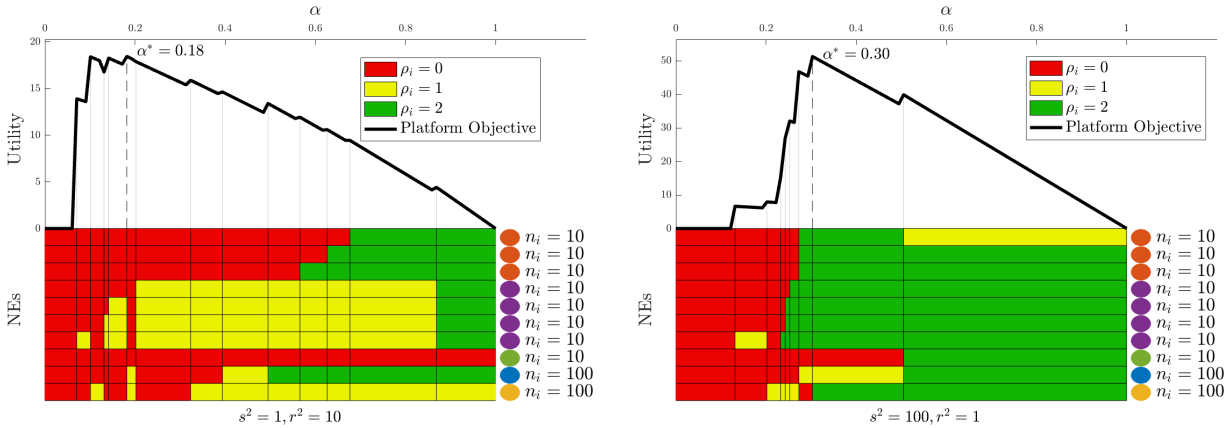
The full *best response function* $\mathrm{BR}(\boldsymbol{\rho}, \alpha) := [\mathrm{BR}_1(\boldsymbol{\rho}_{-1}, \alpha), \ldots, \mathrm{BR}_N(\boldsymbol{\rho}_{-N}, \alpha)]^T$ is a vector that collects all users' individual best response functions. The "fixed points" of the best response at a given α constitute the pure-strategy NEs:

$$\mathrm{NE}(\alpha) := \{\boldsymbol{\rho} : \boldsymbol{\rho} = \mathrm{BR}(\boldsymbol{\rho}, \alpha)\}. \tag{18}$$

Fig 6 depicts a particular fixed point of the best response function (therefore a NE) in our federated mean estimation example. Each of the $N = 10$ users in assigned one of 5 unique sensitivity functions c_i, which are shown in the figure.

Solving the Fair Data Acquisition Problem The platform's goal is to set α such that it maximizes the total amount of utility it receives. Since the NE set depends on α the platform has some control over the behavior of the users. The mechanism design problem in this case reduces to:

$$\begin{aligned} \underset{\alpha, \boldsymbol{\rho}}{\text{maximize}} \quad & (1-\alpha)U(\boldsymbol{\rho}) \\ \text{subject to} \quad & \boldsymbol{\rho} \in \mathrm{NE}(\alpha). \end{aligned} \tag{19}$$



(a) Homogeneous

(b) Heterogeneous

Figure 7: **(a) and (b)** The top of both sub-figures plot the optimization landscape for the platform design problem equation 19.The solid black line depicts the partial solution to equation 19, where we optimize over $\boldsymbol{\rho}$ for fixed α. Precisely that is: $\max_{\boldsymbol{\rho}}(1-\alpha)U(\boldsymbol{\rho}),\ \boldsymbol{\rho} \in \mathrm{NE}(\alpha)$. The bottom of each plot depicts the optimizing $\arg\max_{\boldsymbol{\rho}}(1-\alpha)U(\boldsymbol{\rho}),\ \boldsymbol{\rho} \in \mathrm{NE}(\alpha)$. $\rho_i = 0$ means that the users send no data to the platform, $\rho_i = 1$ means the users securely aggregate their data with other users, and $\rho_i = 2$ means the users send all their data to the platform. Each row of bars corresponds to a unique user i. On the right-hand side, the amount of data each user has n_i is given, as well as the color-code corresponding to the c_i of each user. By looking at the same color plot in Fig 6, the c_i for the given user can be found. If there are multiple NEs that achieve the maximum utility, one is shown arbitrarily.
(a) Is in a homogeneous regime. In this example it turns out that the optimal α^* for the platform is such that both of the users with $n_i = 100$ choose $\rho_i = 1$, while the other users choose $\rho_i = 0$, showing the importance of the amount of data, as opposed to higher privacy options.
(b) Is in a more heterogeneous setting. We find that the optimal α^* ensures that most users choose $\rho_i = 2$, except one user with high privacy sensitivity that chooses $\rho_i = 0$, and one user that chooses $\rho_i = 1$.

Note that this is exactly equation 2, except we have used the fact that $\alpha(\boldsymbol{\rho})U(\boldsymbol{\rho}) = \sum_i \phi_i(\boldsymbol{\rho})$. Solving equation 19 is a daunting task, due to the complex structure of the constraint, however, it is numerically tractable in this case. The key idea to efficiently solve equation 19 is to exploit symmetries in the utility function. For instance, if $\boldsymbol{\rho}'$ is constructed from $\boldsymbol{\rho}$ by permuting the values ρ_i, ρ_j where $n_i = n_j$, then $U(\boldsymbol{\rho}) = U(\boldsymbol{\rho}')$. This allows us compute $\phi_i(\boldsymbol{\rho})$ for a much smaller number of representative $\boldsymbol{\rho}$, rather than a full combinatorial search. Despite producing the same utility, $\boldsymbol{\rho}$ and $\boldsymbol{\rho}'$ may have very different stability properties. To deal this this, we can implement an efficient tree search to identify the NEs within each group. Once we have an efficient algorithm for computing NEs, we conduct a grid search to determine the optimal α. Note that for an arbitrary problem, finding the equilibria can be a challenging task, and thus the more difficult it is to compute the equilibria, the more difficult it will be to solve the design problem in equation 2. We provide a full description of our solution in Appendix D.

Fig 7 shows the numerical solution to equation 19 for two different choices of s^2 and r^2. Fig 7a is a more homogeneous setting and we observe that the optimal α^* sets the payment just high enough so that both of the users with $n_i = 100$ choose to participate at $\rho_i = 1$, while all other users choose $\rho_i = 0$. Essentially, the platform identifies the two users with $n_i = 100$, and focused on collecting their data, rather than the data of the other 8 users. This configuration collects a majority of the data for relatively little payment. Since the data is very homogeneous anyways, the platform is not losing much utility by allowing those two users to choose $\rho_i = 1$ rather than $\rho_i = 2$. As the platform increases α, more users choose to participate, eventually with many of them choosing $\rho_i = 2$, however, the benefit is minimal, and it is outweighed by the extra payments that are required to achieve that outcome. Fig 7b covers a more heterogeneous setting. In this setting, we see that α^* is higher, and a larger fraction of the total utility is paid to users. For this α^* only two of the users do not choose $\rho_i = 2$. There is also one user with $\rho_i = 1$, and one with $\rho_i = 0$ that have

a higher privacy sensitivities. In the more heterogeneous setting, allowing users to choose a private option is more costly, so the extra payment to ensure more user choose $\rho_i = 2$ is worthwhile in this regime.

Other Formulations We conclude this section by remarking that this is just one particular formulation of a federated learning problem where users have privacy choice. Another interesting formulation can be found in Aldaghri et al. (2023), which comes from the literature on personalized federated learning Li et al. (2021). In this formulation, all users join the federated learning process, but some users can choose between a private option, where their data is protected via differential privacy, or a standard non-private option. Exploring fair incentives in this, and other models could be an interesting direction for future work.

5 Fairness Constraints: Data Acquisition

In the previous section, we considered a complex model, numerically studying the equilibria of the users based on fair payments from the platform, as well as how the platform optimally chooses α. In this section, we consider a tractable model and theoretically study how the optimal α^* changes based on the privacy sensitivity of users, under a fairness framework based on Theorem 2. This section addresses this problem by investigating the incentives of a platform designing a mechanism under the constraint of fairness.

Consider $N \geq 2$ users each with identical statistical marginal contribution, i.e., for any i, j we have $S \subseteq [N] \backslash \{i, j\}$, $U(\boldsymbol{\rho}_{S \cup \{i\}}) = U(\boldsymbol{\rho}_{S \cup \{j\}})$. The platform is restricted to making fair payments satisfying axioms (B.i-iii) with the additional constraint that $\alpha(\boldsymbol{\rho}) = \alpha \in [0, 1]$. Users choose one of two available privacy levels $\rho_i \in \mathcal{E}^N$, with $\mathcal{E} = \{\rho_1', \rho_2'\}$ and $\rho_2' > \rho_1'$. We can write the utility of the user i as

$$u(\rho_i, \boldsymbol{\rho}_{-i}) = \alpha \phi(\rho_i; \boldsymbol{\rho}_{-i}) - c \mathbb{1} \{\rho_i = \rho_2'\}. \tag{20}$$

Users gain utility from incentives provided by the platform but incur a cost of c if they choose the less private option. For now, we assume c is the same for all users; later we discuss the case where c is different. Note that we can drop the index of ϕ_i due to the assumption of equal marginal contribution. To enrich the problem, we allow users to employ a mixed strategy denoted by $\mathbf{p} = [p, (1-p)]^T$, where users choose the ρ_1' with probability p and ρ_2' with probability $1 - p$. This is justified because we expect users to repeatedly interact with platforms and sample from their mixed strategy and ultimately converge to their expected utility.

The platform is also trying to maximize the fraction of the total expected utility $U(\mathbf{p}) := \mathbb{E}_{\boldsymbol{\rho} \sim \mathbf{p}} [U(\boldsymbol{\rho})]$ that it keeps as in equation 3. The platform's goal is to choose a payment value α such that it optimizes:

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & (1-\alpha) U(\mathbf{p}^*(\alpha)) \\ \text{subject to} \quad & \mathbf{p}^*(\alpha) \in \text{NE}(\alpha). \end{aligned} \tag{21}$$

The objective is simplified compared to equation 3 by exploiting the pseudo-efficiency axiom, which says that the sum of payments is α times the total utility. The constraint in equation 21 implicitly encodes the user behavior governed by equation 20, and will change with the privacy sensitivity c. Theorem 3 characterizes the solution of equation 21 for different values of c. To make equation 21 amenable to insightful analysis, we make some mild assumptions.

Assumption 1. *The utility U is monotone: $\boldsymbol{\rho}_S^{(2)} \geq \boldsymbol{\rho}_S^{(1)} \implies U(\boldsymbol{\rho}_S^{(2)}) > U(\boldsymbol{\rho}_S^{(1)}) \ \forall S \subseteq [N]$.*

Assumption 2. *The utility U has diminishing returns. Let $n_{private}(\boldsymbol{\rho}_S)$ represent the number of elements of $i \in S$ such that $\rho_i = \rho_1'$, i.e., the number of users choosing the higher privacy option. Furthermore, define $\Delta_i U(\boldsymbol{\rho}_S) := U(\boldsymbol{\rho}_S^{(i+)}) - U(\boldsymbol{\rho}_S)$, where $\boldsymbol{\rho}_S^{(i+)}$ is equal to $\boldsymbol{\rho}_S$ except $\rho_i^{(i+)} = \rho_2'$. In other words, $\Delta_i U(\boldsymbol{\rho}_S)$ is the marginal increase in utility when the ith user switches to the lower privacy option. Then U satisfies:*

$$n_{private}(\boldsymbol{\rho}_S^{(1)}) \geq n_{private}(\boldsymbol{\rho}_S^{(2)}) \implies \Delta_i U(\boldsymbol{\rho}^{(1)}) > \Delta_i U(\boldsymbol{\rho}^{(2)}). \tag{22}$$

It is helpful to define the *expected relative payoff*, where the expectation is taken with respect to the actions of the other players. When all other users choose a mixed strategy $\mathbf{p}$, the expected relative payoff is defined as:

$$\gamma(p) := \phi(\rho_2'; \mathbf{p}) - \phi(\rho_1'; \mathbf{p}) = \mathbb{E}_{\substack{\rho_j \sim \mathbf{p} \\ j \neq i}} [\phi(\rho_2'; \boldsymbol{\rho}_{-i}) - \phi(\rho_1'; \boldsymbol{\rho}_{-i})]. \tag{23}$$

For convenience, we have defined γ in terms of the scalar p, rather than the vector $\mathbf{p} = [p,\ (1-p)]^T$. This quantity represents the expected gain in incentive (normalized to make it invariant to α) if a user switches to a less private level from the more private level given everyone else plays the mixed strategy $\mathbf{p}$.

Theorem 3. *Consider a binary privacy level game with N users and a platform. If U satisfies Assumptions 1 and 2, and the platform payments are fair as defined in Theorem 2 with constant α then the optimal α^* can be divided into three regimes depending on c. The boundaries of these regions are $\gamma_{max} := \max_p \gamma(p)$ and some $c_{th} < \gamma_{max}$ such that:*

1. *When $c > \gamma_{max}$, $\alpha^* = 0$ is the maximizer of 21.*
2. *When $c_{th} < c < \gamma_{max}$ then α^* is the minimizing $\alpha \in [0,1]$ such that $p^*(\alpha) \in \gamma^{-1}(c/\alpha)$.*
3. *When $c < c_{th}$: α^* is the smallest $\alpha \in [0,1]$ such that $p(\alpha) = 0$, where*

$$c_{th} = \max\left\{ c \,\middle|\, \frac{1 - c/\gamma_{min}}{1-\alpha} - \frac{U(p^*(\alpha))}{U(0)} \geq 0 \ \ \forall \alpha \leq c/\gamma_{min} \right\}. \tag{24}$$

Theorem 3 can be interpreted as follows. If privacy sensitivity is above γ_{max} for the given task, it is not worth the effort of the platform to participate. On the other hand, if privacy sensitivity is less than c_{th}, the platform should set α to be as small as possible, while still ensuring that all users choose the low privacy setting. Finally, if privacy sensitivities lie somewhere in between, α^* should be chosen based on the γ function, and generally will lead to a mixed strategy with some proportion of users choosing each of the two options.

Comparison to other works Two key novelties of our work is that we (1) consider a constraint of fairness and (2) have users choose a privacy level, rather than report their privacy sensitivity. This is different from Fallah et al. (2022), and Cummings et al. (2023), which rely on incentive compatibility, and have users report their privacy parameters. In Fallah et al. (2022), a computationally efficient algorithm is proposed for computing user payments and privacy levels to assign users. Both of these works consider a mean estimation problem, where users have i.i.d. samples, and so also have the "equal marginal contribution" assumption that we have. Distinct from our model, users have an additional term in their utility where they benefit from reduced error in the estimation problem. These works focus on maximizing the platform utility, and it is very clear that the payments deviate significantly from the fair ones that satisfy the fairness axioms. Hu & Gong (2020) is perhaps the work most relevant to ours. They consider an incentive design problem where the platform fixes the total sum of payments R and the amount each user receives is proportional to their privacy level ρ_i, which the users choose. This proportional scheme, while potentially viewed as a type of fairness, does not satisfy our axioms. For a particular utility function, they develop a computationally efficient algorithm to compute the equilibrium privacy levels ρ_i based on the privacy sensitivities of the users and the total sum of payments R. In all of these works, users have a linear privacy sensitivity function with rate c_i. Though this seems different from our binary privacy problem, there is a direct correspondence here since we allow mixed strategies, so in expectation, our sensitivity is also reduced to a linear function of the mixed strategy: i.e., $\mathbb{E}\left[c_i \mathbb{1}\left\{\rho_i = \rho_2'\right\}\right] = c_i \Pr(\rho_i = \rho_2')$.

5.1 Mechanism Design: Mean Estimation Example

Let's look at the utility function from equation 11, and fair payments that we calculated from Theorem 2 in equation 10. In this case there are $N = 2$ users, and we will assume each user has a sensitivity function as in equation 20. This setting satisfies the conditions of Theorem 3, and so we can use our above result to characterize the optimal α^* to equation 21 for a range of different c values. That is, as the privacy sensitivity parameter c changes, how is the optimal strategy of the platform impacted? Fig. 8 depicts the solution to equation 21. The

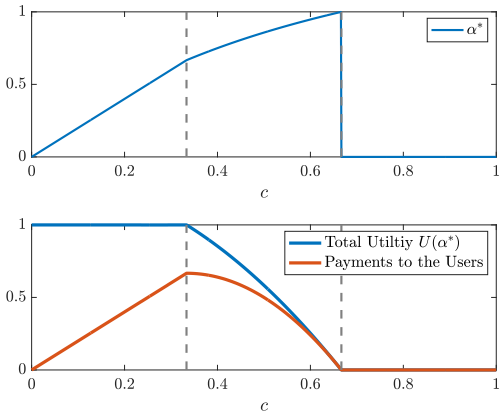


Figure 8: Utility of users and platform when platform solves equation 21. The solution has three separate regions as predicted by Theorem 3.

top plot shows the optimal α^* vs. c, while the bottom plot shows the total utility and the fraction of utility paid to each user at the optimal point α^* for a range of $c \in [0, 1]$.

As predicted by Theorem 3, we find that the solution is clearly divided into three regions. Equation 24 tells us that $c_{th} = \frac{1}{3}$ and $\gamma_{max} = \frac{2}{3}$, matching our observations in Fig. 8. In the first region when $c \leq \frac{1}{3}$ the privacy sensitivity of the users is low, and the platform is able to capture most of the utility for itself, paying less of it out to the users. In this region, $\alpha^* = 2c$, growing linearly with the privacy sensitivity. We also see that throughout this regime, the total utility is maximized, as predicted by the theory. In the region where $c \in [\frac{1}{3}, \frac{2}{3}]$, optimal α^* is no longer growing linearly, and now grows as $\alpha^* = \frac{6c}{3c+2}$ no longer have enough incentive to always choose the less private option, the total utility also begins to decrease, meaning less utility is available for incentives. These factors lead to a decrease in total utility in this region. As α^* continues to increase towards 1. Once $\alpha^* = 1$ at $c = \frac{2}{3}$, the platform is getting no utility, so may as well choose $\alpha^* = 0$. Finally, for $c \geq \frac{2}{3}$, the platform no longer attempts to incentivize the users, and the total utility and payments fall to zero with $\alpha^* = 0$.

For this particular example, it is possible to analytically solve the NE constraint in equation 21, exact analytic expressions for the curves in Fig. 8 are given in Appendix B.4.1.

5.2 Considering Different Privacy Sensitivities

The computational burden in solving equation 21 is in characterizing the constraint, since the objective reduces to a one-dimensional optimization over $\alpha \in [0, 1]$. In the previous section, with the knowledge that the game is symmetric, we are able to easily characterize the equilibria as a function of α. If the c_i's are all different, for arbitrary utility functions, the problem essentially reduces to finding the equilibria in a general game. To make this tractable, we will need some assumptions. In Hu & Gong (2020), the specific choice of utility function and payments makes computation of the equilibrium tractable. If we have only two groups of users with different c_i that act together, and a finite privacy space, we can appeal to tools for enumerating equilibria in matrix games (Avis et al., 2010). In this case if the privacy space is also binary, then the equilibria have an analytical solution, which we provide in Appendix E. Like the symmetric case, there are 3 cases for each of the two users as well as corresponding thresholds that depend on c_1 and c_2 respectively, resulting in 9 total cases. For example, in the case where payment is below the threshold of both users, neither participate at the low-privacy level, when the payment is high enough both participate at the low privacy level, and for the remaining intermediate cases, either only one user chooses the low privacy option, or there is some asymmetric mixed strategy. Below, we numerically investigate this case:

This problem differs from equation 21 because the equilibrium is governed by asymmetric users. For example, if user 1 and user 2 have privacy sensitivity c_1 and c_2 respectively, we have

$$u_1(\mathbf{p}_1, \mathbf{p}_2) = \mathbf{p}_1^T \mathbf{\Phi}_1^{(2)} \mathbf{p}_2 - [0\ c_1]^T \mathbf{p}_1, \quad u_2(\mathbf{p}_1, \mathbf{p}_2) = \mathbf{p}_1^T \mathbf{\Phi}_2^{(2)} \mathbf{p}_2 - [0\ c_2]^T \mathbf{p}_2. \tag{25}$$

Consider a setting where there are only two users (these can be thought of as representing two *groups* of users) with utility function u_1 and u_2 listed above. Thus, when the platform is trying to optimize it's own utility, it must take into consideration that these two groups will play different strategies.

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \mathbf{p}_1^T \mathbf{U} \mathbf{p}_2 - (1-\alpha) \mathbf{p}_1^T \mathbf{U} \mathbf{p}_2 \\ \text{subject to} \quad & (\mathbf{p}_1, \mathbf{p}_2) \in \text{NE}(\alpha). \end{aligned} \tag{26}$$

Fig. 9 plots the results of simulating the solution of 26. It shows that there is one region when c_1 and c_2 are both small and close together ($< 1/3$), the platform chooses α to collect data from both users. If the difference is large, even in this region, the users may be asymmetrically engaged. When $c_1 > c_2 > 1/3$, the platform chooses α such that only user 2 chooses to participate, even if the difference is very small, and vice versa if $c_2 > c_1 > 1/3$, as before, when $c_1, c_2 > 2/3$ the sensitivity to too high and the platform can no longer offer enough payment to the users.

Broader Impact Statement One of the unique defining characteristics of data is that its generation process is inherently distributed, so no single entity exists to advocate for data sellers. In the past, platforms

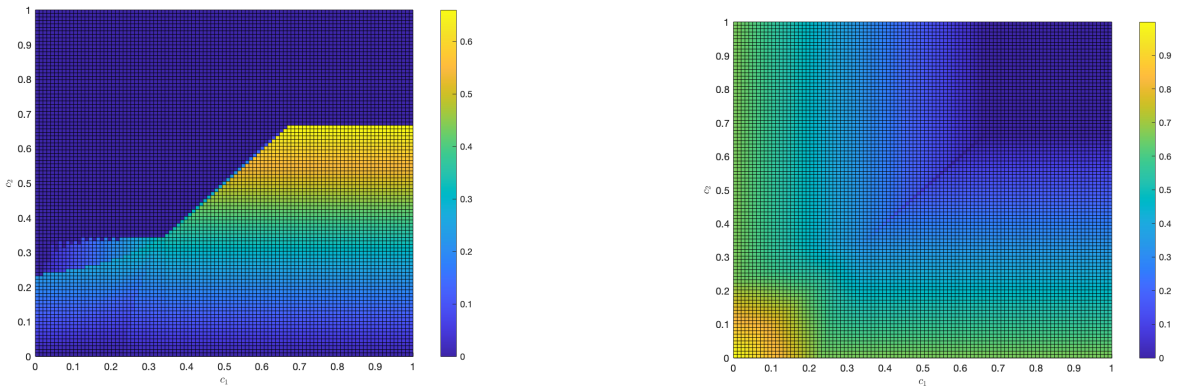


Figure 9: (Left) The payments to user 2 from the platform for a range of c_1, c_2. (Right) The platform's share of utility for the optimal α^* payments for a range of values c_1, c_2.

have been able to extract data from users, often with little to no compensation in return. As public consciousness around privacy changes, a nuanced relationship around privacy between platforms and users must develop. Transparency and understanding the value of user data is an important step in empowering regulators, consumers, and platforms.

- Users making strategic decisions about when they share their data stand to gain from incentives.
- For regulators, understanding the amount of value that flows through the interactions between platforms can enable better policies around data. Frameworks like those discussed in Theorem 1 and 2 can be a starting point in understanding exactly how much this value is.
- For platforms, understanding which data tasks are economically viable, and how they allocate incentive is important. Our discussion in Section 5, and our three regimes help shed light on this.

6 Conclusion

This paper introduces two formal definitions of fair payments in the context of acquisition of private data. The first treats the users and the platform together and uses axioms like those of the Shapley value to determine a unique fair distribution of utility. In the second, we define a notion of fairness between the users only, leading to a definition of fairness that admits a range of values, of which the platform is free to choose the most favorable. By formulating a federated mean estimation problem, we show that heterogeneous users can have significantly different contributions to the overall utility, and that a fair incentive, according to our second notion, must take into account the amount of data, privacy level as well as the degree of heterogeneity. We formulate and solve the fairness-constrained mechanism design problem in this federated mean estimation problem, and also find that data heterogeneity and user properties play an important role in the solution.

While previous literature has investigated how platforms should design incentives for users in order to optimize its utility, the definitions of fairness we propose offers another important way to evaluate the fairness of these mechanisms. This is a critical step towards future research in ensuring that data acquisition mechanisms are *both* fair for users and efficient for platforms.

Though we provide a characterization of optimal fair mechanisms when privacy sensitivity is the same across users, designing mechanisms and developing theories that scale up these solutions to deal with platform that interact with large and diverse groups of users is critical. Additionally, users may come and go, as their sensitivities may change over time. Understanding how fluctuating users alter the model is of great practical significance. Furthermore, there is subjectivity in the choice of axioms, and other choices may lead to meaningful notions of fairness worthy of study. We have also assumed a non-divisible and transferable utility, but in many cases, users are paid for their data in the form of access to services. Investigating the impact of this will also be important for the practical application of a comprehensive theory for fairness.