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

Justin S. Kang
Department of EECS
University of California, Berkeley
Berkeley, CA 94704
justin_kang@berkeley.edu

Ramtin Pedarsani
Department of Electrical and Computer Engineering
University of California, Santa Barbara
Santa Barbara, 93106
ramtin@ece.ucsb.edu

Kannan Ramchandran
Department of EECS
University of California, Berkeley
Berkeley, CA 94704
kannanr@berkeley.edu

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

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 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 level 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 user privacy level options:

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

International Workshop on Federated Learning in the Age of Foundation Models in Conjunction with NeurIPS 2023 (FL@FM-NeurIPS'23). This workshop does not have official proceedings and this paper is non-archival.

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, or data statistics.

iii) A non-private option, where the platform can collect any relevant data from a user device.

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 what the platform is paying?

A major shortcoming of the current understanding of data value is that in many cases, it fails to explicitly consider a critical factor in an individual's decision to share data—privacy. This work puts forth a rigorous notion of the fair value of data that explicitly includes privacy and makes use of the axiomatic framework of the *Shapley value* from game theory (Shapley, 1952). Furthermore, we also ask: *What happens when platforms are required to pay users fairly?* The two key sections of this work, and their main contributions are organized as follows:

Section 3 introduces an axiomatic notion of fairness, placing restrictions on the relative amounts distributed to the players. We then present a FL case study of the proposed notion studying how payments change as a function of privacy level, amount of data, and degree of heterogeneity.

Section 4 explores the platform incentive design problem when the platform is restricted to those fair payments. Theorem 2 establish that under certain conditions, there are three distinct regimes for the optimal incentives depending on the common privacy sensitivity of the users.

1.1 Relevant Literature

Privacy and Fairness Currently, popular forms of privacy include federated learning (Kairouz et al., 2021) and differential privacy (DP) (Dwork, 2008; Bun and 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. In Jia et al. (2019), Ghorbani and 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. We take this idea further and explicitly includes privacy in the definition of the fair value of data.

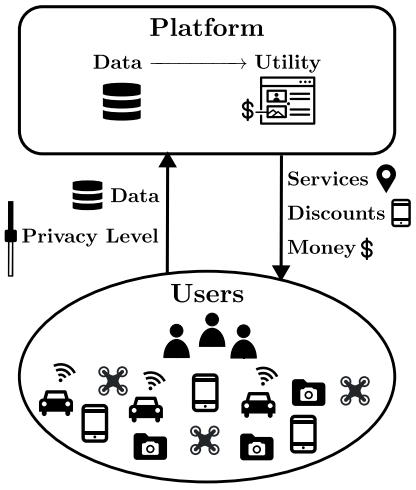


Figure 1: Users generate data with phones, cameras, vehicles, and drones. This data goes to the platform but requires some level of privacy. The platform uses this data to generate utility, often by using the data for learning tasks. In return, the platform may provide the users with payments in the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

Optimal Data Acquisition One line of literature studies *data acquisition*, where platforms attempt to collect data from privacy conscious users. Ghosh and Roth (2011) study 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. Hu and Gong (2020) goes beyond linear estimation to consider FL, where each user has a unique privacy sensitivity function and the platform pays them via a proportional scheme. Karimireddy et al. (2022) consider mechanisms that ensure users don't "free-load". Roth and Schoenebeck (2012); Chen et al. (2018); Chen and Zheng (2019); Cummings et al. (2023) also follow this data acquisition framework.

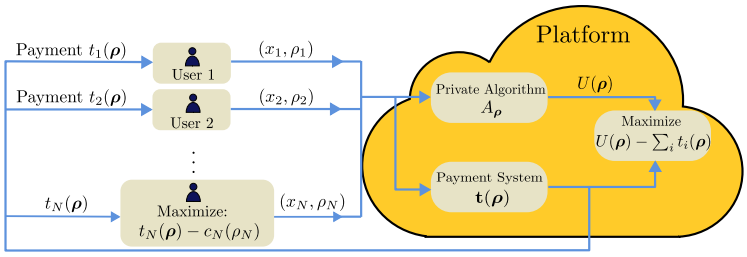


Figure 3: Users send data x_i and privacy level ρ_i to the central platform in exchange for payments $t_i(\rho_i; \boldsymbol{\rho}_{-i})$. The central platform extracts utility from the data at a given privacy level and optimizes incentives to maximize the difference between the utility and the sum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

2 PROBLEM SETTING

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.
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. The lack of rigour 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.

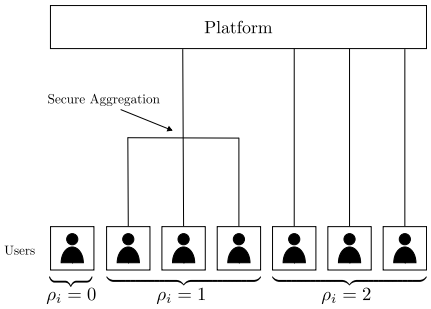


Figure 2: Users choose between three levels of privacy. $\rho_i = 0$:, users send no data, $\rho_i = 1$: model is securely combined with other users who also choose $\rho_i = 1$, $\rho_i = 2$: users send their relevant information directly to the platform.

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 obfuscated and only transmitted in part (perhaps via FL) and finally if $\rho_i = 2$, the users send a sufficient statistic for their data to the platform. The platform applies a $\boldsymbol{\rho}$-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.

The platform generates a transferable and divisible utility $U(\boldsymbol{\rho})$ from the user data and 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 motivates 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:

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

From the perspective of the platform, the goal is to design the payments $t_i(\rho_i; \boldsymbol{\rho}_{-i})$ such that it maximizes the difference between the utility it receives and the payments made to the players i.e., $U(\boldsymbol{\rho}) - \mathbb{1}^T \mathbf{t}(\boldsymbol{\rho})$. This is depicted in Fig 3. One way to formulate this problem is to consider

maximizing this difference at equilibrium points:

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

where we have used the shorthand $f(\mathcal{P}) = \mathbb{E}_{\boldsymbol{\rho}\sim\mathcal{P}}[f(\boldsymbol{\rho})]$. $\text{NE}(\mathbf{t})$ denotes the set of Nash Equilibrium strategies induced by $\mathbf{t}$, which is the vector with payment function t_i at index i. 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 4, we consider a **fairness constraint**. Model limitations are discussed in Appendix A.

3 Axiomatic Fairness with Privacy

What is a fair way to distribute utility back to the users as incentives? In this section, we view the users as a coalition, pooling their resources to generate utility (Appendix B, considers what happens if we include the platform in the coalition). 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. In our definition of fairness, we are interested in the intrinsic value of the data. That is, not the *market value* that users are willing to sell for (potentially depressed), but rather, how much of the utility generated comes from the data. 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. The answer to this fairness question turns out to be connected to the celebrated Shapley value (Shapley, 1952). Following an axiomatic approach to fairness, the Shapley value describes how to fairly divide utility among a coalition. In this section we develop an axiomatic Shapley value-based approach to fairness for users providing private data to platforms.

Axiomatic Fairness Due to asymmetry between the platform and the users, it makes sense to discuss fairness between users. We can consider the following axioms on the fair value of each user ϕ_i:

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

ii) **(Pseudo-Efficiency)** 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 $0 \leq \alpha(\boldsymbol{\rho}) \leq 1$.

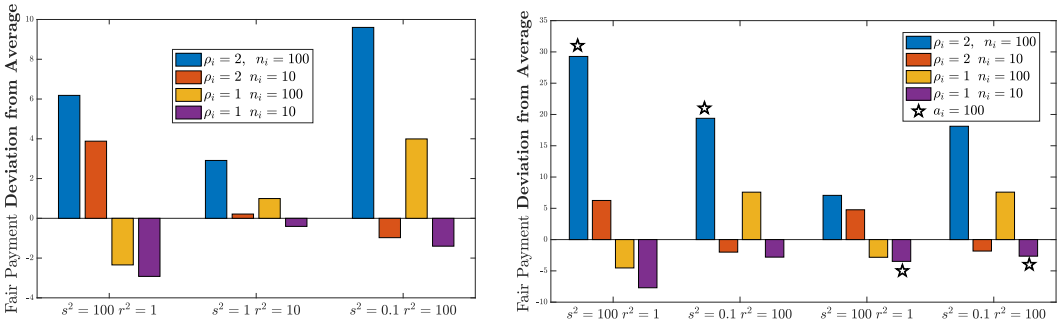
iii) **(Additivity)** Let $\phi_i(\boldsymbol{\rho})$ be the value of users for the utility function U, under the $\boldsymbol{\epsilon}$-private algorithm $A_{\boldsymbol{\rho}}$. Let V be a separate utility function, also based on the output of the algorithm $A_{\boldsymbol{\epsilon}}$, and let $\phi_i'(\boldsymbol{\rho})$ be the utility of the users with respect to V. Then under the utility $U + V$, the value of user i is $\phi_i(\boldsymbol{\rho}) + \phi_i'(\boldsymbol{\rho})$.

Theorem 1. *Let $\phi_i(\boldsymbol{\rho})$ satisfying axioms* 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* ii*, ϕ_i takes the form:*

$$\phi_i(\boldsymbol{\rho}) = \frac{\alpha(\boldsymbol{\rho})}{N} \sum_{S\subseteq[N]\backslash\{i\}} \frac{1}{\binom{N-1}{|S|}} \left(U(\boldsymbol{\rho}_{S\cup\{i\}}) - U(\boldsymbol{\rho}_S)\right). \tag{3}$$

It may seem that the computational complexity of equation 3 is $N\,|\mathcal{E}|^N$, but this is really only true for a worst-case exact computation. In practice, U typically has some kind of structure that makes the problem much more tractable. In Ghorbani and Zou (2019), Jia et al. (2019), Wang and Jia (2023) and Lundberg and Lee (2017) special structures are used to compute these types of sums with significantly lower complexities, particularly in cases where the U is related to the accuracy of a deep network.

Example: Fair Incentives in Federated Learning Recently, Donahue and Kleinberg (2021) consider a setting where heterogeneous users voluntarily opt-in to federation. We now use Theorem 1 to answer: *how much less should the platform pay a user that chooses to federate with others as compared to one that provides full access to their data?* Let each user $i \in [N]$ have a unique mean and variance $(\theta_i, \sigma_i^2) \sim \Theta$, where Θ is some global joint distribution. User i draws n_i samples i.i.d. from its local distribution $\mathcal{D}_i(\theta_i, \sigma_i^2)$. To motivate this example, 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,



(a) (b)

Figure 4: Plots of *difference from the average* utility per user $U(\boldsymbol{\rho})/N$ for each of the four different types of users. There are $N = 10$ 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, $N_0 = 1$ users chooses to not participate ($\rho_i = 0$). Furthermore, each has n_i data samples. Without loss of generality, we take $\alpha(\boldsymbol{\rho}) = 1$, and the results of this section can be scaled accordingly. In (a) all users have equal importance $a_i = 1$, in (b) here is one user i with $a_i = 100$ (indicated with a star), while all other users $j \neq i$ have $a_j = 1$.

by serving the best advertisements possible to each user. Fig. 2 summarizes our FL formulation with a 3-level privacy space $\mathcal{E} = \{0, 1, 2\}$. Let $s^2 = \mathrm{Var}(\theta_i)$ and $r^2 = \mathbb{E}[\sigma_i^2]$. When $\rho_i = 2$, user i provides its local estimator $\hat{\theta}_i$ directly to the platform. When $\rho_i = 1$, user i's local estimator is securely aggregated with all other users that choose this same privacy to produce $\hat{\theta}^f$, which the platform uses to construct its estimate $\hat{\theta}_i^p$. The goal of the platform is to construct estimators $\hat{\theta}_i^p$ that minimize the expected mean squared-error (and thus utility U, defined below) of each estimate, while respecting the privacy level vector $\boldsymbol{\rho}$:

$$\mathrm{EMSE}_i(\boldsymbol{\rho}) := \mathbb{E}\left[\left(\hat{\theta}_i^p(\boldsymbol{\rho}) - \theta_i\right)^2\right] \quad U(\boldsymbol{\rho}) := \sum_{i=1}^{n} a_i \log\left(\frac{(r^2 + 2s^2)}{\mathrm{EMSE}_i(\boldsymbol{\rho})}\right). \tag{4}$$

a_i represents the relative importance of each user. Since some users may spend more than others, the platform may care more about their θ_i more accurately, adding another layer of heterogeneity. also note we have defined $\mathrm{EMSE}_i(\mathbf{0}) := r^2 + 2s^2$.

Fig 4a 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 heterogeneous environment and a user's data will not be helpful for estimating the other θ_is. Users choosing $\rho_i = 2$ are paid the most, since at least their own data 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. On the right side, we have a regime where $s^2 = 0.1$ and $r^2 = 100$, which is more homogeneous. 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, which interpolates between the two extremes. In Fig 4b each set of graphs has exactly one user has $a_i = 100$, making it 100 times more important than the others. Looking at the two leftmost sets of bars in Fig 4b we see that when user i with $\rho_i = 2$ and $n_i = 100$ is the most important one user i receives most of the benefit in the heterogeneous case, but in the homogeneous case, other users also benefit. Another key point is the similarity between the second and fourth set of graphs. This tells an interesting story: when users are homogeneous, regardless of which user has $a_i = 100$, it is those users with large n_i that will benefit.

4 Fairness Constraints: Data Acquisition

We now use our concrete definition of fairness to constrain the platform to a class of fair payments. The platform can choose the fraction of utility that it keeps α, but the incentives it provides to users must be distributed in a fair way. Consider N 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 (i-iii) with the additional constraint that

$\alpha(\boldsymbol{\rho}) = \alpha \in [0, 1]$. Users choose one of two available privacy levels $\rho_i \in \mathcal{E}$, with $\mathcal{E} = \{\rho_1', \rho_2'\}$ and $\rho_2' > \rho_1'$. We can write the utility of the user i as

$$u(\rho_i, \boldsymbol{\rho}_{-i}) = \alpha\phi(\rho_i; \boldsymbol{\rho}_{-i}) - c\mathbb{1}\left\{\rho_i = \rho_2'\right\}. \tag{5}$$

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 $1 - p$. The platform is also trying to maximize the fraction of the total expected utility $U(\mathbf{p}) := \mathbb{E}_{\boldsymbol{\rho}\sim\mathbf{p}}\left[U(\boldsymbol{\rho})\right]$ that it keeps as in equation 2. The platform's goal is to choose a payment value α such that it optimizes:

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & (1-\alpha)U(\mathbf{p}^*(\alpha)) \\ \text{subject to} \quad & \mathbf{p}^*(\alpha) \in \text{NE}(\alpha). \end{aligned} \tag{6}$$

The objective is simplified compared to equation 2 by exploiting the pseudo-efficiency axiom, which says that the sum of payments is α times the total utility. The constraint in equation 6 implicitly encodes the user behavior governed by equation 5, and will change with the privacy sensitivity c.

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:

$$\gamma(p) := \phi(\rho_2'; \mathbf{p}) - \phi(\rho_1'; \mathbf{p}) = \mathbb{E}_{\substack{\rho_j\sim\mathbf{p} \\ j\neq i}}\left[\phi(\rho_2'; \boldsymbol{\rho}_{-i}) - \phi(\rho_1'; \boldsymbol{\rho}_{-i})\right]. \tag{7}$$

Theorem 2. *Consider a binary privacy level game with N users and a platform. If U satisfies Assumptions 1: monotonicity and 2: diminishing returns, and the platform payments are fair as defined in Theorem 1 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 6.*
2. *When $c_{th} < c < \gamma_{max}$ then α^* is the smallest $\alpha \in [0, 1]$ 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*

$$c_{th} = \max\left\{c \,\middle|\, \frac{1 - c/\gamma_{min}}{1-\alpha} - \frac{U(p^*(\alpha))}{U(0)} \geq 0 \;\; \forall \alpha < c/\gamma_{min}\right\}. \tag{8}$$

Details of the assumptions can be found in Appendix D.4 and E. Theorem 2 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. Appendix G provides some discussion non-homogeneous privacy sensitivity.

Mechanism Design Example Fig. 5 depicts the solution to equation 6, for a DP-Baysian estimation problem (details in Appendix C) meeting the conditions of Theorem 2. As predicted, the solution is clearly divided into three regions. Equation 8 tells us that $c_{th} = \frac{1}{3}$ and $\gamma_{max} = \frac{2}{3}$, matching our observations in Fig. 5. In the first region when $c \leq \frac{1}{3}$ the platform is able to capture most of the utility for itself, paying less of it out to the users. We also see that throughout this regime, the total utility is maximized, as predicted by the theory. For $c \in [\frac{1}{3}, \frac{2}{3}]$, the total utility begins to decrease, as users no longer have enough incentive to always choose the less private option. Finally, for $c \geq \frac{2}{3}$, the platform no longer attempts to incentivize the users, and the total utility falls to zero.

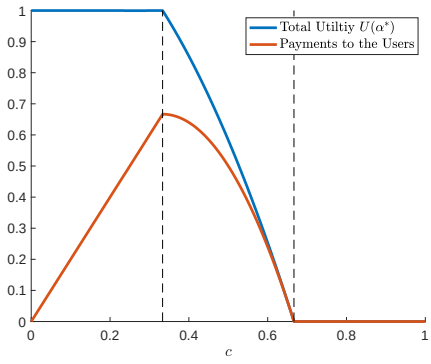


Figure 5: Utility of users and platform when platform solves equation 6. The solution has three separate regions as predicted by Theorem 2.

5 Conclusion

This paper introduces a formal definition of fair payments in the context of the acquisition of private data. 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 must take into account the amount of data, privacy level as well as the degree of heterogeneity. Theorem 2 provides us with interesting insights about how a platform would design payments under a fairness constraint.

This work also opens the door to new questions: How do we design mechanisms that consider fairness with heterogeneous privacy sensitivities with an arbitrary number of users? How do we efficiently compute fair values? Are their other meaningful notions of fairness worthy of study? How do we consider the fact that incentives are often non-monetary? Answering these questions will only become more important as data continues to play an increasing role in our economy.

Acknowledgements

Justin Kang and Kannan Ramchandran are supported by NSF EAGER: SaTC grant number 2232146. Ramtin Pedarsani is supported by NSF grants 2003035 and 2236483.

References

Avis, D., Rosenberg, G. D., Savani, R., and von Stengel, B. (2010). Enumeration of nash equilibria for two-player games. *Economic Theory*, 42(1):9–37.

Bun, M. and Steinke, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In Hirt, M. and Smith, A., editors, *Theory of Cryptography*, pages 635–658, Berlin, Heidelberg. Springer Berlin Heidelberg.

Chaudhuri, S. and Courtade, T. A. (2023). Mean estimation under heterogeneous privacy: Some privacy can be free.

Chen, Y., Immorlica, N., Lucier, B., Syrgkanis, V., and Ziani, J. (2018). Optimal data acquisition for statistical estimation. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, EC '18, page 27–44, New York, NY, USA. Association for Computing Machinery.

Chen, Y. and Zheng, S. (2019). Prior-free data acquisition for accurate statistical estimation. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 659–677.

Cheng, S.-F., Reeves, D. M., Vorobeychik, Y., and Wellman, M. P. (2004). Notes on equilibria in symmetric games. In *Proceedings of the 6th International Workshop On Game Theoretic And Decision Theoretic Agents GTDT*.

Cummings, R., Elzayn, H., Pountourakis, E., Gkatzelis, V., and Ziani, J. (2023). Optimal data acquisition with privacy-aware agents. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pages 210–224, Los Alamitos, CA, USA. IEEE Computer Society.

Donahue, K. and Kleinberg, J. (2021). Model-sharing games: Analyzing federated learning under voluntary participation. In *2021 AAAI Conference on Artifical Intellegence*.

Dwork, C. (2008). Differential privacy: A survey of results. In Agrawal, M., Du, D., Duan, Z., and Li, A., editors, *Theory and Applications of Models of Computation*, pages 1–19, Berlin, Heidelberg. Springer Berlin Heidelberg.

Fallah, A., Makhdoumi, A., Malekian, A., and Ozdaglar, A. (2022). Optimal and differentially private data acquisition: Central and local mechanisms.

Ghorbani, A., Kim, M., and Zou, J. (2020). A distributional framework for data valuation. In III, H. D. and Singh, A., editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3535–3544. PMLR.

Ghorbani, A. and Zou, J. (2019). Data shapley: Equitable valuation of data for machine learning. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2242–2251. PMLR.

Ghosh, A. and Roth, A. (2011). Selling privacy at auction. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, EC '11, page 199–208, New York, NY, USA. Association for Computing Machinery.

Guan, Z., Lv, Z., Du, X., Wu, L., and Guizani, M. (2019). Achieving data utility-privacy tradeoff in internet of medical things: A machine learning approach. *Future Generation Computer Systems*, 98:60–68.

Hart, S. and Mas-Colell, A. (1989). Potential, value, and consistency. *Econometrica*, 57(3):589–614.

Hu, R. and Gong, Y. (2020). Trading data for learning: Incentive mechanism for on-device federated learning. In *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pages 1–6.

Jia, R., Dao, D., Wang, B., Hubis, F. A., Hynes, N., Gürel, N. M., Li, B., Zhang, C., Song, D., and Spanos, C. J. (2019). Towards efficient data valuation based on the shapley value. In Chaudhuri, K. and Sugiyama, M., editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1167–1176. PMLR.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., Rouayheb, S. E., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konecný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Qi, H., Ramage, D., Raskar, R., Raykova, M., Song, D., Song, W., Stich, S. U., Sun, Z., Suresh, A. T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F. X., Yu, H., and Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2):1–210.

Karimireddy, S. P., Guo, W., and Jordan, M. I. (2022). Mechanisms that incentivize data sharing in federated learning.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 4768–4777, Red Hook, NY, USA. Curran Associates Inc.

Roth, A. and Schoenebeck, G. (2012). Conducting truthful surveys, cheaply. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, page 826–843, New York, NY, USA. Association for Computing Machinery.

Shapley, L. S. (1952). *A Value for N-Person Games*. RAND Corporation, Santa Monica, CA.

Wang, J. T. and Jia, R. (2023). Data banzhaf: A robust data valuation framework for machine learning. In Ruiz, F., Dy, J., and van de Meent, J.-W., editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 6388–6421. PMLR.

Xu, Z., Collins, M., Wang, Y., Panait, L., Oh, S., Augenstein, S., Liu, T., Schroff, F., and McMahan, H. B. (2023). Learning to generate image embeddings with user-level differential privacy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7969–7980.

A Model Limitations

Known sensitivity functions To solve equation 2, the platform requires the privacy sensitivity c_i of each user, and our solution in Section 4 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 4 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 2 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 1, 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. Privacy-utility trade-offs are also well studied in a range of different areas (Xu et al., 2023; Guan et al., 2019).

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.

B 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 $a \in \{0, 1\}$ represent the action of the platform. Let $a = 1$ when the platform chooses to join the coalition, and $a = 0$ 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:

$$U(a, \boldsymbol{\rho}) := \begin{cases} U(\boldsymbol{\rho}) & a = 1 \\ 0 & a = 0 \end{cases}, \quad [\boldsymbol{\rho}_S]_i := \begin{cases} \rho_i & i \in S \\ 0 & \text{else} \end{cases}. \tag{9}$$

Let $\phi_p(a, \boldsymbol{\rho})$ and $\phi_i(a, \boldsymbol{\rho})$, $i \in [N]$ represent the "fair" amount of utility awarded to the platform and each user i respectively, given a 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 and Mas-Colell (1989) show that these values are unique and well defined if they satisfy the following three axioms:

B.i) **(Fairness)** For $i, j \in [N] : U(a, \boldsymbol{\rho}_{S\cup\{i\}}) = U(a, \boldsymbol{\rho}_{S\cup\{j\}}) \ \forall S \subset [N]\backslash\{i, j\} \implies \phi_i(a, \boldsymbol{\rho}) = \phi_j(a, \boldsymbol{\rho})$.
In addition, for any user $i \in [N]$, $U(1, \boldsymbol{\rho}_{S\cup\{i\}}) - U(1, \boldsymbol{\rho}_S) = 0 \ \ \forall S \subset [N]\backslash\{i\} \implies \phi_i(a, \boldsymbol{\rho}) = 0$.

B.ii) **(Efficiency)** The sum of values is the total utility $U(a, \boldsymbol{\rho}) = \phi_p(a, \boldsymbol{\rho}) + \sum_i \phi_i(a, \boldsymbol{\rho})$.

B.iii) **(Additivity)** Let $\phi_p(a, \boldsymbol{\rho})$ and $\phi_i(a, \boldsymbol{\rho})$ 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 $\phi'_p(a, \boldsymbol{\rho})$ and $\phi'_i(a, \boldsymbol{\rho})$ be the utility of the platform and individuals with respect to V. Then under the utility $U + V$, the value of user i is $\phi_i(a, \boldsymbol{\rho}) + \phi'_i(a, \boldsymbol{\rho})$ and the value of the platform is $\phi_p(a, \boldsymbol{\rho}) + \phi'_p(a, \boldsymbol{\rho})$.

Theorem 3. *Let $\phi_p(a, \boldsymbol{\epsilon})$ and $\phi_i(a, \boldsymbol{\epsilon})$ satisfying axioms* (A.i-iii) *represent the portion of total utility awarded to the platform and each user i from utility $U(a, \boldsymbol{\epsilon})$. Then they are unique and take the form:*

$$\phi_p(a, \boldsymbol{\rho}) = \frac{1}{N+1} \sum_{S \subseteq [N]} \frac{1}{\binom{N}{|S|}} U(a, \boldsymbol{\rho}_S), \tag{10}$$

$$\phi_i(a, \boldsymbol{\rho}) = \frac{1}{N+1} \sum_{S \subseteq [N] \setminus \{i\}} \frac{1}{\binom{N}{|S|+1}} \left(U(a, \boldsymbol{\rho}_{S \cup \{i\}}) - U(a, \boldsymbol{\rho}_S) \right). \tag{11}$$

Theorem 3 is proved in Appendix D.2. We now consider a simple setting where we can apply this result.

C DP-Bayes Estimation Example

This section describes an example of the fair values applied to a particular estimation problem. Example 1 provides provides the fair values according to Theorem 3 while Example 2 considers Theorem 1.

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:

$$A(\mathbf{X}) = \mathbf{w}(\boldsymbol{\epsilon})^T \mathbf{X} + Z, \tag{12}$$

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 12. Then, we define:

$$A_{\boldsymbol{\epsilon}} = \underset{A \in \mathcal{A}_{\text{lin}}(\boldsymbol{\epsilon})}{\arg\min}\, \mathbb{E}[L(A(\mathbf{X}), \mu)] \qquad r(\boldsymbol{\epsilon}) = \mathbb{E}[L(A_{\boldsymbol{\epsilon}}(\mathbf{X}), \mu)]. \tag{13}$$

In words, $A_{\boldsymbol{\epsilon}}$ is an $\boldsymbol{\epsilon}$-DP estimator of the form equation 12, 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 (exact calculations are available in Appendix D.1). Consider the example of a binary privacy space $\mathcal{E} = \{0, \infty\}$. By equation 27, the utility can be written in matrix form as:

$$\mathbf{U} = \begin{bmatrix} 0 & 2/3 \\ 2/3 & 1 \end{bmatrix}. \tag{14}$$

Example 1. Note from equation 10 and equation 11, 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 10 and equation 11, we find that the fair allocations of the utility are given by:

$$\boldsymbol{\Phi}_p = \begin{bmatrix} 0 & 1/3 \\ 1/3 & 5/9 \end{bmatrix}, \; \boldsymbol{\Phi}_1^{(1)} = \begin{bmatrix} 0 & 1/3 \\ 0 & 2/9 \end{bmatrix}, \; \boldsymbol{\Phi}_2^{(1)} = \begin{bmatrix} 0 & 0 \\ 1/3 & 2/9 \end{bmatrix}. \tag{15}$$

Example 2. Consider the utility function defined in equation 14, for the $N = 2$ user mean estimation problem with $\mathcal{E} = \{0, \infty\}$. By Theorem 1 the fair allocation satisfying (i-iii) must be of the form:

$$\boldsymbol{\Phi}_1^{(2)} = \mathbf{A} \odot \begin{bmatrix} 0 & 2/3 \\ 0 & 1/2 \end{bmatrix}, \quad \boldsymbol{\Phi}_2^{(2)} = \mathbf{A} \odot \begin{bmatrix} 0 & 0 \\ 2/3 & 1/2 \end{bmatrix}, \; \mathbf{A} = \mathbf{A}^T, \; 0 \leq [\mathbf{A}]_{ij} \leq 1. \tag{16}$$

D Missing Proofs

D.1 Proof of Equation 27

In this section, we present the calculations required to arrive at the utility values in equation 27. First let's treat the trivial case of $\epsilon_1 = 0$, $\epsilon_2 = 0$. The optimal $\boldsymbol{\epsilon}$-DP estimator is simply the optimal Bayes estimator with no data, i.e., the prior mean. Let us define this estimator as $\hat{\mu}_{(0,0)} = 0$. Its risk function is

$$R(\mu, \hat{\mu}_{(0,0)}) = \mathbb{E}\left[L(\hat{\mu}_{(0,0)}, \mu) \mid \mu\right] = \mu^2. \tag{17}$$

The Bayes risk of $\hat{\mu}_{(0,0)}$ is the expectation of this quantity taken using our prior:

$$r([0,0]) = \mathbb{E}\left[\mu^2\right] = \frac{1}{12}. \tag{18}$$

Next, consider the case where user i chooses privacy level $\epsilon_1 = \epsilon' > 0$, and the other user chooses $\epsilon_2 = 0$. In this case the estimator depends on X_1, $\hat{\mu}_{(\epsilon',0)} = w_1 X_1 + Z$. Then the risk function is:

$$R(\mu, \hat{\mu}_{(\epsilon',0)}) = \mathbb{E}\left[(w_1 X_1 + Z - \mu)^2 \mid \mu\right] = \left(\mu + \frac{1}{2}\right)\left(\mu - \frac{w_1}{2}\right)^2 + \left(-\mu + \frac{1}{2}\right)\left(\mu + \frac{w_1}{2}\right)^2 + \frac{2}{\eta^2}. \tag{19}$$

Now taking the expectation with respect to our prior over μ, we have:

$$\mathbb{E}\left[R(\mu, \hat{\mu}_{(\epsilon',0)})\right] = \frac{1}{12}\left(3w_1^2 - 2w_1 + 1\right) + \frac{2}{\eta^2}, \tag{20}$$

here η is the inverse scale parameter for Z. Note that equation 20 is minimized when η is maximized. The $\boldsymbol{\epsilon}$-DP condition enforces the constraint $\eta \leq \frac{\epsilon'}{w_1}$. This constraint will be met with equality for the optimal w_1. The optimal $w_1^* = \frac{1}{3 + \frac{24}{\epsilon'^2}}$. Thus, we have:

$$\hat{\mu}_{(\epsilon',0)} = \frac{1}{3 + \frac{24}{\epsilon'^2}} X_1 + Z, \quad Z \sim \text{Laplace}\left(\frac{\epsilon'}{3\epsilon'^2 + 24}\right), \tag{21}$$

and the resulting Bayes risk is:

$$r([\epsilon', 0]) = r([0, \epsilon']) = \frac{1}{12}\left(1 - \frac{1}{3 + \frac{24}{\epsilon'^2}}\right). \tag{22}$$

For the case with $\epsilon_1 = \epsilon_2 = \epsilon'$ we can repeat the same process by defining $\hat{\mu}_{(\epsilon',\epsilon')} = w_1 X_1 + w_2 X_2 + Z$. By symmetry, we must have $w_1 = w_2$, so we drop the index. Then the risk function and its expectation are:

$$R(\mu, \hat{\mu}_{(\epsilon',\epsilon')}) = 2\left(\mu + \frac{1}{2}\right)\left(-\mu + \frac{1}{2}\right)\mu^2 + \left(\mu + \frac{1}{2}\right)^2 (w - \mu)^2 + \left(-\mu + \frac{1}{2}\right)^2 (\mu + w)^2 + \frac{2}{\eta}^2 \tag{23}$$

$$\mathbb{E}\left[R(\mu, \hat{\mu}_{(\epsilon',\epsilon')})\right] = \frac{1}{12}(8w^2 - 4w + 1) + \frac{2}{\eta^2}. \tag{24}$$

By a similar argument to the previous case, the Bayes optimal estimator and the corresponding Bayes risk is:

$$\hat{\mu}_{(\epsilon',\epsilon')} = \frac{1}{4 + \frac{12}{\epsilon'^2}}(X_1 + X_2) + Z, \quad Z \sim \text{Laplace}\left(\frac{\epsilon'}{4\epsilon'^2 + 12}\right), \tag{25}$$

$$r([\epsilon', \epsilon']) = \frac{1}{12}\left(1 - \frac{1}{2 + \frac{6}{\epsilon'^2}}\right). \tag{26}$$

Finally letting $U(\boldsymbol{\epsilon}) = c_1 r(\boldsymbol{\epsilon}) + c_2$. Take $U(\mathbf{0}) = 0 \implies c_1 = -12c_2$. And $\max_{\boldsymbol{\epsilon}} U(\boldsymbol{\epsilon}) = 1 \implies c_1 = 24(1 - c_2)$. Simplifying gives us our desired result:

$$\mathbf{U} = \begin{bmatrix} U([0,0]^T) & U([0,\epsilon']^T) \\ U([\epsilon',0]^T) & U([\epsilon',\epsilon']^T) \end{bmatrix} = \begin{bmatrix} 0 & 2\left(3 + \frac{24}{(\epsilon')^2}\right)^{-1} \\ 2\left(3 + \frac{24}{(\epsilon')^2}\right)^{-1} & \left(1 + \frac{3}{(\epsilon')^2}\right)^{-1} \end{bmatrix} \tag{27}$$

□

D.2 Proof of Theorem 3 and Theorem 1

We will begin with the proof of Theorem 1, which is standard and follows the typical proof of the Shapley value. We begin by proving $\phi_i(\boldsymbol{\rho})$ as defined in equation 3 satisfies axioms (B.i-iii). First assume $U(\boldsymbol{\rho}_{S\cup\{i\}}) = U(\boldsymbol{\rho}_{S\cup\{j\}}) \;\; \forall S \subset [N]\backslash\{i,j\}$, then:

$$\begin{aligned}
\phi_i(\boldsymbol{\rho}) &= \frac{\alpha(\boldsymbol{\rho})}{N} \sum_{S\subseteq[N]\backslash\{i\}} \frac{U(\boldsymbol{\rho}_{S\cup\{i\}}) - U(\boldsymbol{\rho}_S)}{\binom{N-1}{|S|}} && (28)\\
&= \frac{\alpha(\boldsymbol{\rho})}{N} \left(\sum_{S\subseteq[N]\backslash\{i,j\}} \frac{U(\boldsymbol{\rho}_{S\cup\{i\}}) - U(\boldsymbol{\rho}_S)}{\binom{N-1}{|S|}} + \sum_{S\subseteq[N]\backslash\{i,j\}} \frac{\big(U(\boldsymbol{\rho}_{S\cup\{j\}\cup\{i\}}) - U(\boldsymbol{\rho}_{S\cup\{j\}})\big)}{\binom{N-1}{|S|+1}} \right) && (29)\\
&= \frac{\alpha(\boldsymbol{\rho})}{N} \left(\sum_{S\subseteq[N]\backslash\{i,j\}} \frac{U(\boldsymbol{\rho}_{S\cup\{j\}}) - U(\boldsymbol{\rho}_S)}{\binom{N-1}{|S|}} + \sum_{S\subseteq[N]\backslash\{i,j\}} \frac{\big(U(\boldsymbol{\rho}_{S\cup\{i\}\cup\{j\}}) - U(\boldsymbol{\rho}_{S\cup\{i\}})\big)}{\binom{N-1}{|S|+1}} \right) && (30)\\
&= \phi_j(\boldsymbol{\rho}), && (31)
\end{aligned}$$

proving axiom (B.i) is satisfied. For the proof that axiom (B.ii) is satisfied, we write:

$$\begin{aligned}
\sum_i \phi_i(\boldsymbol{\rho}) &= \frac{\alpha(\boldsymbol{\rho})}{N} \sum_i \sum_{S\subseteq[N]\backslash\{i\}} \frac{U(\boldsymbol{\rho}_{S\cup\{i\}}) - U(\boldsymbol{\rho}_S)}{\binom{N-1}{|S|}} && (32)\\
&= \frac{\alpha(\boldsymbol{\rho})}{N} \left(\sum_i \sum_{S\subseteq[N]\backslash\{i\}} \frac{U(\boldsymbol{\rho}_{S\cup\{i\}})}{\binom{N-1}{|S|}} - \sum_i \sum_{S\subseteq[N]\backslash\{i\}} \frac{U(\boldsymbol{\rho}_S)}{\binom{N-1}{|S|}} \right) && (33)\\
&= \alpha(\boldsymbol{\rho})U(\boldsymbol{\rho}) + \frac{\alpha(\boldsymbol{\rho})}{N} \left(\sum_i \sum_{\substack{S\subseteq[N]\backslash\{i\}\\ |S|<N-1}} \frac{U(\boldsymbol{\rho}_{S\cup\{i\}})}{\binom{N-1}{|S|}} - \sum_i \sum_{S\subseteq[N]\backslash\{i\}} \frac{U(\boldsymbol{\rho}_S)}{\binom{N-1}{|S|}} \right) && (34)\\
&= \alpha(\boldsymbol{\rho})U(\boldsymbol{\rho}) + \frac{\alpha(\boldsymbol{\rho})}{N} \left(\sum_i \sum_{\substack{S\subseteq[N]\\ i\in S\\ |S|<N-1}} \frac{U(\boldsymbol{\rho}_S)}{\binom{N-1}{|S|-1}} - \sum_{\substack{S\subseteq[N]\\ |S|\leq N-1}} \frac{(N-|S|)U(\boldsymbol{\rho}_S)}{\binom{N-1}{|S|}} \right) && (35)\\
&= \alpha(\boldsymbol{\rho})U(\boldsymbol{\rho}) + \frac{\alpha(\boldsymbol{\rho})}{N} \left(\sum_{\substack{S\subseteq[N]\\ |S|\leq N-1}} \frac{|S|\,U(\boldsymbol{\rho}_S)}{\binom{N-1}{|S|-1}} - \sum_{\substack{S\subseteq[N]\\ |S|\leq N-1}} \frac{(N-|S|)U(\boldsymbol{\rho}_S)}{\binom{N-1}{|S|}} \right) && (36)\\
&= \alpha(\boldsymbol{\rho})U(\boldsymbol{\rho}), && (37)
\end{aligned}$$

thus proving axiom (B.ii) is satisfied. Finally, we note that (B.iii) is satisfied by linearity. Next, we establish the uniqueness of equation 3. To prove uniqueness, we take an approach that is standard in the literature where we define the unanimity game, show the uniqueness of the $\phi_i(\boldsymbol{\rho})$ in that case, and then argue that uniqueness follows from additivity (B.iii).

Define the unanimity utility, indexed by some $T \subseteq [N]$:

$$U_T(\boldsymbol{\rho}) = \begin{cases} 1 & \text{if } T \subseteq \text{supp}(\boldsymbol{\rho}) \\ 0 & \text{if } else. \end{cases} \tag{38}$$

$\{U_T\}_{T\subseteq[N]}$ form a linear basis for utility function such that any utility U can be represented uniquely by a set of values $\{b_T\}_{T\subseteq[N]}$. In addition, by direct application of the axioms, it is easy to see that for the unanimity utility, the fair allocation $\phi_i^{(T)}(\boldsymbol{\rho})$ is unique and is of the form:

$$\phi_i^{(T)}(\boldsymbol{\rho}) = \begin{cases} \frac{\alpha(\boldsymbol{\rho})}{T} & \text{if } i \in T \\ 0 & \text{if } else. \end{cases} \tag{39}$$

Thus, for any utility U, the fair value is represented uniquely by $\sum_{T\subseteq[N]} b_T \phi_i^{(T)}(\boldsymbol{\rho})$, since this value is unique, it must be equivalent to equation 3.

Now we consider the proof of Theorem 3. By a similar argument to the above, we can establish that:

$$\phi_p(a,\boldsymbol{\rho}) = \frac{1}{N+1}\sum_{S\subseteq[N]} \frac{U(a,\boldsymbol{\rho}_S)-U(0,\boldsymbol{\rho}_S)}{\binom{N}{|S|}} \tag{40}$$

as well as:

$$\phi_i(a,\boldsymbol{\rho}) = \frac{1}{N+1}\sum_{\substack{S\subseteq[N]\setminus\{i\}\\ a'\in\{0,a\}}} \frac{1}{\binom{N}{|S|+\mathbb{1}(a'=1)}}\left(U(a',\boldsymbol{\rho}_{S\cup\{i\}})-U(a',\boldsymbol{\rho}_S)\right) \tag{41}$$

$$\tag{42}$$

Applying the definition $U(0,\boldsymbol{\rho})=0$ we have

$$\phi_p(a,\boldsymbol{\rho}) = \frac{1}{N+1}\sum_{S\subseteq[N]} \frac{U(a,\boldsymbol{\rho}_S)}{\binom{N}{|S|}} \tag{43}$$

$$\phi_i(a,\boldsymbol{\rho}) = \frac{1}{N+1}\sum_{S\subseteq[N]\setminus\{i\}} \frac{1}{\binom{N}{|S|+1}}\left(U(a,\boldsymbol{\rho}_{S\cup\{i\}})-U(a,\boldsymbol{\rho}_S)\right), \tag{44}$$

completing the proof.

D.3 Error Computation for Section 3

In this section we prove Proposition 4 and 5 from which exact error expressions follow.

Proposition 4. *For the federated mean estimation problem described in Section 3, the expected mean-squared error is given by:*

$$\mathbb{E}\left[\left(\hat{\theta}_i^p-\theta_i\right)^2\right] =$$

$$r^2\left(\sum_{j=1}^{N_2} w_{ij}^2\cdot\frac{1}{n_j}+\frac{1}{N_1}w_{i0}^2\frac{1}{\bar{n}}\right)+s^2\left(\sum_{\substack{j=1\\ j\neq i}}^{N_2} w_{ij}^2+\frac{1}{N_1^2}\sum_{\substack{j=N_2+1\\ j\neq i}}^{N_2+N_1} w_{i0}^2+\left(\sum_{\substack{j=1\\ j\neq i}}^{N_2} w_{ij}+\frac{1}{N_1}\sum_{\substack{j=N_2+1\\ j\neq i}}^{N_2+N_1} w_{i0}\right)^2\right), \tag{45}$$

where $\bar{n}=\left(\frac{1}{N_1}\sum_{j=N_2+1}^{N_1+N_2}\frac{1}{n_j}\right)^{-1}$.

Proof. Consider an estimator of the form $\hat{\theta}_i^p=\sum_{i=1}^{N} v_{ij}\hat{\theta}_j$, where user j has n samples, and θ_j is the local model of user j. By Theorem 4.2 of Donahue and Kleinberg (2021), the error can be written as:

$$\mathbb{E}\left[\left(\hat{\theta}_i^p-\theta_i\right)^2\right] = r^2\sum_{j=1}^{N} v_{ij}^2\cdot\frac{1}{n_j}+s^2\left(\sum_{j\neq i} v_{ij}^2+\left(\sum_{j\neq i} v_{ij}\right)^2\right) \tag{46}$$

For $j=1,\ldots,N_2$, we have $v_{ij}=w_{ij}$. For $j=N_2+1,\ldots,N_2+N_1$, we have $v_{ij}=\frac{w_{i0}}{N_1}$. Finally, for $j>N_1+N_2$, we have $v_{ij}=0$. Thus the first term can be written as:

$$r^2\sum_{j=1}^{N} v_{ij}^2\cdot\frac{1}{n_j} = r^2\left(\sum_{j=1}^{N_2} w_{ij}^2\frac{1}{n_j}+\sum_{j=N_2+1}^{N_2+N_1}\frac{1}{n_j}\left(\frac{w_{i0}}{N_1}\right)^2\right) \tag{47}$$

$$= r^2\left(\sum_{j=1}^{N_2} w_{ij}^2\frac{1}{n_j}+\frac{1}{N_1}w_{i0}^2\frac{1}{\bar{n}}\right). \tag{48}$$

Making these same substitutions to $\sum_{j\neq i} v_{ij}^2$ and $\sum_{j\neq i} v_{ij}$ yields the desired result. □

Proposition 5. *The error expression equation 45 is minimized if $\rho_i = 0$ with weights:*

$$w_{i0} = \frac{N_1}{N_1 + N_2 \frac{V_0}{\bar{V}}}, \quad w_{ij} = \frac{V_0/V_j}{N_1 + N_2 \frac{V_0}{\bar{V}}}. \tag{49}$$

If $\rho_i = 1$ equation 45 is minimized by:

$$w_{i0} = \frac{N_1}{N_1 + N_2 \frac{V_0}{\bar{V}}} + \frac{N_2}{N_1 + N_2 \frac{V_0}{\bar{V}}} \frac{s^2}{\bar{V}}, \tag{50}$$

$$w_{ij} = \frac{V_0/V_j}{N_1 + N_2 \frac{V_0}{\bar{V}}} - \frac{1}{N_1 + N_2 \frac{V_0}{\bar{V}}} \frac{s^2}{V_j}. \tag{51}$$

Finally, if $\rho_i = 2$, equation 45 is minimized by:

$$w_{i0} = \frac{N_1}{N_1 + N_2 \frac{V_0}{\bar{V}}} - \frac{N_1}{N_1 + N_2 \frac{V_0}{\bar{V}}} \frac{s^2}{V_i}, \tag{52}$$

$$w_{ij} = \frac{V_0/V_j}{N_1 + N_2 \frac{V_0}{\bar{V}}} - \frac{V_0/V_j}{N_1 + N_2 \frac{V_0}{\bar{V}}} \frac{s^2}{V_i} \tag{53}$$

$$w_{ii} = \frac{V_0/V_i}{N_1 + N_2 \frac{V_0}{\bar{V}}} + \frac{N_1 + N_2 \frac{V_0}{\bar{V}} - \frac{V_0}{V_i}}{N_1 + N_2 \frac{V_0}{\bar{V}}} \frac{s^2}{V_i} \tag{54}$$

Proof. First we will consider the case where $\rho_i = 1$. Considering the point where the derivative of equation 45 with respect to $w_{ik}, k \geq 1$ is equal to zero gives:

$$\frac{2r^2}{n_k} w_{ik} - \frac{2r^2}{\bar{n} N_1} \left(1 - \sum_{j=1}^{N_2} w_{ij}\right) + s^2 \left(2w_{ik} - 2\frac{N_1 - 1}{N_1^2} \left(1 - \sum_{j=1}^{N_2} w_{ij}\right) + \frac{2}{N_1^2} \left(N_1 - 1 + \sum_{j=1}^{N_2} w_{ij}\right)\right) = 0, \tag{55}$$

$$\left(\frac{r^2}{n_k} + s^2\right) w_{ik} = \left(\frac{r^2}{\bar{n}} + s^2\right) \frac{w_{i0}}{N_1} - \frac{s^2}{N_1}. \tag{56}$$

It is easily verified from the second derivative that solving this equation gives us the unique minimum of equation 45. For ease of notation, define $V_k := \left(\frac{r^2}{n_k} + s^2\right)$ and $V_0 := \left(\frac{r^2}{\bar{n}} + s^2\right)$, $\bar{V} = \left(\frac{1}{N_2} \sum_{k=1}^{N_2} \frac{1}{V_k}\right)^{-1}$. Thus, we have:

$$w_{ik} = \frac{V_0 \frac{w_{i0}}{N_1} - \frac{s^2}{N_1}}{V_k}. \tag{57}$$

Noting that $w_{i0} + \sum_{j=1}^{N_2} w_{ij} = 1$, we have:

$$w_{i0} + \frac{N_2}{N_1} \frac{V_0}{\bar{V}} w_{i0} - \frac{N_2}{N_1} \frac{s^2}{\bar{V}} = 1, \tag{58}$$

$$w_{i0} = \frac{N_1}{N_1 + N_2 \frac{V_0}{\bar{V}}} + \frac{N_2}{N_1 + N_2 \frac{V_0}{\bar{V}}} \frac{s^2}{\bar{V}}, \tag{59}$$

$$w_{ij} = \frac{V_0/V_j}{N_1 + N_2 \frac{V_0}{\bar{V}}} - \frac{1}{N_1 + N_2 \frac{V_0}{\bar{V}}} \frac{s^2}{V_j}. \tag{60}$$

This completes the proof for those users i such that $\rho_i = 1$. When $\rho_i = 2$, the gradient condition with respect to $k \geq 1, k \neq i$ is:

$$w_{ik} V_k = \frac{V_0}{N_1} w_{i0}, \tag{61}$$

and similarly, the gradient condition when $k = i$ is:

$$w_{ii} V_i + w_{i0} \frac{N_2 V_0}{N_1 \bar{V}} + \frac{s^2}{V_i} = 1. \tag{62}$$

Combining these together gives our desired result. $\rho_i = 0$ □

D.4 Proof of Theorem 2

We begin by formally stating the 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:*

$$n_{private}(\boldsymbol{\rho}_S^{(1)}) \geq n_{private}(\boldsymbol{\rho}_S^{(2)}) \implies \Delta_i U(\boldsymbol{\rho}^{(1)}) > \Delta_i U(\boldsymbol{\rho}^{(2)}). \tag{63}$$

The symmetric Nash equilibria of our game is characterized Cheng et al. (2004) by the minimizers of

$$\min_p \sum_{s \in \mathcal{E}} [u(s,p) - u(p,p)]_+^2 , \tag{64}$$

where $u(s,p)$ is the utility a user when they choose privacy level $\rho_i = s$, and all other users play mixed strategy $\mathbf{p}$, and $u(p,p) = \mathbb{E}_{s\sim\mathbf{p}}[u(s,p)]$. Since our action space is binary, there are only two terms in this sum. Applying the definition of u and writing out both terms of this sum yields:

$$\begin{aligned}
\sum_{s \in \mathcal{E}} [u(s,p) - u(p,p)]_+^2 &= [u(\rho_1,p) - u(p,p)]_+^2 + [u(\rho_2,p) - u(p,p)]_+^2 && (65)\\
&= [c(1-p) - \alpha(\phi(p,p) - \phi(\rho_1,p))]_+^2 + [c(1-p) - \alpha(\phi(p,p) - \phi(\rho_2,p))]_+^2 && (66)\\
&= [(1-p)(c - \alpha\gamma(p))]_+^2 + [-p(c - \alpha\gamma(p))]_+^2 , && (67)
\end{aligned}$$

where we define $\gamma(p) := \phi(\rho_2,p) - \phi(\rho_1,p)$. γ is an important quantity in this problem that described the relative increase in payment a user receives for choosing a higher privacy level when the other users choose mixed strategy $\mathbf{p}$. In general, to say something about the equilibria, we must say something about γ. We can now use Assumptions 1 and 2, as well as the definition of $\phi(\cdot;\cdot)$ to establish properties of γ. First we show $\gamma(p) \geq 0$ using monotonicity of U:

$$\begin{aligned}
\gamma(p) &= \phi(\rho_2,p) - \phi(\rho_1,p), && (68)\\
&= \mathbb{E}_{\substack{\rho_j\sim\mathbf{p}\\ \rho_i=\rho_2'}} \left[\frac{1}{N} \sum_{S\subseteq[N]\setminus\{i\}} \frac{1}{\binom{N-1}{|S|}} \left(U(\boldsymbol{\rho}_{S\cup\{i\}}) - U(\boldsymbol{\rho}_S) \right) \right] \\
&\qquad - \mathbb{E}_{\substack{\rho_j\sim\mathbf{p}\\ \rho_i=\rho_1'}} \left[\frac{1}{N} \sum_{S\subseteq[N]\setminus\{i\}} \frac{1}{\binom{N-1}{|S|}} \left(U(\boldsymbol{\rho}_{S\cup\{i\}}) - U(\boldsymbol{\rho}_S) \right) \right], && (69)\\
&= \frac{1}{N} \sum_{S\subseteq[N]\setminus\{i\}} \frac{1}{\binom{N-1}{|S|}} \mathbb{E}_{\substack{\rho_j\sim\mathbf{p}\\ j\neq i}} \left[U(\boldsymbol{\rho}_{S\cup\{i\}}^{(i+)}) - U(\boldsymbol{\rho}_{S\cup\{i\}}^{(i-)}) \right] \geq 0. && (70)
\end{aligned}$$

In equation 69 we have used the definition of the fair value from Theorem 1, and in equation 70, we have simplified the expression, exchanged the sum and expectation, and used the fact that the expectation of a non-negative random variable is non-negative.

Next, we will show that under Assumption 2 (and our assumption of equal marginal contribution) we also have $\gamma'(p) \geq 0$. Assume $p_2 > p_1$, and let $b(n,p) = \binom{N}{n} p^i (1-p)^{N-i}$:

$$\gamma(p_2) - \gamma(p_1) = \frac{1}{N} \sum_{S\subseteq[N]\setminus\{i\}} \frac{1}{\binom{N-1}{|S|}} \left(\mathbb{E}_{\substack{\rho_j\sim\mathbf{p}_2\\ j\neq i}} \left[U(\boldsymbol{\rho}_{S\cup\{i\}}^{(i+)}) - U(\boldsymbol{\rho}_{S\cup\{i\}}^{(i-)}) \right] - \mathbb{E}_{\substack{\rho_j\sim\mathbf{p}_1\\ j\neq i}} \left[U(\boldsymbol{\rho}_{S\cup\{i\}}^{(i+)}) - U(\boldsymbol{\rho}_{S\cup\{i\}}^{(i-)}) \right] \right) \tag{71}$$

$$= \frac{1}{N} \sum_{S\subseteq[N]\setminus\{i\}} \frac{1}{\binom{N-1}{|S|}} \sum_{n=0}^{N} (b(n,p_2) - b(n,p_1)) \Delta_i U(\boldsymbol{\rho}(n)) \quad \text{s.t. } n_{private}(\boldsymbol{\rho}(n)) = N - n \tag{72}$$

Now note that $b(n, p_2) - b(n, p_1)$ is zero-mean, and decreasing, furthermore, $\Delta_i U(\boldsymbol{\rho}(n))$ is non-negative and non-increasing. Let n^* represent the smallest value of n such that $b(n, p_2) - b(n, p_1)$ is negative. Then we have:

$$\Delta_i U(\boldsymbol{\rho}(n)) = \sum_{n=0}^{n^*-1} \left(b(n, p_2) - b(n, p_1)\right) \Delta_i U(\boldsymbol{\rho}(n)) + \sum_{n=n^*}^{N} \left(b(n, p_2) - b(n, p_1)\right) \Delta_i U(\boldsymbol{\rho}(n)) \tag{73}$$

$$\geq \left(\sum_{n=0}^{n^*-1} b(n, p_2) - b(n, p_1) \right) \left(\Delta_i U(\boldsymbol{\rho}(n^* - 1)) - \Delta_i U(\boldsymbol{\rho}(n^*))\right) \tag{74}$$

$$\geq 0. \tag{75}$$

With the knowledge that $\gamma(p) \geq 0$ and $\gamma'(p) \geq 0$ we can compute p^* for three distinct cases. Defining $\gamma_{max} := \max_p \gamma(p)$ and $\gamma_{min} := \min_p \gamma(p)$, we have:

Case 1 $c - \alpha\gamma_{max} > 0$:

$$\sum_{s \in \mathcal{E}} [u(s, p) - u(p, p)]_+^2 = [(1 - p)(c - \alpha\gamma(p))]_+^2 \tag{76}$$

Since this quantity is non-negative, it is clearly minimized when $p^* = 1$, where it is exactly 0. Furthermore, since $c - \alpha\gamma_{max} > 0$ is satisfied with strict inequality, it is the unique minimizer.

Case 2 $c/\alpha \in [\gamma_{min}, \gamma_{max}]$:

$$\sum_{s \in \mathcal{E}} [u(s, p) - u(p, p)]_+^2 = [(1 - p)(c - \alpha\gamma(p))]_+^2 + [-p(c - \alpha\gamma(p))]_+^2 , \tag{77}$$

In the above case, this is minimized when $p^* \in \gamma^{-1}(c/\alpha)$.

Case 3 $c - \alpha\gamma_{min} < 0$:

$$\sum_{s \in \mathcal{E}} [u(s, p) - u(p, p)]_+^2 = [-p(c - \alpha\gamma(p))]_+^2 , \tag{78}$$

In the above case, the expression is minimized when $p^* = 0$. To summarize, we have:

$$p^*(\alpha) = \begin{cases} 1 & \text{if } \alpha < \frac{c}{\gamma_{max}} \\ \gamma^{-1}(c/\alpha) & \text{if } \alpha \in [\frac{c}{\gamma_{max}}, \frac{c}{\gamma_{min}}] \\ 0 & \text{if } \alpha > \frac{c}{\gamma_{min}} \end{cases} . \tag{79}$$

This establishes that the Nash equilibrium is cleanly separated into three regions. From this fact, we are able to show that the optimal strategy of the platform is also separated into three regions. We consider a platform that solves the following problem, where we define $U(p) := \mathbb{E}_{\rho_i \sim \mathbf{p}} [U(\boldsymbol{\rho})]$:

$$\min_{\alpha} (1 - \alpha) U(p^*(\alpha)), \tag{80}$$

Clearly, when privacy sensitivity is large, specifically, when $c \geq \gamma_{max}$ then $\alpha^* = 0$ is the optimal solution, since $p^*(\alpha) = 1$ for all $\alpha < 1$, and for $\alpha > 1$ the objective becomes negative.

Alternatively, when c is very small, we can determine the optimal value as follows. We first note that Assumption 1 implies that $U(p)$ is a decreasing function of p. Thus the condition for $\alpha^* = \frac{c}{\gamma_{min}}$ is:

$$\frac{1 - c/\gamma_{min}}{1 - \alpha} > \frac{U(p^*(\alpha))}{U(0)} \quad \forall \alpha < c/\gamma_{min}. \tag{81}$$

Since the left-hand side takes value $\frac{1}{1-\alpha}$ at $c = 0$, while the right-hand side is 1, as well as the fact that both sides are continuous, by the Intermediate Value Theorem, (and our previous, which implies that for c large enough this condition does not hold), there is some minimum c_{th}, where this condition fails. Thus we conclude, there are three regions:

(1) a region where $c \leq c_{th}$ is small, and α^* is the smallest α such that $p^* = 0$, (2) an intermediate region where a symmetric mixed strategy is played, and (3) a region where $c \geq \gamma_{max}$, and $\alpha^* = 0, p^* = 1$.

E Monotonicity of Utility

When beginning this work, the dearth of algorithms that supported heterogeneous privacy constraints surprised us, given the increasing number of privacy options available to users. All of the algorithms that did exist were provably sub-optimal Hu and Gong (2020), or placed constraints on privacy parameters to prove approximate optimality Fallah et al. (2022). In both of these works, the pathology of the algorithm leads to error that is not monotonically decreasing in $\boldsymbol{\rho}$. For DP-based notions of privacy, which both of the aforementioned works are, one can prove that an optimal error must be monotonic. This observation inspired a recent work that studies a *saturation* phenomenon Chaudhuri and Courtade (2023). Similar ideas can also be found in Cummings et al. (2023). The idea is that an optimal algorithm will sometimes give users that choose a large ϵ_i more privacy than they asked for, to ensure that it still efficiently uses information from users j with $\epsilon_j \ll \epsilon_i$.

F 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 contibution" 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 and 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., $\mathbb{E}\left[c_i \mathbb{1}\left\{\rho_i = \rho_2'\right\}\right] = c_i \mathrm{Pr}(\rho_i = \rho_2')$.

G Different Privacy Sensitivities

The computational burden in solving equation 6 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 and 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 equlibira 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 H. Similar to 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 6 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

$$u_1(\mathbf{p}_1, \mathbf{p}_2) = \mathbf{p}_1^T \boldsymbol{\Phi}_1^{(2)} \mathbf{p}_2 - [0\ c_1]^T \mathbf{p}_1, \quad u_2(\mathbf{p}_1, \mathbf{p}_2) = \mathbf{p}_1^T \boldsymbol{\Phi}_2^{(2)} \mathbf{p}_2 - [0\ c_2]^T \mathbf{p}_2. \tag{82}$$

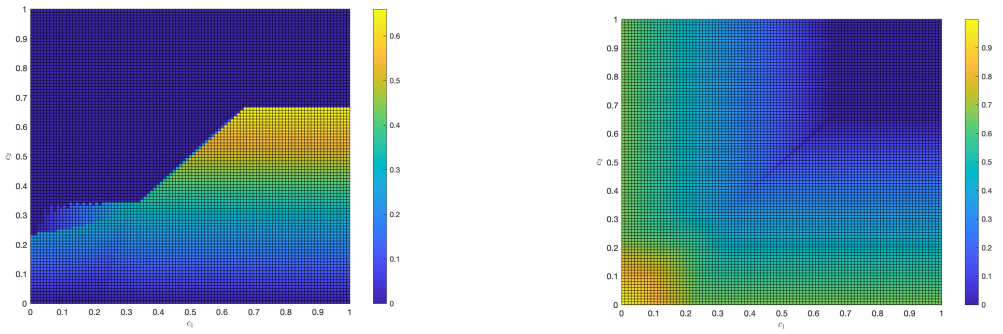


Figure 6: (Left) The payments to user 2 from the platform for a range of c_1, c_2. (Right) The platform's share of utility for the optimal α^* payments for a range of values c_1, c_2.

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.

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \mathbf{p}_1^T \mathbf{U} \mathbf{p}_2 - (1-\alpha)\mathbf{p}_1^T \mathbf{U} \mathbf{p}_2 \\ \text{subject to} \quad & (\mathbf{p}_1, \mathbf{p}_2) \in \text{NE}(\alpha). \end{aligned} \tag{83}$$

Fig. 6 plots the results of simulating the solution of 83. 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.

H Equilibria for Binary Privacy Level with Two Different Privacy Sensitivities

Let $\mathbf{p} = [p\ (1-p)]^T$ be the mixed strategy of user 1 and let $\mathbf{q} = [q\ (1-q)]^T$ be the mixed strategy of user 2. When they play these respective strategies, the utility of user 1 is:

$$\begin{aligned} u_1(\mathbf{p}, \mathbf{q}) &= pq\alpha\phi(\rho_1', \rho_2') + p(1-q)\alpha\phi(\rho_1', \rho_2') + (1-p)q\alpha\phi(\rho_2', \rho_1') + (1-p)(1-q)\alpha\phi(\rho_2', \rho_2') + -c_1(1-p) \\ &= p\alpha\phi(\rho_1', \mathbf{q}) + (1-p)\alpha\phi(\rho_2', \mathbf{q}) - c_1(1-p) \\ &= p\left(c_1 - \alpha\gamma(q)\right) + \alpha\phi(\rho_2', \mathbf{q}) - c_1. \end{aligned}$$

By a symmetric argument, we also have that

$$u_2(\mathbf{p}, \mathbf{q}) = q\left(c_2 - \alpha\gamma(p)\right) + \alpha\phi(\rho_2', \mathbf{p}) - c_2. \tag{84}$$

We are interested in characterizing the best response maps:

$$\text{BR}_1(\mathbf{q}; \alpha) = \arg\max_{\mathbf{p}} u_1(\mathbf{p}, \mathbf{q}) \quad \text{BR}_2(\mathbf{p}; \alpha) = \arg\max_{\mathbf{q}} u_2(\mathbf{p}, \mathbf{q}), \tag{85}$$

since their intersection characterize the set of NEs.

We begin with finding an analytic expression for $\text{BR}_1(\mathbf{q}; \alpha)$, which, we will break into three distinct cases:

Case 1: $c_1 - \alpha\gamma_{max} > 0$

In this case, the constant factor in front of p is always positive (invoking the monotonicity and non-negativity we proved in the previous section under the assumptions), thus the best response is:

$$\text{BR}_1(\mathbf{q}; \alpha) = [1\ 0]^T\ \forall \alpha < \frac{c_1}{\gamma_{max}}. \tag{86}$$

Case 2: $c_1 - \alpha\gamma_{min} < 0$

In this case, by a similar argument to before, the constant factor in front of p is always negative, thus the best response is:

$$\mathrm{BR}_1(\mathbf{q};\alpha) = [0\ 1]^T\ \forall \alpha > \frac{c_1}{\gamma_{min}}. \tag{87}$$

Case 3: $\alpha \in \left[\frac{c_1}{\gamma_{max}}, \frac{c_1}{\gamma_{min}}\right]$

In this case, the sign of the factor in front of p changes with $\mathbf{q}$. We can write the best response piecewise as:

$$\mathrm{BR}_1(\mathbf{q};\alpha) = \begin{cases} [1\ 0]^T & \text{if } c_1 - \alpha\gamma(q) > 0 \\ \{[a\ b]^T : a, b \geq 0,\ a+b=1\} & \text{if } c_1 - \alpha\gamma(q) = 0 \\ [0\ 1]^T & \text{if } c_1 - \alpha\gamma(q) < 0 \end{cases} \tag{88}$$

This same analysis can be applied to $\mathrm{BR}_2(\mathbf{p};\alpha)$. The NE is characterized by the sets where these two maps intersect. The following table summarize the equilibria p^*, q^*, written as scalars for readability.

	$\alpha \leq \frac{c_1}{\gamma_{max}}$	$\alpha \in \left[\frac{c_1}{\gamma_{max}}, \frac{c_1}{\gamma_{min}}\right]$	$\alpha > \frac{c_1}{\gamma_{min}}$
$\alpha \leq \frac{c_2}{\gamma_{max}}$	$(1,1)$	$(0,1)$	$(0,1)$
$\alpha \in \left[\frac{c_2}{\gamma_{max}}, \frac{c_2}{\gamma_{min}}\right]$	$(1,0)$	$\left\{(1,0),(0,1),\left(\gamma^{-1}\left(\frac{c_1}{\alpha}\right), \gamma\left(\frac{c_2}{\alpha}\right)\right)\right\}$	(0,1)
$\alpha > \frac{c_2}{\gamma_{min}}$	(1,0)	(1,0)	(0,0)

When α is below the threshold for the two users (the top left entry), both c_1 and c_2 are too small for it to be worthwhile for the users to participate at the lower privacy option. Conversely, if α is above the threshold for both users, then both users choose the less private option. When neither of these extremes occur the results are more nuanced.

I 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 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 similar to those discussed in Theorem 3 and 1 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 4, and our three regimes help shed light on this.