Algorithmic Interactions With Strategic Users: Incentives, Interplay, and Impact

by

Alireza Fallah

B.S., Sharif University of Technology (2017)
S.M., Massachusetts Institute of Technology (2019)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2023

©2023 Alireza Fallah. All rights reserved.
The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Authored by: Alireza Fallah
Department of Electrical Engineering and Computer Science
July 31, 2023

Certified by: Asuman Ozdaglar
MathWorks Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by: Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Algorithmic Interactions With Strategic Users: Incentives, Interplay, and Impact

by

Alireza Fallah

Submitted to the Department of Electrical Engineering and Computer Science
on July 31, 2023, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The societal challenges posed by machine learning algorithms are becoming increasingly important, and to effectively study them, it is crucial to incorporate the incentives and preferences of users into the design of algorithms. In many cases, algorithms are solely designed based on the platform's objectives, without taking into account the potential misalignment between the platform's goals and the interests of users.

This thesis presents frameworks for studying the interactions between a platform and strategic users. The central objective of the platform is to estimate a parameter of interest by collecting users' data. However, users, recognizing the value of their data, demand privacy guarantees or compensations in exchange for sharing their information. The thesis delves into various aspects of this problem, including the estimation task itself, the allocation of privacy guarantees, and the potential vulnerabilities of these guarantees to the platform's power.

In particular, in the first part of this thesis, we formulate this question as a Bayesian-optimal mechanism design problem, in which an individual can share her data in exchange for a monetary reward but at the same time has a private heterogeneous privacy cost which we quantify using differential privacy. We consider two popular data market architectures: central and local. In both settings, we establish minimax lower bounds for the estimation error and derive (near) optimal estimators for given heterogeneous privacy loss levels for users. Next, we pose the mechanism design problem as the optimal selection of an estimator and payments that elicit truthful reporting of users' privacy sensitivities. We further develop efficient algorithmic mechanisms to solve this problem in both privacy settings. Moreover, we investigate the case that users have heterogeneous sensitivities for two types of privacy losses corresponding to local and central privacy measures.

In the second part, we study a different aspect of the data market design: the optimal choice of architecture from both users' and the platform's point of view. The platform collects data from users by means of a mechanism that could partially protect users' privacy. We prove that a simple shuffling mechanism, whereby individual data is fully anonymized with some probability, is optimal from the viewpoint of users. We also develop a game-theoretic model of data sharing to study the impact of this shuffling mechanism on the platform's behavior and users' utility. In particular, we uncover an intriguing phenomenon that highlights the fragility of provided privacy guarantees: as the value of pooled data rises for users, the platform can exploit this opportunity to decrease the provided privacy guarantee, ultimately leading to reduced user welfare at equilibrium.

Thesis Supervisor: Asuman Ozdaglar
Title: Professor of Electrical Engineering and Computer Science

> "It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment. When I have clarified and exhausted a subject, then I turn away from it, in order to go into darkness again; the never-satisfied man is so strange if he has completed a structure, then it is not in order to dwell in it peacefully, but in order to begin another."
>
> *Karl Friedrich Gauss*
>
> *Letter to Bolyai, 1808*

Acknowledgments

I am deeply grateful to all those who have supported and guided me throughout my journey in completing my Ph.D. First and foremost, I would like to express my heartfelt gratitude to my advisor, Asu Ozdaglar, for her unwavering support and constant enthusiasm about research. Her mentorship and inspiration have been invaluable to me throughout my years of pursuing my Ph.D. I am truly fortunate to have had such an amazing advisor who always encouraged me to explore new topics and problems.

I would also like to extend my thanks to the members of my thesis committee, Daron Acemoglu and Costis Daskalakis. I am enormously grateful to Daron for his exceptional mentorship, teaching me how to delve into the details of mathematical theorems while keeping the broader picture in mind. And I deeply appreciate Costis for his invaluable feedback and comments, which have helped me improve my work.

Throughout my Ph.D. journey, I have been incredibly fortunate to work with outstanding collaborators and mentors. My sincere appreciation goes to Ali Makhdoumi, Azarakhsh Malekian, and Aryan Mokhtari for their remarkable guidance and support, providing me with valuable insights and feedback during our collaborations. I am also deeply grateful to my other collaborators, including Julien Fageot, Serhat Aybat, Mert Gürbüzbalaban, Sarath Pattathil, Thibaut Horel, Theo Diamandis, Yonina Eldar, Farzan Farnia, Andrea Giometto, Daniel Huttenlocher, Francesca Parise, Umut Simsekli, and Lingjiong Zhu, for the opportunities to work together. Additionally, I would like to express my sincere thanks to the members of the Apple Machine Learning Privacy team, specifically Kunal Talwar, John Duchi, Omid Javidbakht, and Hilal Asi, with whom I had the privilege of working. Furthermore, I want to acknowledge Kristian Georgiev and Evan Vogelbaum, whom I had the honor of mentoring during their time as undergraduate researchers in our group.

I am forever indebted to the vibrant and supportive MIT community, which has made my Ph.D. experience joyful and fulfilling over the past six years. In particular, I would like to thank other former and current members of our group, including Kaiqing Zhang, Amirhossein Reisizadeh, Nuri Denizcan Vanli, Emily Meigs, James Siderius, Hannah Li, Chanwoo Park, Omar Bennouna, Charles Lyu, Kihyun Kim, and Jiawei Zhang. A special mention goes to Sarath, my officemate for the past five years, with whom I enjoyed countless discussions on various topics, from research to everyday life. I am also grateful to my friends at LIDS and CSAIL, including Sohil Shah, Arjun Balasingam, Vishrant Tripathi, Abin Shah, Xinzhe Fu, Julia Gaudio, Ezra Tal, Charlie Yun, and many others who have enriched my academic and social life at MIT. I am also thankful for the Sidney-Pacific (SidPac) graduate community, which embraced me and warmly welcomed me upon my arrival at MIT and the US. In particular, I would like to thank my co-chairs at SidPac, Neil Gaikwad, Anupam Jena, Gurrein Madan, and Yamin Arefeen.

My appreciation extends to the dedicated staff at LIDS, CSAIL, and the broader EECS department, including Roxana Hernandez, Rachel Wright, Brian Jones, Rachel Cohen, Rich Lay, Francisco Jaimes, Shyla Putrevu, Deborah Goodwin, Janet Fischer, Kathleen McCoy, and Lynne Dell, for their assistance with all administrative tasks, which have been essential in ensuring a smooth academic journey.

I would like to acknowledge the unwavering support of my Iranian friends at MIT and beyond. In particular, I am deeply thankful to Ali Fahimniya, Mehrdad Khani, Zahra Hejrati, Amir Tohidi, Farnaz Jahanbakhsh, Mina Dalirrooyfard, Ali Vakilian, Arman Rezaee, Tina Torkaman, Sajjad Mohammadi, Farzam Ebrahimnejad, and Mehran Bahmani for their friendship and support.

Lastly, I express my heartfelt gratitude to my family. To my parents, Hamid and Elahe, without whom I would not have accomplished anything, I owe the greatest debt. Their love, encouragement, and belief in me have been the driving force behind my accomplishments. And to my wife, Maryam, who has been a constant pillar of support and inspiration, I am forever grateful. Her unwavering belief in me and her presence through the ups and downs of life have been invaluable.

Once again, I am deeply grateful for each and every person who has been a part of this remarkable experience, including those whom I might have inadvertently missed acknowledging here. Together, you have made my Ph.D. journey a truly joyful and amazing one, and I cannot thank you all enough for that.

Grants and funding: My Ph.D. research was generously supported by the Apple Scholars in AI/ML Ph.D. fellowship, the MathWorks Engineering fellowship, and the Siebel Scholarship. In addition, my research was also supported by the following grants: NSF grant CNS-1239054, Connection Science Center funding, Army Research Office grant W911NF-18-1-0407, and Air Force Research Laboratory grant FA8750-19-2-1000.

Contents

1 Introduction 15

1.1 Problems and Contributions 15

1.2 Related Literature 20

2 Optimal Deferentially Private Data Acquisition 25

2.1 Introduction 25

2.2 Differential Privacy and Platform's Estimation Problem 28

2.2.1 Central and Local Differential Privacy 29

2.2.2 (Near) Optimal Estimation With Central Differential Privacy 32

2.2.3 Optimal Estimation With Local Differential Privacy 36

2.3 Data Acquisition Mechanism With Privacy Guarantees 37

2.3.1 Payment Identity 40

2.3.2 Reformulating the Platform's Problem 42

2.4 Privacy-Concerned Data Acquisition in the Central Privacy Setting 42

2.4.1 Characterization of the Optimal Central Privacy Data Acquisition Mechanism 43

2.4.2 Computing the Optimal Privacy Loss Function 45

2.5 Privacy-Concerned Data Acquisition in the Local Privacy Setting 47

2.5.1 Characterization of the Optimal Local Privacy Data Acquisition Mechanism 48

2.5.2 Computing the Optimal Privacy Loss Function 50

2.6 Data Acquisition With Central Versus Local Differential Privacy 52

2.6.1 An Illustrative Example 53

2.6.2 Optimal Privacy Loss Levels in the Central Versus Local Setting . . . 56
2.7 Conclusion and Discussions 57
2.8 Proofs and Additional Results 60
2.8.1 Additional Results and Details 91

3 Bridging Central and Local Differential Privacy Mechanisms 97
3.1 Introduction 97
3.2 Privacy Measure and the Minimax Optimal Estimator 100
3.2.1 Local and Central Privacy Losses 100
3.2.2 Minimax Optimal Estimator 102
3.3 Data Acquisition Mechanism With Two-Part Privacy Guarantees 105
3.3.1 Utility of the Users and the Platform 105
3.3.2 Platform's Problem 108
3.4 From the Mechanism Design Problem to an Optimization Problem 109
3.5 Finding the (Approximately) Optimal Two-Part Data Acquisition Mechanism 110
3.5.1 Optimality of Having Only Local Noises in the Gaussian Mechanism . 111
3.5.2 Computing the Optimal Privacy Loss Function 112
3.5.3 A Numerical Example 114
3.6 Conclusion 117
3.7 Proofs 117

4 How Good Are Privacy Guarantees? Platform Architecture and Violation of User Privacy 135
4.1 Introduction 135
4.2 Environment 138
4.2.1 Mask-Shuffle Mechanism and Its Optimality 140
4.2.2 Proof of Theorem 4.1 143
4.2.3 The Game Between the Platform and Users 145
4.3 Equilibrium 146
4.4 Characterization 147
4.5 Platform Choice of Mechanism: Pivot Vs. Mask-Shuffle Mechanisms 149

4.6 Conclusion 151

4.7 Proofs 154

List of Figures

2-1 (a) the central setting and (b) the local setting. In the local setting, in contrast to the central setting, the users privatize their data before sharing with the platform. . . . 31

2-2 The interaction between the users and the platform in the central privacy setting. . . . 43

2-3 The interaction between the users and the platform in the local privacy setting. 48

2-4 (a) the variance in the central setting, (b) the variance in the local setting, and (c) the variance in the local minus the central setting for two users with $\textsc{var} = 1/4$ and uniform privacy sensitivities over $[1, 2]$ as a function of the privacy sensitivities (c_1, c_2). . . . 54

2-5 (a) the analyst's objective in the central setting, (b) the analyst's objective in the local setting, and (c) the analyst's objective in the local minus the central setting for two users with $\textsc{var} = 1/4$ and uniform privacy sensitivities over $[1, 2]$ as a function of the privacy sensitivities (c_1, c_2). . . . 54

2-6 (a) user 1's optimal privacy loss in the central setting, (b) user 1's optimal privacy loss in the local setting, and (c) the difference between user 1's optimal privacy loss in the local setting and the central setting as a function of (c_1, c_2). Here, we have $\textsc{var} = 1/4$ and the privacy sensitivities are uniform over $[1, 2]$. 55

3-1 The interaction between the users and the platform in the two-part private data acquisition. . . . 103

3-2 Weight of user 1's data in the optimal mechanism as a function of c_1 and c_2 for two users with $\gamma = 1$, $\alpha = 2$, and $\textsc{var} = 1/4$. . . . 115

3-3 (a) the platform's objective and (b) the platform's estimation error as a function of (c_1, c_2) for two users with $\gamma = 1$, $\alpha = 2$, and $\textsc{var} = 1/4$. 116

3-4 User 1's expected utility a function of c_1 for two users with $\gamma = 1$, $\alpha = 2$, and $\textsc{var} = 1/4$. 116

3-5 (a) user one's expected utility and (b) the platform's expected objective as a function of γ for two users with $\alpha = 2$ and $\textsc{var} = 1/4$. 116

4-1 (a) the mask-shuffle mechanism (b) the partial shuffler. 140

Chapter 1

Introduction

1.1 Problems and Contributions

Machine Learning (ML) algorithms are shaping every aspect of our lives, including work, healthcare, education, transportation, and communication. Though the promise of these new technologies is evident, the challenges they create are equally fundamental. In particular, it is crucial to understand the implications of these emerging technologies on society and address any ethical, legal, or social issues that may arise in their interactions with society.

Many algorithms, particularly those involving user participation, are often designed with the goal of optimizing certain objectives or achieving desirable outcomes for the platforms. However, overlooking user incentives and strategic behavior can have unintended consequences due to potential conflicting interests. More specifically, in many settings, there is a misalignment between the what platforms want and what user utilities or social objectives are.

Privacy concerns serve as a notable example, as the success of machine learning algorithms heavily relies on the collection and utilization of vast amounts of user data. Most online platforms' business model is partly based on acquiring and harvesting users' data. These data are often put to a multitude of uses, including improvements of algorithms, learning about underlying parameter (for example, in health care applications), and targeting individualized ads to users.

Each user's data is informative about an underlying state (e.g., some health condition of the population or its severity) and about their own private characteristics and preferences (that can be used for targeted intrusive ads). While users may tolerate or even value some of these data applications, they may wish to prevent others and protect some degree of privacy (for example, against intrusive ads). In fact, a survey in 2021 Lucas et al. [2021] found out that 86% of United States population find data privacy as a growing concern and 68% say they are concerned by the level of data collected by businesses.

Consequently, we need to adjust the design of data markets to include privacy-preserving mechanisms. In this regard, in Chapter 2, we consider a data market with users who ask for different levels of privacy guarantees to share their data, leading to the following key practical question:

Given users' heterogeneous privacy demands, how do we decide on privacy allocations to different users?

We answer this question and study the impact of data market architecture on the design of mechanisms for purchasing data from privacy-sensitive strategic users. We do so by considering two data market architectures: *central* and *local*. In the central setting, users trust the platform and share their raw non-private data with it. Yet, they require the platform's output, i.e., the estimator, to be *private*. On the other hand, the local setting is more conservative: users do not trust the platform and make their data private before even sharing it with the platform. We also use *differential privacy* to quantify the privacy loss of users to characterize its trade-offs with the benefits of data. Roughly speaking, differential privacy provides an upper bound on how sensitive the output of an algorithm is to an individual's data.

We consider a model in which a user, in addition to her private data, has a *heterogeneous privacy sensitivity* that is unknown to the platform and represents her cost per unit differential privacy loss. We design an *incentive-compatible* mechanism to elicit privacy sensitivity correctly. We frame this as a *mechanism design* problem in which an individual can share her data in exchange for a monetary reward or services. Individuals participate in the mechanism by reporting their privacy sensitivities and sharing their data. This mechanism returns three elements: (i) an *optimal* estimator, (ii) the privacy guarantee allocated to each user, and (iii) the compensation each user receives for her privacy loss. Thus, the mechanism endogenously determines the privacy loss levels as a function of both users' sensitivities and how their data is used in the estimation problem of the platform. Next, we also develop *efficient algorithms* to solve the resulting mechanism design problem in both privacy settings. Our mechanism in the central setting can be implemented in *quasi-linear time*, and it admits a *Polynomial Time Approximation Scheme (PTAS)* in the local setting.

Another key contribution of our work is characterizing the *optimal* estimator for given privacy loss levels desired by users. While this question has been answered in the homogeneous setting, i.e., when privacy levels are equal, our work is the first to answer it in the general heterogeneous case. To do so, we first establish minimax lower bounds for the estimation error using Le Cam's method from the statistics literature. We then provide nearly tight upper bounds that match our lower bounds and enable the choice of optimal estimator.

Finally, our work also compares central and local architectures from the perspective of both the users and the platform. We show that the platform prefers the central setting, meaning that her optimal utility under central differential privacy is always (weakly) larger. Intuitively, this holds since the platform receives non-private data in the central case, as opposed to the local case in which users make their data private before sharing it. However, the users' preference between central and local settings depends on their privacy sensitivities: if a user has a low privacy sensitivity, i.e., she cares more about learning the parameter of interest rather than her privacy, she would prefer the central setting. In contrast, users who have high privacy sensitivity would favor the local architecture.

Central and local architectures correspond to privacy losses at two different stages of data acquisition. The local privacy loss is due to the leakage of a user's information when she shares her data with the platform, and the central privacy loss is due to the released estimate by the platform to the public. One natural question is whether we can allow users to report possibly different privacy sensitivities for these two types of privacy losses, depending on their level of trust in the platform. In chapter 3, we consider a model of utility for users in which their privacy loss is a heterogeneous combination of the local and central privacy losses. We accordingly derive the privacy allocations at both local and central stages and the compensation to each user as the solution of this new mechanism design formulation.

The results of these two chapters are based on joint work with Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar Fallah et al. [2022a,b].

In the aforementioned chapters, we fixed the central and local architectures to primarily study the heterogeneity aspect of privacy demands in the data market, i.e., when users have different and unknown (to the platform) privacy sensitivities. Furthermore, we used monetary compensations to incentivize users to report their privacy sensitivities truthfully. In Chapter 4, we study a different aspect of the data market design: the optimal choice of architecture from both users' and the platform's point of view.

The platform collects data from users by means of a *mechanism* that could partially protect users' privacy. We develop a game-theoretic model of data sharing to study the choice of this mechanism in the space of all possible mappings from users' data to the platform's input. In selecting the mechanism, the platform chooses the balance between data harvesting and privacy to convince users to share their data (we assume that the platform can commit to an algorithm to achieve the desired balance). We establish a number of main results in this environment:

First, from the users' point of view, the optimal privacy-preserving mechanism takes a simple “mask-shuffling" form, whereby a user's data is revealed to the platform with some probability (and fully masked and anonymized with the complementary probability). Then, all the data go through a *shuffler* and a random permutation of the original set of datapoints is revealed to the platform. We establish that such a mask-shuffling mechanism provides the best privacy guarantee to users for any given amount of learning about the underlying state. This type of shuffling is attractive from the viewpoint of users because it maintains information about the underlying common state but ensures that the platform learns much less about the individual in expectation. From our modeling viewpoint, the mechanism is attractive because it enables us to characterize the extent of privacy guarantees by the probability with which the platform commits to shuffle the data of users.

Next, we utilize the mask-shuffle mechanism and frame the interaction between the platform and users as a two-stage game. In the first stage, the platform determines the probability of offering shuffling to the users. In the second stage, with this information, users independently decide on the probability of sharing (or masking) their data. We characterize the Bayesian-Stackelberg equilibrium of this game and present several comparative statics results. Specifically, we demonstrate that as users become more interested in learning the shared information, the platform offers a lower shuffling guarantee at equilibrium since users require less persuasion to disclose their data.

Moreover, we show that, an increase in the importance of the underlying common state makes users worse off and the platform better off. This is a paradoxical result, since holding the platform's algorithm fixed, users' welfare would have increased (for example, because they learn more about the underlying health condition). However, recognizing this, the platform then relaxes privacy guarantees so much that its profits increases significantly and user welfare decreases. Put differently, when learning about the underlying state becomes more important for users, the platform can exploit this preference to tilt things for its own benefit.

Finally, we demonstrate that the platform has an incentive to deviate from the optimal mask-shuffling mechanisms preferred by the users. We identify a set of "pivot mechanisms" that allow the platform to make individual privacy dependent on the choices of other users. By designing a pivot mechanism that guarantees not to use any user data if any user decides not to share their data, the platform can exploit user preferences towards the common state. As a result, the user's cost of not sharing her data increases, making her more likely to share her data, even if she values her privacy. We also demonstrate that more continuous versions of pivot mechanisms can achieve the same outcome. This result underscores the fact that self-regulation by platforms may not be sufficient to ensure user privacy and highlights the challenges in protecting individual privacy in the face of collective benefits.

Taken together, our results highlight various facets of excessive platform power, even when platforms can design and commit to algorithms for protecting user privacy. The results of this chapter are based on joint work with Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar Acemoglu et al. [2023].

The results of this thesis shed light on the intricate dynamics between algorithms, strategic users, and data market design. By addressing the variety of preferences, optimizing privacy-preserving mechanisms, and understanding the interplay between users and platforms, we aim to mitigate the challenges and negative consequences arising from the widespread use of data. Ultimately, the responsible integration of ML algorithms into society requires continuous examination, refinement, and consideration of diverse perspectives.

1.2 Related Literature

Private data acquisition literature: Our work in Chapters 2 and 3 builds on the growing literature on optimal data acquisition from strategic privacy conscious users. Several of these papers use differential privacy to quantify the cost users incur when sharing their data Ghosh and Roth [2011], Nissim et al. [2012], Nissim et al. [2014]. A pioneering paper in this literature is Ghosh and Roth [2011], which consider designing a mechanism for collecting data from users that explicitly experience a cost for privacy loss. Ghosh and Roth [2011] assume that each user has a private bit and a heterogeneous privacy loss parameter and the platform's goal is to estimate the sum of user's data by using a differentially private and dominant strategy truthful mechanism. This paper considers both the case when the user data and privacy parameter are independent (as in our case) and when they are correlated. For the independent case, their mechanism results in providing a single privacy level to all users whose data are collected (because of their worst case view with a focus on dominant strategy truthful mechanisms and lack of distributional assumptions on user data or cost parameters). For the correlated case, Ghosh and Roth [2011] provide an impossibility result for the existence of a truthful and individually rational mechanism. Several papers build on Ghosh and Roth [2011], extending it to take it or leave it offers Ligett and Roth [2012], and strengthening the impossibility results Nissim et al. [2014].

Another line of work tackles the open question posed by Ghosh and Roth [2011] on whether a model with distributional assumption on users' costs and Bayesian mechanism design approach could be used to develop optimal mechanism for collecting data with privacy guarantees. Roth and Schoenebeck [2012], Chen et al. [2018], and Chen and Zheng [2019] followed this approach using a randomized mechanism in which user's data is used with a probability that depends on the reported privacy costs of the users.[1] These papers do not

[1] This is different from our mechanism in which payments and resulting privacy losses depend on the reported privacy sensitivity of all users.

use differential privacy to model privacy costs, but rather use a menu of probability-price pairs to control the privacy loss and compensation for each user.

Another noteworthy paper in this literature is Cummings et al. [2015], which consider data purchase from users that provide different levels of data accuracy (variance) and may strategically price access to their data. The variance in the data can represent uncertainty in data quality or intentionally added noise in order to guarantee privacy. This paper does not impose a functional form for the privacy loss in terms of a differential privacy parameter and instead allows for a flexibility in offering a menu of different variance levels (or equivalently, arbitrary costs for each level independently).

Our work differs from these works by assuming prior information on user privacy sensitivities, and focusing on characterizing the optimal Bayesian incentive compatible mechanism. We further assume that user data are drawn from the same underlying distribution. This allows the platform to put more weight on the data of a user with lower price sensitivity, leading to different privacy levels for participating users. Another important distinction of our model is our assumption that users derive utility from the accuracy of the estimation outcome which changes the privacy allocation of the optimal mechanism. Finally our work considers different privacy architectures, central and local, and explores the different privacy guarantees provided by an optimal mechanism under these different architectures. Prior to our work Cummings et al. [2022] has considered a setting in which the users benefit from a better estimation outcome. They consider a linear estimator with Laplace additive noise and show how it allows for heterogeneous privacy guarantees to different users in the central model. We depart from this paper by establishing the (near) optimal estimator, considering strategic users in reporting their privacy costs, and studying both central and local settings and their comparison (see also Pai and Roth [2013] for a survey).

In our work, as well as the above papers, the platform can verify the data of users. A different stream of this literature considers a setting in which individuals have the ability to misreport their information Perote and Perote-Pena [2003], Dekel et al. [2010], Meir et al. [2012], Ghosh et al. [2014], Cai et al. [2015], Liu and Chen [2016, 2017].

Our work also relates to the literature that consider privacy aware mechanism design and selling strategies such as McSherry and Talwar [2007], Nissim et al. [2012], Abernethy et al.

[2019], Lei et al. [2020], Chen et al. [2021a], and Chen et al. [2021b]. In particular, Chen et al. [2021a] consider a dynamic personalized pricing problem with unknown nonparametric demand models under data privacy protection, while Lei et al. [2020], Chen et al. [2021a], and Han et al. [2021] consider parametric demand models. We note that both our research question and results are different from these papers. In particular, we study the design of optimal mechanisms for collecting data from strategic users with privacy concerns while these papers consider demand learning for personalized pricing under privacy concerns in an online learning framework and provide (tight) bounds on the regret of the optimal algorithm.

In addition, our work relates to the literature that studies the problem of choosing the proper level of differential privacy given the goal of protecting individuals' privacy such as Lee and Clifton [2011], Hsu et al. [2014], and Mehner et al. [2021]. We depart from this line of work by studying the endogenous choice of differential privacy levels based on individuals' privacy sensitivity and their interactions with a platform.

The literature on data market and platform behavior: Our work, especially Chapter 4, also relates to the emerging literature on the social dimension of data and online platform behavior, for example, Acemoglu et al. [2022] and Bergemann et al. [2020]. Bergemann et al. [2020] consider a setting in which a (trusted) data intermediary collects users' data and resells them to a platform. They show that the data externality, whereby a user's data is predictive of others, can reduce the intermediary's cost of acquiring the data. Acemoglu et al. [2022] consider a more general, though reduced-form, data externality and establish that this externality also reduces the value of data to both users and the platform. As a result, data externalities depress data prices and amplify inefficiencies. Relatedly, Ichihashi [2020] considers the interactions between a privacy-concerned user and a platform, where the user's activity reveals private information (see also Fainmesser et al. [2022] for a similar model). These papers do not consider general privacy-preserving mechanisms.

More broadly, our work is also related to the literature on data collection and sharing. Hörner and Skrzypacz [2016] study the design of mechanisms for selling data, while Goldfarb and Tucker [2011], Bergemann and Bonatti [2015], Montes et al. [2019], and Jagabathula et al. [2020] investigate how individual private information can be used to improve resource allocation. Competition implications of online data sharing and technologies have been explored in, among others, Bimpikis et al. [2021] and Gur et al. [2019]. Bergemann and Bonatti [2015] study the problem of selling cookies for targeted advertisement and study how the price of data changes with the reach of the dataset and the fragmentation of data sales. Fu et al. [2022] study data collection and privacy in recommendation systems. Other works on information-sharing and market structure include Li [2002], Li and Zhang [2008], Ha and Tong [2008], Shang et al. [2015], Foster et al. [2016], Lobel and Xiao [2017], Bimpikis et al. [2019], Candogan and Drakopoulos [2020], Immorlica et al. [2020], Hu et al. [2020], Ashlagi et al. [2020], Anunrojwong et al. [2021], Besbes and Mouchtaki [2021], and Ashlagi et al. [2021] (see Bergemann and Bonatti [2019] for a survey).

The differential privacy literature: Our work in Chapters 2 and 3 relates to the literature on differential privacy. Initiated by the work of Dwork et al. [2006a,b], differential privacy has emerged as a popular framework in computer science and engineering for characterizing the privacy leakage of data oriented algorithms. Our work, in particular, is related to the private mean estimation which has been studied extensively over the past decade Duchi et al. [2013], Barber and Duchi [2014], Karwa and Vadhan [2017], Asoodeh et al. [2021], Kamath et al. [2019, 2020], Cummings et al. [2021], Acharya et al. [2021]. Additionally, our work in Chapter 3 uses the Rényi differential privacy introduced by Bun and Steinke [2016] and Mironov [2017].

Chapter 2

Optimal Deferentially Private Data Acquisition

2.1 Introduction

The data of billions of people around the world are used every day for improving search algorithms, recommendations on online platforms, personalized advertising, and the design of new drugs, services and products. With rapid advances in machine learning (ML) algorithms and further growth in data collection, these practices will become only more widespread in the years to come. However, a common concern with many of these data-intensive applications centers on privacy — as a user's data is harnessed, more and more information about her behavior and preferences are uncovered and potentially utilized by platforms and advertisers.

A popular solution to the tension between privacy costs and benefits of data is to use methods such as differential privacy in order to limit the extent to which an individual's data is uncovered and exploited. The basic idea of differential privacy is to provide an upper bound on how sensitive the output of an algorithm (e.g., the vector of recommendations from an online site) is to an individual's data. Although differential privacy methods are already used by many of the tech companies, including, Apple, Google and Microsoft (see, e.g., Erlingsson et al. [2014] and Ding et al. [2017]), a key practical question remains: how do we decide how much privacy an individual will obtain? Imagine, for example, that two individuals have similar data, but one is very privacy conscious, while the other one does not think that she has any concerns of privacy. It is natural to provide different privacy levels for these two individuals when acquiring their data, but exactly how?

This chapter is an attempt to answer this key question and study the impact of data market architecture on the design of mechanisms for purchasing data from privacy sensitive strategic users. We consider a platform interested in estimating an underlying parameter using data collected from users. While users benefit from the outcome of the estimation, they are cognizant of the privacy losses they will incur and hence might be discouraged from sharing their data. User data come from some underlying population distribution where its mean is given by the parameter of interest. We formulate this question as a mechanism design problem, in which an individual can share her data in exchange for a monetary reward or services, but at the same time has a heterogeneous privacy sensitivity that represents her cost per unit privacy loss. We assume a known prior on user's privacy sensitivity (which is independent of the data distribution). While an individual's data is difficult to manipulate, her privacy preferences are easier to falsify (if monetary rewards were increasing in how privacy conscious individual is, then she might prefer to misrepresent this information). Individuals participate in the mechanism by reporting their privacy sensitivities and sharing their data. This mechanism simultaneously determines an "optimal" estimator, compensation for the users, and privacy losses an individual will incur. Thus, the mechanism endogenously determines the privacy loss levels as a function of both user sensitivities and also how their data is used in the estimation problem of the platform.

We consider two popular differential privacy settings for providing privacy guarantees for the users: central and local. In the central privacy setting, we require the output of the estimation process to be differentially private with respect to each individual's data. In the local privacy setting, we impose a differential privacy requirement with respect to the individual data of each user. Before formulating the optimal mechanism design problem, we derive optimal estimators for given heterogeneous privacy loss levels for users in the two privacy settings. We establish minimax lower bounds for the estimation error and use these bounds to characterize the form of the optimal estimator with central and local privacy guarantees. In particular, in the central setting we show that, for a given vector of privacy losses, a linear estimator that combines a (properly designed) weighted average of the users' data points and a Laplace noise achieves the (near) optimal estimation error among all estimators that can achieve the desired privacy losses. In addition, in the local setting, we show that, for a given vector of privacy losses, first adding a Laplace noise to the data of each user and then taking a weighted average of the users' data points achieves the optimal estimation error.

In the second part of the chapter, we formulate the Bayesian-optimal mechanism design problem where the objective of the platform is to minimize the sum of the estimation error and total payment for the users. We first provide a characterization of the optimal payment as a function of the reported privacy sensitivities. This is closely related to the payment identity in Myerson's optimal auction design problem (Myerson [1981]), but differs in that the reported privacy sensitivities of other users impacts a user's utility not only through her privacy loss level and payment but also through the overall estimation error (all users benefit from a lower estimation error). We then focus our attention to linear estimators (which were shown to be optimal for differentially private estimation given exogenous privacy loss levels). We show that under some regularity conditions on the distribution of privacy sensitivities, the problem of finding the optimal privacy levels can be cast as the solution to a non-convex optimization problem. In both settings, we first reformulate the platform's problem in terms of designing a pair of weight and privacy loss functions. These functions map the vector of reported privacy sensitivities to a vector of privacy losses for users and a vector of weights in the linear estimator of the platform, respectively. In the central setting, we use the structure of the problem to derive an efficient score-based algorithm for implementing our mechanism in time $\mathcal{O}(n \log n)$. In the local setting, we develop a Polynomial Time Approximation Scheme (PTAS) to solve the platform's problem.

In the last section, we compare the central and local differential privacy settings and establish that the platform achieves a (weakly) higher utility in the central privacy setting than that in the local one. This is because the local setting provides a stronger privacy guarantee and hence increases the final estimator's variance, which in turn reduces the platform's utility. We also illustrate that, for a given vector of privacy sensitivities, the privacy loss level allocated to a user in the optimal local data acquisition mechanism is not necessarily higher than the central setting, and in fact, it can be strictly lower (providing better privacy guarantees).

From a technical point of view, our first technical contribution is deriving the minimax optimal private mean estimator for heterogeneous differential privacy levels. Prior to our work, the optimal estimation has been studied only for homogeneous differential privacy levels (see, e.g., Duchi et al. [2013], Dwork et al. [2014], Barber and Duchi [2014]). Utilizing this optimal estimator, we demonstrate how existing mechanism design tools can be applied to our setting, resulting in a point-wise optimization approach using virtual values to find the optimal mechanism. Our second technical contribution involves developing efficient algorithms specifically tailored for solving non-convex point-wise optimization problems that arise in private data acquisition. This differs from the conventional mechanism design setting, where the optimal mechanism can be obtained by solving a linear program. In terms of the structural aspect, our problem deviates from the classic mechanism design, where the optimal allocation typically follows a threshold rule. Instead, in our problem, the optimal differential privacy level exhibits a continuous dependence on privacy sensitivity.

The rest of the chapter proceeds as follows. Section 2.2 presents the setting, describes central and local differential privacy, and provides near optimal minimax estimator with heterogeneous privacy losses. In Section 2.3, we establish how the platform's mechanism design problem turns into a point-wise optimization problem over the privacy losses. In Section 2.4, we characterize the optimal privacy loss levels in the central privacy setting and find a polynomial time algorithm to find them. In Section 2.5, we characterize the optimal privacy losses in the local privacy setting and establish that it admits a PTAS. Section 2.6 compares the central and the local privacy settings. Section 2.7 concludes, while the last section includes the omitted proofs from the text.

2.2 Differential Privacy and Platform's Estimation Problem

We consider a *platform* interested in estimating an underlying parameter $\theta \in \mathbb{R}$ by collecting relevant data from a set of *users* denoted by $\mathcal{N} = \{1, \ldots, n\}$. Each user $i \in \mathcal{N}$ has some personal data $X_i \in \mathcal{X}$ which is informative about θ. We assume that $X_i = \theta + Z_i$, where $(Z_1, \ldots, Z_n)$ are independent and identically distributed mean zero random variables with a variance denoted by $\textsc{var}$.[1] Throughout the chapter, for simplicity, we assume $|Z_i| \leq \frac{1}{2}$ for all $i \in \mathcal{N}$.[2]

Users share their data with the platform since a more accurate estimate of parameter θ is useful for their objective (e.g., identifying a treatment from collecting individual medical records). However, sharing of individual data raises privacy concerns which users are cognizant of. Failure to address these privacy concerns would discourage users from sharing their data. We model the privacy demand of users as a maximum privacy loss they can tolerate. We use the notion of differential privacy to combine optimal estimation with such privacy guarantees.

In the next section, we assume the privacy loss level each user is willing to accept is given and derive (near-)optimal estimators that achieve these levels using different privacy guarantees. In particular, the central setting provides a privacy guarantee in terms of how user data impacts the final estimate of the platform, whereas the more restrictive local setting seeks a guarantee for the individual data shared by each user.

In section 2.4, we endogenize the choice of the privacy loss levels by assuming a privacy sensitive user utility.

2.2.1 Central and Local Differential Privacy

We first formalize the differential privacy framework we use to quantify guarantees on privacy demand of users. We focus on two settings, known as *central* and *local* differential privacy. In the central case, we assume that the users trust the platform to share their data and require a privacy guarantee for user data by limiting its impact on the output of the analyst's estimation problem. In the local case, we assume a more restrictive privacy demand on the individual data shared by each user.

[1] The assumption that Z_i's are independent and have the same variance is reasonable in the context of estimation from a population and is made to simplify the notation and analysis. Our characterization of the optimal data acquisition mechanism readily extends to a setting with correlated users' data with different variances.

[2] This is without loss of generality and the analysis extends to an arbitrary bound on $|Z_i|$'s by properly adjusting the estimator used by the platform.

We start with the definition of central differential privacy which slightly generalizes the standard definition in Dwork et al. [2006a,b] by allowing different levels of privacy loss for each user.[3]

Definition 2.1 (Central differential privacy)**.** *Let $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n \in \mathbb{R}_+^n$. Assume $\mathcal{S}, \mathcal{S}' \in \mathcal{X}^n$ are two datasets that differ in the i-th component (which represents user i's data). A randomized algorithm $\mathcal{A} : \mathcal{X}^n \to \mathbb{R}$ is $\boldsymbol{\varepsilon}$-centrally differentially private if for all measurable sets $\mathcal{W}$ in $\mathbb{R}$,*

$$\mathbb{P}(\mathcal{A}(\mathcal{S}) \in \mathcal{W}) \leq e^{\varepsilon_i} \; \mathbb{P}(\mathcal{A}(\mathcal{S}') \in \mathcal{W}).$$

This definition implies that the algorithm's output changes with probability at most e^{ε_i} when the data of user i changes. In particular, Definition 2.1 is equivalent to $e^{-\varepsilon_i} \leq \frac{\mathbb{P}(\mathcal{A}(\mathcal{S}) \in \mathcal{W})}{\mathbb{P}(\mathcal{A}(\mathcal{S}') \in \mathcal{W})} \leq e^{\varepsilon_i}$ for all $i \in \mathcal{N}$, neighboring $\mathcal{S}, \mathcal{S}' \in \mathcal{X}^n$, which only differ in user i's data, and all measurable sets $\mathcal{W}$ in $\mathbb{R}$. Therefore, ε_i can be interpreted as a variable that captures the maximum *privacy loss* that Algorithm $\mathcal{A}$ ensures for user i: the smaller ε_i is, the less of an impact user i's data has on the output of Algorithm $\mathcal{A}$, implying a lower privacy loss (or equivalently a higher privacy guarantee) for user i's data.

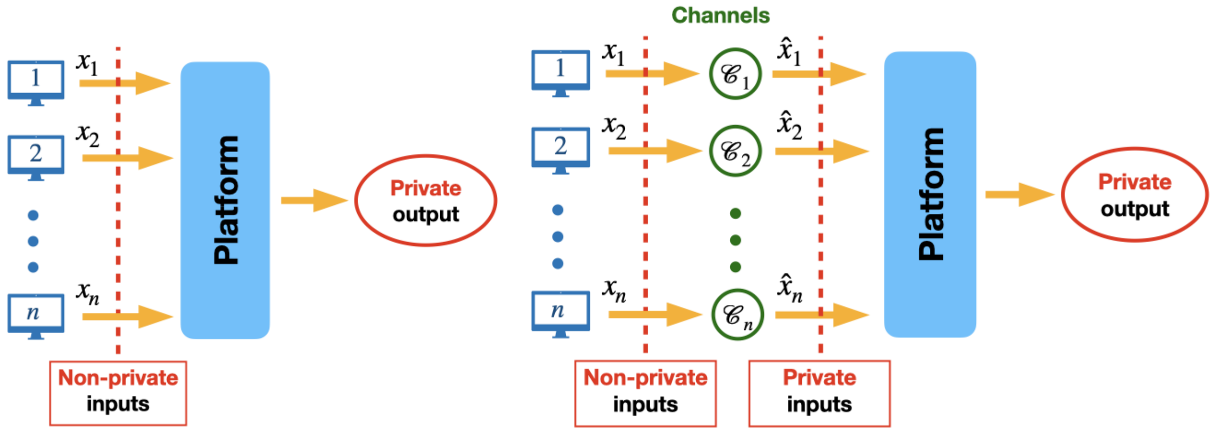
Local differential privacy considers the setting where the users do not trust the platform with their data. The users therefore first produce a private version of their data through a mapping before sharing it with the platform. Building on the literature on differential privacy, we refer to this mapping as a *channel* (see, e.g., Duchi et al. [2013]) and define it to be locally differentially private as follows.

Definition 2.2 (Local differential privacy)**.** *A randomized channel $\mathcal{C} : \mathcal{X} \to \mathbb{R}$ is ε-locally differentially private if for any $x, x' \in \mathbb{R}$ and all measurable sets $\mathcal{W}$ in $\mathbb{R}$,*

$$\mathbb{P}(\mathcal{C}(x) \in \mathcal{W}) \leq e^{\varepsilon} \mathbb{P}(\mathcal{C}(x') \in \mathcal{W}).$$

Let $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n \in \mathbb{R}_+^n$. An algorithm $\mathcal{A} : \mathcal{X}^n \to \mathbb{R}$ is $(\varepsilon_i)_{i=1}^n$-locally differentially private if it

[3] This extension is in line with the literature that introduced *personalized* or *heterogeneous* differential privacy, where each user can have a different privacy loss Jorgensen et al. [2015], Alaggan et al. [2015], Niu et al. [2021]. In our setting users have different preferences for their privacy which motivates our definition with heterogeneous privacy loss levels.



(a) (b)

Figure 2-1: (a) the central setting and (b) the local setting. In the local setting, in contrast to the central setting, the users privatize their data before sharing with the platform.

takes $(\mathcal{C}_i(x_i))_{i=1}^n$ *as input (as opposed to* $(x_i)_{i=1}^n$ *itself), where* $\mathcal{C}_i$ *is an* ε_i*-locally differentially private channel.*

It is worth noting that an $(\varepsilon_i)_{i=1}^n$-locally differentially private algorithm is $(\varepsilon_i)_{i=1}^n$-centrally differentially private according to Definition 2.1 as well (see Dwork et al. [2014, Observation 12.1].) Figure 2-1a and 2-1b depict central and local differential privacy architectures, respectively.

A point worth mentioning is that in the local privacy setting, the data is privatized directly on the user side, giving users control over its implementation. Unlike the central setting, the local setting does not rely on the platform credibly delivering the promised privacy level.

In both central and local cases, the basic mechanism to ensure privacy is adding *fine-tuned noise*. As we establish next, a Laplace mechanism which adds a zero-mean Laplace noise to the variable of interest is optimal, and therefore we adopt this throughout (we make the optimality statement precise in this section). Recall that the density of a mean-zero (one-dimensional) Laplace distribution with parameter η, denoted by Laplace(η), is given by

$$p(z) = \frac{1}{2\eta} \exp(-|z|/\eta) \quad \text{for all } z \in \mathbb{R}$$

and its variance is given by $2\eta^2$. The following lemma characterizes the differential privacy guarantees obtained by a Laplace mechanism.

Lemma 2.1 (Dwork et al. [2014]). *Consider a real-valued function $f : \mathcal{X}^n \to \mathbb{R}$ and let W be a Laplace noise with parameter $1/\varepsilon$, i.e., $W \sim \text{Laplace}(1/\varepsilon)$. Then, $\mathcal{A}(\boldsymbol{x}) := f(\boldsymbol{x}) + W$ for any $\boldsymbol{x} \in \mathcal{X}^n$, is $(\varepsilon L_i(f))_{i=1}^n$-centrally differentially private, where $L_i(f)$ is the sensitivity of f with respect to the i-th coordinate, and is given by*

$$L_i(f) := \sup \left\{ |f(\boldsymbol{x}) - f(\boldsymbol{x}')| : \textit{for all } \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X} \textit{ that only differ in the } i\textit{-th coordinate} \right\}. \tag{2.1}$$

Next, we consider the following problem in both the central and the local differential privacy settings: Assume that the desired privacy level of users, i.e., ε_i for user i, is given to the platform. What is the optimal choice of the estimator in terms of expected square error?

To answer this question, we first provide minimax lower bounds for private mean-estimation problem under both central and local definitions of differential privacy (given in Definitions 2.1 and 2.2, respectively). We then prove that a linear estimator with Laplace mechanism achieves those lower bounds up to a logarithmic factor. While the private mean estimation problem has been extensively studied when privacy levels across all users are equal Duchi et al. [2013], Dwork et al. [2014], Barber and Duchi [2014], to the best of our knowledge, it has not been studied in our setting where the privacy levels of users are heterogeneous.

2.2.2 (Near) Optimal Estimation With Central Differential Privacy

Let $\mathcal{P}$ be a family of distributions, defined over the sample space $\mathcal{X}$. Our goal is to estimate the mean $\theta : \mathcal{P} \to \mathbb{R}$ where $\theta(P) = \mathbb{E}_{X \sim P}[X]$ for any $P \in \mathcal{P}$. We let $X_1, \cdots, X_n$ be n independent and identically distributed samples that are drawn from $P \in \mathcal{P}$ and $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n$ be the privacy levels. In the central setting, an estimator $\hat{\theta}(X_1, \cdots, X_n)$ is a real-valued measurable function over $\mathcal{X}^n$ which estimates $\theta(P)$. We define $\mathcal{Q}_c(\boldsymbol{\varepsilon})$ as the class of $\boldsymbol{\varepsilon}$-centrally differentially private estimators, according to Definition 2.1. With this notation in hand, the minimax estimation error is given by

$$\mathcal{L}_c(\mathcal{P}, \theta, \boldsymbol{\varepsilon}) := \inf_{\hat{\theta} \in \mathcal{Q}_c(\boldsymbol{\varepsilon})} \sup_{P \in \mathcal{P}} \mathbb{E}_{(X_i \sim P)_{i=1}^n, \hat{\theta}} \left[\left| \hat{\theta}(X_{1:n}) - \theta(P) \right|^2 \right], \tag{2.2}$$

where the expectation is taken over the randomness in both samples $X_{1:n}$ and the estimator. The supremum in the above expression is the worst-case estimation error over all distributions of the data points. Therefore, given that the platform does not know the distribution of the data points, the infimum outputs the $\boldsymbol{\varepsilon}$-centrally differentially private estimator that minimizes this worst-case estimation error.

Our goal is to provide a lower bound on the minimax rate defined above and prove that such lower bound can be (almost) achieved by linear estimators with Laplace mechanism. To do so, let us first, formally define this class of estimators. Given the data of users $x_1, \cdots, x_n$, a linear estimator with Laplace mechanism is in the form of

$$\hat{\theta} = \sum_{i=1}^n w_i x_i + \text{Laplace}(1/\eta), \tag{2.3}$$

where w_i is the weight that the estimator allocates to the data of user i with $\sum_{i=1}^n w_i = 1$. Given this estimator, the following lemma shows that the data of each user is centrally differentially private.

Lemma 2.2. *The estimator $\hat{\theta}$ given in (2.3) is $(w_i \eta)_{i=1}^n$-centrally differentially private.*

This lemma directly follows from Lemma 2.1. The proof of this lemma as well as other omitted proofs are presented in the Section 2.8.

We next establish a lower bound for the estimation error in the central setting and prove that a linear estimator with Laplace mechanism (almost) achieves the lower bound.

Theorem 2.1. *Let $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n$ and, without loss of generality, suppose $\varepsilon_1 \leq \cdots \leq \varepsilon_n \leq 1$. Also, let $\mathcal{P}^*$ be the family of distributions P such that $|X| \leq \frac{1}{2}$ almost surely.*[4] *There exists*

[4] The choice of upper bound 1/2 is without loss of generality and is made to guarantee the length of the support is bounded by 1, simplifying the equations.

a (universal) positive constant c_l such that[5]

$$\mathcal{L}_c(\mathcal{P}^*, \theta, \boldsymbol{\varepsilon}) \geq c_l \left(\max_{k \in \{0,1,\cdots,n\}} \frac{1}{n - k + (\sum_{i=1}^{k} \varepsilon_i)^2} \wedge 1 \right). \tag{2.4}$$

Moreover, there exists an $\boldsymbol{\varepsilon}$-centrally differentially private linear estimator $\hat{\theta}$ and a (universal) constant c_u such that

$$\mathbb{E}_{(X_i \sim P)_{i=1}^n, \hat{\theta}} \left[\left| \hat{\theta}(X_{1:n}) - \theta(P) \right|^2 \right] \leq c_u \max_{k \in \{0,1,\cdots,n\}} \frac{\log(n+1)}{n - k + (\sum_{i=1}^{k} \varepsilon_i)^2}, \tag{2.5}$$

for any $P \in \mathcal{P}^*$.

We prove the lower bound by using the Le Cam's method Yu [1997] that reduces the problem of finding lower bounds to a hypothesis testing problem between two distributions. More specifically, using this technique, we need to bound the change in the distribution of estimator's output, i.e., the distribution of $\hat{\theta}(X_{1:n})$, when the underlying data distribution changes. To bound the change in the distribution, we first notice that bounding the change in the distribution by using a single distance between the distributions does not immediately give us the desired bound. We circumvent this challenge by using a combination of two well-known distances between two distributions: Total Variation (TV) and Kullback–Leibler (KL).

We establish the upper bound by constructing a linear estimator in the form of (2.3) that achieves the desired bound. Note that, by Lemma 2.2, to have an $(\varepsilon_i)_{i=1}^n$-centrally differentially private estimator, we should have

$$\eta w_i \leq \varepsilon_i \text{ for all } i. \tag{2.6}$$

An interesting and somewhat counter-intuitive observation is that the above constraints are not necessarily all binding for the optimal estimator. In other words, the optimal estimator might end up providing higher privacy levels than reported for certain users. This means that we might achieve a lower variance for the estimator by guaranteeing better privacy

[5] For any $x, y \in \mathbb{R}$, we let $x \wedge y$ denote $\min\{x, y\}$.

levels (i.e., lower ε_i's) for certain users. The main reason for this structure is that, if we keep all the constraints active while some users ask for less privacy, this might lead to putting *too much weight* on their data. In fact, the optimal estimator in the proof of Theorem 2.1 is built by capping the weight that we assign to the data of a portion of users with the highest ε_i's, i.e., users with the lowest privacy restrictions. Let us elaborate this matter with an example. Suppose $(\varepsilon_i)_{i=1}^n$ are given as

$$\varepsilon_1 = \cdots = \varepsilon_{\lfloor n-\sqrt{n}\rfloor} = \frac{1}{\sqrt{n}}, \quad \varepsilon_{\lfloor n-\sqrt{n}\rfloor+1} = \cdots = \varepsilon_n = 1. \tag{2.7}$$

As shown in the proof of Theorem 2.1, the linear estimator

$$\hat{\theta} = \sum_{i=1}^{n} \frac{1}{n} x_i + \text{Laplace}\left(\frac{1}{\sqrt{n}}\right), \tag{2.8}$$

achieves the variance $\mathcal{O}(\frac{1}{n})$ which matches the lower bound, and hence it is optimal. Moreover, this estimator is $\frac{1}{\sqrt{n}}$-centrally differentially private with respect to every user's data, meaning it guarantees a much better level of privacy for users $\lfloor n-\sqrt{n}\rfloor + 1$ to n. Now let us see what happens if we consider the linear estimator that keeps all the constraints active:

$$\hat{\theta} = \sum_{i=1}^{n} \frac{\varepsilon_i}{\sum_{j=1}^n \varepsilon_j} x_i + \text{Laplace}\left(\frac{1}{\sum_{j=1}^n \varepsilon_j}\right). \tag{2.9}$$

The variance of this estimator is

$$\mathbb{E}[|\hat{\theta}-\theta|^2] = \frac{2}{(\sum_{j=1}^n \varepsilon_j)^2} + \sum_{i=1}^{\lfloor n-\sqrt{n}\rfloor} \frac{1/n}{(\sum_{j=1}^n \varepsilon_j)^2}\text{VAR} + \sum_{i=\lfloor n-\sqrt{n}\rfloor+1}^{n} \frac{1}{(\sum_{j=1}^n \varepsilon_j)^2}\text{VAR}.$$

Note that, $\sum_{j=1}^n \varepsilon_j \approx 2\sqrt{n}$, and hence, the first two terms in the right-hand side of the above expression are $\mathcal{O}(\frac{1}{n})$. However, the third term is $\Omega(\frac{1}{\sqrt{n}})$. This leads to the total variance being $\Omega(\frac{1}{\sqrt{n}})$, and thus, this estimator is suboptimal.

2.2.3 Optimal Estimation With Local Differential Privacy

Here, we consider the local differential privacy setting. In this setting, and for any i, instead of observing X_i, the platform observes $\hat{X}_i := \mathcal{C}_i(X_i)$, where $\mathcal{C}_i : \mathcal{X} \to \hat{\mathcal{X}}$ is a ε_i-locally differentially private channel. Hence, the estimator $\hat{\theta}$ would be defined over $\hat{\mathcal{X}}^n$ and would be cast as $\hat{\theta}(\hat{X}_{1:n})$. Also, $\mathcal{Q}_l(\boldsymbol{\varepsilon})$ denotes the class of mechanisms $\mathcal{M} : \mathcal{X}^n \to \hat{\mathcal{X}}^n$ where $\mathcal{M}(X_1, \cdots, X_n) = (\mathcal{C}_i(X_i))_{i=1}^n$, with $\mathcal{C}_i$ being an ε_i-locally differentially private channel. Under local differential privacy, the minimax rate is defined as

$$\mathcal{L}_l(\mathcal{P}, \theta, \boldsymbol{\varepsilon}) := \inf_{\hat{\theta}, \mathcal{M} \in \mathcal{Q}_l(\boldsymbol{\varepsilon})} \sup_{P \in \mathcal{P}} \mathbb{E}_{(X_i \sim P)_{i=1}^n, \hat{\theta}} \left[\left| \hat{\theta}(\hat{X}_{1:n}) - \theta(P) \right|^2 \right], \tag{2.10}$$

where the expectation is taken over the randomness in both samples $X_{1:n}$ and the estimator. Again, the supremum in the above expression is the worst-case estimation error over all distributions of the data points. Therefore, given that the platform does not know the distribution of the data points, the infimum outputs the $\boldsymbol{\varepsilon}$-locally differentially private estimator that minimizes this worst-case estimation error.

In this case, the linear estimator with Laplace mechanism is defined as follow: User i releases an ε_i-locally differentially private version of x_i, denoted by $\hat{x}_i$, using Laplace mechanism, i.e., $\hat{x}_i = x_i + \text{Laplace}(1/\varepsilon_i)$. Using these private data points, we form the following estimate

$$\hat{\theta} = \sum_{i=1}^{n} w_i \hat{x}_i, \tag{2.11}$$

where w_i is the weight that the estimator allocates to the private data of user i with $\sum_{i=1}^n w_i = 1$. We next establish a lower bound for the estimation error in the local setting and prove that a linear estimator with Laplace mechanism achieves the lower bound.

Theorem 2.2. *Let $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n$ with $\varepsilon_i \leq 1$ for all i. Also, let $\mathcal{P}^*$ be the family of distributions P such that $|X| \leq \frac{1}{2}$ almost surely. There exists a (universal) positive constant ℓ_l such that*

$$\mathcal{L}_l(\mathcal{P}^*, \theta, \boldsymbol{\varepsilon}) \geq \ell_l \left(\frac{1}{\sum_{i=1}^n \varepsilon_i^2} \wedge 1 \right). \tag{2.12}$$

Moreover, there exists an $\boldsymbol{\varepsilon}$-locally differentially private linear estimator $\hat{\theta}$ and a universal constant ℓ_u such that

$$\mathbb{E}_{(X_i \sim P)_{i=1}^n, \hat{\theta}} \left[\left| \hat{\theta}(X_{1:n}) - \theta(P) \right|^2 \right] \leq \frac{\ell_u}{\sum_{i=1}^n \varepsilon_i^2}, \tag{2.13}$$

for any $P \in \mathcal{P}^*$.

Similar to the proof of Theorem 2.1, we prove the lower bound by using the Le Cam's method. To establish the upper bound, similarly, we construct a linear estimator that achieves the lower bound up to a constant factor.

2.3 Data Acquisition Mechanism With Privacy Guarantees

In this section, we endogenize the choice of the privacy loss levels by assuming a utility function that captures different privacy sensitivities. In particular, each user $i \in \mathcal{N}$ has a type or *privacy sensitivity* $c_i \in \mathbb{R}_+$ that represents the per unit cost of privacy loss for user i. We assume each c_i is independently drawn from a publicly known distribution with cumulative distribution function $F_i(\cdot)$ and probability density function $f_i(\cdot)$. We also let $\mathbf{c} = (c_1, \ldots, c_n)$ denote the vector of privacy sensitivities. The privacy sensitivity of each user is their private information.

We consider a mechanism whereby individuals participate by sharing their data and reporting their privacy sensitivities.[6] While users can misrepresent their privacy sensitivities, they have no capability to manipulate their data (e.g., their data is collected by the analyst when they participate or can be verified). Depending on the reported sensitivity, the analyst provides a compensation for the user in exchange for her data. This compensation may be a direct monetary payment or it may be an implicit transfer, for example, in the form of some good or service the analyst provides to the user to acquire her data. The mechanism designer simultaneously determines the privacy loss levels (which were assumed given in the

[6] From here on, we will use the terms mechanism designer and platform interchangeably.

previous section) and a differentially private estimator based on users' data that achieves these levels.

Given this interaction, we next specify a data acquisition mechanism with privacy guarantees on users' data.

Definition 2.3 (Private data acquisition mechanism)**.** We call the tuple $(\hat{\theta}, \boldsymbol{\varepsilon}, \mathbf{t})$ a *private data acquisition mechanism* where

1. $\hat{\theta} : \mathcal{X}^n \times \mathbb{R}_+^n \to \mathbb{R}$ is a (centrally or locally) differentially private estimator that maps acquired user data $\mathbf{x} = (x_i)_{i=1}^n$ and privacy losses $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n$ to an estimate $\hat{\theta}(\mathbf{x}, \boldsymbol{\varepsilon})$.[7]

2. For all $i \in \mathcal{N}$, $\varepsilon_i : \mathbb{R}_+^n \to \mathbb{R}_+$ is a function that maps privacy sensitivities $\mathbf{c}$ to a privacy loss for user i, $\varepsilon_i(\mathbf{c})$, with $\boldsymbol{\varepsilon}(.) = (\varepsilon_i(\cdot))_{i=1}^n$.

3. For all $i \in \mathcal{N}$, $t_i : \mathbb{R}_+^n \to \mathbb{R}_+$ is a function that maps privacy sensitivities $\mathbf{c}$ to a payment for user i, $t_i(\mathbf{c})$, with $\boldsymbol{t}(.) = (t_i(\cdot))_{i=1}^n$.

The above functions are assumed to be differentiable, with their derivatives being Riemann integrable. The minimax optimal estimators derived in Subsections 2.2.2 and 2.2.3 meet these assumptions.

We will study mechanisms with estimators that provide both central and local differential privacy guarantees (see Definitions 2.1 and 2.2) and use the notations $\hat{\theta}_{\text{central}}$ and $\hat{\theta}_{\text{local}}$ to highlight the distinction.

Each user that participates in a private data acquisition mechanism $(\hat{\theta}, \boldsymbol{\varepsilon}, \mathbf{t})$ shares her data with the platform leading to a lower estimation error. Users derive benefit from accessing this more accurate estimate (e.g., representing a new medical treatment that is of value for all users), but incur a privacy cost proportional to their privacy sensitivity c_i. Throughout, we find it more convenient to work with cost instead of utility. In particular, we model the user's cost from participation by the mean square error of the platform's estimate $\hat{\theta}$ and her privacy cost by $c_i\boldsymbol{\varepsilon}(\mathbf{c})$. Hence, the cost function of a user i with type c_i who reports c_i'

[7] We assume x_i is removed from $\mathbf{x}$ if user i does not participate in the mechanism.

is given by

$$\text{COST}(c_i', c_i; \boldsymbol{\varepsilon}, \mathbf{t}, \hat{\theta}) = \mathbb{E}_{\mathbf{c}_{-i}} \left[\text{MSE}(c_i', \mathbf{c}_{-i}; \boldsymbol{\varepsilon}, \hat{\theta}) + c_i \varepsilon_i(\mathbf{c}_{-i}, c_i') - t_i(\mathbf{c}_{-i}, c_i'))\right], \tag{2.14}$$

where the first term is the expected mean squared error of the estimator given by

$$\text{MSE}(c_i', \mathbf{c}_{-i}; \boldsymbol{\varepsilon}, \hat{\theta}) = \mathbb{E}_{\mathbf{x}} \left[|\hat{\theta}(\mathbf{x}, \boldsymbol{\varepsilon}) - \theta|^2\right].$$

Note that the privacy losses $\boldsymbol{\varepsilon}$ depends on reported privacy sensitivities $(c_i', \mathbf{c}_{-i})$, therefore we make the dependence of the mean square error on $(c_i', \mathbf{c}_{-i})$ explicit in our notation. The second term of (2.14) represents the privacy cost that the user incurs, and the third term is the payment that the user receives.

A user $i \in \mathcal{N}$ that does not participate in the mechanism does not compromise her privacy, but neither gets compensation nor enjoys the benefit of a reduced mean square error (arising from an estimate based on a collection of users' data). Therefore, the cost of a nonparticipating user becomes the mean square error of her "best" estimate of parameter θ based on her data alone, $\hat{\theta}(X_i)$, given by

$$\mathbb{E}_{X_i} \left[|\hat{\theta}(X_i) - \theta|^2\right] = \mathbb{E}_{X_i} \left[|X_i - \theta|^2\right] = \text{VAR}. \tag{2.15}$$

For a given $\hat{\theta}(\cdot)$, the goal of the platform is to minimize an objective function given by

$$\mathbb{E}_{\mathbf{c}} \left[\text{MSE}(\mathbf{c}, \boldsymbol{\varepsilon}, \hat{\theta}) + \sum_{i=1}^{n} t_i(\mathbf{c})\right],$$

over the choices of $\varepsilon_i(\cdot)$ and $t_i(\cdot)$ for all $i \in \mathcal{N}$. In the platform's objective, the first term is the mean square error of estimator $\hat{\theta}$ given reported types and resulting privacy losses $\boldsymbol{\varepsilon}$, i.e.,

$$\text{MSE}(\mathbf{c}, \boldsymbol{\varepsilon}, \hat{\theta}) = \mathbb{E}_{\mathbf{x}} \left[|\hat{\theta}(\mathbf{x}, \boldsymbol{\varepsilon}) - \theta|^2\right].$$

The second term is the total compensation the analyst provides to the users for truthfully reporting their privacy sensitivities and acquiring their data. In Section 2.8 we establish that, similar to the classical mechanism design setting, *revelation principle* holds and therefore the platform can focus on direct revelation mechanisms where individuals reporting their type truthfully is a (Bayesian Nash) equilibrium. *Incentive compatibility* constraints formalize this equilibrium outcome by imposing that user i has no incentive to misrepresent her type when others report truthfully (i.e., reporting her type correctly is a Bayesian Nash equilibrium of the underlying incomplete information game). Similarly, *individual rationality* constraints ensure that the platform does not make users worse off by participating in the mechanism. Together with these constraints, the mechanism designer's optimization problem can be written as

$$\min_{\boldsymbol{\varepsilon}(\cdot),\mathbf{t}(\cdot)} \quad \mathbb{E}_{\mathbf{c}}\left[\mathrm{MSE}(\mathbf{c},\boldsymbol{\varepsilon},\hat{\theta})+\sum_{i=1}^{n} t_i(\mathbf{c})\right] \tag{2.16}$$

$$\textsc{cost}(c_i,c_i;\boldsymbol{\varepsilon},\mathbf{t},\hat{\theta}) \le \textsc{cost}(c_i',c_i;\boldsymbol{\varepsilon},\mathbf{t},\hat{\theta}) \quad \text{for all } i\in\mathcal{N}, c_i, c_i' \tag{2.17}$$

$$\textsc{cost}(c_i,c_i;\boldsymbol{\varepsilon},\mathbf{t},\hat{\theta}) \le \textsc{var} \quad \text{for all } i\in\mathcal{N}, c_i, \tag{2.18}$$

where the constraints in (2.17) and (2.18) represent the incentive compatibility and the individual rationality constraints, respectively.[8]

2.3.1 Payment Identity

For a given estimator $\hat{\theta}$, the platform decision comprises the privacy loss functions $\boldsymbol{\varepsilon}(\cdot)$ and the payment functions $\mathbf{t}(\cdot)$. We next identify the payment as a function of the privacy loss functions. In this regard, we define the *interim* quantities

$$t_i(c_i)=\mathbb{E}_{\mathbf{c}_{-i}}\left[t(c_i,\mathbf{c}_{-i})\right] \text{ and } \varepsilon_i(c_i)=\mathbb{E}_{\mathbf{c}_{-i}}\left[\varepsilon_i(c_i,\mathbf{c}_{-i})\right] \text{ for all } i\in\mathcal{N}, c_i.$$

Proposition 2.1. *For a given estimator $\hat{\theta}:\mathcal{X}^n\times\mathbb{R}_+^n\to\mathbb{R}$, a central or local privacy data acquisition mechanism $(\hat{\theta},\boldsymbol{\varepsilon},\mathbf{t})$ satisfies incentive compatibility (2.17) and individual*

[8] We assume that the variance and the payments both appear with the same coefficient in the platform's objective. Our analysis readily extends to a setting with differing coefficients.

rationality (2.18) *if and only if*

$$t_i(c_i) = \mathbb{E}_{\mathbf{c}_{-i}}\left[\mathrm{MSE}(\mathbf{c}, \boldsymbol{\varepsilon}, \hat{\theta})\right] - \mathrm{VAR} + c_i\varepsilon_i(c_i) + \int_{z=c_i}^{\infty} \varepsilon_i(z)dz + d_i, \tag{2.19}$$

for some constant $d_i \geq 0$*, and* $\varepsilon_i(z)$ *is non-increasing (or equivalently, is weakly decreasing) in* z *for all* $i \in \mathcal{N}$*.*

Proposition 2.1 determines the payment in terms of the privacy loss functions. This proposition is closely related to the payment identity in classical mechanism design (see Myerson [1981]) and in particular single-dimensional mechanism design. In particular, by evaluating the first order condition corresponding to the incentive compatibility constraint (2.17), we establish that this constraint holds if and only if

$$t_i(c_i) =$$
$$t_i(0) + \mathbb{E}_{\mathbf{c}_{-i}}\left[\mathrm{MSE}(c_i, \mathbf{c}_{-i}, \boldsymbol{\varepsilon}, \hat{\theta})\right] - \mathbb{E}_{\mathbf{c}_{-i}}\left[\mathrm{MSE}(0, \mathbf{c}_{-i}, \boldsymbol{\varepsilon}, \hat{\theta})\right] + c_i\varepsilon_i(c_i) - \int_{z=0}^{c_i} \varepsilon_i(z)dz$$

and $\varepsilon_i(z)$ is weakly decreasing in z. We then use the above expression in the individual rationality constraint (2.18) and prove

$$t_i(0) = \mathbb{E}_{\mathbf{c}_{-i}}\left[\mathrm{MSE}(0, \mathbf{c}_{-i}, \boldsymbol{\varepsilon}, \hat{\theta})\right] - \mathrm{VAR} + \int_{z=0}^{\infty} \varepsilon_i(z)dz + d_i$$

for some $d_i \geq 0$. Equation (2.19) follows from the previous two expressions. It is worth noting that, for a central or local privacy data acquisition mechanism $(\hat{\theta}, \boldsymbol{\varepsilon}, \mathbf{t})$ to be optimal, we must have $d_i = 0$ in (2.19).

In concluding this subsection, we want to highlight that our benchmark for individual rationality (given in (2.18)) is that the users will not benefit from the platform's estimate if they do not participate. If we consider an alternative benchmark in which the users benefit from the platform's estimator even if they do not participate, then the payments increase, and the platform's cost decreases. However, as we show in Section 2.8, our characterization of the optimal privacy levels that will follow remains unchanged.

2.3.2 Reformulating the Platform's Problem

We next use Proposition 2.1 to reformulate the platform's problem in terms of only the privacy loss functions and the *virtual costs*, defined as

$$\psi_i(c) = c + \frac{F_i(c)}{f_i(c)}, \qquad \text{for all } i \in \mathcal{N}, c \in \text{supp}(f),$$

where the support of $f(\cdot)$ is defined as $\text{supp}(f) = \{c \in \mathbb{R}_+ \, : \, f(c) \neq 0\}$.

Proposition 2.2. *For a given estimator $\hat{\theta} : \mathcal{X}^n \times \mathbb{R}_+^n \to \mathbb{R}$, the optimal privacy loss in the central or local privacy data acquisition mechanism is the solution of*

$$\min_{\boldsymbol{\varepsilon}(\cdot)} \quad \mathbb{E}_{\mathbf{c}} \left[(n+1)\text{MSE}(\mathbf{c}, \boldsymbol{\varepsilon}, \hat{\theta}) + \sum_{i=1}^{n} \varepsilon_i(\mathbf{c}) \psi_i(c_i) \right] - n\text{VAR} \tag{2.20}$$

$$\varepsilon_i(z) = \mathbb{E}_{\mathbf{c}_{-i}} \left[\varepsilon_i(z, \mathbf{c}_{-i}) \right] \text{ \textit{is weakly decreasing in} } z \text{ \textit{for all} } i \in \mathcal{N}. \tag{2.21}$$

Proposition 2.2 is an analogue of Myerson's reduction of mechanism design to virtual welfare maximization, adapted to our data acquisition setting (Myerson [1981]), and it follows from invoking Proposition 2.1.

2.4 Privacy-Concerned Data Acquisition in the Central Privacy Setting

In the rest of the chapter, we will focus on linear estimators, which we showed to be near optimal for given privacy loss levels. Our goal in this section is to address the analyst's mechanism design problem in the central privacy setting for the near optimal choice of estimator found in Section 2.2.2:

$$\hat{\theta}_{\text{central}}(x_1, \ldots, x_n) := \sum_{i=1}^{n} w_i(\mathbf{c}) x_i + \text{Laplace}\left(\frac{1}{\eta}\right) \tag{2.22}$$

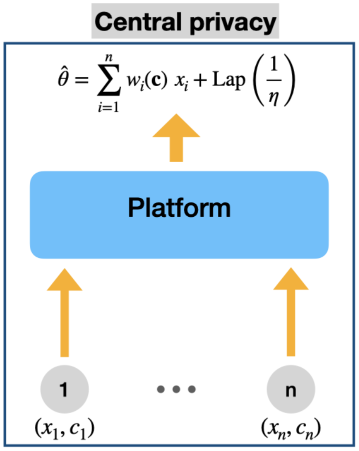

Central privacy
$\hat{\theta} = \sum_{i=1}^{n} w_i(\mathbf{c})\, x_i + \mathrm{Lap}\left(\frac{1}{\eta}\right)$
Platform
1
n
(x_1, c_1)
(x_n, c_n)

Figure 2-2: The interaction between the users and the platform in the central privacy setting.

such that

$$\sum_{i=1}^{n} w_i(\mathbf{c}) = 1, \text{ and } \eta w_i(\mathbf{c}) \leq \varepsilon_i(\mathbf{c}) \text{ for all } i \in \mathcal{N}.$$

Figure 2-2 depicts the interaction between the platform and the users in the central privacy setting and when the platform is using the above (near) optimal choice of estimator.

2.4.1 Characterization of the Optimal Central Privacy Data Acquisition Mechanism

Our next theorem characterizes the optimal privacy loss function in the central privacy setting under the following assumption.

Assumption 2.1. For any $i \in \mathcal{N}$, the *virtual cost* $\psi_i(c) = c + \frac{F_i(c)}{f_i(c)}$ is increasing in c.

Assumption 2.1 is standard in mechanism design and in particular for procurement auctions which is closer to our setting. It resembles the regularity condition adopted in mechanism design literature and holds for a variety of distributions and in particular for distributions with log-concave density functions such as uniform, exponential, and normal (see, e.g., Rosling [2002]).

Theorem 2.3. *Suppose Assumption 2.1 holds. For any reported vector of privacy sensitivities $(c_1, \ldots, c_n)$, the optimal privacy loss level in the central privacy data acquisition* *mechanism is* $\varepsilon_i^*(\mathbf{c}) = y_i^*$ *for* $i \in \mathcal{N}$*, where* $(y_1^*, \ldots, y_n^*)$ *is the optimal solution of*

$$\begin{aligned} \min_{\mathbf{y}} \frac{n+1}{\left(\sum_{j=1}^n y_j\right)^2} & \left(2 + \sum_{i=1}^n y_i^2 \text{ VAR}\right) + \sum_{i=1}^n \psi_i(c_i) y_i \\ & \textit{s.t. } y_i \geq 0, \textit{ for all } i \in \mathcal{N}. \end{aligned} \tag{2.23}$$

Moreover, for all $i \in \mathcal{N}$ *the weight of user* i*'s data in the platform's estimator is* $\frac{y_i^*}{\sum_{j=1}^n y_j^*}$.

Before providing the proof idea of this theorem, let us highlight the difference between our characterization and that of classic mechanism design (e.g., Myerson [1981] or the procurement counterpart). In classic mechanism design, the designer's problem becomes a linear optimization. In our setting, however, the designer's problem is a non-linear and non-convex optimization. An important implication of this distinction is that, contrary to classic mechanism design where the optimal mechanism typically involves a threshold rule, in this case, the optimal privacy loss level is not a threshold strategy. Instead, it is a continuous function that depends on the privacy sensitivity.

To prove Theorem 2.3, we first note that the mean square error of the linear estimator $\hat{\theta}_{\text{central}}$ in (2.22) is given by

$$\text{MSE}(\boldsymbol{c}, \boldsymbol{\varepsilon}, \hat{\theta}_{\text{central}}) = \frac{2}{\eta^2} + \sum_{i=1}^n w_i(\mathbf{c})^2 \text{VAR}.$$

We next plug the above characterization into Proposition 2.2, and note that, if we drop the constraint (2.21) (which is $\varepsilon_i(c_i) = \mathbb{E}_{\mathbf{c}_{-i}}\left[\varepsilon_i(c_i, \mathbf{c}_{-i})\right]$ being weakly decreasing in c_i), it suffices to solve the following pointwise optimization problem

$$\begin{aligned} \min_{\boldsymbol{\varepsilon}(\mathbf{c}), \mathbf{w}(\mathbf{c}), \eta} \frac{2(n+1)}{\eta^2} & + \sum_{i=1}^n (n+1) \text{VAR } w_i(\mathbf{c})^2 + \sum_{i=1}^n \psi_i(c_i) \varepsilon_i(\mathbf{c}) \\ & \text{s.t. } \boldsymbol{\varepsilon}_i(\mathbf{c}) \geq 0, \text{ for all } i \in \mathcal{N} \\ & \sum_{i=1}^n w_i(\mathbf{c}) = 1 \\ & \eta w_i(\mathbf{c}) \leq \varepsilon_i(\mathbf{c}) \text{ for all } i \in \mathcal{N}. \end{aligned} \tag{2.24}$$

We next focus on solving the above problem. To do so, we establish that the constraints in (2.24) are binding in the optimal solution and therefore this problem is equivalent to the optimization problem (2.23) given in Theorem 2.3 statement. Finally, we conclude the proof by showing that the solution to this pointwise optimization satisfies the aforementioned constraint (2.21) that we dropped. More specifically, we show that the i-th component of the optimal solution of (2.23), under Assumption 2.1, is weakly decreasing in c_i.

The characterization of Theorem 2.3 leads to the following observation:

Corollary 2.1. *Suppose Assumption 2.1 holds. For any reported vector of privacy sensitivities* $(c_1, \dots, c_n)$*, in the optimal central data acquisition mechanism, we have* $\varepsilon_i^*(\mathbf{c}) \geq \varepsilon_j^*(\mathbf{c})$ *for all* $i, j \in \mathcal{N}$ *such that* $\psi_i(c_i) < \psi_j(c_j)$*.*

This corollary states the intuitive fact that in the optimal central data acquisition mechanism, users with higher virtual privacy sensitivities have lower (i.e., better) privacy loss levels.

2.4.2 Computing the Optimal Privacy Loss Function

The implementation of the optimal central privacy data acquisition mechanism involves solving problem (2.23), which is a non-convex program. We next develop a score-based method that efficiently solves problem (2.23).

To guide the analysis, without loss of generality, we assume $\psi_1(c_1) \leq \cdots \leq \psi_n(c_n)$, and define $\psi_{n+1}(c_{n+1}) = \infty$. We first rewrite problem (2.23) by introducing a variable for the summation of y_i's as follows

$$\min_{S \geq 0} \min_{\mathbf{y}} \frac{n+1}{S^2} \left(2 + \sum_{i=1}^{n} y_i^2 \ \mathrm{VAR} \right) + \sum_{i=1}^{n} \psi_i(c_i) y_i \tag{2.25}$$

$$\text{s.t.} \ \sum_{i=1}^{n} y_i = S \tag{2.26}$$

$$y_i \geq 0, \ \text{for all } i \in \mathcal{N}.$$

For a given S, the optimization over $\mathbf{y}$ is a convex program. Using Karush–Kuhn–Tucker

Algorithm 1: Computing the optimal privacy loss in the central setting

```
Input: The vector of privacy sensitivities $(c_1, \dots, c_n)$
Sort the terms $\{\psi_i(c_i)\}_i$. Without loss of generality, let us assume
    $$\psi_1(c_1) \leq \cdots \leq \psi_n(c_n);$$
Let $B_0 = \tilde{B}_0 = 0$;
for $i = 1$ to $n$ do
    Let
        $$A_i = \frac{i}{2(n+1)\text{VAR}}, \quad B_i = B_{i-1} + \frac{\psi_i(c_i)}{2(n+1)}, \quad \tilde{B}_i = \tilde{B}_{i-1} + \frac{\psi_i(c_i)^2}{2(n+1)\text{VAR}};$$
    Let
        $$OBJ_i(\lambda) = 2(n+1)\left(\lambda A_i - B_i\right)^2 + \frac{A_i \lambda^2 - \tilde{B}_i}{2\left(\lambda A_i - B_i\right)^2};$$
    Let
        $$\lambda_i^* = \arg\min_{\lambda} \; OBJ_i(\lambda) \text{ s.t. } \psi_i(c_i) \leq \lambda \leq \psi_{i+1}(c_{i+1})$$
        with the convention $\psi_{n+1}(c_{n+1}) = \infty$;
end
Let $i^* = \arg\max_i \text{OBJ}_i(\lambda_i^*)$;
Output: The optimal solution is given by
    $$y_j^* = 0 \text{ for } j > i^* \text{ and } y_j^* = \frac{\lambda_{i^*}^* - \psi_i(c_i)}{2(n+1)\text{VAR}\left(\lambda_{i^*}^* A_{i^*} - B_{i^*}\right)^2} \text{ for } j \leq i^*.$$
```

(KKT) condition (see e.g. Bertsekas [1997]), the solution to this optimization problem is[9]

$$(y_1, \dots, y_n) = \left(\left(\frac{(\lambda - \psi_1(c_1))S^2}{2(n+1)\text{VAR}} \right)^+, \dots, \left(\frac{(\lambda - \psi_n(c_n))S^2}{2(n+1)\text{VAR}} \right)^+ \right), \tag{2.27}$$

where λ is such that

$$\sum_{i=1}^{n} \left(\frac{(\lambda - \psi_1(c_1))S^2}{2(n+1)\text{VAR}} \right)^+ = S. \tag{2.28}$$

Using this relation, we can write S as a function of λ which allows us to rewrite the minimiza-

[9] For any $x \in \mathbb{R}$, we let x^+ denote $\max\{x, 0\}$.

tion problem (2.25) over $\lambda \in [\psi_1(c_1), \infty]$ rather than S. We solve this resulting minimization problem by finding the optimal λ in the interval $[\psi_i(c_i), \psi_{i+1}(c_i)]$ for all $i = 1, \ldots, n$ and then selecting the λ with the lowest objective function. Algorithm 1 summarizes the above procedure and the following proposition states the formal result:

Proposition 2.3. *For any vector of reported privacy sensitivities $(c_1, \ldots, c_n)$, Algorithm 1 finds the optimal privacy loss levels in the optimal central data acquisition mechanism (i.e., the solution of problem (2.23)) in time $\mathcal{O}(n \log n)$.*

Algorithm 1 needs sorting n elements which requires time $\mathcal{O}(n \log n)$. We also prove that each iteration of the for loop can be done in time $\mathcal{O}(1)$, establishing that the overall running time of Algorithm 1 is $\mathcal{O}(n \log n)$.

As depicted in Algorithm 1, the virtual cost of each user determines whether the data of that user is included in the final estimator of the platform. In particular, there exists a threshold $\bar{\psi}$ such that only the data of users whose virtual cost $\psi_i(c_i)$ is below $\bar{\psi}$ are used in the estimator of the platform. This feature of the optimal data acquisition mechanism is reminiscent of the classical optimal mechanism of Myerson [1981] with one important difference though: unlike the classical mechanism design in which the item gets allocated to a single user, here the data of multiple users are being used and that the weight of each user's data depends on her virtual cost and the entire profile of virtual costs.

2.5 Privacy-Concerned Data Acquisition in the Local Privacy Setting

In the local differential privacy setting, each user i shares a differentially private version of her data with the platform who then combines them to form an estimator for the underlying parameter. In particular, first the user reports her privacy sensitivity that determines both the payment to the user and the variance of the noise to be added to the user's data. The platform then collects the "transformed data" of the users and combines them to form an estimation of the underlying parameter. The difference between this setting and the central privacy setting is that the data that each user shares with the platform is already

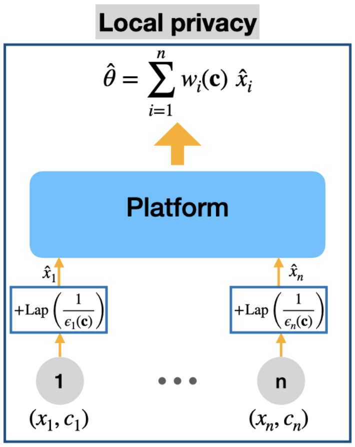

Local privacy
$\hat{\theta} = \sum_{i=1}^{n} w_i(\mathbf{c})\ \hat{x}_i$
Platform
$\hat{x}_1$
$\hat{x}_n$
$+\mathrm{Lap}\left(\frac{1}{\varepsilon_1(\mathbf{c})}\right)$
$+\mathrm{Lap}\left(\frac{1}{\varepsilon_n(\mathbf{c})}\right)$
1
n
(x_1, c_1)
(x_n, c_n)

Figure 2-3: The interaction between the users and the platform in the local privacy setting.

differentially private. As a result, the final estimator of the platform is also differentially private (composition property of differential privacy). Therefore, the platform does not need to transform its estimator to make it differentially private and her only estimation task is finding an unbiased estimator with minimum bias. Our goal in this section is to address the analyst's mechanism design problem in the local privacy setting for the optimal choice of estimator found in Section 2.2.3:

$$\hat{\theta} = \sum_{i=1}^{n} w_i \hat{x}_i, \qquad \text{where } \hat{x}_i = x_i + \text{Laplace}(1/\varepsilon_i) \text{ for all } i \in \mathcal{N}. \tag{2.29}$$

Figure 2-3 depicts the interaction between the users and the platform in the local privacy setting.

2.5.1 Characterization of the Optimal Local Privacy Data Acquisition Mechanism

Our next theorem characterizes the optimal mechanism in the local privacy setting under Assumption 2.1.

Theorem 2.4. *Suppose Assumption 2.1 holds. For any reported vector of privacy sensitivities $(c_1, \ldots, c_n)$, the optimal privacy loss level in the local privacy data acquisition is*

$\varepsilon_i^*(\mathbf{c}) = y_i^*$ for $i \in \mathcal{N}$, where $(y_1^*, \dots, y_n^*)$ is the optimal solution of

$$\min_{\mathbf{y}} \quad \frac{n+1}{\sum_{i=1}^n \frac{1}{\text{VAR}+\frac{2}{y_i^2}}} + \sum_{i=1}^n \psi_i(c_i) y_i \tag{2.30}$$
$$\textit{s.t. } y_i \geq 0 \textit{ for all } i \in \mathcal{N}.$$

Moreover, for all $i \in \mathcal{N}$, the weight of user i's data in the platform estimator is proportional to

$$\frac{1}{\text{VAR} + \frac{2}{y_i^{*2}}}.$$

To prove Theorem 2.4, we first note that, for a given vector of privacy sensitivities $\mathbf{c}$, the mean square of the linear estimator given in (2.29) is

$$\text{MSE}(\boldsymbol{c}, \boldsymbol{\varepsilon}, \hat{\theta}_{\text{local}}) = \sum_{i=1}^n w_i(\mathbf{c})^2 \left(\text{VAR} + \frac{2}{\varepsilon_i(\mathbf{c})^2} \right).$$

Similar to the proof of Theorem 2.3, we drop the constraint (2.21), and consider the following pointwise optimization:

$$\begin{aligned} \min_{w_i(\mathbf{c}), \varepsilon_i(\mathbf{c})} \quad & \sum_{i=1}^n w_i(\mathbf{c})^2 \left((n+1)\text{VAR} + \frac{2(n+1)}{\varepsilon_i(\mathbf{c})^2} \right) + \psi_i(c_i)\varepsilon_i(\mathbf{c}) \\ & \sum_{i=1}^n w_i(\mathbf{c}) = 1 \\ & w_i(\mathbf{c}) \geq 0, \varepsilon_i(\mathbf{c}) \geq 0 \text{ for all } i \in \mathcal{N}. \end{aligned} \tag{2.31}$$

We next note that the optimization over weights $(w_i(\mathbf{c}))_{i=1}^n$ subject to (2.31) is a quadratic optimization problem that, for a given $(\varepsilon_i(\mathbf{c}))_{i=1}^n$, and we can solve explicitly. In particular, $w_i(\mathbf{c})$ is proportional to

$$\frac{1}{\text{VAR} + \frac{2}{\varepsilon_i(\mathbf{c})^2}} \text{ for all } i \in \mathcal{N}.$$

Plugging in these weights, the rest of the proof follows similar to the proof of Theorem 2.3.

The characterization of Theorem 2.4 leads to the following observation:

Corollary 2.2. *Suppose Assumption 2.1 holds. For any reported vector of privacy sensitivities $(c_1, \dots, c_n)$, in the optimal local data acquisition mechanism, we have $\varepsilon_i^*(\mathbf{c}) \geq \varepsilon_j^*(\mathbf{c})$ for all $i, j \in \mathcal{N}$ such that $\psi_i(c_i) < \psi_j(c_j)$.*

This corollary, which is analogous to Corollary 2.1, states a similar fact in the local setting: in the optimal local data acquisition mechanism, users with higher virtual privacy sensitivity have lower privacy loss levels (better privacy guarantees).

2.5.2 Computing the Optimal Privacy Loss Function

The implementation of the optimal mechanism involves solving problem (2.30), which is a non-convex problem. Thus, using algorithms such as gradient descent might lead to finding a saddle point or a local minima rather than the global minimum. However, in what follows, we present an algorithm that takes advantage of the problem's structure and establishes that finding the global minima admits a Polynomial Time Approximation Scheme (PTAS).

To guide the analysis, without loss of generality, we assume $\psi_1(c_1) \leq \cdots \leq \psi_n(c_n)$. Letting $(y_1^*, \dots, y_n^*)$ be the optimal solution of (2.30), the first order condition implies that there exists $\lambda \in \mathbb{R}_+$ such that

$$\frac{4y_i^*}{\left(2 + \text{VAR} y_i^{*2}\right)^2} = \frac{\psi_i(c_i)}{n+1}\lambda^2, \text{ for all } y_i^* \neq 0.$$

We first prove that if there exists $i \in \{1, \dots, n\}$ such that $y_i^* = 0$, then we have $y_j^* = 0$ for $j > i$. We also establish that, for such i, we have[10]

$$y_j^* = y_j^{(h)}(\lambda) \text{ for } j \leq i-1 \text{ and } y_i^* \in \{y_i^{(l)}(\lambda), y_i^{(h)}(\lambda)\}, \tag{2.32}$$

where for any $\lambda \in \mathbb{R}_+$, $y_i^{(l)}(\lambda)$ and $y_i^{(h)}(\lambda)$ are the smallest and the largest solutions of

$$\frac{4z}{\left(\text{VAR} z^2 + 2\right)^2} = \frac{\psi_i(c_i)}{n+1}\lambda^2.$$

[10] Equation (2.32) holds when $\psi_i(c_i) > \psi_{i-1}(c_{i-1})$. In the proof of Proposition 2.4, presented in Section 2.8, we provide the detail for the case $\psi_i(c_i) = \psi_{i-1}(c_{i-1})$ as well.

```
Algorithm 2: Computing the optimal privacy loss in the local setting
Input: The vector of privacy sensitivities $(c_1, \dots, c_n)$ and $\epsilon \in \mathbb{R}_+$
Sort the terms $\psi_i(c_i)$, and without loss of generality, let us assume
        $\psi_1(c_1) \le \cdots \le \psi_n(c_n).$
for $i = 1$ to $n$ do
    Let $\Delta$ be the maximum Lipschitz parameter of functions $\frac{n+1}{\lambda}$, $y_j^{((h))}(\lambda)$, and
    $y_j^{(l)}(\lambda)$ over $\lambda \in [\underline{y}_i, \bar{y}_i]$;
    Find
        $\lambda_i \in \mathrm{Grid}(i, \frac{\epsilon}{\Delta}) = \left\{ k\frac{\epsilon}{\Delta} : \ k = \lfloor \underline{y}_i \frac{\Delta}{\epsilon} \rfloor, \dots, \lceil \bar{y}_i \frac{\Delta}{\epsilon} \rceil \right\}$
    as the solution of
        $\min_{\lambda \in \mathrm{Grid}(i, \frac{\epsilon}{\Delta})} \min \left\{ \frac{n+1}{\lambda} + \sum_{j=1}^{i} \psi_j(c_j) y_j^{(h)}(\lambda), \frac{n+1}{\lambda} + \sum_{j=1}^{i-1} \psi_j(c_j) y_j^{(h)}(\lambda) + \psi_j(c_j) y_j^{(l)}(\lambda) \right\};$
    Let $\mathrm{OBJ}_i$ be the objective evaluated at $y_j^{(i)} = y_j^{(h)}(\lambda_i)$ for $j \le i-1$, $y_j^{(i)} = 0$ for
    $j \ge i+1$, and $y_i^{(i)} = y_i^{(h)}(\lambda_i)$ if the optimal solution of the above optimization is
    the first term and $y_i^{(i)} = y_i^{(l)}(\lambda_i)$, otherwise;
end
Output: Letting $i^* = \arg\min_{i \in \mathcal{N}} \mathrm{OBJ}_i$, the approximate solution is $(y_1^{(i^*)}, \dots, y_n^{(i^*)})$.
```

Therefore, the platform's problem becomes finding the optimal i and the optimal λ. We search for the optimal i by considering all elements of $\mathcal{N}$. We also search over the optimal λ by considering a grid search. To form a grid for the possible optimal values of λ, we establish the following upper bound and lower bound on the optimal λ:

$$\bar{y}_i = y^{(h)}\left(\left(\frac{(n+1)3\sqrt{3}}{\psi_{i-1}(c_{i-1})8\sqrt{2\mathrm{VAR}}} \right)^{1/2} \right) \quad \text{and } \underline{y}_i = \frac{n}{\mathrm{VAR} + \left(\frac{\sqrt{2}n\left(\sum_{j=1}^{n} \psi_j(c_j)\right)}{(n+1)} \right)^{2/3}}.$$

Algorithm 2 summarizes the above procedure and the following proposition states the formal result:

Proposition 2.4. *For any vector of reported privacy sensitivities* $(c_1, \dots, c_n)$ *and* $\epsilon > 0$, *Algorithm 2 finds privacy loss levels for the local data acquisition mechanism whose cost* *(i.e., the platform's objective) is at most* $1+\epsilon$ *of the optimal cost in time* $\text{poly}(n, \frac{1}{\epsilon})$.[11]

Notice that the approximation factor in Proposition 2.4 depends on the underlying parameters and therefore we have a Polynomial Time Approximation Scheme (PTAS) for finding the optimal privacy loss levels. Also, the output of Algorithm 2 satisfies $y_i \geq y_j$ when $\psi_i(c_i) \leq \psi_j(c_j)$ and therefore, as shown in Proposition 2.1, is implementable.

We conclude this section by highlighting that computing the payment function (2.19) necessitates integrating over the privacy loss levels $\varepsilon_i(\cdot)$, which does not have an explicit characterization in our setting. However, in Section 2.8.1, we demonstrate that this integral (and, therefore, the payment function) can be approximated to achieve any desired level of accuracy ϵ. Consequently, this approximation yields an ϵ-approximate incentive compatibility (ϵ-IC) mechanism, where the incentive compatibility constraint (2.17) is violated by at most ϵ. The concept of ϵ-IC has been previously employed in the literature (see, e.g., the literature review of Balseiro et al. [2022]). In Section 2.8.1, we provide the formal definition of ϵ-IC and outline how the payment function (2.19) can be approximated to achieve ϵ-IC.

2.6 Data Acquisition With Central Versus Local Differential Privacy

In this section, we compare the performance of the optimal data acquisition mechanism in the central and local privacy settings.

First, let us consider a case in which there is no restriction on the estimator, i.e., the estimator does not need to be a linear combination of users' data with a Laplace mechanism. In this case, finding the optimal value of platform's objective function in the central (local) differential privacy setting is equivalent to solving problem (2.16) for all centrally (locally) differentially private estimators. Note that, as stated in Section 2.2, any $\boldsymbol{\varepsilon}$-locally differentially private estimator is $\boldsymbol{\varepsilon}$-centrally differentially private as well. As a result, the platform's optimal objective in the central privacy setting is always weakly smaller than her optimal objective in the local privacy setting. This is because the minimization problem in

[11] $\text{poly}(\cdot)$ denotes a function that is polynomial in its inputs.

the central setting is solved over a weakly larger set of estimators. Next, we show that the same result holds even if we restrict our focus to the class of linear estimators.

Proposition 2.5. *Let $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n$. For any $\boldsymbol{\varepsilon}$-locally differentially private linear estimator:*

$$\hat{\theta}_{\text{local}} = \sum_{i=1}^{n} w_i \hat{x}_i \quad \hat{x}_i = x_i + \mathit{Laplace}(1/\varepsilon_i),$$

with $\sum_{i=1}^n w_i = 1$, there exists a $\boldsymbol{\varepsilon}$-differentially private linear estimator $\hat{\theta}_{\text{central}}$ such that

$$\mathbb{E}[|\hat{\theta}_{\text{central}} - \theta|^2] \leq \mathbb{E}[|\hat{\theta}_{\text{local}} - \theta|^2]. \tag{2.33}$$

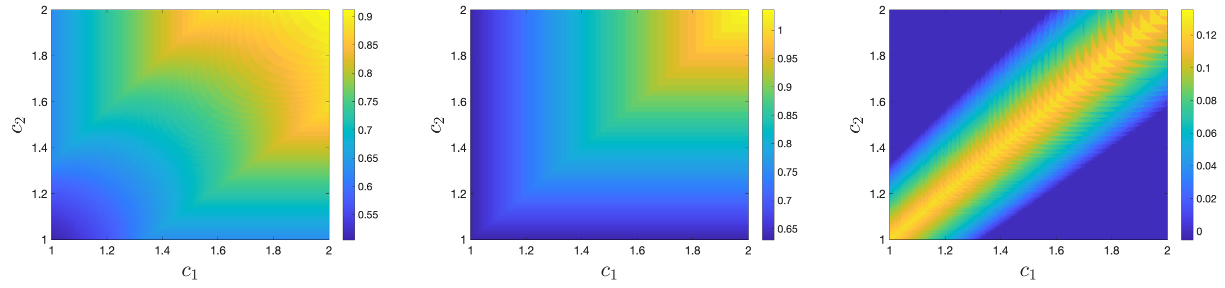
This result consequently implies that, for any locally differentially private linear estimator, there exists a centrally differentially private linear estimator which delivers the same privacy loss levels with (weakly) lower estimation error. By keeping privacy loss levels unchanged, the privacy cost and the payments will also remain unchanged in the platform's objective. Hence, Proposition 2.5 implies that the platform's optimal objective function under central differential privacy setting is weakly smaller than her optimal objective function under local differential privacy. The following corollary formally states this observation.

Corollary 2.3. *For any reported vector of privacy sensitivities $(c_1, \ldots, c_n)$, the optimal solution of the local privacy optimization problem (2.30) is not smaller than the optimal solution of the central privacy optimization problem (2.23).*

2.6.1 An Illustrative Example

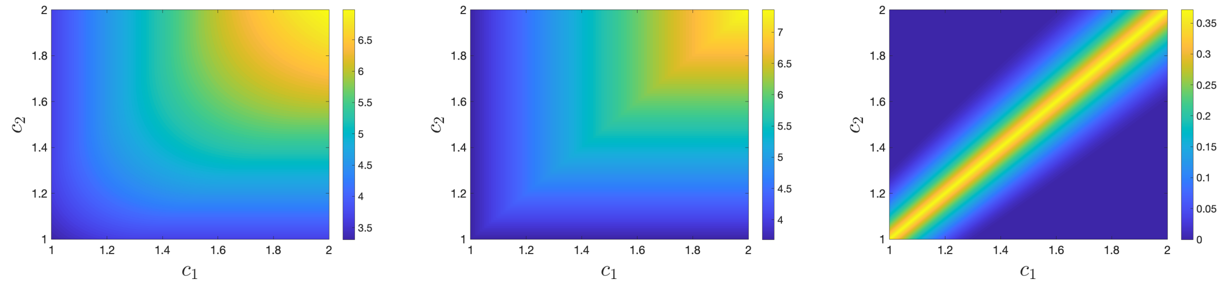
We next illustrate the difference between the performance of our proposed central privacy mechanism and our proposed local privacy mechanism in the context of a simple example. We consider two users with uniform privacy sensitivities drawn from $[1, 2]$ (so that the virtual privacy sensitivity of user $i \in \{1, 2\}$ becomes $2c_i - 1$ for $c_i \in [1, 2]$) and $\text{VAR} = 1/4$.

Figures 2-4a, 2-4b, and 2-4c depict the variance of the estimator for the central setting, the local setting, and their difference, respectively for all pairs of privacy sensitivities (c_1, c_2). We observe that the variance in the central setting is always weakly larger than the variance



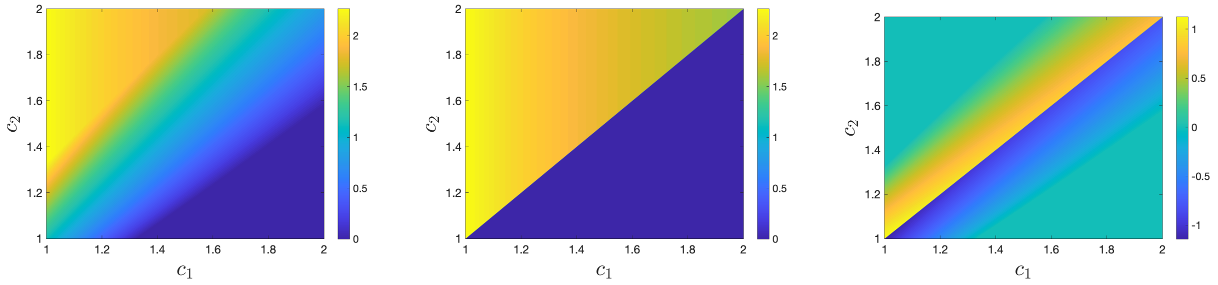
(a) (b) (c)

Figure 2-4: (a) the variance in the central setting, (b) the variance in the local setting, and (c) the variance in the local minus the central setting for two users with VAR $= 1/4$ and uniform privacy sensitivities over $[1, 2]$ as a function of the privacy sensitivities (c_1, c_2).



(a) (b) (c)

Figure 2-5: (a) the analyst's objective in the central setting, (b) the analyst's objective in the local setting, and (c) the analyst's objective in the local minus the central setting for two users with VAR $= 1/4$ and uniform privacy sensitivities over $[1, 2]$ as a function of the privacy sensitivities (c_1, c_2).



(a)

(b)

(c)

Figure 2-6: (a) user 1's optimal privacy loss in the central setting, (b) user 1's optimal privacy loss in the local setting, and (c) the difference between user 1's optimal privacy loss in the local setting and the central setting as a function of (c_1, c_2). Here, we have $\text{VAR} = 1/4$ and the privacy sensitivities are uniform over $[1, 2]$.

in the local setting. This is because the local setting provides a stronger privacy guarantee and will hurt the variance of the final estimator. We also observe that when there is a large discrepancy between the two privacy sensitivities, the variance of the central and the local settings are equal. This is because the platform obtains all of its data from only one of the users and therefore central and local setting become identical. Further, when the two costs are very close to each (i.e., $c_1 \approx c_2$), the platform's weight for the data of each of the users in the estimator become close to each other. This implies that the variance of the local and the central setting become very close to each other.

Figures 2-5a, 2-5b, and 2-5c depict the platform's objective for the central setting, the local setting, and their difference, respectively for all pairs of privacy sensitivities (c_1, c_2). We observe that the cost in the central setting is always weakly smaller than the cost in the local setting. This is again because the local setting provides a stronger privacy guarantee and will hurt the platform's objective. When there is a large discrepancy between the two privacy sensitivities (i.e. $|c_1 - c_2| \approx 1$), the objective of the central and the local settings are equal. This is because the platform obtains all of its data from only one of the users and therefore central and local setting become identical.

Figures 2-6a and 2-6b depict the optimal allocation of user 1 in the central and local settings, and Figure 2-6c depicts the optimal allocation of user 1 in the local setting minus the central setting. We observe that, the privacy loss level of a user in the local setting can be lower (i.e., better privacy) than the central setting. We next provide a formal statement for this observation in the context of an example.

2.6.2 Optimal Privacy Loss Levels in the Central Versus Local Setting

Here, we illustrate that the privacy loss level of a user in the optimal local data acquisition mechanism can be smaller than the central setting, i.e., the optimal mechanism in the local setting may provide strictly better privacy guarantees to a user compared to the central setting. To simplify the notation, we work with the virtual cost rather than the privacy sensitivity. Note that this is without loss of generality as we do not pose any assumption on the virtual cost (other than Assumption 2.1).

Proposition 2.6. *Assume the virtual costs of users $1, \cdots, n-1$ are all equal to ψ_1. We also denote the virtual cost of user n by ψ_n. Then, for*[12]

$$\psi_n \in \left[\psi_1 + \Theta\left(\frac{1}{n^{2/3}}\right), \psi_1 + \Theta\left(\frac{1}{n^{1/3}}\right)\right],$$

the optimal privacy loss level of user n in the local setting is zero, while her optimal privacy level in the central setting is non-zero.

Proposition 2.6 follows by comparing the optimal solutions of the non-convex programs (2.23) and (2.30). Note that, since the local privacy is a more stringent requirement, one may expect that, in the optimal local mechanism, the delivered privacy guarantees to users are worse (higher privacy loss levels) compared to the central setting. This proposition shows that may not be the case. To gain an intuition, note that, as the privacy sensitivity of user n, i.e., ψ_n, increases, her privacy loss level, in both central and local settings, goes to zero. Recall that in the central estimator (2.2.2), the privacy loss level of user n is denoted as $w_n \eta$. To achieve a near-zero privacy loss level, we have two possibilities: either η approaches zero (which results in a larger estimation error due to the Laplace noise variance being $2/\eta^2$), or w_n must approach zero. Similarly, in the local estimator, user n's privacy loss level is

[12] $f(n) = \Theta(g(n))$ means that there exist n_0 and constants c_1, c_2 such that for $n \geq n_0$, we have $c_1 g(n) \leq f(n) \leq c_2 g(n)$.

represented by ε_n, and if this term approaches zero, w_n must also approach zero; otherwise, the estimation error will be considerably large. In summary, to deliver such small privacy loss, in the optimal central and local mechanisms, the platform must allocate zero weight to user n's data. Otherwise, the added noise in the platform's estimator makes the estimation error unbounded.

In the local setting, each user first maps her data to a private version and then shares it with the platform, meaning that, by definition, the privacy loss level of a user only depends on this mapping and not the platform's estimator. This in turn implies that the reallocation of the weights will not impact the privacy loss levels delivered to other users and hence will not change their compensations. In the central setting, however, decreasing user n's weight in the optimal estimator, increases the allocated weight to other users' data (because the sum of the wights add up to one). This in turn increases their allocated privacy loss levels and hence their compensation. Therefore, in the central setting the platform is more reluctant to give up on user n's data and increase other users' allocated weights, which is what we establish in Proposition 2.6.

2.7 Conclusion and Discussions

We study the design of mechanisms for acquiring data from users with privacy concerns who also benefit from a lower estimation error. We consider two architectures: (i) central privacy setting in which users share their data with the platform, incur some privacy loss and get compensated for their loss. The platform then combines the data of users and outputs an estimator that guarantees the promised heterogeneous privacy loss to each user; and (ii) local privacy setting in which users share a differentially private version of their data with the platform, incur some privacy loss and get compensated for their loss. The platform then combines the data of users and outputs an estimator.

In both cases, we first establish that a linear estimator with proper weights and added Laplace noise achieves the nearly optimal minimax bound, which is of independent interest. Building on this characterization, we then optimally solve the corresponding mechanism design problem for both settings. In the central privacy setting, we establish a polynomial time score-based algorithm that finds the optimal privacy loss levels. In the local privacy setting, however, we establish a Polynomial Time Approximation Scheme (PTAS) for finding the optimal privacy losses.

Finally, we compare the performance of the central and the local architectures. In particular, we show that the platform's utility in the central privacy setting is always higher than in the local privacy setting. But, there is no dominance in terms of the optimal privacy loss level, i.e., depending on her privacy sensitivity, a user may have a higher or lower privacy loss in the central setting compared to the local setting.

In our analysis, we utilized a set of simplifying assumptions to aid our investigation. Here, we would like to underline these assumptions, furnish reasons for their use within the context of our application, and outline potential avenues for future exploration. As an illustrative example, we consider the purchase of medical data by emerging companies such as Hu-manity.co, where users receive compensation for sharing their medical information.

- **Verifiability of data:** We made the assumption that while users have the ability to misrepresent their privacy sensitivity, they cannot falsify the actual data itself. This assumption is applicable in scenarios where users sell the "rights" to their data, and the platform collects data, such as in the context of Hu-manity.co, where medical data from users is gathered.

- **Independence of data and privacy sensitivity:** We have assumed that there is no relation between users' data and their privacy sensitivity. In the context of the Hu-manity.co application, this implies that users become aware of their privacy sensitivity before their actual medical data (i.e., realized data) is revealed. This assumption arises from situations where users are unable to collect/process the data themselves. However, we recognize that in other applications, such as the sharing of financial data, this assumption may not hold. Without this assumption, there may be a sample bias in the data collected by the platform, which would require correction. We leave exploring this direction as an interesting future avenue of research.

- **Extensions in estimation models:** We focused on the private mean estimation task from a population, but it would be interesting to extend our results to more complex estimation models. Here are potential extensions: estimating a multi-dimensional underlying parameter denoted as θ and estimating an underlying parameter θ when customers have data (X_i, Y_i) defined as $Y_i = X_i'\theta + Z_i$.

It is worth noting that extending our results to these scenarios requires establishing the equivalent counterparts of Theorems 2.1 and 2.2. This involves finding the (minimax) optimal estimator while considering heterogeneous differential privacy concerns. Once such an estimator is obtained (or when the minimax optimality of the estimator is not a concern), our results on the characterization of the mechanism continue to hold. This means that similar to the derivations presented in Section 2.3, the platform's problem revolves around solving a point-wise optimization problem. However, it is important to note that, similar to our current setting, this problem can also be non-convex, necessitating the development of an efficient algorithm for its solution.

- **Additive user utility:** In our model, each user's utility is determined as the payment received minus the mean squared error (MSE) of the estimator, minus the privacy sensitivity multiplied by their level of differential privacy. This utility form assumes two key assumptions. First, it assumes additivity and that the privacy sensitivity is directly multiplied by the privacy level. While these assumptions simplify the derivations, it is worth noting that, similar to the classic mechanism design setting, all the results extend as long as the user utility is quasi-linear. In other words, the utility is a function of the MSE, privacy sensitivity, and privacy level and is subtracted by the payment. As an example, it is possible for the MSE and payment to have user-specific known coefficients.

  Second, the model assumes that there is only one privately known user parameter, which is privacy sensitivity. If the users have privately known weights for either the MSE or payment, the problem becomes a multi-dimensional mechanism design whose study is beyond the scope of this work.

- **Trusting the platform:** In the central setting, we assumed there is trust between users and the platform: users share their data with the platform, relying on the platform to handle the data responsibly and deliver the promised privacy level without exploiting it for other purposes (this form of credibility is present in data acquisition mechanisms and not classic auctions that are studied in Akbarpour and Li [2020]). This concern regarding platform credibility motivated us to also study the local privacy setting. In this setting, data is privatized directly on the user side, granting users control over the implementation of privacy measures. This local structure has been effectively implemented by various tech companies, such as Apple, which has incorporated local differential privacy techniques into their data handling processes (see, e.g., Apple).

2.8 Proofs and Additional Results

Proof of Lemma 2.2

Since $|Z_i| \leq \frac{1}{2}$, the difference of every two realizations of X_i would be bounded by one. Therefore, the sensitivity of $\hat{\theta}$ to x_i, defined in (2.1), is given by $w_i(\boldsymbol{c})$. Hence, Lemma 2.1 immediately implies the result. ■

Proof of Theorem 2.1

For the sake of subsequent analysis, we find it helpful to recall the definition of two well-known distribution distances. Let P and Q be two distributions, defined over a probability space $(\Omega, \mathcal{F})$. Then,

- the total variation (TV) distance is denoted by $\|P - Q\|_{\mathrm{TV}}$, and is given by

$$\|P - Q\|_{\mathrm{TV}} := \sup_{A \in \mathcal{F}} |P(A) - Q(A)| = \frac{1}{2} \int_{\Omega} |dP - dQ|.$$

- when P is absolutely continuous with respect to Q, the Kullback–Leibler (KL) distance is denoted by $D_{\mathrm{KL}}(P, Q)$, and it is given by

$$D_{\mathrm{KL}}(P, Q) := \int \log \frac{dP}{dQ} dP.$$

One inequality that we particularly find it helpful for analysis is the Pinsker's inequality which states that

$$\|P - Q\|_{\mathrm{TV}}^2 \leq \frac{1}{2} D_{\mathrm{KL}}(P, Q). \tag{2.34}$$

To prove the lower bound (2.4), we use Le Cam's method Yu [1997] which is a well-known technique in deriving minimax lower bounds. The main idea of Le Cam's method is reducing the estimation problem to a testing problem. More formally, let P_1 and P_2 be two distributions in $\mathcal{P}_k$ with

$$\gamma := \frac{1}{2} |\theta(P_1) - \theta(P_2)|. \tag{2.35}$$

Furthermore, for $j \in \{1, 2\}$, let Q_j be the marginal distribution of $\hat{\theta}$, given that the samples $X_{1:n}$ are all drawn from P_j, i.e.,

$$Q_j(A) = \int_{\mathbb{R}^n} \mathbb{P}(\hat{\theta}(x_{1:n}) \in A) dP_j^n(x_{1:n}), \tag{2.36}$$

for any measurable set A. Then, by Le Cam's method, we have (see Barber and Duchi [2014] for more details)

$$\mathcal{L}_c(\mathcal{P}_k, \theta, \boldsymbol{\varepsilon}) \geq \gamma^2 \left(\frac{1}{2} - \frac{1}{2} \|Q_1 - Q_2\|_{\mathrm{TV}} \right). \tag{2.37}$$

Next, we bound $\|Q_1 - Q_2\|_{\mathrm{TV}}$ in the following lemma.

Lemma 2.3. *Let P_1 and P_2 be two distributions in $\mathcal{P}$ such that P_1 is absolutely continuous with respect to P_2. Consider Q_1 and Q_2 as defined in (2.36). Then, for any $k \in \{0, 1, \cdots, n\}$,*

$$\|Q_1 - Q_2\|_{TV} \leq 2 \|P_1 - P_2\|_{TV} \sum_{i=1}^{k} (e^{\varepsilon_i} - 1) + \sqrt{\frac{n-k}{2} D_{KL}(P_1, P_2)}. \tag{2.38}$$

We prove this lemma at the end of this section. Now, using this lemma, let us complete the proof of (2.4). Let $\delta \in [0, 1/2]$, and define P_1 and P_2 as

$$P_1(-1/2) = P_2(1/2) = \frac{1+\delta}{2}, \quad P_1(1/2) = P_2(-1/2) = \frac{1-\delta}{2}. \tag{2.39}$$

Obviously $P_1, P_2 \in \mathcal{P}^*$. Also, for $i \in \{1, 2\}$, $\mathbb{E}_{P_i}[X] = (-1)^i \delta/2$, and hence $\gamma = \delta/2$. In addition, by definition, we have

$$\|P_1 - P_2\|_{\mathrm{TV}} = \frac{1}{2} \times 2 \times \left(\frac{1+\delta}{2} - \frac{1-\delta}{2} \right) = \delta,$$
$$D_{\mathrm{KL}}(P_1, P_2) = \delta \log \frac{1+\delta}{1-\delta} \leq 3\delta^2,$$

where the last inequality holds for $\delta \in [0, 1/2]$. Hence, by Lemma 2.3, along with the fact that $e^{\varepsilon_i} - 1 \leq 2\varepsilon_i$ for $\varepsilon_i \leq 1$, we have that for

$$\|Q_1 - Q_2\|_{\mathrm{TV}} \leq 4\delta \sum_{i=1}^{k} \varepsilon_i + \delta \sqrt{\frac{3(n-k)}{2}}.$$

Therefore, using (2.37), we obtain

$$\mathcal{L}_c(\mathcal{P}_k, \theta, \boldsymbol{\varepsilon}) \geq \frac{\delta^2}{8} \left(1 - \delta \left[4 \sum_{i=1}^{k} \varepsilon_i + \sqrt{\frac{3(n-k)}{2}} \right] \right). \tag{2.40}$$

Choosing

$$\delta = \left(8 \sum_{i=1}^{k} \varepsilon_i + \sqrt{6(n-k)} \right)^{-1} \wedge \frac{1}{2},$$

implies

$$\mathcal{L}_c(\mathcal{P}_k, \theta, \boldsymbol{\varepsilon}) \geq \frac{1}{16} \left(\left(8 \sum_{i=1}^{k} \varepsilon_i + \sqrt{6(n-k)} \right)^{-2} \wedge \frac{1}{4} \right). \tag{2.41}$$

Using inequality $(x+y)^2 \leq 2(x^2+y^2)$ with $x = 8 \sum_{i=1}^{k} \varepsilon_i$ and $y = \sqrt{6(n-k)}$ completes the proof of (2.4).

Next, we show the upper bound (2.5). First note that, since $|X| \leq 1/2$ almost surely for any $P \in \mathcal{P}$, we have $\theta(P) \leq 1/2$ for any $P \in \mathcal{P}$. Hence, $\hat{\theta} = 0$, which is a linear estimator in the form of (2.3) with $w_i = 0$ for all i and $\eta = \infty$, leads to $\mathbb{E}[|\hat{\theta} - \theta|^2] \leq 1/4$. Hence, it suffices to find a linear estimator $\hat{\theta}$ with Laplace mechanism such that

$$\mathbb{E}[|\hat{\theta} - \theta|^2] \leq \mathcal{O}(1) \log(n+1) \max_{j} \frac{1}{n - j + (\sum_{i=1}^{j} \varepsilon_i)^2}. \tag{2.42}$$

To do so, let k^* be the largest $k \in \{0, 1, \cdots n-1\}$ such that

$$\varepsilon_{n-k} > \frac{1}{\sqrt{k+1}}, \tag{2.43}$$

if such k exists. Now, we consider two cases:

- First, assume such k does not exists. Then, consider linear estimator

$$\hat{\theta} = \sum_{i=1}^{n} \frac{\varepsilon_i}{\eta} x_i + \text{Laplace}\left(\frac{1}{\eta}\right),$$

with

$$\eta = \sum_{i=1}^{n} \varepsilon_i.$$

In this case, it is straightforward to see

$$\mathbb{E}[|\hat{\theta} - \theta|^2] \leq \frac{\sum_{i=1}^n \varepsilon_i^2 + 2}{\eta^2} = \frac{\sum_{i=1}^n \varepsilon_i^2 + 2}{(\sum_{i=1}^n \varepsilon_i)^2}, \tag{2.44}$$

where we the fact that the variance is bounded by one since the absolute value of the random variable almost surly bounded by one. Next, note that, since (2.43) does not hold for any k, we have $\varepsilon_{n-k} \leq \frac{1}{\sqrt{k+1}}$ for any k, which implies that

$$\sum_{i=1}^{n} \varepsilon_i^2 \leq \sum_{i=1}^{n} \frac{1}{i} = \mathcal{O}(1)\log(n+1).$$

Plugging this relation into (2.44) implies

$$\mathbb{E}[|\hat{\theta} - \theta|^2] \leq \mathcal{O}(1)\log(n+1)\frac{1}{(\sum_{i=1}^n \varepsilon_i)^2}.$$

This completes the proof of (2.42) since

$$\frac{1}{(\sum_{i=1}^n \varepsilon_i)^2}$$

is

$$\frac{1}{n-j+(\sum_{i=1}^{j}\varepsilon_i)^2}$$

with $j=n$.

- Now assume there exists at least one k that satisfies (2.43) (and hence the aforementioned k^* is well-defined.) As a result, we have

$$\varepsilon_1 \leq \frac{1}{\sqrt{n}}, \cdots, \varepsilon_{n-k^*-1} \leq \frac{1}{\sqrt{k^*+2}}, \tag{2.45}$$

$$\varepsilon_n \geq \cdots \varepsilon_{n-k^*} > \frac{1}{\sqrt{k^*+1}}. \tag{2.46}$$

In this case, consider the following linear estimator

$$\hat{\theta} = \sum_{i=1}^{n-k^*-1} \frac{\varepsilon_i}{\eta} x_i + \sum_{i=n-k^*}^{n} \frac{1/\sqrt{k^*+1}}{\eta} x_i + \text{Laplace}\left(\frac{1}{\eta}\right),$$

with

$$\eta = \sum_{i=1}^{n-k^*-1} \varepsilon_i + \frac{k^*+1}{\sqrt{k^*+1}} = \sum_{i=1}^{n-k^*-1} \varepsilon_i + \sqrt{k^*+1}.$$

First, by Lemma 2.2, this estimator is

$$(\varepsilon_1, \cdots, \varepsilon_{n-k^*-1}, \frac{1}{\sqrt{k^*+1}}, \cdots, \frac{1}{\sqrt{k^*+1}})$$

differentially private, and hence, due to (2.46), it is $(\varepsilon_i)_{i=1}^n$-differentially private as well. Second, using the fact that variance is bounded by one, we have

$$\begin{aligned}\mathbb{E}[|\hat{\theta}-\theta|^2] &\leq \frac{\sum_{i=1}^{n-k^*-1} \varepsilon_i^2 + \frac{k^*+1}{k^*+1} + 2}{\eta^2} \\ &\leq \frac{\sum_{i=k^*+2}^{n} \frac{1}{i} + 3}{(\sum_{i=1}^{n-k^*-1} \varepsilon_i + \sqrt{k^*+1})^2},\end{aligned} \tag{2.47}$$

where the last inequality follows from (2.45) and definition of η. To complete the proof, note that, the numerator of (2.47) is upper bounded by $\mathcal{O}(1)\log(n+1)$ and its denominator is lower bounded by

$$\left(\sum_{i=1}^{n-k^*-1} \varepsilon_i\right)^2 + k^* + 1 = \left(\sum_{i=1}^{j} \varepsilon_i\right)^2 + n - j \text{ with } j = n - k^* - 1.$$

Hence, (2.42) holds in this case as well. ■

Proof of Lemma 2.3

Let $\tilde{Q}$ be the marginal distribution of $\hat{\theta}$ given that $X_1, \cdots, X_k$ are drawn from P_1 and $X_{k+1}, \cdots, X_n$ are drawn from P_2, i.e.,

$$\tilde{Q}(A) = \int_{\mathbb{R}^n} \mathbb{P}(\hat{\theta}(x_{1:n}) \in A) dP_1^k(x_{1:k}) dP_2^{n-k}(x_{k+1:n}). \tag{2.48}$$

Note that, we have

$$\|Q_1 - Q_2\|_{\text{TV}} \leq \|Q_1 - \tilde{Q}\|_{\text{TV}} + \|\tilde{Q} - Q_2\|_{\text{TV}}. \tag{2.49}$$

The idea is to bound the two terms on the right hand side separately. In particular, we show

$$\|Q_1 - \tilde{Q}\|_{\text{TV}}^2 \leq \frac{n-k}{2} D_{\text{KL}}(P_1, P_2), \tag{2.50}$$

$$\|\tilde{Q} - Q_2\|_{\text{TV}} \leq 2\|P_1 - P_2\|_{\text{TV}} \sum_{i=1}^{k} (e^{\varepsilon_i} - 1). \tag{2.51}$$

If we show these two bounds, then plugging them into (2.49) will show Lemma 2.3.

- We start by showing (2.50). First, note that, by data processing inequality, we have

$$\|Q_1 - \tilde{Q}\|_{\text{TV}} \leq \|P_1^n - P_1^k P_2^{n-k}\|_{\text{TV}}.$$

Next, using Pinsker's inequality (2.34), we obtain

$$\begin{aligned}\|P_1^n - P_1^k P_2^{n-k}\|_{\text{TV}}^2 &\leq \frac{1}{2} D_{\text{KL}}(P_1^n, P_1^k P_2^{n-k}) \\ &\leq \frac{n-k}{2} D_{\text{KL}}(P_1, P_2),\end{aligned}$$

where the second inequality follows from the chain rule for KL-divergence. This completes the proof of (2.50).

- Next, we show (2.51). By total variation distance definition, it suffices to show that, for any measurable set A, $|\tilde{Q}(A) - Q_2(A)|$ is upper bounded by the right hand side of (2.38). To see this, first, note that we have

$$\begin{aligned}\tilde{Q}(A) - Q_2(A) &= \int_{\mathbb{R}^n} \mathbb{P}(\hat{\theta}(x_{1:n}) \in A)(dP_1^k(x_{1:k}) - dP_2^k(x_{1:k}))dP_2^{n-k}(x_{k+1:n}) \\ &= \int_{\mathbb{R}^{n-k}} \Delta(x_{k+1:n})dP_2^{n-k}(x_{k+1:n}),\end{aligned} \tag{2.52}$$

where

$$\Delta(x_{k+1:n}) := \int_{\mathbb{R}^k} \mathbb{P}(\hat{\theta}(x_{1:n}) \in A)(dP_1^k(x_{1:k}) - dP_2^k(x_{1:k})). \tag{2.53}$$

To show (2.51), it suffices to show

$$|\Delta(x_{k+1:n})| \leq 2\|P_1 - P_2\|_{\mathrm{TV}} \sum_{i=1}^{k}(e^{\varepsilon_i} - 1). \tag{2.54}$$

To do so, first, note that $dP_1^k(x_{1:k}) - dP_2^k(x_{1:k})$ can be cast as

$$dP_1^k(x_{1:k}) - dP_2^k(x_{1:k}) = \sum_{i=1}^{k} dP_2^{i-1}(x_{1:i-1})\left(dP_1(x_i) - dP_2(x_i)\right)dP_1^{k-i}(x_{i+1:k}).$$

Plugging this into (2.53), we obtain

$$\Delta(x_{k+1:n}) = \sum_{i=1}^{k} \int_{\mathbb{R}^n} \mathbb{P}(\hat{\theta}(x_{1:n}) \in A)dP_2^{i-1}(x_{1:i-1})\left(dP_1(x_i) - dP_2(x_i)\right)dP_1^{k-i}(x_{i+1:k}). \tag{2.55}$$

Let $x_{1:n}^i$ be a vector similar to $x_{1:n}$, except on i-th coordinate, where x_i is replaced by x_i'. Note that,

$$\int_{\mathbb{R}^k} \mathbb{P}(\hat{\theta}(x_{1:n}^i) \in A)dP_2^{i-1}(x_{1:i-1})\left(dP_1(x_i) - dP_2(x_i)\right)dP_1^{k-i}(x_{i+1:k}) = 0.$$

Hence, we could write (2.55) as

$$\Delta(x_{k+1:n}) = \tag{2.56}$$
$$\sum_{i=1}^{k} \int_{\mathbb{R}^k} \left(\mathbb{P}(\hat{\theta}(x_{1:n}) \in A) - \mathbb{P}(\hat{\theta}(x^i_{1:n}) \in A)\right) dP_2^{i-1}(x_{1:i-1}) \left(dP_1(x_i) - dP_2(x_i)\right) dP_1^{k-i}(x_{i+1:k}).$$

Hence, we have

$$|\Delta(x_{k+1:n})| \leq \tag{2.57}$$
$$\sum_{i=1}^{k} \int_{\mathbb{R}^k} \left|\mathbb{P}(\hat{\theta}(x_{1:n}) \in A) - \mathbb{P}(\hat{\theta}(x^i_{1:n}) \in A)\right| dP_2^{i-1}(x_{1:i-1}) \left|dP_1(x_i) - dP_2(x_i)\right| dP_1^{k-i}(x_{i+1:k}).$$

Note that, by differential privacy definition, we have

$$(e^{-\varepsilon_i} - 1)\mathbb{P}(\hat{\theta}(x^i_{1:n}) \in A) \leq \mathbb{P}(\hat{\theta}(x_{1:n}) \in A) - \mathbb{P}(\hat{\theta}(x^i_{1:n}) \in A) \leq (e^{\varepsilon_i} - 1)\mathbb{P}(\hat{\theta}(x^i_{1:n}) \in A),$$

which implies

$$\left|\mathbb{P}(\hat{\theta}(x_{1:n}) \in A) - \mathbb{P}(\hat{\theta}(x^i_{1:n}) \in A)\right| \leq (e^{\varepsilon_i} - 1)\mathbb{P}(\hat{\theta}(x^i_{1:n}) \in A).$$

Plugging this into (2.57), we obtain

$$|\Delta(x_{k+1:n})| \leq \tag{2.58}$$
$$\sum_{i=1}^{k} (e^{\varepsilon_i} - 1) \int_{\mathbb{R}^k} \mathbb{P}(\hat{\theta}(x^i_{1:n}) \in A) dP_2^{i-1}(x_{1:i-1}) \left|dP_1(x_i) - dP_2(x_i)\right| dP_1^{k-i}(x_{i+1:k}).$$

Finally, note that

$$\begin{aligned}
&\int_{\mathbb{R}^k} \mathbb{P}(\hat{\theta}(x^i_{1:n}) \in A) dP_2^{i-1}(x_{1:i-1}) \left|dP_1(x_i) - dP_2(x_i)\right| dP_1^{k-i}(x_{i+1:k}) \\
&= \int_{\mathbb{R}^{k-1}} \mathbb{P}(\hat{\theta}(x^i_{1:n}) \in A) dP_2^{i-1}(x_{1:i-1}) dP_1^{k-i}(x_{i+1:k}) \int_{\mathbb{R}} \left|dP_1(x_i) - dP_2(x_i)\right| \\
&\leq 2\|P_1 - P_2\|_{\mathrm{TV}}, \qquad (2.59)
\end{aligned}$$

where the last inequality follows from the fact that

$$\int_{\mathbb{R}^{k-1}} \mathbb{P}(\hat{\theta}(x^i_{1:n}) \in A) dP_2^{i-1}(x_{1:i-1}) dP_1^{k-i}(x_{i+1:k})$$

is bounded by 1. Plugging (2.59) into (2.58) completes the proof of (2.54). ■

Proof of Theorem 2.2

To show the lower bound (2.12), we again use the Le Cam's method. Here, for $j \in \{1, 2\}$, we define Q_j to be the marginal distribution of $\mathcal{M}$, given that the samples $X_{1:n}$ are all drawn from P_j, i.e.,

$$Q_j(A) = \int_{\mathbb{R}^n} \mathbb{P}(\mathcal{M}(x_{1:n}) \in A) dP_j^n(x_{1:n}), \tag{2.60}$$

for any measurable set $A \subset \mathbb{R}^n$. Then, again, by Le Cam's method, we have (see Duchi et al. [2013] for more details)

$$\mathcal{L}_l(\mathcal{P}_k, \theta, \boldsymbol{\varepsilon}) \geq \gamma^2 \left(\frac{1}{2} - \frac{1}{2} \|Q_1 - Q_2\|_{\mathrm{TV}} \right), \tag{2.61}$$

where γ is given by (2.35). A slight extension of Corollary 1 in Duchi et al. [2013] implies

$$\|Q_1 - Q_2\|_{\mathrm{TV}}^2 \leq \frac{1}{4} \left(D_{\mathrm{KL}}(Q_1, Q_2) + D_{\mathrm{KL}}(Q_2, Q_1) \right) \leq \|P_1 - P_2\|_{\mathrm{TV}}^2 \sum_{i=1}^{n} (e^{\varepsilon_i} - 1)^2.$$

Hence, for $\varepsilon_i \leq 1$, using $e^{\varepsilon_i} - 1 \leq 2\varepsilon_i$, we have

$$\|Q_1 - Q_2\|_{\mathrm{TV}} \leq 2 \|P_1 - P_2\|_{\mathrm{TV}} \sqrt{\sum_{i=1}^{n} \varepsilon_i^2}.$$

Taking P_1 and P_2 similar to (2.39), and using (2.61), we have

$$\mathcal{L}_l(\mathcal{P}_k, \theta, \boldsymbol{\varepsilon}) \geq \frac{1}{8} \delta^2 \left(1 - 2\delta \sqrt{\sum_{i=1}^{n} \varepsilon_i^2} \right). \tag{2.62}$$

Choosing

$$\delta = \frac{1}{4\sqrt{\sum_{i=1}^n \varepsilon_i^2}} \wedge \frac{1}{2}$$

completes the proof of (2.12).

To show the upper bound (2.13), we form the following estimator

$$\hat{\theta} = \sum_{i=1}^n \frac{\varepsilon_i^2}{\sum_{j=1}^n \varepsilon_j^2} \hat{x}_i, \text{ where } \hat{x}_i = x_i + \text{Laplace}\left(\frac{1}{\varepsilon_i}\right). \tag{2.63}$$

Clearly this estimator is $(\varepsilon_i)_{i=1}^n$-locally differentially private. Next, note that

$$\mathbb{E}\left[|\hat{\theta} - \theta|^2\right] = \sum_{i=1}^n \frac{\varepsilon_i^4}{(\sum_{j=1}^n \varepsilon_j^2)^2} \text{Var}(\hat{x}_i) = \sum_{i=1}^n \frac{\varepsilon_i^4}{(\sum_{j=1}^n \varepsilon_j^2)^2} \left(\text{Var}(X) + \frac{1}{\varepsilon_i^2}\right). \tag{2.64}$$

Using the fact that $\frac{1}{\varepsilon_i^2} \geq 1 \geq \text{Var}(X)$, we obtain

$$\mathbb{E}\left[|\hat{\theta} - \theta|^2\right] \leq 2\sum_{i=1}^n \frac{\varepsilon_i^4}{(\sum_{j=1}^n \varepsilon_j^2)^2} \cdot \frac{1}{\varepsilon_i^2} = 2\sum_{i=1}^n \frac{\varepsilon_i^2}{(\sum_{j=1}^n \varepsilon_j^2)^2} = \frac{2}{\sum_{j=1}^n \varepsilon_j^2}, \tag{2.65}$$

which completes the proof. ■

Proof of Proposition 2.1

Letting

$$h_i(c) = \mathbb{E}_{\mathbf{c}_{-i}}\left[\text{MSE}(c, \mathbf{c}_{-i}, \boldsymbol{\varepsilon}, \hat{\theta})\right],$$

$$t_i(c) = \mathbb{E}_{\mathbf{c}_{-i}}\left[t(c, \mathbf{c}_{-i})\right],$$

and

$$\varepsilon_i(c) = \mathbb{E}_{\mathbf{c}_{-i}}\left[\varepsilon_i(c, \mathbf{c}_{-i})\right],$$

we can write the incentive compatibility constraint as

$$h_i(c_i) + c_i\varepsilon_i(c_i) - t_i(c_i) \leq h_i(c_i') + c_i\varepsilon_i(c_i') - t_i(c_i').$$

Taking derivative of the right-hand side with respect to c_i' and evaluating the derivative at $c_i' = c_i$ and equating it to zero leads to

$$h_i'(c_i) + c_i\varepsilon_i'(c_i) - t_i'(c_i) = 0.$$

By taking the integral of this expression we obtain

$$t_i(c_i) = t_i(0) + \int_{z=0}^{c_i} (h_i'(z) + z\varepsilon_i'(z))\, dz = t_i(0) + h_i(c_i) - h_i(0) + c_i\varepsilon_i(c_i) - \int_{z=0}^{c_i} \varepsilon_i(z)dz. \tag{2.66}$$

We next show that the payment in (2.66) together with a weakly decreasing $\varepsilon_i(z)$ guarantees that the incentive compatibility constraint. To see this, we consider two possibilities depending on whether c_i' is larger or smaller than c_i:

- For $c_i' \geq c_i$: by plugging in the payment in (2.66) the incentive compatibility constraint becomes equivalent to

  $$\varepsilon_i(c_i')(c_i - c_i') \geq \int_{z=c_i'}^{c_i} \varepsilon_i(z)dz,$$

  which holds because $\varepsilon_i(z)$ is weakly decreasing in z.

- For $c_i' \leq c_i$: again, by plugging in the payment in (2.66) the incentive compatibility constraint becomes equivalent to

  $$\varepsilon_i(c_i')(c_i - c_i') \leq \int_{z=c_i}^{c_i'} \varepsilon_i(z)dz,$$

  which, again, holds because $\varepsilon_i(z)$ is weakly decreasing in z. This completes one direction of the proof.

To see the other direction, notice that the first order condition of the incentive compatibility implies (2.66). Finally notice that the incentive compatibility implies

$$h_i(c_i) + c_i\varepsilon_i(c_i) - t_i(c_i) \leq h_i(c_i') + c_i\varepsilon_i(c_i') - t_i(c_i').$$

and

$$h_i(c_i') + c_i'\varepsilon_i(c_i') - t_i(c_i') \leq h_i(c_i) + c_i'\varepsilon_i(c_i) - t_i(c_i).$$

Taking summation of these two equations results in

$$\left(\varepsilon_i(c_i) - \varepsilon_i(c_i')\right)(c_i - c_i') \leq 0,$$

which shows that $\varepsilon_i(\cdot)$ is weakly decreasing.

We next consider the individual rationality constraint. Using (2.66), we can rewrite this constraint as

$$t_i(0) \geq h_i(0) - \text{VAR} + \int_{z=0}^{c_i} \varepsilon_i(z)dz \quad \text{for all } c_i \tag{2.67}$$

which means it only needs to hold for $c_i = \infty$. Hence, we could cast $t_i(0)$ as

$$t_i(0) = h_i(0) - \text{VAR} + \int_{z=0}^{\infty} \varepsilon_i(z)dz + d_i$$

for some nonnegative constant d_i. Plugging this back in (2.66) results in

$$t_i(c_i) = h_i(c_i) - \text{VAR} + c_i\varepsilon_i(c_i) + \int_{z=c_i}^{\infty} \varepsilon_i(z)dz + d_i.$$

This completes the proof. ■

Proof of Proposition 2.2

Using the payment identity in Proposition 2.1, we obtain

$$
\begin{aligned}
\mathbb{E}_{c_i}\left[t_i(c_i)\right] &= \mathbb{E}[\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\hat{\theta})] - \text{VAR} + \mathbb{E}_{c_i}[c_i\varepsilon_i(c_i)] + \mathbb{E}_{c_i}\left[\int_{z=c_i}^{\infty}\varepsilon_i(z)dz\right] \\
&= \mathbb{E}[\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\hat{\theta})] - \text{VAR} \\
&\quad + \int_{\mathbf{z}_{-i}}\int_{z_i}\left(z_i\varepsilon_i(z_i,\mathbf{z}_{-i}) + \int_{y_i=z_i}^{\infty}\varepsilon_i(y_i,\mathbf{z}_{-i})dy_i\right) f_i(z_i)dz_i f_{-i}(\mathbf{z}_{-i})d\mathbf{z}_{-i} \\
&= \mathbb{E}[\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\hat{\theta})] - \text{VAR} + \int_{\mathbf{z}_{-i}}\int_{z_i} z_i\varepsilon_i(z_i,\mathbf{z}_{-i})f_i(z_i)dz_i f_{-i}(\mathbf{z}_{-i})d\mathbf{z}_{-i} \\
&\quad + \int_{\mathbf{z}_{-i}}\int_{z_i=0}^{\infty}\int_{y_i=z_i}^{\infty}\varepsilon_i(y_i,\mathbf{z}_{-i})dy_i f_i(z_i)dz_i f_{-i}(\mathbf{z}_{-i})d\mathbf{z}_{-i} \\
&\overset{(a)}{=} \mathbb{E}[\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\hat{\theta})] - \text{VAR} + \int_{\mathbf{z}_{-i}}\int_{z_i} z_i\varepsilon_i(z_i,\mathbf{z}_{-i})f_i(z_i)dz_i f_{-i}(\mathbf{z}_{-i})d\mathbf{z}_{-i} \\
&\quad + \int_{\mathbf{z}_{-i}} f_{-i}(\mathbf{z}_{-i})d\mathbf{z}_{-i}\int_{y_i=0}^{\infty}\varepsilon_i(y_i,\mathbf{z}_{-i})dy_i\int_{z_i=0}^{y_i} f_i(z_i)dz_i \\
&= \mathbb{E}[\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\hat{\theta})] - \text{VAR} + \int_{\mathbf{z}_{-i}}\int_{z_i} z_i\varepsilon_i(z_i,\mathbf{z}_{-i})f_i(z_i)dz_i f_{-i}(\mathbf{z}_{-i})d\mathbf{z}_{-i} \\
&\quad + \int_{\mathbf{z}_{-i}} f_{-i}(\mathbf{z}_{-i})d\mathbf{z}_{-i}\int_{y_i=0}^{\infty}\varepsilon_i(y_i,\mathbf{z}_{-i})dy_i F_i(y_i) \\
&\overset{(b)}{=} \mathbb{E}[\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\hat{\theta})] - \text{VAR} + \int_{\mathbf{z}}\left(z_i + \frac{F_i(z_i)}{f_i(z_i)}\right)\varepsilon_i(\mathbf{z})f(\mathbf{z})d\mathbf{z}, \qquad (2.68)
\end{aligned}
$$

where (a) follows from changing the order of the integrals and (b) follows by a change of variable from y_i to z_i. Substituting equation (2.68) in the platform's objective function, results in

$$
\mathbb{E}_{\mathbf{c}}\left[\mathrm{MSE}(\mathbf{c},\boldsymbol{\varepsilon},\hat{\theta}) + \sum_{i=1}^{n} t_i(\mathbf{c})\right] = \mathbb{E}_{\mathbf{c}}\left[(n+1)\mathrm{MSE}(\mathbf{c},\boldsymbol{\varepsilon},\hat{\theta}) + \sum_{i=1}^{n}\psi(c_i)\varepsilon_i(\mathbf{c})\right] - n\text{VAR}.
$$

Finally, note that, by using Proposition 2.1, the payment identity guarantees a privacy level function is decreasing if $\varepsilon_i(\cdot)$ is decreasing. This completes the proof. ■

Proof of Theorem 2.3

Since the variance of the data points are VAR, the variance of the estimator given in (2.22) (i.e., the mean square error) is

$$\frac{2}{\eta^2} + \sum_{i=1}^{n} w_i(\mathbf{c})^2 \text{VAR}.$$

By plugging this expression into the characterization of Proposition 2.2, we see that for any vector of reported privacy costs $(c_1, \ldots, c_n)$, the point-wise optimization problem becomes

$$\begin{aligned} \min_{\boldsymbol{\varepsilon}(\mathbf{c}),\mathbf{w}(\mathbf{c}),\eta} \; & \frac{2(n+1)}{\eta^2} + \sum_{i=1}^{n} (n+1)\text{VAR } w_i(\mathbf{c})^2 + \psi_i(c_i)\varepsilon_i(\mathbf{c}) \\ & \text{s.t. } \boldsymbol{\varepsilon}_i(\mathbf{c}) \geq 0, \text{ for all } i \in \mathcal{N} \\ & \sum_{i=1}^{n} w_i(\mathbf{c}) = 1 \\ & \eta w_i(\mathbf{c}) \leq \varepsilon_i(\mathbf{c}) \text{ for all } i \in \mathcal{N}. \end{aligned}$$

In the optimal solution we must have $\eta w_i(\mathbf{c}) = \varepsilon_i(\mathbf{c})$ for all $i \in \mathcal{N}$. Therefore, the above optimization is equivalent to

$$\begin{aligned} \min_{\boldsymbol{\varepsilon}(\mathbf{c}),\mathbf{w}(\mathbf{c}),\eta} \; & \frac{2(n+1)}{\eta^2} + \sum_{i=1}^{n} (n+1)\text{VAR } \frac{\boldsymbol{\varepsilon}_i^2(\mathbf{c})}{\eta^2} + \psi_i(c_i)\varepsilon_i(\mathbf{c}) \\ & \text{s.t. } \boldsymbol{\varepsilon}_i(\mathbf{c}) \geq 0, \text{ for all } i \in \mathcal{N} \\ & \sum_{i=1}^{n} \frac{\boldsymbol{\varepsilon}_i(\mathbf{c})}{\eta} = 1, \end{aligned}$$

which in turn, by letting $\boldsymbol{\varepsilon}_i(\mathbf{c}) = y_i$ for all $i \in \mathcal{N}$, is equivalent to

$$\begin{aligned} \min_{\mathbf{y}} \; & \frac{2(n+1)}{\left(\sum_{j=1}^{n} y_j\right)^2} + \sum_{i=1}^{n} (n+1)\text{VAR} \left(\frac{y_i}{\sum_{j=1}^{n} y_j} \right)^2 + \psi_i(c_i) y_i \\ & \text{s.t. } y_i \geq 0, \text{ for all } i \in \mathcal{N}. \end{aligned} \tag{2.69}$$

The corresponding payment to user i is given by

$$\mathbb{E}_{\mathbf{c}_{-i}}\left[\mathrm{MSE}(c_i, \mathbf{c}_{-i}; \boldsymbol{\varepsilon}, \hat{\theta}_{\text{central}})\right] - \mathrm{VAR} + c_i \mathbb{E}_{\mathbf{c}_{-i}}\left[\varepsilon_i(c_i, \mathbf{c}_{-i})\right] + \int_{x=c_i}^{\infty} \mathbb{E}_{-i}\left[\varepsilon_i(x, \mathbf{c}_{-i})\right] dx.$$

By invoking Proposition 2.1, this payment and allocation satisfy the incentive compatibility and the individual rationality constraints provided that the optimal privacy level function is weakly decreasing in the reported privacy cost which we prove next.

Let $(y_1, \ldots, y_n)$ be the solution of optimization problem (2.69) for $c_1, \ldots, c_n$. Now, suppose we increases one of the c_i's, which, without loss of generality, we assume is the first one. Let $c_1' > c_1$ and $c_i' = c_i$ for $i = 2, \ldots, n$ and suppose $y_1', \ldots, y_n'$ is the corresponding optimal solution of optimization problem (2.69). The optimality condition implies that

$$\begin{aligned}
&\frac{2(n+1)}{\left(\sum_{j=1}^{n} y_j^2\right)^2} + \frac{(n+1)}{\left(\sum_{j=1}^{n} y_j^2\right)^2} \sum_{i=1} y_i^2 \mathrm{VAR} + \psi_i(c_i) y_i \\
&\le \frac{2(n+1)}{\left(\sum_{j=1}^{n} {y_j'}^2\right)^2} + \frac{(n+1)}{\left(\sum_{j=1}^{n} {y_j'}^2\right)^2} \sum_{i=1} {y_i'}^2 \mathrm{VAR} + \psi_i(c_i) y_i'
\end{aligned}$$

and

$$\begin{aligned}
&\frac{2(n+1)}{\left(\sum_{j=1}^{n} {y_j'}^2\right)^2} + \frac{(n+1)}{\left(\sum_{j=1}^{n} {y_j'}^2\right)^2} \sum_{i=1} {y_i'}^2 \mathrm{VAR} + \psi_i(c_i') y_i' \\
&\le \frac{2(n+1)}{\left(\sum_{j=1}^{n} {y_j}^2\right)^2} + \frac{(n+1)}{\left(\sum_{j=1}^{n} {y_j}^2\right)^2} \sum_{i=1} {y_i}^2 \mathrm{VAR} + \psi_i(c_i') y_i
\end{aligned}$$

Taking summation of both sides of these equations and using the fact that $c_i = c_i'$ for $i = 2, \ldots, n$, we obtain

$$(y_1 - y_1')(\psi_1(c_1) - \psi_1(c_1')) \le 0.$$

Assumption 2.1 and the above inequality establishes that the solution of problem (2.69) is weakly decreasing in the privacy cost.

Finally, notice that if the platform pays user i

$$-\text{VAR} + c_i \varepsilon_i^*(\mathbf{c}) + \int_{c_i} \varepsilon_i^*(z, \mathbf{c}_{-i}) dz + \frac{2}{\left(\sum_{j=1}^n \varepsilon_j^*(\mathbf{c})\right)^2} + \sum_{i=1}^n \left(\frac{\varepsilon_i(\mathbf{c})}{\sum_{j=1}^n \varepsilon_j(\mathbf{c})}\right)^2 \text{VAR}.$$

the expected payment becomes the same as the characterization of Proposition 2.1. This completes the proof. ■

Proof of Corollary 2.1

If $\psi_i(c_i) = \psi_j(c_j)$, then by swapping the i-th and j-th components of the solution the objective remains the same and therefore we can always let $y_i^* \geq y_j^*$. Now, suppose $\psi_i(c_i) < \psi_j(c_j)$. We let

$$\tilde{y}_\ell = \begin{cases} y_\ell^* & \ell \neq i, j \\ y_i^* & \ell = j \\ y_j^* & \ell = i. \end{cases}$$

The difference of the objective function evaluated at $(\tilde{y}_1, \ldots, \tilde{y}_n)$ and $(y_1^*, \ldots, y_n^*)$ becomes

$$(\psi_i(c_i) - \psi_i(c_j)) \left(y_j^* - y_i^*\right) \geq 0$$

where the inequality follows from the optimality condition. Inequality $\psi_i(c_i) < \psi_j(c_j)$, implies that $y_i^* \geq y_j^*$, proving the corollary. ■

Proof of Proposition 2.3

Problem (2.23) is the same as

$$\begin{aligned}\min_{S\geq 0}\min_{\mathbf{y}} \frac{n+1}{S^2}&\left(2+\sum_{i=1}^{n} y_i^2 \text{ VAR}\right)+\sum_{i=1}^{n}\psi_i(c_i)y_i\\ \text{s.t. } &\sum_{i=1}^{n} y_i = S\\ &y_i \geq 0, \text{ for all } i\in\mathcal{N}.\end{aligned}$$

Let us consider the optimization over $\mathbf{y}$ for a given S. The Lagrangian of this problem is

$$\sum_{i=1}^{n}\frac{(n+1)\text{VAR}}{S^2}y_i^2+\sum_{i=1}^{n}\psi_i(c_i)y_i-\lambda\left(\sum_{i=1}^{n}y_i-S\right)-\sum_{i=1}^{n}\mu_i y_i.$$

The KKT conditions imply that the optimal solutions $(y_1^*,\ldots,y_n^*)$, λ^* and μ_i^* satisfy

$$2(n+1)\frac{\text{VAR}}{S^2}y_i^*+\psi_i(c_i)-\lambda^*-\mu_i^*=0 \text{ for all } i\in\mathcal{N}.$$

If $y_i^*>0$, we have $\mu_i^*=0$ and therefore

$$y_i^*=\frac{\left(\lambda^*-\psi_i(c_i)\right)S^2}{2(n+1)\text{VAR}}.$$

If $y_i^*=0$, we have $\mu_i^*=\psi_i(c_i)-\lambda^*\geq 0$. Hence, if, we define

$$y(\lambda)=\left(\left(\frac{(\lambda-\psi_1(c_1))S^2}{2(n+1)\text{VAR}}\right)^+,\ldots,\left(\frac{(\lambda-\psi_n(c_n))S^2}{2(n+1)\text{VAR}}\right)^+\right) \text{ for any } \lambda, \tag{2.70}$$

$(y_i^*)_{i=1}^n$ would be equal to $y(\lambda^*)$. Next, define

$$S(\lambda)=\sum_{i=1}^{n}y_i(\lambda). \tag{2.71}$$

We can see that the function $S(\lambda)$ is increasing in λ and that λ^* is such that $S(\lambda^*)=S$. Once we find λ^*, (2.70) gives the optimal solution (subject to $\sum_{i=1}^{n}y_i=S$). To find λ^* we first sort the terms $\{\psi_i(c_i)\}_i$ in $\mathcal{O}(n \log n)$. Without loss of generality, let us assume

$$\psi_1(c_1) \leq \cdots \leq \psi_n(c_n).$$

We let i^* be the smallest element of $\mathcal{N}$ for which

$$S(\psi_j(c_j)) \geq S.$$

If no such element exists, then we let $i^* = n+1$. Therefore, for any $S > 0$, there exists $i^* > 1$ and $\lambda^* \in [\psi_{i^*-1}(c_{i^*-1}), \psi_{i^*}(c_{i^*})]$ such that in the optimal solution we have

$$y_i = 0 \text{ for } i > i^* \text{ and } y_i = \frac{(\lambda^* - \psi_i(c_i))\, S^2}{2(n+1)\text{VAR}} \text{ for } i \leq i^*,$$

with the convention that $\psi_{n+1}(\cdot) = \infty$. Also, using (2.70) and (2.71), S and λ^* satisfy the following relation

$$\sum_{i=1}^{i^*} \frac{(\lambda^* - \psi_i(c_i))S^2}{2(n+1)\text{VAR}} = S$$

which results in

$$\frac{1}{S} = \lambda^* A_{i^*} - B_{i^*},$$

with

$$A_{i^*} = \frac{i^*}{2(n+1)\text{VAR}}, B_{i^*} = \sum_{i=1}^{i^*} \frac{\psi_i(c_i)}{2(n+1)\text{VAR}}.$$

Hence, in this case, the original optimization for the given S can be cast as

$$\begin{aligned}
&\frac{2(n+1)}{S^2} + \frac{n+1}{S^2}\sum_{i=1}^{i^*}\left(\frac{(\lambda-\psi_i(c_i))S^2}{2(n+1)\text{VAR}}\right)^2\text{VAR} + \sum_{i=1}^{i^*}\frac{(\lambda-\psi_i(c_i))S^2}{2(n+1)\text{VAR}}\psi_i(c_i) && (2.72)\\
&= \frac{2(n+1)}{S^2} + \frac{S^2}{4(n+1)\text{VAR}}\sum_{i=1}^{i^*}(\lambda^2-\psi_i^2(c_i)) && (2.73)\\
&= 2(n+1)\left(\lambda A_{i^*}-B_{i^*}\right)^2 + \frac{1}{\left(\lambda A_{i^*}-B_{i^*}\right)^2}\left(\frac{A_{i^*}}{2}\lambda^2-\frac{\tilde{B}_{i^*}}{2}\right), && (2.74)
\end{aligned}$$

with

$$\tilde{B}_{i^*} := \sum_{i=1}^{i^*}\frac{\psi_i(c_i)^2}{2(n+1)\text{VAR}}.$$

Note that, as S moves from zero to infinity, λ^* also moves from $\psi_1(c_1)$ to infinite. Hence, instead of minimizing (2.72) over S, we could minimize (2.74) over λ. To do so, it suffices to solve

$$\begin{aligned}
\min_{\lambda}\ & 2(n+1)\left(\lambda A_{i^*}-B_{i^*}\right)^2 + \frac{1}{\left(\lambda A_{i^*}-B_{i^*}\right)^2}\left(\frac{A_{i^*}}{2}\lambda^2-\frac{\tilde{B}_{i^*}}{2}\right) && (2.75)\\
& \lambda \geq \psi_{i^*}(c_{i^*})\\
& \lambda \leq \psi_{i^*+1}(c_{i^*+1}),
\end{aligned}$$

for $i^* \in \{2, \cdots, n+1\}$ and pick the one with minimum value. As the last step, we establish that (2.75) can be solved in time $\mathcal{O}(1)$ which implies that the total optimization problem can be solved in time $\mathcal{O}(n)$.

To do so, note that the objective function of (2.75) can be written as

$$\begin{aligned}
&2(n+1)\left(\lambda A_{i^*}-B_{i^*}\right)^2 + \frac{1}{\left(\lambda A_{i^*}-B_{i^*}\right)^2}\left(\frac{A_{i^*}}{2}\lambda^2-\frac{\tilde{B}_{i^*}}{2}\right)\\
&= \frac{2(n+1)\left(\lambda A_{i^*}-B_{i^*}\right)^4 + \frac{A_{i^*}}{2}\lambda^2 - \frac{\tilde{B}_{i^*}}{2}}{\left(\lambda A_{i^*}-B_{i^*}\right)^2}. && (2.76)
\end{aligned}$$

One can see that the derivative of (2.76) is in the form of a polynomial of degree four divided by $\left(\lambda A_{i^*}-B_{i^*}\right)^3$. Hence, the derivative of (2.76) has at most four roots and they all can be characterized using the formulas for roots of a degree four polynomial. Therefore, to find the solution of (2.75), it suffices to compare the value of the objective function (2.76) at endpoints of the constraint interval $[\psi_{i^*}(c_{i^*}), \psi_{i^*+1}(c_{i^*+1})]$ and those roots of the derivative that lie within this interval. These are at most six points and thus the optimization problem (2.75) can be solved in time $\mathcal{O}(1)$. ■

Proof of Theorem 2.4

Using the payment identity in Proposition 2.1, we obtain

$$\begin{aligned}
\mathbb{E}_{c_i}\left[t_i(c_i)\right] =& \mathbb{E}[\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\mathbf{w})] - \mathrm{VAR} + \mathbb{E}_{c_i}[c_i\varepsilon_i(c_i)] + \ \mathbb{E}_{c_i}\left[\int_{z=c_i}\varepsilon_i(z)dz\right] \\
=& \mathbb{E}[\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\mathbf{w})] - \mathrm{VAR} \\
&+ \int_{\mathbf{z}_{-i}}\int_{z_i}\left(z_i\varepsilon_i(z_i,\mathbf{z}_{-i}) + \int_{y_i=z_i}\varepsilon_i(y_i,\mathbf{z}_{-i})dy_i\right)f_i(z_i)dz_i f_{-i}(\mathbf{z}_{-i})d\mathbf{z}_{-i} \\
\overset{(a)}{=}& \mathbb{E}[\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\mathbf{w})] - \mathrm{VAR} \\
&+ \int_{\mathbf{z}_{-i}}\int_{z_i}\left(z_i\varepsilon_i(z_i,\mathbf{z}_{-i}) + \varepsilon_i(z_i,\mathbf{z}_{-i})\frac{F_i(z_i)}{z_i}\right)f_i(z_i)dz_i f_{-i}(\mathbf{z}_{-i})d\mathbf{z}_{-i} \\
=& \mathbb{E}[\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\mathbf{w})] - \mathrm{VAR} + \int_{\mathbf{z}}\left(z_i + \frac{F_i(z_i)}{z_i}\right)\varepsilon_i(\mathbf{z})f(\mathbf{z})d\mathbf{z},
\end{aligned} \tag{2.77}$$

where (a) follows from changing the order of the integrals. Moreover, we have

$$\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\mathbf{w}) = \sum_{i=1}^{n}\mathrm{VAR}w_i^2(\mathbf{c}) + \sum_{i=1}^{n}\frac{2w_i^2(\mathbf{c})}{\varepsilon_i^2(\mathbf{c})}. \tag{2.78}$$

Substituting equations (2.77) and (2.78) in the platform's objective function, results in

$$\begin{aligned}
&\mathbb{E}_{\mathbf{c}}\left[\mathrm{MSE}(\mathbf{c};\boldsymbol{\varepsilon},\mathbf{w}) + \sum_{i=1}^{n}t_i(\mathbf{c})\right] = \\
&\mathbb{E}_{\mathbf{c}}\left[\sum_{i=1}^{n}w_i^2(\mathbf{c})(n+1)\mathrm{VAR} + (n+1)\frac{2w_i^2(\mathbf{c})}{\varepsilon_i^2(\mathbf{c})} + \psi(c_i)\varepsilon_i(\mathbf{c})\right] - n\mathrm{VAR}.
\end{aligned}$$

For any vector of reported privacy costs $(c_1, \ldots, c_n)$, we solve the point-wise optimization problem:

$$\begin{aligned}\min_{\mathbf{w},\mathbf{y}} \quad & \sum_{i=1}^{n} w_i^2 \left((n+1)\textsc{var} + \frac{2(n+1)}{y_i^2} \right) + \psi_i(c_i) y_i \\ & \sum_{i=1}^{n} w_i = 1 \\ & w_i \geq 0 \text{ for all } i \in \mathcal{N}.\end{aligned}$$

Let us first minimize the objective over w_i's. Using Cauchy-Schwarz inequality, for any given $\mathbf{y}$ we have

$$\left(\sum_{i=1}^{n} w_i^2 \left((n+1)\textsc{var} + \frac{2(n+1)}{y_i^2} \right) \right) \left(\sum_{i=1}^{n} \frac{1}{(n+1)\textsc{var} + \frac{2(n+1)}{y_i^2}} \right) \geq \left(\sum_{i=1}^{n} w_i \right)^2 = 1$$

and therefore the minimum of $\sum_{i=1}^{n} w_i^2 \left((n+1)\textsc{var} + \frac{2(n+1)}{y_i^2} \right)$ becomes

$$\frac{1}{\sum_{i=1}^{n} \frac{1}{(n+1)\textsc{var} + \frac{2(n+1)}{y_i^2}}}.$$

with solution

$$w_i = \frac{1}{\left((n+1)\textsc{var} + \frac{2(n+1)}{y_i^2} \right) \sum_{i=1}^{n} \frac{1}{(n+1)\textsc{var} + \frac{2(n+1)}{y_i^2}}} \text{ for all } i \in \mathcal{N}.$$

Therefore, the point-wise optimization problem becomes

$$\min_{\mathbf{y}} \frac{1}{\sum_{i=1}^{n} \frac{1}{(n+1)\textsc{var} + \frac{2(n+1)}{y_i^2}}} + \sum_{i=1}^{n} \psi_i(c_i) y_i.$$

A similar argument to that of Theorem 2.3 establishes that, under Assumption 2.1, the optimal x_i is weakly decreasing in c_i and therefore the corresponding payment, noise variance, and weight function satisfy the incentive compatibility and the individual rationality. ■

Proof of Proposition 2.4

Without loss of generality we assume

$$\psi_1(c_1) \leq \cdots \leq \psi_n(c_n).$$

We make use of the following two lemmas in this proof.

Lemma 2.4. *Suppose Assumption 2.1 holds. For any reported vector of privacy sensitivities $(c_1, \ldots, c_n)$, in the optimal local data acquisition mechanism, we have $\varepsilon_i^*(\mathbf{c}) \geq \varepsilon_j^*(\mathbf{c})$ for all $i, j \in \mathcal{N}$ such that $\psi_i(c_i) < \psi_j(c_j)$.*

Proof of Lemma 2.4: If $\psi_i(c_i) = \psi_j(c_j)$, then by swapping the i-th and j-th components of the solution the objective remains the same and therefore we can always let $y_i^* \geq y_j^*$. Now, suppose $\psi_i(c_i) < \psi_j(c_j)$. We let

$$\tilde{y}_\ell = \begin{cases} y_\ell^* & \ell \neq i, j \\ y_i^* & \ell = j \\ y_j^* & \ell = i. \end{cases}$$

The difference of the objective function evaluated at $(\tilde{y}_1, \ldots, \tilde{y}_n)$ and $(y_1^*, \ldots, y_n^*)$ becomes

$$\left(\psi_i(c_i) - \psi_i(c_j)\right)\left(y_j^* - y_i^*\right) \geq 0$$

where the inequality follows from the optimality condition. Inequality $\psi_i(c_i) < \psi_j(c_j)$, implies that $y_i^* \geq y_j^*$, proving the corollary. ■

Lemma 2.5. *For any $\lambda \in \mathbb{R}_+$ and any $\psi_i(c_i)$ the equation*

$$\frac{4z}{\left(\text{VAR}z^2 + 2\right)^2} = \frac{\psi_i(c_i)}{n+1}\lambda^2 \tag{2.79}$$

either has no solution or at most two solutions in $\mathbb{R}_+$. Furthermore:

(a) The solutions can be found in time $\mathcal{O}(1)$.

(b) *The smallest solution is strictly increasing in $\psi_i(c_i)$ and the largest solution is strictly decreasing in $\psi_i(c_i)$.*

Proof of Lemma 2.5: The derivative of the function $\frac{4z}{(\text{VAR}z^2+2)^2}$ with respect to z is

$$\frac{4}{(2+\text{VAR}z^2)^2}\left(1-\frac{4\text{VAR}z^2}{(2+\text{VAR}z^2)}\right),$$

which is positive if and only if $z \leq \sqrt{\frac{2}{3\text{VAR}}}$. Therefore, the function is zero at $z=0$, increases to $\frac{3\sqrt{3}}{8\sqrt{2\text{VAR}}}$ at $z=\sqrt{\frac{2}{3\text{VAR}}}$ and then decreases to 0 as $z \to \infty$. Therefore, either there is no solution or there are at most two solutions. To see the proof of part (a), note that finding the solutions of (2.79) is equivalent to finding the roots of a degree four polynomial that can be characterized using the formulas for roots of a degree four polynomial. Finally, to see the proof of part (b) notice that

$$\frac{4z}{(\text{VAR}z^2+2)^2}$$

is strictly increasing for $z \leq \sqrt{\frac{2}{3\text{VAR}}}$. The smallest solution of (2.79) is the intersection of this function over $z \leq \sqrt{\frac{2}{3\text{VAR}}}$ with the function level $\frac{\psi_i(c_i)}{n+1}\lambda^2$. As $\psi_i(c_i)$ increases, the intersecting z strictly increases. Further, the largest solution of (2.79) is the intersection of this function $z \geq \sqrt{\frac{2}{3\text{VAR}}}$ with the function level $\frac{\psi_i(c_i)}{n+1}\lambda^2$. As $\psi_i(c_i)$ increases, the intersecting z strictly decreases. ■

When (2.79) has two solutions, we let

$$y_i^{(l)}(\lambda) \text{ and } y_i^{(h)}(\lambda)$$

be the smallest and the largest solutions, respectively. If (2.79) has one solution then the above two solutions coincide.

We now proceed with the proof of the proposition. The KKT condition for problem

(2.30) implies that when $y_i^* \neq 0$, then

$$\frac{4y_i^*}{\psi_i(c_i)\left(2+\text{VAR}y_i^{*2}\right)^2} = \frac{1}{n+1}\left(\sum_{j=1}^{n}\frac{1}{\text{VAR}+\frac{2}{y_j^{*2}}}\right)^2. \tag{2.80}$$

We let

$$\lambda = \sum_{j=1}^{n}\frac{1}{\text{VAR}+\frac{2}{y_j^{*2}}}. \tag{2.81}$$

By using Lemma 2.4, we know that if there exists $i^* \in \{1,\dots,n\}$ such that $y_{i^*} = 0$, then we have $y_i^* = 0$ for $i > i^*$. For such i^*, by using Lemma 2.5, we know that for $i \leq i^*$, $y_i^* \in \{y_i^{(l)}(\lambda), y_i^{(h)}(\lambda)\}$.

We need to find the optimal λ and the corresponding optimal solution. In this regard, we search over all $i^* \in \{1,\dots,n\}$ and then find the optimal λ such that for $i > i^*$ we have $y_i^* = 0$ and for $i \leq i^*$, we have $y_i^* \in \{y_i^{(l)}(\lambda), y_i^{(h)}(\lambda)\}$.

Claim 1: Consider $i^* \in \{1,\dots,n\}$ and an optimal solution such that for $i > i^*$ we have $y_i^* = 0$ and for $i \leq i^*$ we have $y_i^* \in \{y_i^{(l)}(\lambda), y_i^{(h)}(\lambda)\}$. If $\psi_{i^*}(c_{i^*}) > \psi_{i^*-1}(c_{i^*-1})$, then for all $i \leq i^*-1$, we have $y_i^* = y_i^{(h)}(\lambda)$.

Proof of Claim 1: To prove this claim, we assume the contrary and reach a contradiction. In particular, suppose $y_i^* = y_i^{(l)}(\lambda)$ for $i \leq i^*-1$. We can write

$$y_i^{(l)}(\lambda) = y_i^* \overset{(a)}{>} y_{i^*}^* \geq y_{i^*}^{(l)}(\lambda)$$

where (a) follows from Lemma 2.4 together with $\psi_i(c_i) < \psi_{i^*}(c_{i^*})$ (In fact, Lemma 2.4 implies $y_i^* \geq y_{i^*}^*$. However, from the proof, one could see that, since $y_{i^*}^* > 0$, the inequality would be strict.) This is a contradiction by invoking part (b) of Lemma 2.5, completing the proof of Claim 1. ■

For the rest of the proof, we assume $\psi_1(c_1) < \cdots < \psi_n(c_n)$. We will cover the case that two or more of the $\psi_i(c_i)$'s are equal at the end. In this case, claim 1 implies that

$$y_i^* = \begin{cases} y_i^{(h)}(\lambda) & i < i^* \\ y_i^{(l)}(\lambda) \text{ or } y_i^{(h)}(\lambda) & i = i^* \\ y_i = 0 & i > i^*. \end{cases}$$

This provides the solution for a given λ. We next show how we can find the (approximately) optimal λ. In this regard, we first establish a lower bound and an upper bound on λ.

Claim 2: Consider $i^* \in \{1, \dots, n\}$ and an optimal solution such that for $i > i^*$ we have $y_i^* = 0$ and for $i \leq i^*$ we have $y_i^* \in \{y_i^{(l)}(\lambda), y_i^{(h)}(\lambda)\}$. The optimal λ satisfies

$$\lambda \in [\underline{y}_{i^*}, \bar{y}_{i^*}],$$

where

$$\bar{y}_{i^*} = y^{(h)}\left(\left(\frac{(n+1)3\sqrt{3}}{\psi_{i^*-1}(c_{i^*-1})8\sqrt{2\text{VAR}}}\right)^{1/2}\right) \text{ and } \underline{y}_{i^*} = \frac{n}{\text{VAR} + \left(\frac{\sqrt{2}n\left(\sum_{i=1}^{n}\psi_i(c_i)\right)}{(n+1)}\right)^{2/3}}.$$

Proof of Claim 2: As we proved in the proof of Corollary 2.1, the maximum of $\frac{4z}{(\text{VAR}z^2+2)^2}$ is $\frac{3\sqrt{3}}{8\sqrt{2\text{VAR}}}$. Therefore, in order to guarantee that (2.79) has a solution we must have

$$\lambda \leq \left(\frac{(n+1)3\sqrt{3}}{\psi_{i^*-1}(c_{i^*-1})8\sqrt{2\text{VAR}}}\right)^{1/2}.$$

We next derive a lower bound on λ. For $y_i = y$, the objective becomes

$$\frac{n+1}{n}\left(\text{VAR} + \frac{2}{y^2}\right) + \left(\sum_{i=1}^{n}\psi_i(c_i)\right)y$$

which is a convex program whose minimum is

$$\frac{n+1}{n}\left(\text{VAR} + 2\left(\frac{2(n+1)}{n\left(\sum_{i=1}^{n}\psi_i(c_i)\right)}\right)^{-2/3}\right) + \left(\sum_{i=1}^{n}\psi_i(c_i)\right)\left(\frac{2(n+1)}{n\left(\sum_{i=1}^{n}\psi_i(c_i)\right)}\right)^{1/3}.$$

Since the objective is

$$\frac{n+1}{\lambda} + \sum_{i=1}^{n} \psi_i(c_i) y_i$$

and $\psi_i(c_i) \geq 0$, the optimal λ is larger than

$$\begin{aligned}\underline{y}_{i^*} =& \frac{n+1}{\frac{n+1}{n}\left(\text{VAR} + 2\left(\frac{2(n+1)}{n\left(\sum_{i=1}^{n}\psi_i(c_i)\right)}\right)^{-2/3}\right) + \left(\sum_{i=1}^{n}\psi_i(c_i)\right)\left(\frac{2(n+1)}{n\left(\sum_{i=1}^{n}\psi_i(c_i)\right)}\right)^{1/3}} \\ =& \frac{n}{\text{VAR} + \left(\frac{\sqrt{2}n\left(\sum_{i=1}^{n}\psi_i(c_i)\right)}{(n+1)}\right)^{2/3}}.\end{aligned}$$

This completes the proof of Claim 2. ■

Equipped with Claims 1 and 2, we next search over the near optimal λ. Letting $\text{Grid}(i^*, \frac{\epsilon}{\Delta})$ be an $\frac{\epsilon}{\Delta}$ grid of $[\underline{y}_{i^*}, \bar{y}_{i^*}]$ where Δ is the maximum Lipschitz parameter for all functions $\frac{n+1}{\lambda}$ and $y_i^{((h))}(\lambda)$ and $y_i^{(l)}(\lambda)$ over $[\underline{y}_{i^*}, \bar{y}_{i^*}]$. With this definition, the following optimization

$$\min_{\lambda \in \text{Grid}(i^*, \frac{\epsilon}{\Delta})} \min \left\{ \frac{n+1}{\lambda} + \sum_{i=1}^{i^*} \psi_i(c_i) y_i^{(h)}(\lambda), \frac{n+1}{\lambda} + \sum_{i=1}^{i_1 - 1} \psi_i(c_i) y_i^{(h)}(\lambda) + \psi_i(c_i) y_i^{(l)}(\lambda) \right\}$$

achieves at most $(1+\epsilon)$ of the optimal objective. Finally, notice that Δ defined above is polynomial in n. Then proof completes by noting that the output of this procedure satisfies the monotonicity property in $\psi_i(c_i)$ because we do a grid search over λ and once we find λ the corresponding y_i's are decreasing in the virtual costs.

Finally, we conclude the proof by discussing the case that two or more of ψ_i's are equal. For simplicity, we consider the case that

$$\psi_1(c_1) < \cdots < \psi_i(c_i) < \psi_{i+1}(c_{i+1}) = \cdots = \psi_{i+k}(c_{i+k}) < \psi_{i+k+1}(c_{i+k+1}) < \cdots < \psi_n(c_n). \tag{2.82}$$

The argument here generalizes to the case when some of ψ_i's are equal on two or more different values.

For (2.82), we need to modify the algorithm when the for loop counter reaches i, i.e., the case that we take $y_1^* = \cdots = y_i^* = 0$ and $y_j^* > 0$ for $j > i$. In this case, Claim 1 would change as follow: For $i + k \leq j \leq i + 1$, we have $y_j^* \in \{y_j^{(l)}(\lambda), y_j^{(h)}(\lambda)\}$, and for $j > i + k$, we have $y_j^* = y_j^{(h)}(\lambda)$. Moreover, since $\psi_{i+1}(c_{i+1}) = \cdots = \psi_{i+k}(c_{i+k})$, we have $y_{i+1}^{(l)}(\lambda) = \cdots = y_{i+k}^{(l)}(\lambda) = y_{i+1:i+k}(\lambda)$ and $y_{i+1}^{(h)}(\lambda) = \cdots = y_{i+k}^{(h)}(\lambda) = y_{i+1:i+k}(\lambda)$.

Hence, we define an inner loop which considers $k + 1$ cases on the number of $\{y_j^*\}_{j=i+1}^{i+k}$ that are equal to $y_{i+1:i+k}(\lambda)$. Also, when this inner for loop ends, the outer loop jumps to $i + k + 1$ instead of $i + 1$, and thus, the total number of iterations still remains bounded by $2n$. ■

Proof of Proposition 2.5

Let

$$\hat{\theta}_{\text{central}} = \sum_{i=1}^{n} w_i x_i + \text{Laplace}(1/\eta),$$

with

$$\eta = \min_i \frac{\varepsilon_i}{w_i}. \tag{2.83}$$

First, note that, by definition, for any i, $\eta w_i \leq \varepsilon_i$. Hence, by Lemma 2.2, this estimator is $\boldsymbol{\varepsilon}$-differentially private. Hence, it suffices to show (2.33) holds. Note that

$$\mathbb{E}[|\hat{\theta}_{\text{central}} - \theta|^2] = \text{VAR} \sum_{i=1}^{n} w_i^2 + \frac{2}{\eta^2},$$
$$\mathbb{E}[|\hat{\theta}_{\text{local}} - \theta|^2] = \text{VAR} \sum_{i=1}^{n} w_i^2 + 2 \sum_{i=1}^{n} \frac{w_i^2}{\varepsilon_i^2}.$$

Comparing the right hand sides, to establish (2.33), we need to show

$$\frac{1}{\eta^2} \leq \sum_{i=1}^{n} \frac{w_i^2}{\varepsilon_i^2}.$$

To do so, note that,

$$\frac{1}{\eta^2} = \left(\frac{1}{\min_i \frac{\varepsilon_i}{w_i}} \right)^2 = \left(\max_i \frac{w_i}{\varepsilon_i} \right)^2,$$

which is clearly upper bounded by $\sum_{i=1}^{n} \frac{w_i^2}{\varepsilon_i^2}$. Thus, the proof is complete. ■

Proof of Proposition 2.6

We first state a more detailed version of Proposition 2.6.

Proposition 2.6. *Assume the virtual costs of users* $1, \cdots, n-1$ *are all equal to* ψ_1. *We also denote the virtual cost of user* n *by* ψ_n. *Denote the optimal privacy loss levels in the central and local settings by* $(\varepsilon_1^{\text{central}}, \cdots, \varepsilon_n^{\text{central}})$ *and* $(\varepsilon_1^{\text{local}}, \cdots, \varepsilon_n^{\text{local}})$, *respectively.*

1. *In the central setting, if* $\varepsilon_n^{\text{central}} = 0$, *then*

$$\psi_n \geq \psi_1 + \frac{1}{(n-1)\text{VAR}} \sqrt[3]{2\psi_1(n+1)^2}. \tag{2.84}$$

2. *In the local setting, there exists a universal constant* κ, *independent of problem's parameters, such that, if*

$$\psi_n \geq \psi_1 + \kappa \left(\frac{\psi^{1/3}}{n^{2/3}\text{VAR}} + \frac{\psi_1}{n} + \frac{\psi^{-1/3}}{n^{4/3}\text{VAR}^2} \right), \tag{2.85}$$

then $\varepsilon_n^{\text{local}} = 0$.

Therefore, there exists $N \in \mathbb{N}$ *such that for* $n \geq N$ *and*

$$\psi_n \in \left[\psi_1 + \kappa \left(\frac{\psi^{1/3}}{n^{2/3}\text{VAR}} + \frac{\psi_1}{n} + \frac{\psi^{-1/3}}{n^{4/3}\text{VAR}^2} \right), \psi_1 + \frac{1}{(n-1)\text{VAR}} \sqrt[3]{2\psi_1(n+1)^2} \right], \tag{2.86}$$

the optimal privacy loss level of user n *in the local setting is zero while her optimal privacy loss level in the central setting in non-zero.*

Proof: First, note that the optimal privacy loss levels in the central and local settings are the solutions of optimization problems (2.23) and (2.30), respectively. We first note that in both cases, without loss of generality, we can assume $\text{VAR} = 1$ by replacing y_i by $y_i\sqrt{\text{VAR}}$ and ψ_i by $\psi_i \text{VAR}^{3/2}$. Therefore, without loss of generality, we could assume $\text{VAR} = 1$ while studying optimization problems (2.23) and (2.30).

We start with the central setting. The characterization of solutions (2.27) implies that $\varepsilon_1^{\text{central}} = \cdots = \varepsilon_{n-1}^{\text{central}}$. Hence, the Lagrangian corresponding to optimization (2.23) can be rewritten as

$$L(y, z, \mu, \nu) := \frac{n+1}{((n-1)y+z)^2}\left(2 + (n-1)y^2 + z^2\right) + (n-1)\psi_1 y + \psi_n z - \mu y - \nu z, \quad (2.87)$$

where y and z denote the central privacy loss level of the first $n-1$ users and the last user, respectively. If $\varepsilon_n^{\text{central}} = 0$, then there exists a tuple (y^*, z^*, μ^*, ν^*) with $z^* = 0$ such that

$$\frac{\partial}{\partial y} L(y^*, z^*, \mu^*, \nu^*) = 0, \quad (2.88)$$

$$\frac{\partial}{\partial z} L(y^*, z^*, \mu^*, \nu^*) = 0. \quad (2.89)$$

Furthermore, since $z^* = 0$ and the optimal cost is finite, we have $y^* > 0$ which implies $\mu^* = 0$. Next, note that,

$$\frac{\partial}{\partial y} L(y, z, \mu, \nu) = -2(n+1)(n-1)\frac{2 + z^2 - yz}{((n-1)y+z)^3} + (n-1)\psi_1 - \mu.$$

Hence, (2.88) along with $z^* = 0$ and $\mu^* = 0$, implies

$$y^* = \frac{1}{n-1}\sqrt[3]{\frac{4(n+1)}{\psi_1}}. \quad (2.90)$$

Also, note that

$$\frac{\partial}{\partial z} L(y, z, \mu, \nu) = 2(n+1)\frac{(n-1)yz - 2 - (n-1)y^2}{((n-1)y+z)^3} + \psi_n - \nu.$$

Thus, (2.89) along with $z^* = 0$ and $\nu^* \geq 0$, implies

$$\psi_n \geq 2(n+1)\frac{2 + (n-1)(y^*)^2}{(n-1)^3(y^*)^3}$$

Plugging (2.90) into this bound completes the proof of (2.84).

To show (2.85), it suffices to show that if $\psi_n > \psi_1$ and $\varepsilon_n^{\text{local}} > 0$, then

$$\psi_n \leq \psi_1 + \mathcal{O}(1)\left(\frac{\psi^{1/3}}{n^{2/3}} + \frac{\psi_1}{n} + \frac{1}{\psi^{1/3}n^{4/3}}\right).$$

To do so, first assume $\psi_n > \psi_1$ and $\varepsilon_n^{\text{local}} > 0$. Then, by Lemma 2.4, we know $\varepsilon_n^{\text{local}} < \varepsilon_i^{\text{local}}$ for any $i \in \{1, \cdots, n-1\}$.[13] Next, by the characterization of solutions (2.32), we know $\varepsilon_1^{\text{local}} = \cdots = \varepsilon_{n-1}^{\text{local}}$. With a slight abuse of notation, we denote the local privacy loss level of $n-1$ first users and the last user by y and z, respectively, with $y^* > z^*$.

From (2.80) in the proof of Proposition 2.4, we know the following two equations hold

$$\frac{4y^*}{\psi_1(2+y^{*2})^2} = \frac{1}{n+1}\left((n-1)\frac{y^{*2}}{2+y^{*2}} + \frac{z^{*2}}{2+z^{*2}}\right)^2, \tag{2.91}$$

$$\frac{4z^*}{\psi_n(2+z^{*2})^2} = \frac{1}{n+1}\left((n-1)\frac{y^{*2}}{2+y^{*2}} + \frac{z^{*2}}{2+z^{*2}}\right)^2. \tag{2.92}$$

We next provide upper and lower bounds on y^*. To do so, first, by replacing $\frac{z^{*2}}{2+z^{*2}}$ by 0, we obtain

$$\frac{4y^*}{\psi_1(2+y^{*2})^2} \geq \frac{(n-1)^2 y^{*4}}{(n+1)(2+y^{*2})^2}$$

which implies

$$y^* \leq \sqrt[3]{\frac{4(n+1)}{\psi_1(n-1)^2}}. \tag{2.93}$$

Second, we note that $\frac{x^2}{x^2+2}$ is an increasing function of x over $(0, \infty)$. Hence, given that $y^* > z^*$, replacing $\frac{z^{*2}}{2+z^{*2}}$ by $\frac{y^{*2}}{2+y^{*2}}$ leads to an upper bound for the right hand side of (2.91). Therefore, we have

$$\frac{4y^*}{\psi_1(2+y^{*2})^2} \leq \frac{n^2 y^{*4}}{(n+1)(2+y^{*2})^2}$$

which implies

$$y^* \geq \sqrt[3]{\frac{4(n+1)}{\psi_1 n^2}}. \tag{2.94}$$

[13] It is worth mentioning that Lemma 2.4, in fact, implies $\varepsilon_n^{\text{local}} \leq \varepsilon_i^{\text{local}}$. However, by reviewing its proof, one could see that the inequality should be strict, given the assumption $\varepsilon_n^{\text{local}} > 0$.

Next, note that, we can rewrite (2.92) as

$$\frac{\sqrt{\frac{4(n+1)}{\psi_n}}\sqrt{z^*} - z^{*2}}{2 + z^{*2}} = (n-1)\frac{y^{*2}}{y^{*2}+2}. \tag{2.95}$$

By replacing the left hand side by the upper bound

$$\frac{1}{2}\sqrt{\frac{4(n+1)}{\psi_n}}\sqrt{z^*},$$

we obtain

$$\frac{1}{2}\sqrt{\frac{4(n+1)}{\psi_n}}\sqrt{z^*} \geq (n-1)\frac{y^{*2}}{y^{*2}+2}. \tag{2.96}$$

Next, we use the fact that $z^* < y^*$ and the upper bound on y^* (2.94) to further upper bound the left hand side of (2.96). In addition, we use the lower bound on y^* 2.93 to lower bound the right hand side of (2.96). Taking these two steps and simplifying the equation leads to the following result:

$$\frac{(4(n+1))^{2/3}}{2\psi_1^{1/6}(n-1)^{4/3}} + \sqrt{\psi_1}\frac{n^{4/3}}{(n-1)^{4/3}} \geq \sqrt{\psi_n}. \tag{2.97}$$

Using this inequality along with,

$$n^{4/3} = (n-1)^{4/3} + \mathcal{O}(n^{1/3}),$$

we obtain

$$\mathcal{O}(1)\left(\frac{1}{\psi_1^{1/6}n^{2/3}} + \frac{\sqrt{\psi_1}}{n}\right) + \sqrt{\psi_1} \geq \sqrt{\psi_n}. \tag{2.98}$$

Using the fact that $\sqrt{\psi_n} - \sqrt{\psi_1} = \frac{\psi_n - \psi_1}{\sqrt{\psi_1}+\sqrt{\psi_n}}$, we can rewrite (2.98) as

$$\psi_1 + \mathcal{O}(1)(\sqrt{\psi_1} + \sqrt{\psi_n})\left(\frac{1}{\psi_1^{1/6}n^{2/3}} + \frac{\sqrt{\psi_1}}{n}\right) \geq \psi_n. \tag{2.99}$$

Upper bounding $\sqrt{\psi_n}$ on the left hand side of (2.99) by using (2.98) completes the proof. ■

2.8.1 Additional Results and Details

Revelation principle for both central and local privacy settings

Suppose the strategy of user i is a function of its privacy cost denoted by $\beta_i(c_i)$. For a given estimator $\hat{\theta}$ and mechanism $(\boldsymbol{\varepsilon}, \mathbf{t})$, the action profile $\{\beta_i(\cdot)\}_{i=1}^n$ is an equilibrium if

$$\begin{aligned}&\mathbb{E}_{\mathbf{c}_{-i}}\left[\text{VAR}(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}}); \hat{\theta}) + \mathbf{c}_{\mathbf{i}}\varepsilon_{\mathbf{i}}(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}})) - \mathbf{t}_{\mathbf{i}}(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}}))\right] \\ &\leq \mathbb{E}_{\mathbf{c}_{-i}}\left[\text{VAR}(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta'_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}}); \hat{\theta}) + \mathbf{c}_{\mathbf{i}}\varepsilon_{\mathbf{i}}(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta'_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}})) - \mathbf{t}_{\mathbf{i}}(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta'_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}}))\right]\end{aligned}$$

for all $i \in \mathcal{N}, c_i, \beta'_i(\cdot)$. By letting $(\tilde{\boldsymbol{\varepsilon}}, \tilde{\mathbf{t}})$ be such that $\tilde{\varepsilon}_i(c_1, \ldots, c_n) = \varepsilon_i(\beta_1(c_1), \ldots, \beta_n(c_n))$ and $\tilde{t}_i(c_1, \ldots, c_n) = t_i(\beta_1(c_1), \ldots, \beta_n(c_n))$, the users will report truthfully and that the platform's objective is the same as the original mechanism. This establishes the revelation principle. ■

Computing the payment function and approximate incentive compatibility

Recall the incentive compatibility (IC) definition (2.17) states

$$\text{COST}(c_i, c_i; \boldsymbol{\varepsilon}, \mathbf{t}, \hat{\theta}) \leq \text{COST}(c'_i, c_i; \boldsymbol{\varepsilon}, \mathbf{t}, \hat{\theta}) \quad \text{for all } i \in \mathcal{N}, c_i, c'_i.$$

The approximate ϵ-IC definition allows for an ϵ violation of the original IC definition, i.e.,

$$\text{COST}(c_i, c_i; \boldsymbol{\varepsilon}, \mathbf{t}, \hat{\theta}) \leq \text{COST}(c'_i, c_i; \boldsymbol{\varepsilon}, \mathbf{t}, \hat{\theta}) + \epsilon \quad \text{for all } i \in \mathcal{N}, c_i, c'_i. \tag{2.100}$$

The following result highlights that in both the central and local settings if we possess an algorithm that provides the estimator and privacy loss allocations for any given vector of privacy sensitivities, we can efficiently compute payment functions that satisfy ϵ-IC. This means that the algorithm ensures approximate incentive compatibility with an error no greater than ϵ.

Lemma 2.6. *Suppose we have an algorithm that returns the estimator and the privacy loss levels for any given vector of privacy sensitivities. Then, for any ϵ, we can return payment functions in polynomial time such that ϵ-IC holds.*

Proof of Lemma 2.6: Recall that the payment function is

$$t_i(\mathbf{c}) = \mathrm{MSE}(\mathbf{c}, \boldsymbol{\varepsilon}, \hat{\theta}) - \textsc{var} + c_i \varepsilon_i(\mathbf{c}) + \int_{z=c_i}^{\infty} \varepsilon_i(z, \mathbf{c}_{-i}) dz.$$

All the terms, except the integral, can be computed based on the algorithm's output on the estimator and privacy allocations for the vector $\mathbf{c}$. The last step is to show that we can approximate the integral efficiently. To do so, we establish that, for any $\mathbf{c}_{-i}$ and δ, there exists $\bar{c}_i$ such that

$$\int_{z=\bar{c}_i}^{\infty} \varepsilon_i(z, \mathbf{c}_{-i}) dz \leq \delta.$$

To show this, first note that, even as c_i increases, there is a fixed upper bound M on the platform's cost in the optimization problems stated in Theorems 2.3 and 2.4, as the platform can always ignore the data of user i, i.e., put $y_i = 0$ in (2.23) and (2.30). As a result, we have

$$\varepsilon_i(z, \mathbf{c}_{-i}) \leq \frac{M}{\psi_i(z)} = \frac{M}{z + \frac{F_i(z)}{f_i(z)}} \leq \frac{M f_i(z)}{F_i(z)}.$$

Note that, for any δ', there exists $\bar{c}_i$ such that $F_i(\bar{c}_i) \geq 1 - \delta'$. Therefore, we have

$$\begin{aligned} \int_{z=\bar{c}_i}^{\infty} \varepsilon_i(z, \mathbf{c}_{-i}) dz &\leq \int_{z=\bar{c}_i}^{\infty} \frac{M f_i(z)}{F_i(z)} dz \\ &\leq \int_{z=\bar{c}_i}^{\infty} \frac{M f_i(z)}{1-\delta'} dz \\ \frac{M}{1-\delta'} \int_{z=\bar{c}_i}^{\infty} f_i(z) &= \frac{M}{1-\delta'} \left(1 - F_i(\bar{c}_i)\right) \leq \frac{M}{1-\delta'} \delta'. \end{aligned}$$

Setting δ' gives us the desired result. Now, suppose that we want an ϵ-approximate of the integral. Given the above result, we can choose $\bar{c}_i$ such that

$$\int_{z=\bar{c}_i}^{\infty} \varepsilon_i(z, \mathbf{c}_{-i}) dz \leq \frac{\epsilon}{2}.$$

Therefore, it suffices to show that we can approximate $\int_{z=c_i}^{\bar{c}_i} \varepsilon_i(z, \mathbf{c}_{-i}) dz$ up to $\epsilon/2$ error. For any δ, let $\mathcal{S}(\delta)$ be a mesh with sub-intervals of size maximum δ from c_i to $\bar{c}_i$, i.e.,

$$\mathcal{S}(\delta) = (I_0, I_1, \cdots, I_M),$$

where

$$c_i = I_0 < I_1 < \cdots < I_M = \bar{c}_i, \quad \text{with } I_j - I_{j-1} \leq \delta.$$

Note that we have

$$\sum_{j=1}^{M} (I_j - I_{j-1}) \inf_{z \in [I_{j-1}, I_j]} \varepsilon_i(z, \mathbf{c}_{-i}) \leq \int_{z=c_i}^{\bar{c}_i} \varepsilon_i(z, \mathbf{c}_{-i}) dz \leq \sum_{j=1}^{M} (I_j - I_{j-1}) \sup_{z \in [I_{j-1}, I_j]} \varepsilon_i(z, \mathbf{c}_{-i}).$$

Notice that, as shown earlier, the optimal $\varepsilon_i(z, \mathbf{c}_{-i})$ is decreasing in z for a fixed $\mathbf{c}_{-i}$. Thus, we can rewrite the above equation as

$$\sum_{j=1}^{M} (I_j - I_{j-1}) \varepsilon_i(I_j, \mathbf{c}_{-i}) \leq \int_{z=c_i}^{\bar{c}_i} \varepsilon_i(z, \mathbf{c}_{-i}) dz \leq \sum_{j=1}^{M} (I_j - I_{j-1}) \varepsilon_i(I_{j-1}, \mathbf{c}_{-i}).$$

Therefore, if we take $\sum_{j=0}^{M-1} (I_j - I_{j-1}) \varepsilon_i(I_j, \mathbf{c}_{-i})$ as an approximate of the integral, its error will be bounded by the difference of the left-hand side and the right-hand side, i.e.,

$$\begin{aligned} \sum_{j=1}^{M} (I_j - I_{j-1}) \varepsilon_i(I_{j-1}, \mathbf{c}_{-i}) - \int_{z=c_i}^{\bar{c}_i} \varepsilon_i(z, \mathbf{c}_{-i}) dz &\leq \sum_{j=1}^{M} (I_j - I_{j-1}) \left(\varepsilon_i(I_{j-1}, \mathbf{c}_{-i}) - \varepsilon_i(I_j, \mathbf{c}_{-i}) \right) \\ &\leq \delta \sum_{j=1}^{M} \left(\varepsilon_i(I_{j-1}, \mathbf{c}_{-i}) - \varepsilon_i(I_j, \mathbf{c}_{-i}) \right) = \delta \left(\varepsilon_i(c_i, \mathbf{c}_{-i}) - \varepsilon_i(\bar{c}_i, \mathbf{c}_{-i}) \right). \end{aligned}$$

As a result, by letting

$$\delta = \frac{\epsilon}{2 \left(\varepsilon_i(c_i, \mathbf{c}_{-i}) - \varepsilon_i(\bar{c}_i, \mathbf{c}_{-i}) \right)},$$

the exprerssion $\sum_{j=0}^{M-1} (I_j - I_{j-1}) \varepsilon_i(I_j, \mathbf{c}_{-i})$ becomes an $\epsilon/2$-approximate of the integral. Also, computing this sum requires solving the allocation problem M times, where each time can be done in polynomial time. Finally, notice that M is less than $\lceil (\bar{c}_i - c_i)/\delta \rceil$ which is order of $\Omega(\frac{1}{\epsilon})$. Hence, the above procedure establishes an FPTAS for finding the payment. ■

Alternative individual rationality constraint

Here, we consider an alternative individual rationality constraint to (2.18) in which the users benefit from the platform's estimator even if they do not participate. In this case, constraint

(2.18) becomes

$$\textsc{cost}(c_i, c_i; \boldsymbol{\varepsilon}^{(n)}, \mathbf{t}^{(n)}, \hat{\theta}^{(n)}) \leq \mathbb{E}_{\mathbf{c}_{-i}} \left[\mathrm{MSE}(\mathbf{c}_{-i}, \boldsymbol{\epsilon}_{-i}^{(n-1)}, \hat{\theta}^{(n-1)}) \right], \tag{2.101}$$

where the right-hand side is the MSE of an estimator with $n-1$ data points of users in $\mathcal{N} \setminus \{i\}$, noting that without participating in the mechanism, the user does not incur any privacy cost but also does not receive any payment. Here, we use the superscript (k) to show that a function has k inputs.

We next highlight how each one of our results extends to this setting. We do not repeat the proofs as they are identical to those presented earlier.

Proposition 1': For a given estimator $\hat{\theta} : \mathcal{X}^n \times \mathbb{R}_+^n \to \mathbb{R}$ defined for all n, a central or local privacy data acquisition mechanism $(\hat{\theta}^{(n)}, \boldsymbol{\varepsilon}^{(n)}, \mathbf{t}^{(n)})$ satisfies incentive compatibility (2.17) and individual rationality (2.101) if and only if

$$\begin{aligned} t_i(c_i) = \mathbb{E}_{\mathbf{c}_{-i}} \left[\mathrm{MSE}(\mathbf{c}, \boldsymbol{\varepsilon}^{(n)}, \hat{\theta}^{(n)}) \right] - \mathbb{E}_{\mathbf{c}_{-i}} \left[\mathrm{MSE}(\mathbf{c}_{-i}, \boldsymbol{\varepsilon}_{-i}^{(n-1)}, \hat{\theta}^{(n-1)}) \right] \\ + c_i \varepsilon_i(c_i) + \int_{z=c_i}^{\infty} \varepsilon_i^{(n)}(z) dz + d_i, \end{aligned} \tag{2.102}$$

for some constant $d_i \geq 0$, and $\varepsilon_i^{(n)}(z)$ is weakly decreasing) in z for all $i \in \mathcal{N}$.

Proposition 2': For a given estimator $\hat{\theta} : \mathcal{X}^n \times \mathbb{R}_+^n \to \mathbb{R}$ defined for all n, the optimal privacy loss in the central or local privacy data acquisition mechanism is the solution of

$$\begin{aligned} \min_{\{\varepsilon_i^{(n)}(\cdot)\}_{i=1}^n} \quad & \mathbb{E}_{\mathbf{c}} \left[(n+1)\mathrm{MSE}(\mathbf{c}, \boldsymbol{\varepsilon}^{(n)}, \hat{\theta}^{(n)}) + \sum_{i=1}^n \varepsilon_i(\mathbf{c}) \psi_i(c_i) \right] \\ & - \sum_{i=1}^n \mathbb{E}_{\mathbf{c}_{-i}} \left[\mathrm{MSE}(\mathbf{c}_{-i}, \boldsymbol{\varepsilon}_{-i}^{(n-1)}, \hat{\theta}^{(n-1)}) \right] \\ & \text{and } \varepsilon_i^{(n)}(z) = \mathbb{E}_{\mathbf{c}_{-i}} \left[\varepsilon_i^{(n)}(z, \mathbf{c}_{-i}) \right] \text{ is weakly decreasing in } z \text{ for all } i \in \mathcal{N}, \end{aligned}$$

where $\epsilon_j^{(n-1)}$ for all $i \in \mathcal{N}$ and $j \in \mathcal{N} \setminus \{i\}$ is the optimal privacy loss levels for users in $\mathcal{N} \setminus \{i\}$.

From the above proposition, it is evident that finding the optimal $\{\varepsilon_i^{(n)}(\cdot)\}_{i=1}^n$ decouples from finding the optimal $\{\varepsilon_i^{(k)}(\cdot)\}_i$ for any other $k < n$. Therefore, for both central and local settings, the characterization of the optimal privacy levels is the same as the ones given in Theorems 2.3 and 2.4, respectively. This in turn implies that Propositions 2.3 and 2.4 continue to hold. The only difference between this setting and our baseline model is that here in order to compute the payments, one needs to solve for the privacy loss levels for both n users and any subset of $n-1$ users. After solving these $n+1$ optimization problems, we can then use Proposition 1' to obtain the payment for n users.

Chapter 3

Bridging Central and Local Differential Privacy Mechanisms

3.1 Introduction

Central and local architectures point to two different types of privacy considerations: (i) local privacy concern that captures their concern about the revealed information about their personal data to the platform when they share data with the platform, and (ii) central privacy concern that captures their concern about the revealed information about their personal data to the public when the platform outputs an estimate (based on users' collected data).

In this chapter, we consider the design of data acquisition mechanisms and ask the following question:

> *What is the optimal data acquisition mechanism when users have heterogeneous privacy concerns regarding access to their raw data and the outcome of the platform's processing?*

Similar to the previous chapter, we have a platform whose goal is to estimate an underlying parameter of interest by collecting data from a set of users $\mathcal{N} = \{1, \ldots, n\}$ who own a noisy version of the underlying parameter. However, in this chapter, we adopt local and central Rényi differential privacy to measure these two types of privacy losses (Mironov [2017], Bun and Steinke [2016]). The reason for choosing Rényi differential privacy as opposed to the classical definition of differential privacy (Dwork et al. [2006a]) is twofold. First, our framework can cover a wide range of information measures by varying the Rényi divergence parameter. Second, it can be achieved by a Gaussian mechanism, which simplifies our analysis while capturing the main tradeoffs in the design of two-part data acquisition mechanisms.

Before formulating the platform's data acquisition problem, we derive optimal estimators for a given vector of heterogeneous local privacy loss levels. In particular, we establish a minimax lower bound for the estimation error and prove that first, privatizing users' data by adding a properly designed Gaussian noise to them and then using a properly designed weighted sum of these privatized data points achieves this lower bound. Equipped with this result, we then turn to the optimal data acquisition mechanism design when the platform uses an estimator that belongs to the class of linear estimators.

We cast this problem as a mechanism design problem as follows. Each user has a heterogeneous preference regarding the importance of the above two privacy concerns. For instance, if a user fully trusts the platform, then the first type of concern lessens, and the main concern would be about the information revealed about her personal data from the platform's estimate. On the other hand, if a user does not trust the platform at all, the first type of concern would be more than the second one. We model such a setting by assuming that each user i has a privacy sensitivity $c_i \in [0, 1]$ that determines the relative weight she puts on the local privacy concern (therefore, $1 - c_i$ is the weight she puts on the central privacy concern). The utility of user i is the payment she receives from the platform (in exchange for sharing her data), minus c_i times her local privacy loss, and again, minus $1 - c_i$ times her central privacy loss. The platform does not know the value of c_i and (knowing its distribution) must design a (Bayesian) data acquisition mechanism to elicit the true privacy sensitivities (that guide the optimal choice of the local and central privacy losses delivered to each user) and optimize its objective.

In particular, the platform designs a *two-part data acquisition mechanism* that comprises a payment scheme, a local privacy guarantee, and a central privacy guarantee as a function of the reported privacy sensitivity of users. The platform's goal is to minimize the sum of the mean estimation error of the underlying parameter and the expected total payment to users while satisfying the *incentive compatibility* and *individual rationality* constraints. Incentive compatibility ensures that users have no incentive to misreport their privacy sensitivity. Individual rationality ensures that the payment to users (and the delivered privacy guarantees) is such that users are willing to share their data with the platform.

The platform's problem is a functional optimization over three functions of the reported privacy sensitivities: payments, local privacy guarantees, and central privacy guarantees. We first establish a payment identity, similar to the classic mechanism design Myerson [1981], that pins down the payment function in terms of the local and central privacy guarantees by using incentive compatibility and individual rationality constraints. This reduces the space of the platform's decision variables. We then show that the platform's problem, in contrast to the classical mechanism design Myerson [1981], can be cast as an optimization problem that minimizes a non-convex objective (which depends on the *virtual cost* of users) for any reported vector of privacy sensitivities. This reformulation significantly reduces the space of decision variables that the platform needs to optimize. However, it still involves solving a non-convex optimization problem. We further use the structural properties of this non-convex optimization and use duality theory to develop a polynomial time algorithm to approximate the platform's problem. More precisely, we prove that the design of the optimal two-part data acquisition mechanism admits a Polynomial Time Approximation Scheme (PTAS).

The contribution of our work is threefold. First, we develop a minimax lower bound when users have heterogeneous local privacy losses and establish that a linear estimator (approximately) achieves this bound. Second, we formulate the design of the two-part data acquisition mechanism as the solution to a point-wise optimization problem and develop an algorithm to approximately find the optimal data acquisition mechanism (despite the fact that the corresponding optimization is non-convex). Third, we develop a modeling framework for data acquisition mechanisms when users have heterogeneous concerns for both local and central privacy losses. Our focus is on a mean estimation problem, but our framework is more general and can encompass other estimation problems.

The rest of the chapter proceeds as follows. In Section 3.2, we introduce our privacy measure and characterize the minimax optimal estimator. In Section 3.3, we introduce the platform and the user's utilities and formulate the platform's mechanism design problem. In Section 3.4, we establish how the platform's mechanism design problem turns into a point-wise (non-convex) optimization problem. In Section 3.5, we establish an algorithm to approximately find the optimal two-part data acquisition mechanism and then provide some illustrative examples. Section 3.6 concludes while the last section includes the deferred proofs.

3.2 Privacy Measure and the Minimax Optimal Estimator

We follow a model similar to the previous chapter, in which a platform is interested in estimating an underlying parameter $\theta \in \mathbb{R}$ by collecting data of n users, indexed by $\mathcal{N} = \{1, \cdots, n\}$. Similarly, for any $i \in \mathcal{N}$, we denote user i's *personal data* by $X_i \in \mathcal{X}$, and we assume X_i is given by $X_i = \theta + Z_i$ where $Z_1, \cdots, Z_n$ are independent and identically distributed zero-mean random variables with variance $\textsc{var}$ (and we again assume $|Z_i| \leq 1/2$ for any $i \in \mathcal{N}$).

3.2.1 Local and Central Privacy Losses

Before formalizing the utilities/objectives of the platform and the users, let us recall the notions of central and local privacy losses that we adopt in this chapter. The local one is the information leaked about a user's personal data to the platform when she shares her data with the platform, and the central one is the information leaked about a user's personal data to the public when she shares her data with the platform, and the platform releases its estimate to the public. To further illustrate that, consider a platform that wants to learn the efficacy of a drug by collecting patients' medical record. In this context, the local privacy loss corresponds to patients not trusting the hospital to keep their medical records private, and the central privacy loss corresponds to patients trusting the hospital but not trusting that the learning outcome, which is the output of the study, will keep their personal data private (e.g., the hospital output may reveal information about their medical record that can be exploited by insurance companies, hurting the users). Depending on how much different users trust the platform, they might care differently about these two privacy losses. For instance, if a user fully trusts the platform, then her main privacy concern would be the central one, while a user who does not trust the platform at all would be more concerned with the local one as the public only observes the aggregated estimate, as opposed to the platform which observes each user's (shared) data separately.

As stated earlier, in this chapter, we use another variant of the differential privacy framework to quantify these privacy losses. More specifically, we work with a popular one in the machine learning and data science literature, named Rényi differential privacy (RDP) Mironov [2017]. Let us first recall the definition of Rényi divergence.

Definition 3.1 (Rényi divergence)**.** *Let P and Q be two distributions over $\mathbb{R}$ with densities p and q. For any $\alpha \in (1, \infty]$, the Rényi α-divergence between P and Q is denoted by $D_\alpha(P||Q)$ and is given by*

$$D_\alpha(P||Q) := \frac{1}{\alpha - 1} \log \int \left(\frac{p(x)}{q(x)}\right)^\alpha q(x)dx.$$

For two random variables X and Y, $D_\alpha(X||Y)$ denotes the α-divergence between their distributions.

Throughout the chapter, we fix the parameter $\alpha > 1$ that we use in the definition of Rényi divergence and quantify the privacy losses. We next define two notions of differential privacy, known as central and local, to capture the two aforementioned types of privacy losses. Local differential privacy corresponds to the privacy loss of a user when she shares her data with the platform through a randomized mapping, known as a *channel*.

Definition 3.2 (local privacy)**.** *Let $\varepsilon \geq 0$ and $\alpha \in (1, \infty]$. A randomized channel $\mathcal{C} : \mathcal{X} \to \mathbb{R}$ is locally (ε, α)-Rényi (differentially) private if for any $x, x' \in \mathcal{X}$,*

$$D_\alpha(\mathcal{C}(x)||\mathcal{C}(x')) \leq \varepsilon.$$

Conceptually, having a smaller ϵ in the above definition implies that the output's distribution almost remains the same when the user's personal data changes, and therefore a limited amount of information leaks about the user's personal data.

Central differential privacy corresponds to the other privacy loss mentioned above. It bounds the change in the distribution of the platform's output, i.e., the released estimate, by changing one user's data. We next provide the formal definition.

Definition 3.3. *Let $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n \in \mathbb{R}_+^n$ and $\alpha \in (1, \infty]$. A randomized algorithm $\mathcal{A} : \mathcal{X}^n \to \mathbb{R}$ is centrally $(\boldsymbol{\varepsilon}, \alpha)$-Rényi (differentially) private if for any two datasets $x_{1:n}, x'_{1:n} \in \mathcal{X}^n$ that only differ in the i-th coordinate (i.e., data of user i),*

$$D_\alpha(\mathcal{A}(x_{1:n}) || \mathcal{A}(x'_{1:n})) \leq \varepsilon_i.$$

3.2.2 Minimax Optimal Estimator

In general, there are many ways to guarantee either local or central Rényi differential privacy. This implies that the space of mechanisms that the platform can employ to deliver privacy is very large. However, one key observation, as we prove in this section, is that the (worst-case) estimation error is minimized over the class of linear estimators with the Gaussian mechanism, defined next.

Definition 3.4 (Linear estimators with Gaussian mechanism). *Let $\boldsymbol{\varepsilon} \in \mathbb{R}_+^n$ and $\mathbf{w} \in \mathbb{R}_+^n$ such that $\sum_{i=1}^n w_i = 1$. A $(\boldsymbol{\varepsilon}, \mathbf{w})$-linear estimator with Gaussian mechanism is*

$$\hat{\theta}(x_1, \ldots, x_n) := \sum_{i=1}^n w_i \hat{x}_i \text{ where } \hat{x}_i = x_i + \mathcal{N}\left(0, \frac{\alpha}{2\varepsilon_i}\right) \text{ for all } i \in \mathcal{N}.$$

Figure 3-1 depicts this class of estimators.

Let us first state the privacy guarantees of a linear estimator with a Gaussian mechanism.

Lemma 3.1. *With a $(\boldsymbol{\varepsilon}, \mathbf{w})$-linear estimator with Gaussian mechanism, the local privacy delivered to user $i \in \mathcal{N}$ is $\varepsilon_i^{(l)} = \varepsilon_i$ and the central privacy delivered to user $i \in \mathcal{N}$ is*

$$\varepsilon_i^{(c)} = \frac{w_i^2}{\sum_{j=1}^n \frac{w_j^2}{\varepsilon_j}}.$$

We nest establish that for a given vector of local privacy losses $(\varepsilon_i^{(l)})_{i=1}^n$, a linear estimator is *optimal* with respect to mean square error. To formalize this statement, we first need

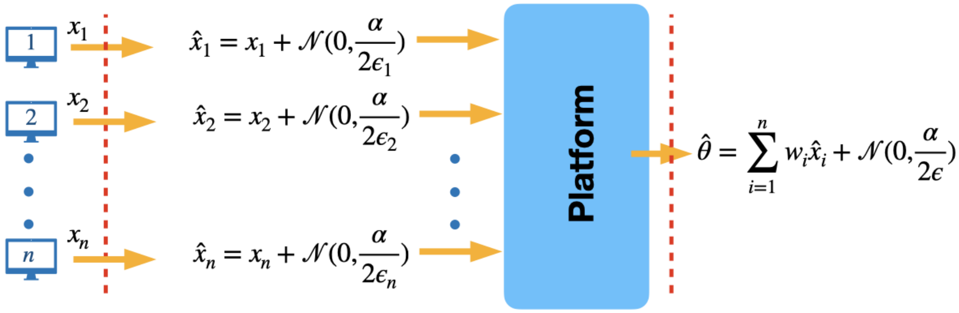

1
2
n
x_1
x_2
x_n
$\hat{x}_1 = x_1 + \mathcal{N}(0, \frac{\alpha}{2\epsilon_1})$
$\hat{x}_2 = x_2 + \mathcal{N}(0, \frac{\alpha}{2\epsilon_2})$
$\hat{x}_n = x_n + \mathcal{N}(0, \frac{\alpha}{2\epsilon_n})$
Platform
$\hat{\theta} = \sum_{i=1}^{n} w_i \hat{x}_i + \mathcal{N}(0, \frac{\alpha}{2\epsilon})$

Figure 3-1: The interaction between the users and the platform in the two-part private data acquisition.

to define the *minimax* estimation error as our notion of optimality. Let $\mathcal{P}$ be a class of distributions over $\mathcal{X}$. For any $P \in \mathcal{P}$, we denote its mean by $\theta(P)$. A locally $(\varepsilon_i^{(l)})_{i=1}^n$- RDP estimator can be cast as $\hat{\theta}((\mathcal{C}_i(x_i))_{i=1}^n)$, where $\mathcal{C}_i(.)$ is the randomized channel corresponding to user i. Let $\mathcal{Q}((\varepsilon_i^{(l)})_{i=1}^n)$ be the class of such locally $(\varepsilon_i^{(l)})_{i=1}^n$- RDP estimators. The minimax estimation error is

$$\mathcal{L}(\mathcal{P}, \mathcal{Q}, (\varepsilon_i^{(l)})_{i=1}^n) := \inf_{\hat{\theta}, \{\mathcal{C}_i\}_{i=1}^n \in \mathcal{Q}((\varepsilon_i^{(l)})_{i=1}^n)} \sup_{P \in \mathcal{P}} \mathbb{E}_{(X_i \sim P)_{i=1}^n, \hat{\theta}}[|\hat{\theta}((\mathcal{C}_i(X_i))_{i=1}^n) - \theta(P)|^2], \tag{3.1}$$

where the expectation is taken over both the randomness of data and the estimator (including private channels). In other words, the optimal estimator is the one that has the lowest worst-case error among all estimators that satisfy the privacy requirements. With this definition in mind, we next state our optimality result.

Theorem 3.1. *Assume $\alpha \geq 2$ and $\varepsilon_i^{(l)} \leq 1$ for all i. Let $\mathcal{P}_1$ be the family of distributions over $[-\frac{1}{2}, \frac{1}{2}]$ and $\mathcal{C}_1, \cdots, \mathcal{C}_n$ be independent channels. Then, there exists a universal constant c such that*

$$\mathcal{L}(\mathcal{P}, \mathcal{Q}, (\varepsilon_i^{(l)})_{i=1}^n) \geq c \min\left\{\frac{1}{\sum_{i=1}^n \varepsilon_i^{(l)}}, 1\right\}.$$

Furthermore, there exists a linear estimator with a Gaussian mechanism such that

$$\mathbb{E}_{(X_i \sim P)_{i=1}^n, \hat{\theta}}[|\hat{\theta}((\mathcal{C}_i(X_i))_{i=1}^n) - \theta(P)|^2] \leq \mathcal{O}(1) \frac{\alpha}{\sum_{i=1}^n \varepsilon_i^{(l)}}.$$

Proof Sketch of Theorem 3.1: The proof of the lower bound uses a technique called Le

Cam's method Yu [1997] that enables us to reduce the lower bounds problem to a hypothesis testing problem. To do so, we first replace the supremum of P over $\mathcal{P}$ by an average over two distributions $P_1, P_2 \in \mathcal{P}$ (which we choose at the end). More formally, we have

$$\sup_{P\in\mathcal{P}} \mathbb{E}_{(X_i\sim P)_{i=1}^n,\hat{\theta}}\left[\left|\hat{\theta}((\mathcal{C}_i(X_i))_{i=1}^n) - \theta(P)\right|^2\right] \geq \\ \frac{1}{2}\sum_{j=1}^{2}\mathbb{E}_{(X_i\sim P_j)_{i=1}^n,\hat{\theta}}\left[\left|\hat{\theta}((\mathcal{C}_i(X_i))_{i=1}^n) - \theta(P_j)\right|^2\right]. \tag{3.2}$$

Now, let Q_j denote the distribution of $\hat{\theta}((\mathcal{C}_i(X_i))_{i=1}^n)$ when $X_1, \cdots, X_n$ are drawn from P_j. Using this notation, we can rewrite the right-hand side of the above inequality as

$$\sup_{P\in\mathcal{P}} \mathbb{E}_{(X_i\sim P)_{i=1}^n,\hat{\theta}}\left[\left|\hat{\theta}((\mathcal{C}_i(X_i))_{i=1}^n) - \theta(P)\right|^2\right] \geq \frac{1}{2}\sum_{j=1}^{2}\mathbb{E}_{Y\sim Q_j}\left[|Y-\theta(P_j)|^2\right]. \tag{3.3}$$

We further lower bound the right-hand side by a term that is proportional to the error of a *nearest neighbor* estimator for a hypothesis testing problem in which a sample of Y is given, and we are asked to determine whether the underlying distribution is P_1 or P_2. More specifically, the nearest neighbor estimator returns $j \in \{1,2\}$ for which $|Y - \theta(P_j)|$ is smaller. Next, we use the Le Cam result, which says that the error of any estimator for the above problem is lower bounded by $\frac{1}{2} - \frac{1}{2}\|Q_1 - Q_2\|_{\mathrm{TV}}$. This is how we reduce the lower bound problem to the problem of developing an *upper bound* for the total variation distance between Q_1 and Q_2.

As the next step, we use the connection between the *total variation distance* and the *Hellinger distance*. Recall that the Hellinger distance between two distributions μ and ν is given by

$$d_{\mathrm{hel}}(\mu,\nu)^2 := \int(\sqrt{d\mu(x)} - \sqrt{d\nu(x)})^2.$$

The key result and our novel contribution in this part is to show the following inequality

$$\|Q_1 - Q_2\|_{\mathrm{TV}}^2 \leq 4\left(\sum_{i=1}^{n}\varepsilon_i^{(l)}\right)\, d_{\mathrm{hel}}(P_1,P_2)^2.$$

The strength of this result is that it bounds the total variation distance between two complex distributions Q_1 and Q_2 by a term that only depends on ε_i's and the Hellinger distance between the two distributions P_1 and P_2 that we are yet to choose. Plugging this back into the lower bound and choosing P_1 and P_2 as two (carefully tuned) Bernoulli distributions completes the proof of the lower bound. ■

3.3 Data Acquisition Mechanism With Two-Part Privacy Guarantees

Here, we first describe the utility functions of the users and the platform and then formulate the platform's optimal data acquisition mechanism.

3.3.1 Utility of the Users and the Platform

As we described earlier, each user suffers from both local and central privacy losses when sharing her data. Each user has a heterogeneous *privacy sensitivity* for these two types of privacy losses. To model such heterogeneity, for each $i \in \mathcal{N}$, we let $c_i \in [0, 1]$ be her *relative local privacy sensitivity*, representing the relative weight that user i assigns to the (per unit cost of) local privacy loss. We also let $1 - c_i$ be her *relative central privacy sensitivity*, representing the relative weight that user i assigns to the (per unit cost of) central privacy loss. Therefore, $c_i \approx 1$ implies that user i suffers mostly from the local privacy loss relative to the central privacy loss. Differently, $c_i \approx 0$ implies that user i suffers mostly from the central privacy loss relative to the local privacy loss. In what follows, we use the term *privacy sensitivity* instead of relative local privacy sensitivity.

For each $i \in \mathcal{N}$, the privacy sensitivity c_i is independently drawn from a publicly known distribution whose support is $[0, 1]$ with cumulative distribution and probability density functions $F_i(\cdot)$ and $f_i(\cdot)$.[1] We also let $\mathbf{c} = (c_1, \ldots, c_n)$ be the vector of privacy sensitivities. The privacy sensitivity of each user is her private information, i.e., the platform does not

[1] We assume that the distributions of the underlying privacy sensitivities are known, while the realized privacy sensitivity of a user is private, and, because of this, we design an incentive compatible mechanism to elicit the true privacy sensitivity. The assumption that the underlying distribution of types is known is common in Bayesian mechanism design. In practice, the platform can learn such distribution from multiple interactions with the user and by using simple mechanisms.

know it. This is because individuals have different views regarding how trustworthy the platform is in protecting their personal data.

The platform's objective is to design a mechanism to collect users' data by paying them to compensate for their privacy losses without knowing the privacy sensitivity of users. To introduce the platform's objective formally, we adopt the formalism of Bayesian mechanism design pioneered by Myerson [1981]. More specifically, the platform designs and announces a payment function, a local privacy loss function, and a central privacy loss function that are mappings from the reported privacy sensitivities of users. The users then report their privacy sensitivities (which may or may not be truthful). Based on the payment function, the platform compensates the users (the compensation could be monetary or some free or discounted service provided to the user). Based on the local and central privacy functions, the platform designs randomized channels and randomized estimation algorithms that deliver guaranteed local and central privacy losses while minimizing the sum of the mean squared error and the total expected payments. Given this interaction, we next formally introduce a data acquisition mechanism with two-part data privacy guarantees.

Definition 3.5 (two-part private data acquisition mechanism)**.** We call a tuple $(\hat{\theta}, \boldsymbol{\varepsilon}^{(l)}, \boldsymbol{\varepsilon}^{(c)}, \mathbf{t})$ a *two-part private data acquisition mechanism* such that:

1. For all $i \in \mathcal{N}$, $\varepsilon_i^{(l)} : \mathbb{R}_+^n \to \mathbb{R}_+$ is a function that maps the vector of privacy sensitivities $\mathbf{c}$ to a local privacy loss for user i, $\varepsilon_i^{(l)}(\mathbf{c})$, with $\boldsymbol{\varepsilon}^{(l)}(\cdot) = (\varepsilon_i^{(l)}(\cdot))_{i=1}^n$.

2. For all $i \in \mathcal{N}$, $\varepsilon_i^{(c)} : \mathbb{R}_+^n \to \mathbb{R}_+$ is a function that maps the vector of privacy sensitivities $\mathbf{c}$ to a central privacy loss for user i, $\varepsilon_i^{(c)}(\mathbf{c})$, with $\boldsymbol{\varepsilon}^{(c)}(.) = (\varepsilon_i^{(c)}(\cdot))_{i=1}^n$.

3. $\hat{\theta} : \mathcal{X}^n \times \mathbb{R}_+^n \times \mathbb{R}_+^n \to \mathbb{R}$ is a centrally $(\boldsymbol{\varepsilon}^{(c)}(\mathbf{c}), \alpha)$-Rényi differentially private estimator that maps acquired locally $(\varepsilon_i^{(l)}(\mathbf{c}), \alpha)$-Rényi differentially private data of user i for $i \in \mathcal{N}$ to an estimate $\hat{\theta}(\mathbf{x}, \boldsymbol{\varepsilon}^{(l)}(\mathbf{c}), \boldsymbol{\varepsilon}^{(c)}(\mathbf{c}))$.

4. For all $i \in \mathcal{N}$, $t_i : \mathbb{R}_+^n \to \mathbb{R}_+$ is a function that maps the vector of privacy sensitivities $\mathbf{c}$ to a payment for user i, $t_i(\mathbf{c})$, with $\boldsymbol{t}(.) = (t_i(\cdot))_{i=1}^n$.

Notice that we have not specified the estimator and the mechanisms that deliver (local and central) Rényi differential privacy. In the rest of this subsection, we introduce the utilities and the platform's problem for general estimators and mechanisms to deliver differential privacy. Later, we focus on linear estimator and Gaussian mechanisms and explicitly solve the platform's problem.

Each user that participates in a two-part private data acquisition mechanism suffers from both local and central privacy losses and needs to be compensated by the platform. In particular, the utility of user i from participation when her privacy sensitivity is c_i and she reports c_i' is given by

$$u_i(\boldsymbol{\varepsilon}^{(l)}(c_i', \mathbf{c}_{-i}), \boldsymbol{\varepsilon}^{(c)}(c_i', \mathbf{c}_{-i}), \mathbf{t}, \hat{\theta}) = \mathbb{E}_{\mathbf{c}_{-i}}[t_i(\mathbf{c}_{-i}, c_i')) - c_i \varepsilon_i^{(l)}(\mathbf{c}_{-i}, c_i') - (1 - c_i)\varepsilon_i^{(c)}(\mathbf{c}_{-i}, c_i')],$$

where the term $t_i(\mathbf{c}_{-i}, c_i'))$ is the payment from the platform, the term $c_i \varepsilon_i^{(l)}(\mathbf{c}_{-i}, c_i')$ is the relative local privacy sensitivity of the user multiplied by her local privacy loss, and the term $(1 - c_i)\varepsilon_i^{(c)}(\mathbf{c}_{-i}, c_i')$ is her relative central privacy sensitivity multiplied by her central privacy loss. A user $i \in \mathcal{N}$ that does not participate in the mechanism neither compromises her privacy nor gets compensation. Therefore, we normalize the utility of a user who does not participate in the mechanism to 0.

The goal of the platform is to minimize the sum of the mean squared error and the overall payment to users. We let $\gamma \in \mathbb{R}_+$ represent the relative weight of the mean estimation error to the payments in the platform's objective.[2] Therefore, the platform's objective is

$$\mathbb{E}_{\mathbf{c}}[\gamma \mathrm{MSE}(\boldsymbol{\varepsilon}^{(l)}(\mathbf{c}), \boldsymbol{\varepsilon}^{(c)}(\mathbf{c}), \hat{\theta}) + \sum_{i=1}^{n} t_i(\mathbf{c})],$$

where the first term is the mean square error of estimator $\hat{\theta}$ given the reported vector of privacy sensitivity and the resulting local and central privacy losses $\boldsymbol{\varepsilon}^{(l)}$ and $\boldsymbol{\varepsilon}^{(c)}$, i.e.,

$$\mathrm{MSE}(\boldsymbol{\varepsilon}^{(l)}(\mathbf{c}), \boldsymbol{\varepsilon}^{(c)}(\mathbf{c}), \hat{\theta}) = \mathbb{E}_{\mathbf{x}}[|\hat{\theta}(\hat{\mathbf{x}}, \boldsymbol{\varepsilon}^{(l)}, \boldsymbol{\varepsilon}^{(c)}) - \theta|^2].$$

Also, each summand of the second term is the compensation that the platform gives to a user to incentivize her to participate and report her privacy sensitivity truthfully.

[2] Notice that by changing the parameter γ, our framework includes a wide range of platform's objectives with differing relative weights between the estimation error and the total payments.

3.3.2 Platform's Problem

In general, the strategy of users can be very complicated. However, as we prove next, similar to the classical mechanism design setting, the revelation principle holds, and therefore the platform can focus on direct revelation mechanisms where individuals reporting their type truthfully is a (Bayesian Nash) equilibrium.

Lemma 3.2. ***[Revelation principle]*** *For any two-part private data acquisition mechanism) and any corresponding user equilibrium, there exists an incentive-compatible two-part private data acquisition mechanism which is equivalent from the point of view of both the platform and the users, when the users tell the truth.*

The above lemma follows from a simple argument, similar to Myerson [1981], but it greatly simplifies the space of mechanisms that the platform needs to consider: there is no loss of generality in focusing on the class of direct incentive compatible mechanisms, meaning the platform's optimization problem can be written as

$$\min_{\boldsymbol{\varepsilon}^{(l)}(\cdot),\boldsymbol{\varepsilon}^{(c)}(\cdot),\mathbf{t}(\cdot)} \quad \mathbb{E}_{\mathbf{c}}[\gamma \mathrm{MSE}(\boldsymbol{\varepsilon}^{(l)}(\mathbf{c}), \boldsymbol{\varepsilon}^{(c)}(\mathbf{c}), \hat{\theta}) + \sum_{i=1}^{n} t_i(\mathbf{c})] \tag{3.4}$$

$$u_i(\boldsymbol{\varepsilon}^{(l)}(\mathbf{c}), \boldsymbol{\varepsilon}^{(c)}(\mathbf{c}), \mathbf{t}, \hat{\theta}) \geq u_i(\boldsymbol{\varepsilon}^{(l)}(c_i', \mathbf{c}_{-i}), \boldsymbol{\varepsilon}^{(c)}(c_i', \mathbf{c}_{-i}), \mathbf{t}, \hat{\theta}) \tag{3.5}$$

$$u_i(\boldsymbol{\varepsilon}^{(l)}(\mathbf{c}), \boldsymbol{\varepsilon}^{(c)}(\mathbf{c}), \mathbf{t}, \hat{\theta}) \geq 0 \quad \text{for all } i \in \mathcal{N}, c_i, \tag{3.6}$$

where the constraints in (3.5) represent the *incentive compatibility.* These constraints guarantee that each user i has no incentive to misrepresent her privacy sensitivity when others report truthfully (reporting truthfully is an equilibrium of the game among the users). Also, the constraints in (3.6) represent *individual rationality*, which ensures that each user receives a non-negative utility from participating in the platform's mechanism and sharing her data.

3.4 From the Mechanism Design Problem to an Optimization Problem

For a given estimator $\hat{\theta}$, the platform's decision comprises the local and central privacy loss functions $\boldsymbol{\varepsilon}^{(l)}(\cdot)$ and $\boldsymbol{\varepsilon}^{(c)}(\cdot)$ together with the payment functions $\mathbf{t}(\cdot)$. We next show that this problem can be equivalently formulated as an optimization problem over the vector of local privacy losses and central privacy losses (as opposed to functions). In the rest of the chapter, we impose the following assumption which is well-known in the mechanism design literature and simplifies the analysis.[3]

Assumption 3.1. *For any user $i \in \mathcal{N}$, the* virtual cost *defined as $\psi_i(c) = c + \frac{F_i(c)}{f_i(c)}$ is increasing in c, where $f_i(\cdot)$ and $F_i(\cdot)$ are probability density and cumulative distribution functions of c_i, respectively.*

The above assumption holds for a wide class of distributions such as the ones with log-concave density functions (e.g., uniform).

Theorem 3.2. *Suppose Assumption 3.1 holds. For a given estimator $\hat{\theta} : \hat{\mathcal{X}}^n \times \mathbb{R}_+^n \times \mathbb{R}_+^n \to \mathbb{R}$, in the optimal two-part data acquisition mechanism, for a given vector of reported privacy sensitivities $\mathbf{c}$, the local and central privacy losses are the solution of*

$$\min_{\{\boldsymbol{\varepsilon}^{(l)}\}_{i=1}^n, \{\boldsymbol{\varepsilon}^{(c)}\}_{i=1}^n} \quad \gamma \mathrm{MSE}(\boldsymbol{\varepsilon}^{(l)}, \boldsymbol{\varepsilon}^{(c)}, \hat{\theta}) + \sum_{i=1}^n \varepsilon_i^{(l)} \psi_i(c_i) + \sum_{i=1}^n \varepsilon_i^{(c)} (1 - \psi_i(c_i)) \tag{3.7}$$

Proof Sketch of Theorem 3.2: We introduce the following *interim functions*

$$t_i(c_i) = \mathbb{E}_{\mathbf{c}_{-i}}[t(c_i, \mathbf{c}_{-i})], \quad \varepsilon_i^{(l)}(c_i) = \mathbb{E}_{\mathbf{c}_{-i}}[\varepsilon_i^{(l)}(c_i, \mathbf{c}_{-i})], \quad \text{and } \varepsilon_i^{(c)}(c_i) = \mathbb{E}_{\mathbf{c}_{-i}}[\varepsilon_i^{(c)}(c_i, \mathbf{c}_{-i})].$$

We first establish a *payment identity* that determines the optimal payment in terms of the optimal local and central privacy losses. In particular, by evaluating the first-order condition corresponding to the incentive compatibility constraint (3.5), we establish that this constraint

[3] Without this assumption, extending the results requires the ironing technique of Myerson [1981].

holds if and only if

$$t_i(c_i) = t_i(0) + \varepsilon_i^{(c)}(c_i) - \varepsilon_i^{(c)}(0) + c_i(\varepsilon_i^{(l)}(c_i) - \varepsilon_i^{(c)}(c_i)) - \int_0^{c_i} (\varepsilon_i^{(l)}(z) - \varepsilon_i^{(c)}(z))dz,$$

and $\varepsilon_i^{(l)}(z) - \varepsilon_i^{(c)}(z)$ is weakly decreasing in z. We then plug this payment identity back into the platform's objective, use the individual rationality constraint, and rewrite the platform's expected utility in terms of the privacy loss functions and the virtual cost of users. This is still a functional optimization problem in terms of $\varepsilon^{(l)}(\cdot)$ and $\varepsilon^{(c)}(\cdot)$. However, we establish that, under Assumption 3.1, we can solve this functional optimization point-wise (i.e., for any given $\mathbf{c}$). ■

Theorem 3.2 highlights the tradeoff in the platform's problem: by decreasing the local privacy loss, the second term of the objective decreases (this term corresponds to the payment to users) while the first term (i.e., the mean squared error) increases. The role of central privacy loss is more nuanced, and there are two cases. If the coefficient $1 - \psi_i(c_i)$ is non-negative, by decreasing the central privacy loss, the third term of the objective decreases while the first term increases. If the coefficient $1 - \psi_i(c_i)$ is negative, increasing the central privacy loss decreases both the third term and the first term. However, we cannot increase the central privacy loss level without limits because the central privacy loss level is always below the local privacy loss level. Therefore, the platform's optimal mechanism should find the "right" balance between these terms.

3.5 Finding the (Approximately) Optimal Two-Part Data Acquisition Mechanism

So far, we have established that the minimax optimal estimator is a linear estimator with a Gaussian mechanism and that the platform's problem, when the virtual costs are monotone, involves solving a point-wise optimization over the local and central privacy guarantees delivered to different users. In this section, we develop an algorithm to solve the platform's problem efficiently.

The following is a direct corollary of Theorem 3.2.

Corollary 3.1. *Suppose Assumption 3.1 holds. For any reported vector of privacy sensitivities* $\mathbf{c}$*, the optimal local privacy loss levels are* $\varepsilon_i^{(l)}(\mathbf{c}) = y_i^*$ *and the optimal central privacy loss levels are*

$$\varepsilon_i^{(c)} = \frac{w_i^{*2}}{\sum_{j=1}^n \frac{w_j^{*2}}{y_j^*}} \text{ where } (w_1^*, \dots, w_n^*) \text{ and } (y_1^*, \dots, y_n^*) \text{ are the optimal solution of}$$

$$\begin{aligned} &\min_{\mathbf{w},\mathbf{y}} \ \mathrm{VAR}\gamma \sum_{i=1}^n w_i^2 + \frac{\gamma\alpha}{2}\sum_{i=1}^n \frac{w_i^2}{y_i} + \sum_{i=1}^n (1-\psi_i(c_i)) \frac{w_i^2}{\sum_{j=1}^n \frac{w_j^2}{y_j}} + \sum_{i=1}^n \psi_i(c_i) y_i \qquad (3.8) \\ &s.t. \ w_i, y_i \geq 0, \text{ for all } i \in \mathcal{N} \text{ and } \sum_{i=1}^n w_i = 1. \end{aligned}$$

Let us highlight the difference between our characterization and that of classic mechanism design (e.g., Myerson [1981]). In classic mechanism design, the designer's problem becomes linear optimization. However, in our setting, the designer's problem is a non-linear and non-convex optimization. This makes the problem of finding the optimal two-part data acquisition mechanism challenging. Before addressing this computational challenge, let us revisit the form of the Gaussian mechanism that we have adopted: the platform adds Gaussian noise locally and then outputs a convex combination of the privatized users' data without adding any noise centrally. More specifically, one may guess that the platform may benefit by having a central noise added to the final output in addition to the local noises. In the following subsection, we establish that there is another Gaussian mechanism for any Gaussian mechanism that only adds local noises and achieves a weakly lower cost.

3.5.1 Optimality of Having Only Local Noises in the Gaussian Mechanism

The platform has the opportunity of adding Gaussian noise to both the personal data of each user and the final estimator and ex-ante one may guess that it is optimal to use both of these instruments. However, as we establish next, interestingly, in the optimal two-part data acquisition mechanism, it is always optimal to only add noises locally.

A linear estimator with a Gaussian mechanism that adds both local and central noises is of the form

$$\hat{\theta}(x_1,\ldots,x_n) := \sum_{i=1}^{n} w_i \hat{x}_i + \mathcal{N}\left(0, \frac{\alpha}{2\varepsilon}\right)$$

$$\text{where } \sum_{i=1}^{n} w_i = 1 \text{ and } \hat{x}_i = x_i + \mathcal{N}\left(0, \frac{\alpha}{2\varepsilon_i}\right) \forall i \in \mathcal{N}.$$

Proposition 3.1. *In the optimal two-part data acquisition mechanism that adopts a linear estimator with a Gaussian mechanism that adds both local and central noises, we have $\varepsilon = \infty$.*

Proposition 3.1 has an important implication in terms of the design of data market architecture when users have both central and local privacy losses: it is optimal to add noise locally! Adding a noise centrally to the final estimator has an advantage because the weights in the final estimator give the platform a lever to deliver heterogeneous central privacy guarantees to users. Despite this advantage, we establish that adding noise centrally is never optimal. This is because the platform prefers to add the noise locally to contribute to both central and local privacy guarantees delivered to users.

3.5.2 Computing the Optimal Privacy Loss Function

The implementation of the optimal two-part private data acquisition mechanism requires solving Problem (3.8), which is a non-convex program. However, we use the structure of the problem to develop a polynomial time algorithm to solve it approximately. To guide the analysis, we first define the auxiliary variable

$$S := \sum_{i=1}^{n} \frac{w_i^2}{y_i}.$$

We first characterize the solution of the optimization problem (3.8) for a given S, and then perform a grid search for S to find an approximate solution to the main problem. In particular, for a fixed S, the Lagrangian of (3.8) is given by

$$\sum_{i=1}^{n} w_i^2 \left(\text{VAR}\gamma + \frac{1-\psi_i(c_i)}{S}\right) + \sum_{i=1}^{n} \psi_i(c_i) y_i + p\left(\sum_{i=1}^{n} \frac{w_i^2}{y_i} - S\right) - q\left(\sum_{i=1}^{n} w_i - 1\right) \\ - \sum_{i=1}^{n} u_i w_i - \sum_{i=1}^{n} v_i y_i, \tag{3.9}$$

where p and q are the Lagrangian multipliers for the constraints $\sum_{i=1}^{n} \frac{w_i^2}{y_i} = S$ and $\sum w_i = 1$, respectively. In addition, for any i, the variables u_i and v_i are the Lagrange multipliers for the constraints $w_i \geq 0$ and $y_i \geq 0$, respectively. Furthermore, using the complementary slackness, for the optimal solution, we have $u_i w_i = v_i y_i = 0$.

By using Karush-Kuhn-Tucker (KKT) conditions, and taking the derivative of (3.9) with respect to y_i and w_i, we can write the primal variables y_i and w_i in terms of the single parameter S as

$$y_i = w_i \sqrt{\frac{p}{\psi_i(c_i)}} \text{ and } w_i = \frac{\frac{q}{2} - \sqrt{\psi_i(c_i)p}}{\text{VAR}\gamma + \frac{1-\psi_i(c_i)}{S}},$$

where p and q as functions of S are given by

$$p = \left(\frac{\sum_{i=1}^{n} \zeta_i \sqrt{\psi_i(c_i)}}{S - \sum_{i=1}^{n} \sqrt{\psi_i(c_i)}\xi_i}\right)^2 \text{ and } q = \frac{2 + 2\sum_j \nu_j \sqrt{\psi_j(c_j)p}}{\sum_j \nu_j}, \tag{3.10}$$

where

$$\nu_i = \frac{1}{\gamma\text{VAR} + \frac{1-\psi_i(c_i)}{S}},\ \zeta_i = \frac{\nu_i}{\sum_j \nu_j},\ \text{and } \zeta_i \left(\sum_{j=1}^{n} \nu_j (\sqrt{\psi_j(c_j)} - \sqrt{\psi_i(c_i)})\right). \tag{3.11}$$

As a consequence, we can write down the optimal w_i's and y_i's (and also the dual variables) for a given S. As the final step of the analysis, we establish lower and upper bounds on S and then perform a grid search to find the approximate optimal solution. In particular, let us denote the objective of Problem (3.35) for $y_i = 1$, $w_i = \frac{1}{n}$, and $\varepsilon = 1$ by M. We establish that an upper bound on S is given by $\bar{S} = \frac{M}{\gamma\alpha/2}$ and a lower bound on S is given by $\underline{S} = \frac{1}{Mn}$. With these notations in mind, the procedure to find the approximately optimal mechanism

Algorithm 3: Computing the optimal two-part private data acquisition mechanism

```
**Input:** The vector of privacy sensitivities $(c_1, \ldots, c_n)$
for $S \in \text{Grid}\left([\underline{S}, \bar{S}], \delta\right)$ do
    Given the definition of $\nu_i$ and $p$ in (3.10) and (3.11), consider the solution
        $$w_i(S) = \frac{\nu_i + \nu_i \sum_j \nu_j \sqrt{\psi_j(c_j)p}}{\sum_j \nu_j} - \nu_i \sqrt{\psi_i(c_i)p}, \quad y_i(S) = w_i(S)\sqrt{\frac{p}{\psi_i(c_i)}}.$$
    Let OBJ$(S)$ be the objective of Problem (3.8) evaluated for this solution.
end
**Output:** $\{y_i(S^*), w_i(S^*)\}_{i=1}^n$, where $(S^*) = \arg\min_{(S)} \text{OBJ}(S)$.
```

is summarized in Algorithm 3.

Theorem 3.3. *For any vector of reported privacy sensitivities and $\epsilon > 0$, Algorithm 3 finds local privacy loss levels and the differentially private linear estimator of the two-part data acquisition mechanism whose cost (i.e., the platform's objective) is at most $1 + \delta$ of the optimal cost in time* $\text{poly}(n, \frac{1}{\delta})$.

Notice that the approximation factor in Theorem 3.3 depends on the underlying parameters. Therefore, we have a Polynomial Time Approximation Scheme (PTAS) for finding the optimal two-part data acquisition mechanism in the class of linear estimators.

3.5.3 A Numerical Example

Here, we give an example with two users to illustrate the performance of the optimal two-part data acquisition mechanism in terms of the guaranteed privacy levels and payments as functions of the reported privacy sensitivities. In particular, let us consider users $i = 1, 2$ and let c_i be uniformly distributed over $[0, 1]$ so that the virtual costs are $\psi_i(c_i) = 2c_i$. We also let $\gamma = 1$, $\alpha = 2$, and $\text{VAR} = 1/4$.

Figure 3-2 shows the weight of user 1's data in the optimal two-part data acquisition mechanism. As we observe, by increasing c_1, user 1 cares more about local privacy, and therefore, the weight of her data in the platform's optimal mechanism decreases. Similarly, by increasing c_2, the platform prefers to predominantly get information from user 1 and therefore increases the weight of user 1's data (and decreases the weight of user 2's data) in

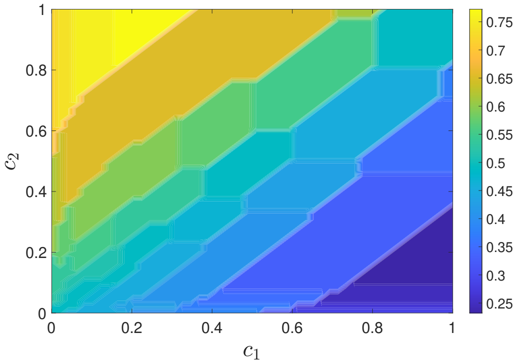


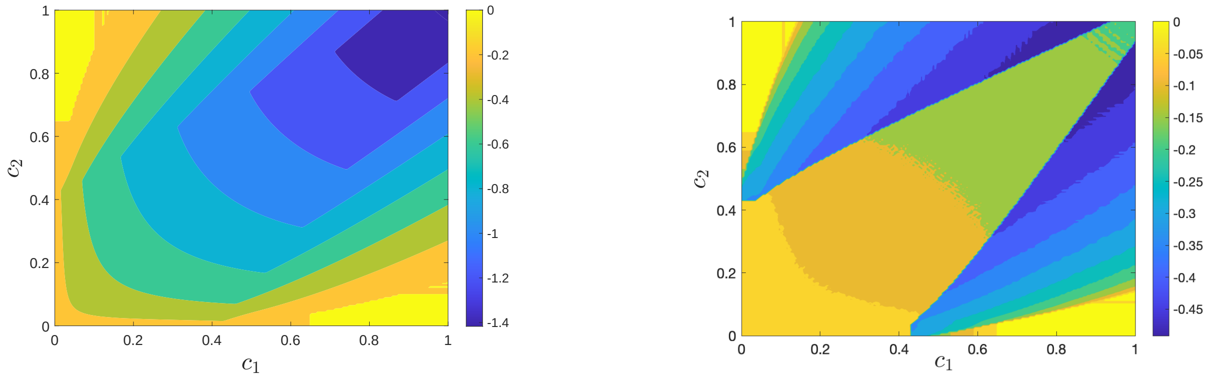
Figure 3-2: Weight of user 1's data in the optimal mechanism as a function of c_1 and c_2 for two users with $\gamma = 1$, $\alpha = 2$, and $\textsc{var} = 1/4$.

the optimal mechanism.

Figure 3-3 shows the platform's objective (i.e., the solution of Problem (3.8)) and its variance as a function of the privacy sensitivities. As we observe, higher privacy sensitivities (which means that users care more about local privacy compared to central privacy) lead to a higher platform cost and a higher estimation error. This is because guaranteeing local privacy is more demanding compared to central privacy.

Figure 3-4 illustrates the expected utility of user 1 (similarly user 1) as a function of her privacy sensitivity. We observe that, unlike classical mechanism design settings, the utility is a continuous function of the user's type (as opposed to a threshold function). Again, we observe that higher privacy sensitivity implies better local privacy, which is more demanding and decreases the user's expected utility.

Figure 3-5 illustrates the platform's expected objective and the user's expected utility in the optimal two-part data acquisition mechanism as a function of γ, the coefficient of the mean-squared error in the platform's objective (note that we have a symmetric setting and therefore the expected utility of users one and two are the same). We observe that as γ increases, the platform's expected objective naturally increases. This is because by increasing γ, the constraints of the platform's problem remain the same while its objective increases. The impact of increasing γ on the user's expected utility is more nuanced. As we increase γ, the platform cares more about the mean-square error and therefore tries to learn the underlying parameter more accurately. This has two opposing effects. First, as a result



(a)

(b)

Figure 3-3: (a) the platform's objective and (b) the platform's estimation error as a function of (c_1, c_2) for two users with $\gamma = 1$, $\alpha = 2$, and VAR $= 1/4$.

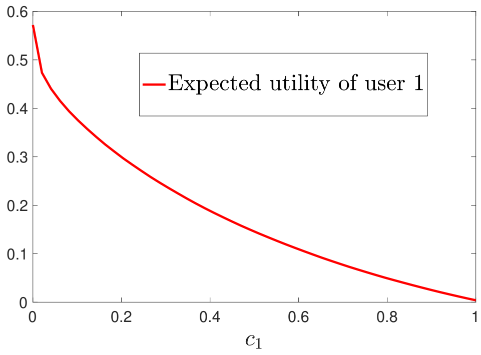
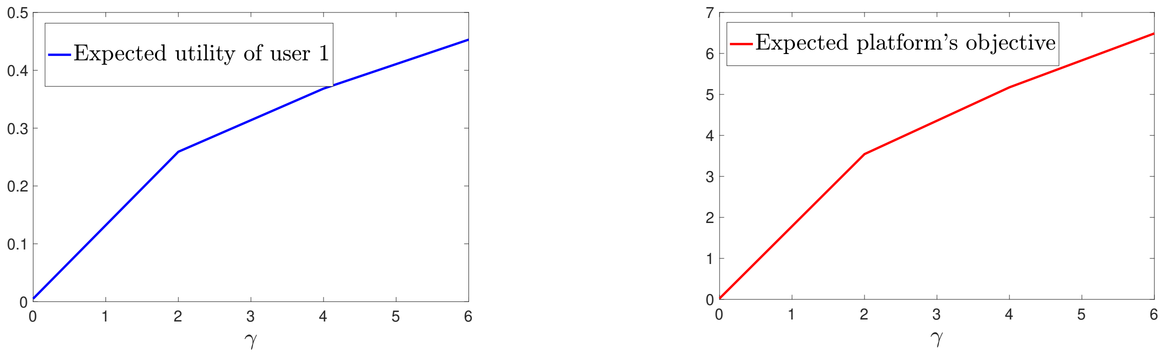


Figure 3-4: User 1's expected utility a function of c_1 for two users with $\gamma = 1$, $\alpha = 2$, and VAR $= 1/4$.



(a)

(b)

Figure 3-5: (a) user one's expected utility and (b) the platform's expected objective as a function of γ for two users with $\alpha = 2$ and VAR $= 1/4$.

of learning the underlying parameter, the user's data will leak more, decreasing the user's utility by increasing their privacy loss. Second, the platform is willing to pay users more to acquire their data more accurately, increasing user utility. As panel (a) of Figure 3-5 shows, the second force dominates, and the user's utility increases as γ increases (meaning when the platform cares more about reducing mean-squared error compared to the payment).

3.6 Conclusion

In this chapter, we develop a unified framework to study the design of data acquisition mechanisms when users have both local and central privacy concerns and are heterogeneous in how they value these two privacy concerns. We use Rényi differential privacy to measure the privacy loss of users and first establish a minimax lower bound that motivates us to focus on linear estimators. We then establish a point-wise optimization problem whose solution fully characterizes the optimal data acquisition mechanism that constitutes a payment scheme to compensate users for their privacy losses, a local privacy guarantee, and a central privacy guarantee all as a function of users' preferences for local and central privacy concerns. We then establish that, even though the corresponding optimization problem is non-convex, the platform's problem admits a Polynomial Time Approximation Scheme. We focused on data acquisition to estimate the mean population. However, our framework is more general and allows for considering other estimates including, for instance, vector estimates and higher moments of the underlying population distribution. In particular, our Theorem 3.2 converts the data acquisition mechanism design problem into a (potentially) non-convex optimization problem.

3.7 Proofs

Proof of Lemma 3.1

The proof follows from the following lemma which is adapted from Mironov [2017].

Lemma 3.3. *For a function $f : \mathcal{X}^n \to \mathbb{R}$, we define its sensitivity with respect to the i-th* *coordinate as*

$$L_i(f) := \sup\left\{|f(x_{1:n}) - f(x'_{1:n})| \ : \ \textit{for all } x_{1:n} \textit{ and } x'_{1:n} \textit{ differing only at } i\textit{-th coordinate}\right\}.$$

For any $\alpha \in (1, \infty]$, $\mathcal{A}(x_{1:n}) = f(x_{1:n}) + W$ *with* $W \sim \mathcal{N}(0, \sigma^2)$ *is* $\left((\frac{\alpha L_i(f)^2}{2\sigma^2})_{i=1}^n, \alpha\right)$*-RDP.*

Proof of Theorem 3.1

We establish the lower bound by using the Le Cam's method Yu [1997] which reduces the lower bound problem to a hypothesis testing problem between two distributions. To prove the lower bound, we need to show that for any $\hat{\theta} \in \mathcal{Q}((\varepsilon_i^{(l)})_{i=1}^n)$, we have

$$\sup_{P \in \mathcal{P}} \mathbb{E}_{(X_i \sim P)_{i=1}^n, \hat{\theta}} \left[\left| \hat{\theta}((\mathcal{C}_i(X_i))_{i=1}^n) - \theta(P) \right|^2 \right] \geq c \min\left(\frac{1}{\sum_{i=1}^n \varepsilon_i^{(l)}}, 1 \right). \tag{3.12}$$

To show this result, we replace $\sup_{P \in \mathcal{P}}$ by an average over two carefully chosen distributions in $\mathcal{P}$. More formally, let P_1 and P_2 be two distributions of choice in $\mathcal{P}$ with $\gamma := \frac{1}{2}|\theta(P_1) - \theta(P_2)|$. Note that

$$\begin{aligned} &\sup_{P \in \mathcal{P}} \mathbb{E}_{(X_i \sim P)_{i=1}^n, \hat{\theta}} \left[\left| \hat{\theta}((\mathcal{C}_i(X_i))_{i=1}^n) - \theta(P) \right|^2 \right] \\ &\quad \geq \frac{1}{2} \sum_{j=1}^2 \mathbb{E}_{(X_i \sim P_j)_{i=1}^n, \hat{\theta}} \left[\left| \hat{\theta}((\mathcal{C}_i(X_i))_{i=1}^n) - \theta(P_j) \right|^2 \right] = \frac{1}{2} \sum_{j=1}^2 \mathbb{E}_{Y \sim Q_j} \left[|Y - \theta(P_j)|^2 \right], \end{aligned} \tag{3.13}$$

where, for any $j \in \{1, 2\}$, Q_j denotes the distribution of $\hat{\theta}((\mathcal{C}_i(X_i))_{i=1}^n)$ when $X_1, \cdots, X_n$ are drawn from P_j. We next lower bound the right-hand side of (3.13) by Markov's inequality

$$\sup_{P \in \mathcal{P}} \mathbb{E}_{(X_i \sim P)_{i=1}^n, \hat{\theta}} \left[\left| \hat{\theta}((\mathcal{C}_i(X_i))_{i=1}^n) - \theta(P) \right|^2 \right] \geq \gamma^2 \frac{1}{2} \sum_{j=1}^2 \mathbb{P}\left(|Y - \theta(P_j)| \geq \gamma \right). \tag{3.14}$$

Now, consider a hypothesis testing problem with the goal of determining whether the underlying distribution is P_1 or P_2, given an observation of Y. One possible approach is to choose $j \in \{1, 2\}$ for which $|Y - \theta(P_j)|$ is smaller. It can be shown that the probability of an incorrect estimate by this approach is upper bounded by $\frac{1}{2} \sum_{j=1}^2 \mathbb{P}\left(|Y - \theta(P_j)| \geq \gamma \right)$ on the right-hand side of (3.14). Furthermore, a seminal result by Le Cam states that the infimum probability of incorrect decision among all possible mappings for the aforementioned hypothesis testing problem is given by $\frac{1}{2} - \frac{1}{2}\|Q_1 - Q_2\|_{\mathrm{TV}}$. Therefore, we obtain the following lower bound

$$\mathcal{L}(\mathcal{P}, \mathcal{Q}, (\varepsilon_i^{(l)})_{i=1}^n) \geq \gamma^2 \left(\frac{1}{2} - \frac{1}{2}\|Q_1 - Q_2\|_{\mathrm{TV}} \right). \tag{3.15}$$

Next, we provide an upper bound on $\|Q_1 - Q_2\|_{\mathrm{TV}}$. To do so, we use the connection between the total variation distance and the Hellinger distance. Hellinger distance has a number of well-known desirable properties. In particular, we use the following two:

- For any two distributions μ_1 and μ_2, we have

$$\|\mu - \nu\|_{\mathrm{TV}} \leq d_{\mathrm{hel}}(\mu, \nu). \tag{3.16}$$

- Let $\mu := \mu_1 \times \cdots \times \mu_n$ and $\nu := \nu_1 \times \cdots \times \nu_n$. Then,

$$d_{\mathrm{hel}}(\mu, \nu)^2 = 2 - 2\prod_{i=1}^n (1 - \frac{1}{2} d_{\mathrm{hel}}(\mu_i, \nu_i)^2). \tag{3.17}$$

Let us go back to the problem of upper bounding $\|Q_1 - Q_2\|_{\mathrm{TV}}$. The following lemma is the key result in our proof:

Lemma 3.4. *Let $\alpha \geq 2$ and suppose $\mathcal{C}(.) : \mathcal{X} - \to \mathbb{R}$ is an (ε, α)-RDP channel. For $j \in \{1, 2\}$, let ν_j be the distribution of $\mathcal{C}(X)$ when $X \sim \mu_j$. Then,*

$$d_{hel}(\nu_1, \nu_2)^2 \leq 2(e^{\varepsilon} - 1)\ d_{hel}(\mu_1, \mu_2)^2.$$

We defer the proof of Lemma 2 to the end of this section. Let us first complete the proof of lower bound using this lemma. Note that, by data processing inequality we have

$$\|Q_1 - Q_2\|_{\mathrm{TV}} \leq \|(\mathcal{C}_i(X_i))_{i=1}^n - (\mathcal{C}_i(\tilde{X}_i))_{i=1}^n\|_{\mathrm{TV}},$$

where $X_i \sim P_1$ and $\tilde{X}_i \sim P_2$. Next, using (3.16) and (3.17) implies

$$\begin{aligned}\|Q_1 - Q_2\|_{\mathrm{TV}}^2 &\leq \|(\mathcal{C}_i(X_i))_{i=1}^n - (\mathcal{C}_i(\tilde{X}_i))_{i=1}^n\|_{\mathrm{TV}}^2 \\ &\leq d_{\mathrm{hel}}((\mathcal{C}_i(X_i))_{i=1}^n, (\mathcal{C}_i(\tilde{X}_i))_{i=1}^n)^2 \\ &= 2 - 2\prod_{i=1}^n (1 - \frac{1}{2} d_{\mathrm{hel}}(\mathcal{C}_i(X_i), \mathcal{C}_i(\tilde{X}_i))^2). \end{aligned} \tag{3.18}$$

Next, note that, by Lemma 3.4, we have

$$d_{\mathrm{hel}}(\mathcal{C}_i(X_i), \mathcal{C}_i(\tilde{X}_i))^2 \leq 2(e^{\varepsilon_i^{(l)}} - 1)\ d_{\mathrm{hel}}(P_1, P_2)^2 \leq 4\varepsilon_i^{(l)}\ d_{\mathrm{hel}}(P_1, P_2)^2,$$

where the last inequality follows from the fact that $\varepsilon_i^{(l)} \leq 1$. Plugging this back into (3.18), we obtain

$$\|Q_1 - Q_2\|_{\mathrm{TV}}^2 \leq 2 - 2\ \prod_{i=1}^n (1 - 2\varepsilon_i^{(l)}\ d_{\mathrm{hel}}(P_1, P_2)^2). \tag{3.19}$$

Next, note that, for nonnegative $y_1, \cdots, y_n$ we have

$$\prod_{i=1}^n (1 - y_i) \geq 1 - \sum_{i=1}^n y_i.$$

To show this, we can first verify it for $n = 2$, and then it is straightforward to show it for any n by induction. Using this inequality with $y_i = 2\varepsilon_i^{(l)}\ d_{\mathrm{hel}}(P_1, P_2)^2$, we can further upper bound (3.19) by

$$\|Q_1 - Q_2\|_{\mathrm{TV}}^2 \leq 4(\sum_{i=1}^n \varepsilon_i^{(l)})\ d_{\mathrm{hel}}(P_1, P_2)^2. \tag{3.20}$$

Plugging this back into (3.15), we have

$$\mathcal{L}(\mathcal{P}, \mathcal{Q}, (\varepsilon_i^{(l)})_{i=1}^n) \geq \gamma^2 \left(\frac{1}{2} - \sqrt{\sum_{i=1}^n \varepsilon_i^{(l)}}\ d_{\mathrm{hel}}(P_1, P_2) \right). \tag{3.21}$$

Next, we define P_1 and P_2 as

$$P_1(-1/2) = P_2(1/2) = \frac{1+2\gamma}{2}, \quad P_1(1/2) = P_2(-1/2) = \frac{1-2\gamma}{2}. \tag{3.22}$$

It is straightforward to verify that $|\theta(P_1) - \theta(P_2)| = 2\gamma$. Moreover, we have

$$d_{\text{hel}}(P_1, P_2)^2 = 2\left(\sqrt{\frac{1+2\gamma}{2}} - \sqrt{\frac{1-2\gamma}{2}}\right)^2 = 2(1 - \sqrt{1-4\gamma^2}) \leq 8\gamma^2,$$

where the last inequality follows from the fact that

$$1 - \sqrt{1-4\gamma^2} = \frac{4\gamma^2}{1+\sqrt{1-4\gamma^2}} \leq 4\gamma^2.$$

Plugging this back into (3.21) implies

$$\mathcal{L}(\mathcal{P}, \mathcal{Q}, (\varepsilon_i^{(l)})_{i=1}^n) \geq \gamma^2 \left(\frac{1}{2} - \sqrt{8\sum_{i=1}^n \varepsilon_i^{(l)}}\ \gamma\right). \tag{3.23}$$

Finally, setting

$$\gamma = \min\left(\frac{1}{4\sqrt{8\sum_{i=1}^n \varepsilon_i^{(l)}}}, \frac{1}{2}\right)$$

completes the proof of lower bound.

To show the upper bound, first, recall that a linear estimator with Gaussian mechanism is in the form of

$$\sum_{i=1}^n w_i \left(x_i + \mathcal{N}\left(0, \frac{\alpha}{2\varepsilon_i^{(l)}}\right)\right). \tag{3.24}$$

The mean square error of this estimator is given by

$$\sum_{i=1}^n w_i^2(\text{VAR} + \frac{\alpha}{2\varepsilon_i^{(l)}}) \leq \alpha \sum_{i=1}^n \frac{w_i^2}{\varepsilon_i^{(l)}},$$

where the last inequality uses the fact that $\alpha \geq 2$ and $\varepsilon_i^{(l)} \leq 1$. Finally, setting

$$w_i = \frac{\varepsilon_i^{(l)}}{\sum_{j=1}^n \varepsilon_j^{(l)}}$$

gives us the desired upper bound.

Proof of Lemma 3.4

Note that

$$\begin{aligned} d_{\text{hel}}(\nu_1, \nu_2)^2 &= \int (\sqrt{\nu_1(z)} - \sqrt{\nu_2(z)})^2 dz = \int \frac{(\nu_1(z) - \nu_2(z))^2}{(\sqrt{\nu_1(z)} + \sqrt{\nu_2(z)})^2} dz \\ &\leq \int \frac{(\nu_1(z) - \nu_2(z))^2}{\nu_1(z) + \nu_2(z)} dz. \end{aligned} \tag{3.25}$$

Note that, for any $j \in \{1, 2\}$, we can cast $\nu_j(z)$ as

$$\nu_j(z) = \int_x \mathcal{C}(z|x) d\mu_j(x).$$

Moreover, for any $x' \in \mathcal{X}$, we have

$$\nu_1(z) - \nu_2(z) = \int_x \mathcal{C}(z|x)(d\mu_1(x) - d\mu_2(x)) = \int_x (\mathcal{C}(z|x) - \mathcal{C}(z|x'))(d\mu_1(x) - d\mu_2(x)).$$

Substituting these into (3.25), we obtain

$$d_{\text{hel}}(\nu_1, \nu_2)^2 \leq \int_z \frac{\left(\int_x (\mathcal{C}(z|x) - \mathcal{C}(z|x'))(d\mu_1(x) - d\mu_2(x))\right)^2}{\int_x \mathcal{C}(z|x)(d\mu_1(x) + d\mu_2(x))} \, dz. \tag{3.26}$$

Next, by Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} &\left(\int_x (\mathcal{C}(z|x) - \mathcal{C}(z|x'))(d\mu_1(x) - d\mu_2(x))\right)^2 \leq \\ &\quad \left(\int_x \mathcal{C}(z|x)(d\mu_1(x) + \mu_2(x))\right) \left(\int_x \frac{(\mathcal{C}(z|x) - \mathcal{C}(z|x'))^2}{\mathcal{C}(z|x)} \frac{(d\mu_1(x) - d\mu_2(x))^2}{d\mu_1(x) + d\mu_2(x)}\right). \end{aligned}$$

Hence, using (3.26) and this inequality, we can further upper bound $d_{\text{hel}}(\nu_1, \nu_2)^2$ by

$$\begin{aligned} d_{\text{hel}}(\nu_1, \nu_2)^2 &\leq \int_z \int_x \frac{(\mathcal{C}(z|x) - \mathcal{C}(z|x'))^2}{\mathcal{C}(z|x)} \frac{(d\mu_1(x) - d\mu_2(x))^2}{d\mu_1(x) + d\mu_2(x)} dz \\ &= \int_x \left[\int_z \frac{(\mathcal{C}(z|x) - \mathcal{C}(z|x'))^2}{\mathcal{C}(z|x)} \; dz\right] \frac{(d\mu_1(x) - d\mu_2(x))^2}{d\mu_1(x) + d\mu_2(x)} \end{aligned} \tag{3.27}$$

where the last equation follows from changing the order of integration using Fubini's theorem. Now, note that the first term on the right-hand side of (3.27) can be cast as

$$\begin{aligned} \int_z \frac{(\mathcal{C}(z|x) - \mathcal{C}(z|x'))^2}{\mathcal{C}(z|x)} \; dz &= \int_z \frac{\mathcal{C}(z|x')^2}{\mathcal{C}(z|x)} \; dz - 2\int_z \mathcal{C}(z|x') \; dz + \int_z \mathcal{C}(z|x) \; dz \\ &= \exp(D_2(\mathcal{C}(x')||\mathcal{C}(x))) - 1. \end{aligned} \tag{3.28}$$

It is known that $D_\alpha(.||.)$ is nondecreasing in α. Thus, using $\alpha \geq 2$, we obtain

$$\int_z \frac{(\mathcal{C}(z|x) - \mathcal{C}(z|x'))^2}{\mathcal{C}(z|x)} \; dz \leq e^{\varepsilon} - 1. \tag{3.29}$$

Plugging this back into (3.27), we have

$$d_{\text{hel}}(\nu_1, \nu_2)^2 \leq (e^{\varepsilon} - 1) \; \int_x \frac{(d\mu_1(x) - d\mu_2(x))^2}{d\mu_1(x) + d\mu_2(x)}. \tag{3.30}$$

To complete the proof, we just need to show that

$$\int_x \frac{(d\mu_1(x) - d\mu_2(x))^2}{d\mu_1(x) + d\mu_2(x)} \leq 2d_{\text{hel}}(\mu_1, \mu_2)^2.$$

To do so, note that

$$\begin{aligned} \int_x \frac{(d\mu_1(x) - d\mu_2(x))^2}{d\mu_1(x) + d\mu_2(x)} &\leq 2 \; \int_x \frac{(d\mu_1(x) - d\mu_2(x))^2}{(\sqrt{d\mu_1(x)} + \sqrt{d\mu_2(x)})^2} \\ &= 2\int_x \left(\sqrt{d\mu_1(x)} - \sqrt{d\mu_2(x)}\right)^2 = 2d_{\text{hel}}(\mu_1, \mu_2)^2. \end{aligned}$$

This concludes the proof of lemma 3.4 and hence the proof of Theorem 3.1. ■

Proof of Lemma 3.2

Consider the strategy of user i as a function of its relative privacy sensitivity shown by $\beta_i(c_i)$. For a given estimator $\hat{\theta}$ and mechanism $(\boldsymbol{\varepsilon}^{(l)}, \boldsymbol{\varepsilon}^{(c)}, \mathbf{t})$, the action profile $\{\beta_i(\cdot)\}_{i=1}^n$ is an equilibrium if

$$\begin{aligned}&\mathbb{E}_{\mathbf{c}_{-i}}\left[t_i(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}})) - \mathbf{c}_{\mathbf{i}}\varepsilon_{\mathbf{i}}^{(\mathbf{l})}(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}})) - \varepsilon_{\mathbf{i}}^{(\mathbf{c})}(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}}))\right] \\ &\geq \mathbb{E}_{\mathbf{c}_{-i}}\left[t_i(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta'_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}})) - \mathbf{c}_{\mathbf{i}}\varepsilon_{\mathbf{i}}^{(\mathbf{l})}(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta'_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}})) - \varepsilon_{\mathbf{i}}^{(\mathbf{c})}(\beta_{-\mathbf{i}}(\mathbf{c}_{-\mathbf{i}}), \beta'_{\mathbf{i}}(\mathbf{c}_{\mathbf{i}})\right]\end{aligned}$$

for all $i \in \mathcal{N}, c_i, \beta'_i(\cdot)$. By letting $(\tilde{\boldsymbol{\varepsilon}}^{(l)}, \tilde{\boldsymbol{\varepsilon}}^{(c)}, \tilde{\mathbf{t}})$ be such that $\tilde{\varepsilon}_i^{(l)}(c_1, \ldots, c_n) = \varepsilon_i^{(l)}(\beta_1(c_1), \ldots, \beta_n(c_n))$, $\tilde{\varepsilon}_i^{(c)}(c_1, \ldots, c_n) = \varepsilon_i^{(c)}(\beta_1(c_1), \ldots, \beta_n(c_n))$, and $\tilde{t}_i(c_1, \ldots, c_n) = t_i(\beta_1(c_1), \ldots, \beta_n(c_n))$, the users will report truthfully and that the platform's objective is the same as the original mechanism. This establishes the revelation principle. ■

Proof of Theorem 3.2

Recall the interim quantities

$$\begin{aligned} t_i(c_i) =& \mathbb{E}_{\mathbf{c}_{-i}}\left[t(c_i, \mathbf{c}_{-i})\right], \\ \varepsilon_i^{(l)}(c_i) =& \mathbb{E}_{\mathbf{c}_{-i}}\left[\varepsilon_i^{(l)}(c_i, \mathbf{c}_{-i})\right], \text{ and} \\ \varepsilon_i^{(c)}(c_i) =& \mathbb{E}_{\mathbf{c}_{-i}}\left[\varepsilon_i^{(c)}(c_i, \mathbf{c}_{-i})\right] \text{ for all } i \in \mathcal{N}, c_i. \end{aligned}$$

Using these quantities, the incentive compatibility constraint becomes

$$t_i(c_i) - c_i\varepsilon_i^{(l)}(c_i) - (1 - c_i)\varepsilon_i^{(c)}(c_i) \geq t_i(c'_i) - c_i\varepsilon_i^{(l)}(c'_i) - (1 - c_i)\varepsilon_i^{(c)}(c'_i).$$

By equating the derivative of the right-hand side with respect to c'_i at c_i to zero, we obtain

$$t'_i(c_i) - c_i\left(\varepsilon_i'^{(l)}(c_i) - \varepsilon_i'^{(c)}(c_i)\right) - \varepsilon_i'^{(c)}(c_i) = 0.$$

This equation gives us the derivative of the payment in terms of the privacy loss levels. By taking the integral of this expression we obtain

$$\begin{aligned} t_i(c_i) &= t_i(0) + \int_0^{c_i} \left(\varepsilon_i'^{(c)}(z) + z\left(\varepsilon_i'^{(l)}(z) - \varepsilon_i'^{(c)}(z) \right) \right) dz \\ &= t_i(0) + \varepsilon_i^{(c)}(c_i) - \varepsilon_i^{(c)}(0) + c_i \left(\varepsilon_i^{(l)}(c_i) - \varepsilon_i^{(c)}(c_i) \right) - \int_0^{c_i} \left(\varepsilon_i^{(l)}(z) - \varepsilon_i^{(c)}(z) \right) dz. \quad (3.31) \end{aligned}$$

We next show that the payment in (3.31) together with a weakly decreasing $\varepsilon_i^{(l)}(z) - \varepsilon_i^{(c)}(z)$ guarantees that the incentive compatibility constraint. To see this, we consider two possibilities depending on whether c_i' is larger or smaller than c_i:

- For $c_i' \geq c_i$, by using the payment in (3.31), the incentive compatibility constraint becomes equivalent to

  $$\left(\varepsilon_i^{(l)}(c_i') - \varepsilon_i^{(c)}(c_i') \right)(c_i - c_i') \geq \int_{c_i'}^{c_i} \left(\varepsilon_i^{(l)}(z) - \varepsilon_i^{(c)}(z) \right) dz,$$

  which holds because $\varepsilon_i^{(l)}(z) - \varepsilon_i^{(c)}(z)$ is weakly decreasing in z.

- For $c_i' \leq c_i$, by using the payment in (3.31), the incentive compatibility constraint becomes equivalent to

  $$\left(\varepsilon_i^{(l)}(c_i') - \varepsilon_i^{(c)}(c_i') \right)(c_i - c_i') \leq \int_{c_i}^{c_i'} \left(\varepsilon_i^{(l)}(z) - \varepsilon_i^{(c)}(z) \right) dz,$$

  which, again, holds because $\varepsilon_i^{(l)}(z) - \varepsilon_i^{(c)}(z)$ is weakly decreasing in z. This completes one direction of the proof.

To see the other direction, notice that using the first order condition for the incentive compatibility constraints, imply (3.31). To see the monotonicity, notice that the incentive compatibility implies

$$\varepsilon_i^{(c)}(c_i) + c_i \left(\varepsilon_i^{(l)}(c_i) - \varepsilon_i^{(c)}(c_i) \right) - t_i(c_i) \leq \varepsilon_i^{(c)}(c_i') + c_i \left(\varepsilon_i^{(l)}(c_i') - \varepsilon_i^{(c)}(c_i') \right) - t_i(c_i').$$

and

$$\varepsilon_i^{(c)}(c_i') + c_i' \left(\varepsilon_i^{(l)}(c_i') - \varepsilon_i^{(c)}(c_i')\right) - t_i(c_i') \leq \varepsilon_i^{(c)}(c_i) + c_i' \left(\varepsilon_i^{(l)}(c_i) - \varepsilon_i^{(c)}(c_i)\right) - t_i(c_i).$$

The summation of these two inequalities yields

$$\left(\left(\varepsilon_i^{(l)}(c_i) - \varepsilon_i^{(c)}(c_i)\right) - \left(\varepsilon_i^{(l)}(c_i') - \varepsilon_i^{(c)}(c_i')\right)\right)(c_i - c_i') \leq 0,$$

that proves $\varepsilon_i^{(l)}(\cdot) - \varepsilon_i^{(c)}(\cdot)$ is weakly decreasing.

We next evaluate the individual rationality constraint. Using (3.31), we can rewrite this constraint as

$$t_i(0) \geq \varepsilon_i^{(c)}(0) + \int_0^{c_i} \left(\varepsilon_i^{(l)}(z) - \varepsilon_i^{(c)}(z)\right) dz \quad \text{for all } c_i. \tag{3.32}$$

Using $\varepsilon_i^{(l)}(z) \geq \varepsilon_i^{(c)}(z)$ for all z, this inequality means that it only needs to hold for $c_i = \infty$. Hence, we could cast $t_i(0)$ as

$$t_i(0) = \varepsilon_i^{(c)}(0) + \int_0^{\infty} \left(\varepsilon_i^{(l)}(z) - \varepsilon_i^{(l)}(z)\right) dz.$$

Plugging this back in (3.31) yields

$$t_i(c_i) = \varepsilon_i^{(c)}(c_i) + c_i \left(\varepsilon_i^{(l)}(c_i) - \varepsilon_i^{(c)}(c_i)\right) + \int_{c_i} \left(\varepsilon_i^{(l)}(z) - \varepsilon_i^{(c)}(z)\right) dz.$$

which is the optimal payment when $\varepsilon^{(l)}(\cdot)$ is decreasing.

With this optimal payment, the expected payment becomes

$$\begin{aligned}
\mathbb{E}_{c_i}\left[t_i(c_i)\right] &= \mathbb{E}_{c_i}[\varepsilon_i^{(c)}(c_i)] + \mathbb{E}_{c_i}\left[c_i\left(\varepsilon_i^{(l)}(c_i) - \varepsilon_i^{(c)}(c_i)\right) + \int_{z=c_i}\left(\varepsilon_i^{(l)}(z) - \varepsilon_i^{(c)}(z)\right)dz\right] \\
&= \mathbb{E}_{c_i}[\varepsilon_i^{(c)}(c_i)] + \int_{\mathbf{z}_{-i}}\int_{z_i}\Bigg(z_i\left(\varepsilon_i^{(l)}(z_i,\mathbf{z}_{-i}) - \varepsilon_i^{(c)}(z_i,\mathbf{z}_{-i})\right) + \\
&\qquad \int_{y_i=z_i}\left(\varepsilon_i^{(l)}(y,\mathbf{z}_{-i}) - \varepsilon_i^{(c)}(y,\mathbf{z}_{-i})\right)dy_i\Bigg) f_i(z_i)dz_i f_{-i}(\mathbf{z}_{-i})d\mathbf{z}_{-i} \\
&\overset{(a)}{=} \mathbb{E}_{c_i}[\varepsilon_i^{(c)}(c_i)] + \int_{\mathbf{z}_{-i}}\int_{z_i}\Bigg(z_i\left(\varepsilon_i^{(l)}(z_i,\mathbf{z}_{-i}) - \varepsilon_i^{(c)}(z_i,\mathbf{z}_{-i})\right) \\
&\qquad + \left(\varepsilon_i^{(l)}(z_i,\mathbf{z}_{-i}) - \varepsilon_i^{(c)}(z_i,\mathbf{z}_{-i})\right)\frac{F_i(z_i)}{f_i(z_i)}\Bigg) f_i(z_i)dz_i f_{-i}(\mathbf{z}_{-i})d\mathbf{z}_{-i} \\
&= \mathbb{E}_{c_i}[\varepsilon_i^{(c)}(c_i)] + \int_{\mathbf{z}}\left(z_i + \frac{F_i(z_i)}{f_i(z_i)}\right)\left(\varepsilon_i^{(l)}(\mathbf{z}) - \varepsilon_i^{(c)}(\mathbf{z})\right)f(\mathbf{z})d\mathbf{z}, \qquad (3.33)
\end{aligned}$$

where (a) follows from changing the order of the integrals. Substituting equation (3.33) in the platform's objective function yields

$$\begin{aligned}
&\mathbb{E}_{\mathbf{c}}\left[\gamma\mathrm{MSE}(\boldsymbol{\varepsilon}^{(l)}(\mathbf{c}),\boldsymbol{\varepsilon}^{(c)}(\mathbf{c}),\hat{\theta}) + \sum_{i=1}^{n} t_i(\mathbf{c})\right] \\
&= \mathbb{E}_{\mathbf{c}}\left[\gamma\mathrm{MSE}(\boldsymbol{\varepsilon}^{(l)}(\mathbf{c}),\boldsymbol{\varepsilon}^{(c)}(\mathbf{c}),\hat{\theta}) + \sum_{i=1}^{n}\psi_i(c_i)\varepsilon_i^{(l)}(\mathbf{c}) + \sum_{i=1}^{n}\left(1-\psi_i(c_i)\right)\varepsilon_i^{(c)}(\mathbf{c})\right].
\end{aligned}$$

Notice that the maximizer of the above objective is the optimal local and central privacy losses, provided that $\varepsilon_i^{(l)}(\cdot) - \varepsilon_i^{(c)}(\cdot)$ is decreasing. For a given privacy sensitivity vector $\mathbf{c}$, let us consider the point-wise minimization given by

$$\min_{\{\boldsymbol{\varepsilon}^{(l)}\}_{i=1}^{n},\{\boldsymbol{\varepsilon}^{(c)}\}_{i=1}^{n}} \quad \gamma\mathrm{MSE}(\boldsymbol{\varepsilon}^{(l)},\boldsymbol{\varepsilon}^{(c)},\hat{\theta}) + \sum_{i=1}^{n}\varepsilon_i^{(l)}\psi_i(c_i) + \sum_{i=1}^{n}\varepsilon_i^{(c)}\left(1-\psi_i(c_i)\right). \qquad (3.34)$$

This point-wise optimization clearly finds the optimal $\varepsilon^{(l)}(\cdot)$ and $\varepsilon^{(c)}(\cdot)$, but the issue is that the corresponding $\varepsilon_i^{(l)}(\cdot) - \varepsilon_i^{(c)}(\cdot)$ may not be decreasing. We next show that under Assumption 3.1 this is always the case.

Let $\{\boldsymbol{\varepsilon}^{(l)}\}_{i=1}^{n}$ and $\{\boldsymbol{\varepsilon}^{(c)}\}_{i=1}^{n}$ be the solution of optimization problem (3.34) for $c_1,\ldots,c_n$. Now, suppose we increase one of the c_i's, which, without loss of generality, we assume is the first one. Let $c'_1 > c_1$ and $c'_i = c_i$ for $i = 2, \ldots, n$ and suppose $\{\boldsymbol{\varepsilon}'^{(l)}\}_{i=1}^n, \{\boldsymbol{\varepsilon}'^{(c)}\}_{i=1}^n$ is the corresponding optimal solution of optimization problem (3.34). The optimality condition implies that

$$\begin{aligned}
&\gamma \mathrm{MSE}(\boldsymbol{\varepsilon}^{(l)}, \boldsymbol{\varepsilon}^{(c)}, \hat{\theta}) + \sum_{i=1}^n \varepsilon_i^{(l)} \psi_i(c_i) + \sum_{i=1}^n \varepsilon_i^{(c)} \left(1 - \psi_i(c_i)\right) \\
&\leq \gamma \mathrm{MSE}(\boldsymbol{\varepsilon}'^{(l)}, \boldsymbol{\varepsilon}'^{(c)}, \hat{\theta}) + \sum_{i=1}^n \varepsilon'^{(l)}_i \psi_i(c_i) + \sum_{i=1}^n \varepsilon'^{(c)}_i \left(1 - \psi_i(c_i)\right)
\end{aligned}$$

and

$$\begin{aligned}
&\gamma \mathrm{MSE}(\boldsymbol{\varepsilon}'^{(l)}, \boldsymbol{\varepsilon}'^{(c)}, \hat{\theta}) + \sum_{i=1}^n \varepsilon'^{(l)}_i \psi_i(c'_i) + \sum_{i=1}^n \varepsilon'^{(c)}_i \left(1 - \psi_i(c_i)\right) \\
&\leq \gamma \mathrm{MSE}(\boldsymbol{\varepsilon}^{(l)}, \boldsymbol{\varepsilon}^{(c)}, \hat{\theta}) + \sum_{i=1}^n \varepsilon_i^{(l)} \psi_i(c'_i) + \sum_{i=1}^n \varepsilon_i^{(c)} \left(1 - \psi_i(c_i)\right)
\end{aligned}$$

The summation of both sides of these inequalities, together with $c_i = c'_i$ for $i = 2, \ldots, n$, results in

$$\left(\left(\varepsilon_1^{(l)} - \varepsilon_1^{(c)}\right) - \left(\varepsilon'^{(l)}_1 - \varepsilon'^{(c)}_1\right)\right) \left(\psi_1(c_1) - \psi_1(c'_1)\right) \leq 0.$$

Assumption 3.1 and the above inequality establishes that the solution of problem (3.34) is weakly decreasing in the privacy sensitivity. ■

Proof of Corollary 3.1

The proof follows by invoking Theorem 3.2 and noting that with

$$\hat{\theta} = \sum_{i=1}^n w_i \left(x_i + \mathcal{N}\left(0, \frac{\alpha}{2\varepsilon_i^{(l)}}\right)\right)$$

we have

$$\varepsilon_i^{(c)} = \frac{w_i^2}{\sum_{j=1}^n \frac{w_j^2}{\varepsilon^{(l)2}_j}}$$

and

$$\mathrm{MSE}(\boldsymbol{\varepsilon}^{(l)}, \boldsymbol{\varepsilon}^{(c)}, \hat{\theta}) = \textsc{var} \sum_{i=1}^{n} w_i^2 + \sum_{i=1}^{n} w_i^2 \frac{\alpha}{2\varepsilon_i^{(l)}}.$$

This completes the proof. ■

Proof of Proposition 3.1

With a Gaussian mechanism that adopts both local and central noises, using a similar argument to that of Theorem 3.2 and Corollary 3.1, the optimal central privacy loss levels are

$$\varepsilon_i^{(c)} = \frac{w_i^{*2}}{\sum_{j=1}^{n} \frac{w_j^{*2}}{y_j^*} + \frac{1}{\varepsilon}},$$

where $(w_1^*, \ldots, w_n^*)$, $(y_1^*, \ldots, y_n^*)$, and ε are the optimal solution of

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{y}, \varepsilon} \quad & \textsc{var}\gamma \sum_{i=1}^{n} w_i^2 + \frac{\gamma\alpha}{2} \sum_{i=1}^{n} \frac{w_i^2}{y_i} + \frac{\alpha\gamma}{2\varepsilon} + \sum_{i=1}^{n} (1 - \psi_i(c_i)) \frac{w_i^2}{\sum_{j=1}^{n} \frac{w_j^2}{y_j} + \frac{1}{\varepsilon}} + \sum_{i=1}^{n} \psi_i(c_i) y_i \\ \text{s.t.} \quad & w_i, y_i \geq 0, \text{ for all } i \in \mathcal{N} \\ & \sum_{i=1}^{n} w_i = 1. \end{aligned}$$

For any solution to the above optimization problem we define the following alternative solution:

$$y_i' = \frac{w_i^2}{\frac{w_i^2}{y_i} + \frac{w_i}{\varepsilon}}, \varepsilon' = \infty, \text{ and } w_i' = w_i \quad \text{for all } i \in \mathcal{N}.$$

We have

$$
\begin{aligned}
&\text{VAR}\gamma \sum_{i=1}^{n} {w_i'}^2 + \frac{\gamma\alpha}{2} \sum_{i=1}^{n} \frac{{w_i'}^2}{y_i'} + \sum_{i=1}^{n} \left(1 - \psi_i(c_i)\right) \frac{{w_i'}^2}{\sum_{j=1}^{n} \frac{{w_j'}^2}{y_j'}} + \sum_{i=1}^{n} \psi_i(c_i) y_i' \\
&\overset{(a)}{=} \text{VAR}\gamma \sum_{i=1}^{n} {w_i}^2 + \frac{\gamma\alpha}{2} \sum_{i=1}^{n} \frac{{w_i}^2}{y_i} + \frac{\alpha\gamma}{2\varepsilon} + \sum_{i=1}^{n} \left(1 - \psi_i(c_i)\right) \frac{{w_i}^2}{\sum_{j=1}^{n} \frac{{w_j}^2}{y_j} + \frac{1}{\varepsilon}} + \sum_{i=1}^{n} \psi_i(c_i) y_i' \\
&\overset{(b)}{\leq} \text{VAR}\gamma \sum_{i=1}^{n} {w_i}^2 + \frac{\gamma\alpha}{2} \sum_{i=1}^{n} \frac{{w_i}^2}{y_i} + \frac{\alpha\gamma}{2\varepsilon} + \sum_{i=1}^{n} \left(1 - \psi_i(c_i)\right) \frac{{w_i}^2}{\sum_{j=1}^{n} \frac{{w_j}^2}{y_j} + \frac{1}{\varepsilon}} + \sum_{i=1}^{n} \psi_i(c_i) y_i,
\end{aligned}
$$

where (a) follows the construction of the new solution and (b) follows from

$$
y_i' = \frac{w_i^2}{\frac{w_i^2}{y_i} + \frac{w_i}{\varepsilon}} \leq y_i \quad \text{for all } i \in \mathcal{N}.
$$

This completes the proof.

Proof of Theorem 3.3

Consider the optimization problem

$$
\begin{aligned}
\min_{\mathbf{w},\mathbf{y}} \;\; & \text{VAR}\gamma \sum_{i=1}^{n} w_i^2 + \frac{\gamma\alpha}{2} \sum_{i=1}^{n} \frac{w_i^2}{y_i} + \sum_{i=1}^{n} \frac{w_i^2}{\sum_{j=1}^{n} \frac{w_j^2}{y_j}} \left(1 - \psi_i(c_i)\right) + \sum_{i=1}^{n} \psi_i(c_i) y_i \qquad (3.35) \\
\text{s.t.} \;\; & w_i, y_i \geq 0, \;\text{for all } i \in \mathcal{N} \\
& \sum_{i=1}^{n} w_i = 1.
\end{aligned}
$$

We can rewrite this optimization problem as

$$\begin{aligned}
\min_{\mathbf{w},\mathbf{y},S} \quad & \text{VAR}\gamma \sum_{i=1}^{n} w_i^2 + \frac{\gamma\alpha}{2} S + \frac{1}{S} \sum_{i=1}^{n} w_i^2 \left(1 - \psi_i(c_i)\right) + \sum_{i=1}^{n} \psi_i(c_i) y_i \\
\text{s.t.} \quad & w_i, y_i \geq 0, \ \text{ for all } i \in \mathcal{N} \\
& \sum_{i=1}^{n} w_i = 1, \\
& \sum_{i=1}^{n} \frac{w_i^2}{y_i} = S.
\end{aligned}$$

Let us fix S. The Lagrangian of this optimization problem becomes

$$\begin{aligned}
& \sum_{i=1}^{n} w_i^2 \left(\text{VAR}\gamma + \frac{1 - \psi_i(c_i)}{S} \right) + \sum_{i=1}^{n} \psi_i(c_i) y_i + p \left(\sum_{i=1}^{n} \frac{w_i^2}{y_i} - S \right) - q \left(\sum_{i=1}^{n} w_i - 1 \right) \\
& - \sum_{i=1}^{n} u_i w_i - \sum_{i=1}^{n} v_i y_i,
\end{aligned}$$

where $u_i, v_i \geq 0$ and for the optimal solution we have $u_i w_i = v_i y_i = 0$, for all i.

Equating the derivative with respect to w_i to zero, yields

$$2w_i \left(\text{VAR}\gamma + \frac{1 - \psi_i(c_i)}{S} \right) + \frac{2pw_i}{y_i} - q = u_i. \tag{3.36}$$

Hence, if $w_i^* > 0$, then $u_i = 0$ which implies

$$2 \left(\text{VAR}\gamma + \frac{1 - \psi_i(c_i)}{S} + \frac{p}{y_i} \right) w_i = q. \tag{3.37}$$

On the other hand, if $w_i^* = 0$, then $u_i = -q \geq 0$.

Equating the derivative with respect to y_i to zero implies

$$\psi_i(c_i) - \frac{pw_i^2}{y_i^2} - v_i.$$

Hence, if $y_i^* = 0$, then $w_i^* = 0$ and $v_i = \psi_i(c_i)$. On the other hand, if $y_i^* > 0$, then $v_i = 0$ and we have

$$w_i = \sqrt{\frac{\psi_i(c_i)}{p}} y_i. \tag{3.38}$$

Now, we claim there is no i for which $w_i^* = 0$ (and hence there is no i for which $y_i^* = 0$). Assume this is not the case, and hence there exists some i_0 for which $w_{i_0}^* = 0$. Therefore, as we established earlier, we have $q = -u_{i_0} \leq 0$. On the other hand, note that there exists j for which $w_j^* > 0$. Hence, for that $y_j^* > 0$ as well. Therefore, using (3.37), along with the fact that $q \leq 0$, we should have

$$\frac{1 - \psi_j(c_j)}{S} + \frac{p}{y_j^*} \leq 0,$$

which implies

$$\frac{\psi_j(c_j) - 1}{S} \geq \frac{p}{y_j^*}.$$

Hence, using $S \geq (w_j^*)^2/y_j^*$,we have

$$\psi_j(c_j) - 1 \geq S\frac{p}{y_j^*} \geq \frac{(w_j^*)^2 p}{y_j^*}.$$

However, since $y_j^* > 0$, by (3.38), the right hand side is equal to $\psi_j(c_j)$, which implies $\psi_j(c_j) - 1 \geq \psi_j(c_j)$ which is a contradiction! As a result, (3.37) and (3.38) hold for all i.

By invoking (3.38) in (3.37), we obtain

$$w_i = \frac{1}{\text{VAR}\gamma + \frac{1-\psi_i(c_i)}{S}} \left(\frac{q}{2} - \sqrt{\psi_i(c_i)p} \right).$$

To simplify the analysis, we define the interim variable

$$\nu_i = \frac{1}{\gamma\text{VAR} + (1 - \psi_i(c_i))/S}. \tag{3.39}$$

Taking summation of the above equation for $i = 1, \ldots, n$ and using $\sum_{i=1}^n w_i = 1$, we obtain

$$\frac{q}{2} = \frac{1 + \sum_j \nu_j \sqrt{\psi_j(c_j)p}}{\sum_j \nu_j},$$

which together with (3.36) results in

$$w_i = \frac{\nu_i}{\sum_j \nu_j} + \frac{\nu_i}{\sum_j \nu_j} \left(\sum_{j=1}^n \nu_j (\sqrt{\psi_j(c_j)p} - \sqrt{\psi_i(c_i)p}) \right). \tag{3.40}$$

Therefore, by using (3.38) and (3.40), once we have S and p, we can find y_i and w_i for all $i \in \mathcal{N}$.

Next, we derive a relation between S and p. Note that (3.40) implies that w_i can be cast as

$$\zeta_i(S) + \xi_i(S)\sqrt{p}$$

with

$$\zeta_i(S) = \frac{\nu_i}{\sum_j \nu_j} \quad \text{and} \quad \xi_i(S) = \frac{\nu_i}{\sum_j \nu_j} \left(\sum_{j=1}^n \nu_j (\sqrt{\psi_j(c_j)} - \sqrt{\psi_i(c_i)}) \right).$$

Using (3.38), we have

$$S = \sum_{i=1}^n \frac{w_i^2}{y_i} = \sum_{i=1}^n w_i \frac{\sqrt{\psi_i(c_i)}}{\sqrt{p}} = \sum_{i=1}^n \frac{\zeta_i(S)\sqrt{\psi_i(c_i)}}{\sqrt{p}} + \sum_{i=1}^n \sqrt{\psi_i(c_i)}\xi_i(S).$$

This implies

$$p = \left(\frac{\sum_{i=1}^n \zeta_i(S)\sqrt{\psi_i(c_i)}}{S - \sum_{i=1}^n \sqrt{\psi_i(c_i)}\xi_i(S)} \right)^2. \tag{3.41}$$

We next show that we can search over a grid to find the approximately optimal S. In this regard, we derive a lower and upper bound on the optimal S.

To do so, first note that the objective function (3.35) is given by

$$\begin{aligned} \text{OBJ } &= \text{VAR}\gamma \sum_{i=1}^n w_i^2 + \frac{\gamma\alpha}{2} S + \frac{1}{S} \sum_{i=1}^n w_i^2 (1 - \psi_i(c_i)) + \sum_{i=1}^n \psi_i(c_i) y_i \\ &\geq \frac{\gamma\alpha}{2} S + \frac{1}{S} \sum_{i=1}^n w_i^2 + \sum_{i=1}^n \psi_i(c_i)(y_i - \frac{w_i^2}{S}). \end{aligned}$$

It is straightforward to see $y_i \geq \frac{w_i^2}{S}$ for all i, and thus, we have

$$\text{OBJ } \geq \frac{\gamma\alpha}{2}S + \frac{1}{S}\sum_{i=1}^{n} w_i^2. \tag{3.42}$$

Using (3.42) along with the fact that Cauchy–Schwarz inequality implies $\sum_{i=1}^{n} w_i^2 \geq 1/n$, we have

$$\text{Optimal objective (OPT) } \geq \frac{\gamma\alpha}{2}S^* + \frac{1}{nS^*}.$$

As a result, we have

$$\frac{\text{OPT}}{\gamma\alpha/2} \geq S^* \geq \frac{1}{\text{OPT}n}. \tag{3.43}$$

Letting $y_i = 1$, $w_i = \frac{1}{n}$, and $\varepsilon = 1$ in the objective of Problem (3.35) gives us an upper bound on the optimal objective OPT. Let us denote this upper bound by M. We have

$$\frac{M}{\gamma\alpha/2} \geq S^* \geq \frac{1}{Mn}. \tag{3.44}$$

Therefore, we obtain an approximate optimal solution by grid search. This provides an $O(\delta)$ optimal solution for the platform's problem because the objective of Problem (3.35) is Lipschitz continuous. ■

Chapter 4

How Good Are Privacy Guarantees? Platform Architecture and Violation of User Privacy

4.1 Introduction

In the last two chapters, we fixed the central and local architectures to primarily study the heterogeneity aspect of privacy demands in the data market, i.e., when users have different and unknown (to the platform) privacy sensitivities. Furthermore, we used monetary compensations to incentivize users to report their privacy sensitivities truthfully. In this chapter, we investigate a different aspect of the data market design: the optimal choice of architecture from both users' and the platform's point of view.

More specifically, we build a model in which each user has a utility consisting of two terms. The first attaches a positive value, with weight α, to the precision of society's (or the platform's) estimate of an underlying common state, θ, based on pooled user data. The second attaches a negative value, with weight β, to the decline in the mean squared error about the individual's own type, which is used by the platform for pricing or ad targeting. For concreteness, we could consider the state θ to correspond to the prevalence of a virus in the population, such as COVID- 19, while the individual type may be whether the user herself has been infected, which she may wish to keep private. We assume that the platform receives positive returns from acquiring more information on both components.

The key decision for users is whether to share their data (or participate in the platform). If the potential cost of privacy violations is large enough, they will choose not to do so, unless adequate privacy guarantees are provided by the platform.

The game between the platform and the users is conceptualized as follows. First, the platform commits to a mechanism for partially preserving user privacy. Second, users decide individually whether to share their data. Then the platform uses the data according to the chosen mechanism, and utilities are realized. We look for a (Bayesian) Stackelberg equilibrium of this game, whereby the platform optimally chooses the mechanism, anticipating the following Bayesian Nash equilibrium.

What makes this game interesting and difficult is the fact that the space of mechanisms that provide privacy guarantees is vast, including partial anonymization, limits on what data can be used for, various ways of adding noise to the data, and differential privacy and related mechanisms.

We define a mechanism as a mapping from the users' data to an output and consider the following pair of quantities in the space of *all possible mechanisms*: leaked information about the underlying common parameter and the sum of leaked information about each user's private data. We then ask the following question:

> *In the space of all mechanisms, is there a mechanism that achieves the Pareto frontier (i.e., leaks the most about the underlying common parameter and the least about users' data)? What is that mechanism?*

We show that a *mask-shuffle mechanism* achieves the Pareto frontier defined above. This proves that from the viewpoint of the users, this mechanism provides the optimal trade-off between the positive and the negative uses of data. Specifically, according to this optimal mechanism, the platform should commit to a probability with which a user's data will be fully anonymized (will be shuffled across users). This type of mask-shuffle is attractive from the users' viewpoint because it maintains information about the underlying state but implies that the platform learns much less about the individual. By choosing the probability of shuffling, the platform can fine-tune the privacy guarantees to users.

Our second result characterizes the (unique and Bayesian) Stackelberg equilibrium under any mask-shuffle mechanism. Additionally, we provide a number of comparative statics of this equilibrium, showing how the extent to which users and the platform care about privacy affects the degree of anonymity. Our comparative statics is more subtle and surprising, highlighting the paradox of platform-provided privacy guarantees. When α (the weight attached to the positive use of pooled data in order to learn the state θ) increases, user utility increases given any mask-shuffle mechanism, and users become more willing to share their data so that it can be pooled with those of others to obtain better estimates of θ. Greater α, however, also means that the platform now chooses lower privacy guarantees. The paradoxical result is that this platform response is powerful enough that, under some conditions we characterize, users end up worse off than they would have been with lower α. We interpret this result as suggesting that platform-provided privacy guarantees are highly imperfect and often insufficient.

Our final result turns to the implications of user privacy preferences on platform choices of data architecture. We prove that the platform has an incentive to deviate from the user-optimal mask-shuffle mechanisms. In particular, we identify a set of *pivot mechanisms* that make individual privacy on the choices of other users, for example, by linking the decision of how much of a user's data to utilize on the sharing decision of other users. We show that the platform can exploit user preferences towards the underlying common state, θ, by designing a pivot mechanism that commits to not utilizing any user data if any one of the users does not share her data. This pivot mechanism makes every user "pivotal" at the margin, meaning that if she decides not to share her data, nothing is learned about θ. Because the user attaches some value to the society learning about θ, the effective cost of not sharing her data increases significantly, and this allows the platform to violate her privacy. We also show that more continuous versions of pivot mechanisms can achieve the same outcome. This result further amplifies our interpretation that self-regulation by platforms is often insufficient to ensure sufficient user privacy.

The rest of the chapter proceeds as follows. Section 4.2 presents the users' and the platform's utility and establishes the optimality of the mask-shuffle mechanism. In Section

4.3, we introduce the equilibrium concept and establish its existence. In Section 4.4, we characterize the equilibrium of the game among the users and the platform and provide some comparative statics for it. In Section 4.5, we establish that the platform has incentives to use mechanisms other than mask-shuffle as opposed to the users. Section 4.6 concludes, while the last section presents the proofs not included in the text.

4.2 Environment

We consider a platform that wishes to collect data from n privacy-aware users denoted by $\mathcal{N} = \{1, \ldots, n\}$. User i's data is represented by $X_i = \theta + Z_i$ where $\theta \sim \mathcal{N}(0,1)$ is a common parameter and $Z_i \sim \mathcal{N}(0,1)$ is user i's private type. We assume both users and the platform drive higher utility from having access to a better estimation of θ. The private type of user i, Z_i, can be used for the platform's benefit, and therefore the platform gains from a better estimation of it while the user suffers a privacy loss. Users and the platform connect through a **mechanism**. Formally, a mechanism $\mathcal{M} : \mathbb{R}^n \to \mathcal{X}$, for some set $\mathcal{X}$, is a randomized algorithm whose input is the users' data, i.e., $x_1, \cdots, x_n$, and its output is received by the platform. The mechanism output is used by the platform to estimate θ. The mechanism output contains information about the underlying parameter θ which leads to a better estimation of this parameter and benefits both the users and the platform. It also reveals information about the private type of users z_i for $i \in \mathcal{N}$ which benefits the platform but harms the users.

Before introducing the utility of the users and the platform we introduce our measure of *revealed information*. Throughout the chapter, we use lower case letters to denote the realization of random variables. Notice that platform's prior on θ and Z_i is $\pi_0 = \mathcal{N}(0,1)$. We denote the platform's posterior on θ and Z_i after observing the mechanism's output by $\pi_\theta(\mathcal{M})$ and $\pi_{Z_i}(\mathcal{M})$, respectively. It can be seen that the best estimator of θ and Z_i's with respect to the mean-squared error, given the mechanism's output, is the mean of the posterior distributions. We define revealed information as the reduction in the mean-squared error from the prior to the posterior, formalized next (in our setting, privacy is ensured when the disclosed information, as defined below, is small. This guarantee is based on an average-case scenario, which differs from the worst-case guarantees provided by differential privacy Dwork et al. [2014].)

Definition 4.1 (Revealed information)**.** *For any mechanism $\mathcal{M}$, revealed information about θ is the reduction in the mean-squared error of θ, i.e.,*

$$\mathcal{I}(\theta \mid \mathcal{M}) = \mathbb{E}\left[\left(\theta - \mathbb{E}_{\theta \sim \pi_0}\left[\theta\right]\right)^2\right] - \mathbb{E}\left[\left(\theta - \mathbb{E}_{\theta \sim \pi_\theta(\mathcal{M})}\left[\theta\right]\right)^2\right],$$

where the expectations are over the randomness in data and the mechanism. Similarly, for any $i \in \mathcal{N}$, revealed information about Z_i is the reduction in the mean-squared error of Z_i, i.e.,

$$\mathcal{I}(Z_i \mid \mathcal{M}) = \mathbb{E}\left[\left(Z_i - \mathbb{E}_{Z_i \sim \pi_0}\left[Z_i\right]\right)^2\right] - \mathbb{E}\left[\left(Z_i - \mathbb{E}_{Z_i \sim \pi_{Z_i}(\mathcal{M})}\left[Z_i\right]\right)^2\right].$$

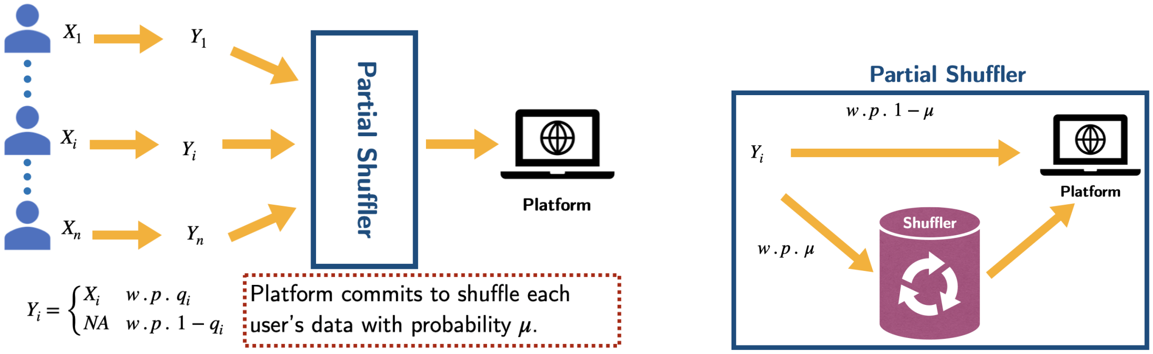
Given the above definition of revealed information, the expected **utility of user** i is given by

$$\mathcal{U}_i(\mathcal{M}) := \alpha\, \mathcal{I}(\theta \mid \mathcal{M}) - \beta\, \mathcal{I}(Z_i \mid \mathcal{M}). \tag{4.1}$$

The first term captures the gain of user i from a better estimation of the underlying parameter θ. For instance, in the context of a medical study, the user gains from a better estimation by the hospital, leading to a more effective drug. The second term captures the loss of learning user i's private data Z_i. Again, in the context of a medical study, the user wants to keep her medical record private. We use parameters α and β that are non-negative as constants to scale the impact of learning the underlying parameter and the user's private data, respectively. In the context of a medical study, again, they capture the relative weight that users assign to a more effective drug versus their privacy loss. The expected **platform's utility** is given by

$$\mathcal{U}_{\text{platform}}(\mathcal{M}) := \mathcal{I}(\theta \mid \mathcal{M}) + \delta \sum_{i=1}^{n} \mathcal{I}(Z_i \mid \mathcal{M}), \tag{4.2}$$

where the first and second terms correspond to the platform's gain from learning θ and users' private type, respectively. Notice that, without loss of generality, we have normalized the impact of learning θ in platform's utility to one and use a non-negative constant δ to scale the impact of learning users' private data in the platform's utility.


X_1
Y_1
X_i
Y_i
X_n
Y_n
Partial Shuffler
Platform
$Y_i = \begin{cases} X_i & w.p.\ q_i \\ NA & w.p.\ 1-q_i \end{cases}$
Platform commits to shuffle each user's data with probability μ.
Partial Shuffler
$w.p.\ 1-\mu$
Y_i
Platform
$w.p.\ \mu$
Shuffler

(a) (b)

Figure 4-1: (a) the mask-shuffle mechanism (b) the partial shuffler.

4.2.1 Mask-Shuffle Mechanism and Its Optimality

The space of mechanisms includes all possible mappings from users' data to an arbitrary space. In principle, this class includes a rich set of mechanisms. Nevertheless, we now establish that user-optimal mechanisms take a relatively simple form, which we call *mask-shuffle mechanisms*. In particular, we prove that the mask-shuffle mechanism achieves the minimum sum of revealed information about private user types, the Z_i's, for a given revealed information about θ.

Definition 4.2 (Mask-shuffle mechanism)**.** *A mask-shuffle mechanism is a pair* $(\mathbf{q}, \mu) \in [0,1]^{n+1}$ *such that:*

1. *The data of each user* $i \in \mathcal{N}$ *is completely hidden from the platform with probability* $1 - q_i$ *(denoted by NA) and is kept with probability* q_i*.*

2. *Letting* Y_i *denote the user* i*'s data after this randomized mapping, the mechanism directly releases each* Y_i *with an independent probability* $1 - \mu$ *and shuffles the rest and releases a permutation of these shuffled* Y_i*'s (i.e.,* $Y_{i_{\sigma(1)}}, \ldots, Y_{i_{\sigma_k}}$ *for some random permutation* σ *where* k *is the number of* Y_i*'s that are shuffled).*

Figure 4-1a illustrates the mask-shuffle mechanism that includes a partial shuffler that shuffles each user's data with probability μ. Figure 4-1b further depicts this partial shuffling element.

Before establishing the optimality of a mask-shuffle mechanism, we explicitly characterize the revealed information in terms of the shuffling parameter μ and the users' action profile $\mathbf{q}$. In what follows, we use the following notation: for any $\mathbf{v} \in [0,1]^k$ and $j \in \{1, \cdots, k\}$, we define

$$S_j(\mathbf{v}) := \sum_{\substack{B \subseteq \{1, \cdots, k\} \\ |B| = j}} \prod_{\ell \in B} v_\ell \prod_{\ell \notin B} (1 - v_\ell). \tag{4.3}$$

This function is also known as the probability density function of Poisson binomial distribution which is the number of heads after k independent coin tosses when the probability of head for coin ℓ is v_ℓ (see, e.g., Wang [1993]).

Proposition 4.1. *For a given $\mu \in [0,1]$ and $\mathbf{q} \in [0,1]^n$, revealed information about θ can be written as*

$$\mathcal{I}(\theta \mid \mathbf{q}, \mu) = \sum_{j=0}^{n} \frac{j}{1+j} S_j(\mathbf{q}).$$

In addition, revealed information about Z_i can be written as

$$\mathcal{I}(Z_i \mid \mathbf{q}, \mu) = (1 - \mu_i) q_i \left(1 - \sum_{k=1}^{n} S_{k-1}(\mathbf{q}_{-i}) \frac{1}{1+k} \right) \tag{4.4}$$
$$+ \sum_{k=1}^{n} \sum_{\substack{B \subseteq \mathcal{N} \\ i \in B, |B| = k}} \sum_{j=1}^{k} \sum_{r=0}^{n-k} \left(\prod_{\ell \in B} \mu_\ell \right) \left(\prod_{\ell \notin B} (1 - \mu_\ell) \right) S_r(\mathbf{q}_{\mathcal{N} \setminus B}) \frac{q_i^2 S_{j-1}^2(\mathbf{q}_{B \setminus i})}{S_j(\mathbf{q}_B)} \frac{1+r}{j(1+(j+r))}$$

where $\mathbf{q}_B := (q_\ell)_{\ell \in B}$.

This result shows that revealed information about θ does not depend on the shuffling parameter μ because irrespective of whether a user's data is shuffled or not the platform can extract the relevant information about θ in this user's data. Revealed information about Z_i, however, depends on the shuffling parameter. In fact, the first term on the right-hand side of (4.4) captures revealed information about Z_i when the data of user i is not shuffled, and the second term corresponds to the case that data of user i is shuffled. It is worth highlighting that, to derive the second term, we first need to characterize the platform's belief on which one of the shuffled data belongs to user i.

We next establish that for a given desired level of revealed information about the common parameter θ, the mask-shuffle mechanism achieves the lowest possible sum of revealed information regarding private types Z_i's. Let us formalize this notion of optimality. Let

$$\mathcal{P} = \left\{ (A, B) : \; A = \mathcal{I}(\theta \mid \mathcal{M}), \; B = \sum_{i=1}^{n} \mathcal{I}(Z_i \mid \mathcal{M}) \text{ for some mechanism } \mathcal{M} \right\}$$

be the set of all pairs of revealed information about θ and revealed information about Z_i's achieved by any mechanism. Let us denote the smallest and largest possible values of A by $\underline{A}$ and $\bar{A}$, respectively. For any $A \in [\underline{A}, \bar{A}]$, the *Pareto frontier* of $\mathcal{P}$ is defined as

$$\left\{ (A, \mathrm{PF}(A)) : \; A \in [\underline{A}, \bar{A}] \right\} \text{ where } \mathrm{PF}(A) = \inf \left\{ B : \; (A, B) \in \mathcal{P} \right\}. \tag{4.5}$$

We next prove that the mask-shuffle mechanism achieves the Pareto frontier of all possible mechanisms.

Theorem 4.1. *For any $A \in [\underline{A}, \bar{A}]$, there exists a mask-shuffle mechanism $\mathcal{M} = (\mathbf{q}, 1)$, for some $q \in [0, 1]$, for which*

$$\mathcal{I}(\theta \mid \mathcal{M}) = A \textit{ and } \sum_{i=1}^{n} \mathcal{I}(Z_i \mid \mathcal{M}) = \mathrm{PF}(A).$$

Theorem 4.1 has two consequences. First, by varying the probability of sharing q from zero to one, revealed information about θ goes from zero to the highest possible level of revelation among all mechanisms. Moreover, for any given revealed information about θ in this range, the lowest possible leakage of users' private information is achieved by a mask-shuffle mechanism with a certain sharing probability q.

In closing, we should highlight that various forms of shuffling have been studied in the differential privacy literature as a technique to boost the provided privacy guarantees (see, e.g., Bittau et al. [2017] and Cheu [2021]). First, our mask-shuffle mechanism is different from simply shuffling all data points as it involves randomly masking some of the user data points and then partially and randomly shuffling them. Second, our analysis reveals the Pareto optimality of a mask-shuffle mechanism in our setting which gives it an important operational justification, unlike shuffling for the purpose of boosting privacy guarantees.

4.2.2 Proof of Theorem 4.1

Here, we present three key lemmas that we use to prove Theorem 4.1. Let us first provide the roadmap of the proof:

- We first prove that for any mechanism, the sum of the revealed information about θ and Z_i's is lower bounded by (a constant fraction of) the revealed information about $\sum_{i=1}^n X_i$. Intuitively, this holds because if a mechanism reveals too much about $\sum_{i=1}^n X_i = n\theta + \sum_{i=1}^n Z_i$, then it must be the case that it reveals information about either θ or Z_i's.

- We then establish that the revealed information about $\sum_{i=1}^n X_i$ is (a constant multiple of) the revealed information about θ. Intuitively, this holds because the conditional distribution of θ given $(X_1, \cdots, X_n)$ depends on $X_1, \cdots, X_n$ only through $\sum_{i=1}^n X_i$. Putting these two lemmas together, we establish a lower bound on the sum of the revealed information about Z_i's in terms of the revealed information about θ. This lower bound characterizes the Pareto frontier of $\mathcal{P}$, defined in (4.5).

- We finally prove that our mask-shuffled mechanism achieves this Pareto frontier.

We next state and prove the above results formally.

Lemma 4.1. *For any mechanism $\mathcal{M}$, we have*

$$\mathcal{I}(\theta \mid \mathcal{M}) + \sum_{i=1}^{n} \mathcal{I}(Z_i \mid \mathcal{M}) \geq \frac{\mathcal{I}\left(\sum_{i=1}^n X_i \mid \mathcal{M}\right)}{n^2 + n} \tag{4.6}$$

and the equality holds for a mask-shuffle mechanism $\mathcal{M} = (\mathbf{q}, 1)$ with any $\mathbf{q} = (q, \ldots, q)$.

Proof sketch: To show this result, we first establish a relation between the revealed information of a random variable and the square of it expectation conditioned on the mechanism $\mathcal{M}$'s output. Using this derivation, we would need to bound the square of conditional expectation of $\sum_{i=1}^n X_i = n\theta + \sum_{i=1}^n Z_i$ by the the square of conditional expectation of θ and Z_i's. To prove such a bound, we use a Cauchy–Schwarz inequality and carefully tailor the weight that we assign to each of the conditional expectations to obtain the tightest bound. We also show that the equality case of the Cauchy–Schwarz inequality holds for the mask-shuffling mechanism by explicitly characterizing the conditional expectations in that case.

Lemma 4.1 establishes a relation among the revealed information about the underlying state θ, the users' type Z_i, and the users' data X_i. We are interested to find a relation between the revealed information about θ and Z_i's. Therefore, the next natural step is to show how the revealed information about $\sum_{i=1}^n X_i$ relates to the revealed information about θ, which is proved in our second lemma.

Lemma 4.2. *For any mechanism $\mathcal{M}$, we have*

$$\mathcal{I}\left(\sum_{i=1}^n X_i \mid \mathcal{M}\right) = (n+1)^2 \ \mathcal{I}(\theta \mid \mathcal{M}). \tag{4.7}$$

Proof sketch: As stated in the previous proof sketch, we know that the revealed information about θ is closely related to the conditional expectation of θ given the mechanism $\mathcal{M}$'s output. To establish the desired result, we show that the conditional expectation of θ and $\sum_{i=1}^n X_i$ only differ by a constant factor.

Deriving this result uses two main observations: (i) the Markov property of the mechanism: given $X_1, \cdots, X_n$, the output of the mechanism $\mathcal{M}$ is independent from θ, and (ii) the conditional distribution of θ given $(X_1, \cdots, X_n)$ only depends on $\sum_{i=1}^n X_i$.

The proof of Theorem 4.1 follows from plugging the relation of Lemma 4.1 into the bound given by Lemma 4.2. In particular, this proves that for any mechanism $\mathcal{M} : \mathbb{R}^n \to \mathcal{X}$, we have

$$\sum_{i=1}^n \mathcal{I}(Z_i \mid \mathcal{M}) \geq \frac{\mathcal{I}(\theta \mid \mathcal{M})}{n}. \tag{4.8}$$

Moreover, equality holds for mask-shuffle mechanism $\mathcal{M} = (\mathbf{q}, 1)$ for any $\mathbf{q} = (q, \ldots, q)$. Therefore, there is an inevitable minimum leakage of users' private information when a mechanism learns θ, and this minimum leakage increases as the mechanism reveals more about θ. Furthermore, the mask-shuffle mechanism has this minimum leakage, i.e., the mask-shuffle mechanism has the lowest possible leakage among all mechanisms that reveal equally about θ. This theorem proves the optimality of the mask-shuffle mechanism from the users' perspective.

The last remaining piece to finish the proof of Theorem 4.1 is to show that mask-shuffle mechanisms of the form $\mathcal{M} = (\mathbf{q}, 1)$ achieve all possible values of revealed information about θ. to see this, notice that by varying q from 0 to 1, the revealed information about θ by $\mathcal{M} = ((q, \cdots, q), 1)$, i.e., $\mathcal{I}(\theta \mid \mathbf{q}, \mu)$ goes from zero to $\frac{n}{n+1}$. The following lemma proves that no other mechanism can reveal more about θ.

Lemma 4.3. *The minimum (i.e., $\underline{A}$) and the maximum (i.e., $\bar{A}$) of $\mathcal{I}(\theta \mid \mathcal{M})$ over all mechanisms are 0 and $\frac{n}{n+1}$, respectively. Moreover, these bounds are achievable for a mask-shuffle mechanism $\mathcal{M} = (\mathbf{q}, 1)$ for some $q \in [0, 1]$.*

Combining Lemmas 4.1, 4.2, and 4.3 proves Theorem 4.1, establishing that the mask-shuffle mechanism achieves the Pareto frontier of revealed information about θ and revealed information about Z_i's.

4.2.3 The Game Between the Platform and Users

As we have seen, a mask-shuffle mechanism consists of a shuffling parameter $\mu \in [0, 1]$ and a vector of sharing probabilities $(q_1, \ldots, q_n)$. Since users own their data, we assume that they directly choose the probability with which their data will be shared with the platform, i.e., each user $i \in \mathcal{N}$ chooses q_i. We refer to $\mathbf{q}$ as the users' action profile. The shuffling parameter μ, on the other hand, is the platform's action: the platform commits to shuffle the data of each user who shares her data with probability $\mu \in [0, 1]$. The timing of the game is as follows:

1. The platform chooses her action μ, specifying the shuffling parameter.
2. Knowing the platform's shuffling parameter, all users simultaneously choose their action, specifying the probability with which they share their information with the shuffler.

The platform and the users choose their actions in an equilibrium that we introduce next.

4.3 Equilibrium

We use the notion of symmetric (Bayesian) Stackelberg equilibrium as our solution concept. Let us first define the user equilibrium for a given platform's action μ.

Definition 4.3 (user equilibrium)**.** *For a given platform's action $\mu \in [0,1]$, a user action profile $\mathbf{q} = (q, \ldots, q)$ is a symmetric Bayesian Nash equilibrium if*

$$\mathcal{U}_i(\mathbf{q}, \mu) \geq \mathcal{U}_i((\mathbf{q}_{-i}, q_i = q'), \mu) \quad \textit{for all } i \in \mathcal{N}, q',$$

where $\mathbf{q}_{-i} = (q_1, \ldots, q_{i-1}, q_{i+1}, \ldots, q_n)$.

We use the notion of symmetric equilibrium to simplify the analysis and to rule out the existence of unintuitive user equilibria. In the rest of the chapter, we adopt the following assumption.

Assumption 4.1. *$\alpha \geq \beta$, where α and β are the weight of the revealed information about the common parameter θ and user's data, respectively, in the user's utility (given in (4.1)).*

Assumption 4.1 focuses attention on the part of the parameter space where there is sufficient value in increasing information about the underlying common state θ. In particular, it rules out the case in which all users choose not to share their information, as we show next:

Proposition 4.2. *Suppose Assumption 4.1 holds.*

1. *For any platform's action $\mu < 1$, there exists $N(\mu)$ such that for $n \geq N(\mu)$ any symmetric user equilibrium is of the form $\mathbf{q} = (q, \ldots, q)$, with $q = \frac{c}{n} + \mathcal{O}(\frac{1}{n^2})$, where c is the unique solution of*

$$\alpha \frac{1 - (c+1)e^{-c}}{c^2} = \beta(1-\mu)\left(1 - \frac{1}{c}\left(1 - \frac{1 - e^{-c}}{c}\right)\right).$$

2. *For platform's action $\mu = 1$, there exists N such that for $n \geq N$, we have the following cases:*

2.1. If $\frac{\alpha}{\beta} \leq 2$, then any intermediary symmetric user equilibrium is of the form $\mathbf{q} = (q, \ldots, q)$, where $q = \frac{\alpha}{2\beta} + \mathcal{O}\left(\frac{\log(n)}{n}\right)$. Also, $\mathbf{q} = (1, \ldots, 1)$ is a user equilibrium.

2.2. If $\frac{\alpha}{\beta} > 2$, then $\mathbf{q} = (1, \ldots, 1)$ is the unique symmetric user equilibrium.

To characterize the symmetric user equilibrium $(q, \cdots, q)$, we let user 1 share her data with probability q_1 and other users share their data with probability q. For $(q, \cdots, q)$ to be a symmetric user equilibrium, we must have that user 1's utility $\mathcal{U}_1(\mathbf{q}, \mu)$ as a function of q_1 is maximized by choosing $q_1 = q$. We solve for such q by considering the first-order conditions and also checking the boundary cases. There are a few points worth mentioning. First, Assumption 4.1 rules out $q_i = 0$ for all $i \in \mathcal{N}$ as an equilibrium. Second, Proposition 4.2 characterizes the users' equilibrium action with a $1/n^2$ precision. Although characterizing the exact constant of the $1/n^2$ term is demanding, in what follows, we prove that this term only affects the lower order terms in the utility functions of the users and the platform.

We next define the Stackelberg equilibrium of the game.

Definition 4.4 (Stackelberg equilibrium)**.** *A pair of (q^e, μ^e) is a symmetric Stackelberg equilibrium if $\mathbf{q}^e = (q^e, \ldots, q^e)$ is a symmetric user equilibrium for μ^e and*

$$\mathcal{U}_{platform}(\mathbf{q}^e, \mu^e) \geq \mathcal{U}_{platform}(\mathbf{q}', \mu'),$$

for any μ' and $\mathbf{q}'$ such that $\mathbf{q}'$ is a symmetric user equilibrium for μ'.

Theorem 4.2. *Suppose Assumption 4.1 holds. There exists a symmetric Stackelberg equilibrium (μ^e, q^e).*

Theorem 4.2 proves the existence of a symmetric Stackelberg equilibrium. In general, such an equilibrium may not be unique. However, in what follows, we prove the properties of the game among the users and the platform that holds for any symmetric Stackelberg equilibrium.

4.4 Characterization

In this section, we characterize the equilibrium and then provide some comparative statics.

Our next theorem proves that for a sufficiently large number of users if α (i.e., the weight of the revealed information about the common parameter θ in the user's utility) is small enough, then the platform's equilibrium shuffling probability is close to 1 (i.e., the platform shuffles almost all the unmasked data points). Conversely, if α is large enough, then the platform's equilibrium shuffling decision is close to 0 (i.e., the platform shuffles almost none of the unmasked data points).

Theorem 4.3. *Suppose $\delta \leq 1$ and Assumption 4.1 holds. For any $\epsilon > 0$, there exists $\underline{\alpha}$ and $\bar{\alpha}$ in $[\beta, \infty)$ and $N^e(\epsilon)$, such that for $n \geq N^e(\epsilon)$ we have:*

1. *If $\alpha \leq \underline{\alpha}$, then $\mu^e \geq 1 - \epsilon$.*

2. *If $\alpha \geq \bar{\alpha}$, then $\mu^e \leq \epsilon$.*

The proof of this theorem relies on the following steps. From Proposition 4.2, for any ϵ there exists $N(\epsilon)$ such that the derivation of Proposition 4.2 holds for $n \geq N(\epsilon)$ and $\mu \leq 1 - \epsilon$. Therefore, to find the optimal choice of μ^e for the platform, we consider two intervals $[0, 1-\epsilon)$ and $[1-\epsilon, 1]$ separately. In particular, we characterize the user equilibrium for any $\mu \in [0, 1-\epsilon)$ by invoking Proposition 4.2, and we find the best choice of shuffling probability for the platform. We also upper bound the platform's utility when the platform chooses $\mu \in [1-\epsilon, 1]$. Putting these two results together, we complete the proof of Theorem 4.3.

To understand the intuition of Theorem 4.3, let us consider what happens when the platform increases the shuffling parameter μ. There are two opposing forces that shape equilibrium decisions. First, for a given user action profile $\mathbf{q}$, the choice of the shuffling parameter μ does not directly change revealed information about θ (as shown in Proposition 4.1) but decreases revealed information about the users' data. Second, increasing the shuffling parameter μ incentivizes the users to share with a higher probability, which increases the platform's utility because it increases both revealed information about θ and about users' data. Theorem 4.3 establishes that for small enough α, the second force dominates and the platform's equilibrium choice is to increase the shuffling parameter very close to 1. For large enough α, on the other hand, the first force dominates and the platform's equilibrium choice is to decrease the shuffling parameter very close to 0.

We next establish our main comparative static result that establishes as α (the weight users attach to information about the underlying, common state θ) increases, they may become worse off. Recall that, holding the privacy mechanism constant, a higher α leads to greater user utility. The next theorem is therefore a paradoxical result on the response of the platform by varying the extent of privacy guarantees.

Theorem 4.4. *Suppose $\delta \leq 1$ and Assumption 4.1 holds. Then, there exists an interval (α_L, α_H) such that the user's utility at equilibrium as a function of α is decreasing over it for sufficiently large n, i.e., for any $\alpha_1 < \alpha_2$ in (α_L, α_H), there exists N such that for any $n \geq N$ the user's utility at equilibrium is larger for $\alpha = \alpha_1$ compared to $\alpha = \alpha_2$.*

We prove that this phenomenon happens when the shuffling probability μ^e at equilibrium starts to decrease from one to zero by increasing α. More precisely, as α increases, the platform takes advantage of the fact that users care more about learning the underlying common state and decreases the probability of shuffling, knowing that users will still share their data. However, the main challenge is that, at the same time, the user's gain from learning the underlying state increases. Nevertheless, we prove that the users' loss from the reduction of the shuffling parameter (and hence the increase of revealed information about their private types Z_i's) dominates their gain from learning the state θ, and hence, the total utility of users decreases.

4.5 Platform Choice of Mechanism: Pivot Vs. Mask-Shuffle Mechanisms

In this section, we characterize the platform's optimal choice of mechanism and establish that platforms will in general choose mechanisms quite different from the mask-shuffle mechanism that is user-optimal, as shown above. Recall that the action of each user such as user i is her sharing probability q_i, and

$$Y_i = \begin{cases} X_i & \text{with probability } q_i \\ \text{NA} & \text{with probability } 1 - q_i \end{cases}$$

is the input of the platform. The platform's action is a mapping from $(Y_1, \dots, Y_n)$ to $\mathcal{X}$ for some set $\mathcal{X}$. The output of the platform's action will then be used to estimate the underlying state θ as well as the private users' data Z_i.

The mask-shuffle mechanism is one particular platform's action, but the space of all platform's actions is very large and it is not clear what would be the platform's optimal choice in this space. Interestingly, we next establish that the optimal platform's action belongs to the following class:

Definition 4.5 (Pivot mechanism)**.** *A pivot mechanism is defined based on a function $\sigma : \mathcal{N} \to \mathbb{R}_+$ such that: when k users share their data, the platform adds a Gaussian noise with zero mean and variance $\sigma^2(k)$ to all users who have shared.*

Intuitively, we refer to these mechanisms as "pivot mechanisms" because they increase the pivotal role of each user, as their sharing decision influences whether the platform can use the data shared by others. A special case of the pivot mechanism that is optimal from the platform's perspective is given below.

Theorem 4.5. *Suppose $\sigma(.)$ satisfies the following condition:*

$$\sigma^2(k-1) \geq \frac{\alpha}{\alpha - \beta} \left(\sigma^2(k) + k + 1\right) \text{ and } \sigma(n) = 0. \tag{4.9}$$

Then, the only symmetric user equilibrium under the pivot mechanism is $q_i = 1$ for all i. Furthermore, the platform's utility under this equilibrium is the maximum platform's utility over all possible mechanisms.

Let us first understand user behavior given such a pivot mechanism. Intuitively, inequality (4.9) ensures that without the user in question sharing her data there will be so much noise added to the data of other users who have shared that estimating the underlying common state, θ, becomes close to impossible for the platform. This is the sense in which the pivot mechanism makes each user pivotal: by refusing to share her data, the user makes it impossible to estimate this underlying state. If α is sufficiently large, as implied by condition (4.9), this is very costly for the user, and she will be convinced to sacrifice her privacy in order to allow the estimation of θ. Given this user behavior, the platform then has a strong incentive to deviate from the user-optimal mask-shuffle mechanism towards such a pivot mechanism.

To clarify the implications of this theorem, we next consider a simple form of this pivot mechanism as a corollary.

Corollary 4.1. *For a pivot mechanism with*

$$\sigma(k) = \begin{cases} 0 & k = n \\ \infty & k < n, \end{cases} \tag{4.10}$$

the unique symmetric user equilibrium is $q_i = 1$ for all i and the platform's utility is $\frac{n(n\delta+1)}{n+1}$, which is the maximum utility over all possible mechanisms.

Under the above pivot mechanism, the platform does not add any noise to users' data so long as they all share. Conversely, the platform "throws away" all users' data even if one of them does not share.

The implications for user utility are dire, however. To see this, we next characterize user welfare under the pivot mechanisms favored by the platform.

Proposition 4.3. *Suppose $\sigma(.)$ satisfies condition (4.9) so that the unique user equilibrium under the pivot mechanism is $q_i = 1$ for all i. The utility of each user is*

$$(\alpha - \beta)\frac{n}{n+1}.$$

4.6 Conclusion

Many platforms deploy data collected from users for a multitude of purposes. Some of these are beneficial to users, for example, when the day-to-day share enables platforms or others to learn more about underlying health conditions or provide better, objective recommendations to them. However, other consequences of extensive data harvesting are potentially very costly for users. Some of those will directly violate their privacy and others will lead to intensive target digital ads. In the extreme, the unregulated sale of individualized data to third parties could be highly problematic for users.

When privacy costs are substantial, users may not be willing to share their data and a shy away from participation in platforms that do not provide explicit guarantees on privacy. This has motivated many platforms to introduce guidelines on how they will treat user data. Despite the growing importance of this problem, we are not aware of systematic studies of how these guarantees are determined and to what extent they subjectively or objectively satisfy user concerns.

This work has taken a first step in the study of this question. We have built a multi-stage model in which users decide whether to share their data based on the privacy-deserving mechanism choices of platforms. Our model captures several salient features of the data-related relationships between platforms and users but is still highly tractable. As a result, we are able to establish several novel results that are of both theoretical interests and provide guidance on the faultiness that exists in private data markets.

Our first result establishes that mask-shuffle mechanism, whereby the user data is fully anonymized with some probability, is Preto optimal, meaning it achieves the minimum information leakage about users' data for any given revealed information about the underlying common parameter. This also implies that it is optimal from the viewpoint of users. With mask-shuffle mechanisms, there exists a unique equilibrium in which the mechanism offered by the platform balances the utility gains from the desirable uses of data with privacy costs for users.

Our second result characterizes the (Bayesian) Stackelberg equilibrium of the game between the platform and the users. This equilibrium concept takes into account that the platform acts first by choosing (committing to) a particular mechanism for privacy preservation (and hence acts like a “Stackelberg leader’ as in the game-theoretic analysis of oligopolistic markets). The label Bayesian refers to the fact that individuals make inferences about how much information will leak about the underlying state and their individual types to the user.

Third and somewhat paradoxically, we show that when the potential utility gains from data pooling increases for users (for example, because data can reveal information about underlying health conditions), users can become worse off. This result is because platforms take advantage of such changes to reduce privacy guarantees so much that user utility declines. This result should be contrasted with what would have happened if the privacy-preserving mechanism was held constant: in this case, user utility would have unambiguously increased because users would have benefited from better deployment of data. The reason why this paradoxical result obtained is because the platform exploits the change in user preferences to reduce privacy guarantees. Our interpretation is that this result highlights the fragility of platform-provided (self-regulated) privacy guarantees.

Finally, we explore the implications of the same forces for platform choice of data architecture. Here, we find that, even more strikingly, platforms have strong incentives to deviate from user-optimal mask-shuffle mechanisms. The reason for this finding is instructive: the platform designs a mechanism (which we refer to as a pivot mechanism) that links whether it can use other users' data to the decision of a marginal user about whether to share her own data. This makes each user pivotal: if they refuse to share their data, it becomes impossible for the platform to use the data of others to estimate the underlying common state (which is valuable for all users). With such pivotal mechanisms, the platform convinces users to sacrifice their privacy, but with significant costs to the welfare of users. This result further amplifies our conclusion that self-regulated privacy guarantees are unlikely to be sufficient for users to obtain high levels of benefit from online platform data architectures.

We view this work as a first step in the analysis of dynamic data markets, when data can be put to a multitude of uses. Several interesting areas remain for future study. First, we assumed that the platform can fully commit to a mechanism, whereas in practice platforms can create ambiguity about how data will be used and deviate from certain promises. The analysis of these issues is more challenging, as it requires an explicit modeling of platform reputation. Second, greater heterogeneity and more diverse uses of data can be introduced into our framework. Third, users typically participate in online platforms over many periods, and thus issues of dynamic data sharing are important in practice. These are also interesting areas for future study. Last but not least, it is important to empirically assess how users react to the prevailing privacy-preserving mechanisms and test some of the implications of this type of approach.

4.7 Proofs

This section includes the omitted proofs from the text and additional results.

Properties of the revealed information measure

Here, for the sake of subsequent analysis, we first generalize the definition of revealed information, provided in Definition 4.1. With slight abuse of notation, we use $\mathcal{I}(.)$ in this case as well.

Definition 4.6. *For any real-valued random variable W and any σ-Field $\mathcal{F}$, the revealed information about W given $\mathcal{F}$ is defined as*

$$\mathcal{I}(W \mid \mathcal{F}) = Variance(W) - \min_{\substack{\tilde{W} \ is \\ \mathcal{F}-measurable}} \mathbb{E}\left[\left(W - \tilde{W}\right)^2\right], \tag{4.11}$$

where the minimization is taken over all random variables $\tilde{W}$ that are $\mathcal{F}$-measurable. In addition, for a random variable H, $\mathcal{I}(W \mid H)$ is defined as $\mathcal{I}(W \mid \sigma(H))$, where $\sigma(H)$ denotes the σ-field generated by H.

It is known [Durrett, 2019, Theorem 4.1.15] that minimum in (4.11) is achieved by choosing $\tilde{W} = \mathbb{E}[W \mid \mathcal{F}]$. We next use this fact to characterize $\mathcal{I}(W \mid \mathcal{F})$.

Lemma 4.4. *Suppose $\mathbb{E}[W^2] < \infty$. Then,*

$$\mathcal{I}(W \mid \mathcal{F}) = \mathbb{E}\left[\mathbb{E}[W \mid \mathcal{F}]^2\right] - \mathbb{E}[W]^2.$$

Proof of Lemma 4.4: Given that minimum in (4.11) is achieved by choosing $\tilde{W} = \mathbb{E}[W \mid \mathcal{F}]$, we should substitute $\tilde{W}$ by $\mathbb{E}[W \mid \mathcal{F}]$ in (4.11). By doing so, we obtain

$$\begin{aligned}\mathcal{I}(W \mid \mathcal{F}) &= \mathbb{E}[W^2] - \mathbb{E}[W]^2 - \mathbb{E}\left[(W - \mathbb{E}[W \mid \mathcal{F}])^2\right] \\ &= 2\mathbb{E}\left[W \, \mathbb{E}[W \mid \mathcal{F}]\right] - \mathbb{E}\left[\mathbb{E}[W \mid \mathcal{F}]^2\right] - \mathbb{E}[W]^2 \\ &= \mathbb{E}\left[\mathbb{E}[W \mid \mathcal{F}]^2\right] - \mathbb{E}[W]^2,\end{aligned}$$

where the last equality follows from the following property of conditional expectation: for any $\mathcal{F}$-measurable random variable H, we have $\mathbb{E}[WH] = \mathbb{E}\left[\mathbb{E}[W \mid \mathcal{F}]\ H\right]$. Here we use it with $H = \mathbb{E}[W \mid \mathcal{F}]$. ■

As a consequence, the following lemma holds.

Lemma 4.5. *Suppose W is a zero-mean random variable with $\mathbb{E}[W^2] < \infty$. Then, for a discrete random variable H, we have*

$$\mathcal{I}(W \mid H) = \sum_{h \in supp(H)} \mathbb{P}(H = h)\mathcal{I}(W \mid H = h).$$

Proof of Lemma 4.5: Using Lemma 4.4, and since W is zero-mean, we have $\mathcal{I}(W \mid H) = \mathbb{E}\left[\mathbb{E}[W \mid \sigma(H)]^2\right]$, where the outer expectation is taken over H. Using the linearity of this expectation, we obtain the desired result. ■

Proof of Proposition 4.1

For θ, note that indices of data points do not matter, since all X_i's have identical distribution. Hence, shuffling does not have any effect on the estimation of θ. More formally, using Lemma 4.5, we have

$$\begin{aligned}\mathcal{I}(\theta \mid \mathbf{q}, \boldsymbol{\mu}) &= \sum_{j=1}^{n} \sum_{\substack{B \subseteq \{1,\cdots,n\} \\ |B|=j}} \prod_{\ell \in B} q_\ell \prod_{\ell \notin B}(1 - q_\ell)\ \mathcal{I}(\theta \mid (X_k)_{k \in B}) \\ &= \sum_{j=1}^{n} \sum_{\substack{B \subseteq \{1,\cdots,n\} \\ |B|=j}} \prod_{\ell \in B} q_\ell \prod_{\ell \notin B}(1 - q_\ell)\ \mathbb{E}\left[\mathbb{E}[\theta \mid (X_k)_{k \in B})]^2\right], \qquad (4.12)\end{aligned}$$

where the second equation follows from Lemma 4.4. Next, we derive $\mathbb{E}\left[\mathbb{E}[\theta \mid (X_k)_{k \in B})]^2\right]$ for any $B \subseteq \{1, \cdots, n\}$. Note that θ and $(X_k)_{k \in B}$ are jointly Gaussian, where the mean of their joint distribution is 0 and the covariance matrix of their joint distribution is given by

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 2 \end{bmatrix}. \tag{4.13}$$

Hence, the distribution of θ given $(X_k)_{k\in B}$ is Gaussian, and its mean is given by

$$\mathbb{E}[\theta \mid (X_k)_{k\in B})] = [1 \cdots 1] \begin{bmatrix} 2 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 2 \end{bmatrix}^{-1} [X_k]_{k\in B}^{\top}. \tag{4.14}$$

Using the Sherman–Morrison formula for the inverse of rank-1 perturbation of a matrix, we can write

$$\begin{bmatrix} 1+1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1+1 \end{bmatrix}^{-1} = \begin{bmatrix} 1-1/\nu & \cdots & -1/\nu \\ \vdots & \ddots & \vdots \\ -1/\nu & \cdots & 1-1/\nu \end{bmatrix}, \tag{4.15}$$

with $\nu = |B| + 1$. Plugging this into (4.14), yields

$$\mathbb{E}[\theta \mid (X_k)_{k\in B})] = \left(1 - \frac{|B|}{\nu}\right) [1 \cdots 1][X_k]_{k\in B}^{\top} = \frac{1}{|B|+1} \sum_{k\in B} X_k. \tag{4.16}$$

Therefore, we have

$$\begin{aligned} \mathbb{E}\left[\mathbb{E}[\theta \mid (X_k)_{k\in B})]^2\right] &= \left(\frac{1}{|B|+1}\right)^2 \mathbb{E}\left[\left(\sum_{k\in B} X_k\right)^2\right] \\ &= \left(\frac{1}{|B|+1}\right)^2 \left(|B|^2 + |B|\right) = \frac{|B|}{|B|+1}. \end{aligned} \tag{4.17}$$

Substituting (4.17) into (4.12) implies

$$\mathcal{I}(\theta \mid \mathbf{q}, \boldsymbol{\mu}) = \sum_{j=1}^{n} \sum_{\substack{B \subseteq \{1,\cdots,n\} \\ |B|=j}} \prod_{\ell \in B} q_\ell \prod_{\ell \notin B} (1 - q_\ell) \, \frac{j}{1+j}, \tag{4.18}$$

which gives us the desired result.

Next, we focus on $\mathcal{I}(Z_i \mid \mathbf{q}, \boldsymbol{\mu})$. Using Lemma 4.5, we can write

$$\begin{aligned}
&\mathcal{I}(Z_i \mid \mathbf{q}, \boldsymbol{\mu}) = (1 - \mu_i) q_i \sum_{k=1}^{n} S_{k-1}(\mathbf{q}_{-i}) \\
&\mathcal{I}(Z_i \mid \mathbf{q}, \boldsymbol{\mu}, \text{user } i \text{ data is shared and not shuffled}, k-1 \text{ other data points are shared}) \\
&+ \sum_{k=1}^{n} \sum_{\substack{B \subseteq \mathcal{N} \\ i \in B, |B|=k}} \sum_{j=1}^{k} \sum_{r=0}^{n-k} \left(\prod_{\ell \in B} \mu_\ell \right) \left(\prod_{\ell \notin B} (1 - \mu_\ell) \right) S_r(\mathbf{q}_{\mathcal{N} \setminus B}) S_j(\mathbf{q}_B) \mathcal{I}(Z_i \mid \mathbf{q}, \boldsymbol{\mu}, \\
&\quad \text{data of set } B \text{ is shuffled}, j \text{ data points in } B \text{ and } r \text{ data points in } \mathcal{N} \setminus B \text{ are shared}),
\end{aligned} \tag{4.19}$$

where the terms of the first summation correspond to the case that the data of user i is not shuffled, and therefore the revealed information about Z_i is non-zero only if user i shares her data. Each term of the summation corresponds to having $k-1$ other data points shared (as we show next, only the number of shared data points matters in the revealed information and not their identity). The second term corresponds to the case that the data of user i is shuffled. In this case, we let B be the set of shuffled data points and we condition the events to having j data points in B and r data points in $\mathcal{N} \setminus B$ being shared. We next find the revealed information in each of these cases.

Finding $\mathcal{I}(Z_i \mid \mathbf{q}, \boldsymbol{\mu},$ user i data is shared and not shuffled$, k - 1$ other data points are shared): Using Lemma 4.4, we need to find the conditional expectation of Z_i. Without loss of generality, we next find the conditional expectation of Z_1 given $X_1, \ldots, X_k$. Notice that the joint distribution of $Z_1, X_1, \ldots, X_k$ is normal with covariance matrix

$$\begin{bmatrix} 1 & 1 & \cdots & 0 \\ 1 & 2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 2 \end{bmatrix}.$$

Therefore, we have

$$\begin{aligned} \mathbb{E}[Z_1 \mid X_1, \ldots, X_k] =& (1, 0, \ldots, 0) \begin{bmatrix} 2 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 2 \end{bmatrix}^{-1} (X_1, \ldots, X_k)^T \\ =& (1, 0, \ldots, 0) \begin{bmatrix} 1 - \frac{1}{\nu_k} & \cdots & -\frac{1}{\nu_k} \\ \vdots & \ddots & \vdots \\ -\frac{1}{\nu_k} & \cdots & 1 - \frac{1}{\nu_k} \end{bmatrix}^{-1} (X_1, \ldots, X_k)^T \\ =& \left(1 - \frac{1}{\nu_k}\right) X_1 - \sum_{\ell=2}^{k} \frac{1}{\nu_k} X_\ell, \end{aligned} \tag{4.20}$$

where $\nu_k = k + 1$. Therefore, we have

$$\begin{aligned} &\mathcal{I}(Z_i \mid \mathbf{q}, \boldsymbol{\mu}, \text{user } i \text{ data is shared and not shuffled}, k-1 \text{ other data points are shared}) \\ &= \mathbb{E}\left[\left(\left(1 - \frac{1}{\nu_k}\right) X_1 - \sum_{\ell=2}^{k} \frac{1}{\nu_k} X_\ell\right)^2\right] = 1 - \frac{1}{1+k}. \end{aligned} \tag{4.21}$$

Finding $\mathcal{I}(Z_i \mid \mathbf{q}, \boldsymbol{\mu},$ data of set B is shuffled$, j$ data points in B and r data points in $\mathcal{N} \setminus B$ are shared): Using Lemma 4.4, we need to find the conditional expectation of Z_i. We can write

$$\begin{aligned}
&\mathbb{E}\left[Z_i \mid \text{ data of set } B \text{ is shuffled}, j \text{ data points in } B \text{ and } r \text{ points in } \mathcal{N} \setminus B \text{ are shared}\right] \\
&\overset{(a)}{=} \mathbb{P}(i \in B \text{ is among those that have shared})\mathbb{E}\Big[Z_i \mid \text{ data of set } B \text{ is shuffled}, j \\
&\quad \text{data points including } i \text{ in } B \text{ and } r \text{ data points in } \mathcal{N} \setminus B \text{ are shared}\Big] \\
&\overset{(b)}{=} \mathbb{P}(i \in B \text{ is among those that have shared}) \sum_{\ell=1}^{j} \frac{1}{j}\mathbb{E}\Big[Z_i \mid \text{ data of set } B \text{ is shuffled}, j \\
&\quad \text{data points in } B \text{ and } r \text{ data points in } \mathcal{N} \setminus B \text{ are shared}, \ell\text{-th one is } i\Big] \\
&\overset{(c)}{=} \mathbb{P}(i \in B \text{ is among those that have shared}) \\
&\sum_{\ell=1}^{j} \frac{1}{j}\left(\left(1 - \frac{1}{\nu_{j+r}}\right)\tilde{X}_\ell - \sum_{t=1, t\neq \ell}^{j+r} \frac{1}{\nu_{j+r}}\tilde{X}_t\right) \\
&\overset{(d)}{=} \mathbb{P}(i \in B \text{ is among those that have shared})\frac{1}{j}\left(\sum_{\ell=1}^{j}\left(1 - \frac{j}{\nu_{j+r}}\right)\tilde{X}_\ell - \sum_{\ell=j+1}^{j+r}\frac{j}{\nu_{j+r}}\tilde{X}_\ell\right) \\
&\overset{(e)}{=} \frac{q_i S_{j-1}(\mathbf{q}_{B\setminus i})}{S_j(\mathbf{q}_B)}\frac{1}{j}\left(\sum_{\ell=1}^{j}\left(1 - \frac{j}{\nu_{j+r}}\right)\tilde{X}_\ell - \sum_{\ell=j+1}^{j+r}\frac{j}{\nu_{j+r}}\tilde{X}_\ell\right),
\end{aligned} \tag{4.22}$$

where (a) holds because if user i does not share, then the revealed information about Z_i is zero, (b) follows from the fact that the shuffled data points have no label and therefore i can be any of them with a uniform probability, (c) follows from a similar argument to that of (4.20), (d) follows from rearranging the terms, and (e) follows from the definition of $S_k(\mathbf{q})$.

Therefore, we have

$$\begin{aligned}
&\mathcal{I}(Z_i \mid \mathbf{q}, \boldsymbol{\mu}, \text{ data of set } B \text{ is shuffled}, j \text{ data points in } B \\
&\qquad \text{and } r \text{ data points in } \mathcal{N} \setminus B \text{ are shared)} \\
&= \frac{q_i^2 S_{j-1}^2(\mathbf{q}_{B\setminus i})}{S_j^2(\mathbf{q}_B)} \frac{1}{j^2} \mathbb{E}\left[\left(\sum_{\ell=1}^{j}\left(1 - \frac{j}{\nu_{j+r}}\right)\tilde{X}_\ell - \sum_{\ell=j+1}^{j+r} \frac{j}{\nu_{j+r}} \tilde{X}_\ell\right)^2\right] \\
&= \frac{q_i^2 S_{j-1}^2(\mathbf{q}_{B\setminus i})}{S_j^2(\mathbf{q}_B)} \frac{1+r}{j(1+(j+r))}.
\end{aligned} \tag{4.23}$$

By using (4.21) and (4.23) in (4.19), we obtain

$$\begin{aligned}
&\mathcal{I}(Z_i \mid \mathbf{q}, \boldsymbol{\mu}) = (1-\mu_i) q_i \sum_{k=1}^{n} S_{k-1}(\mathbf{q}_{-i}) \left(1 - \frac{1}{1+k}\right) \\
&+ \sum_{k=1}^{n} \sum_{\substack{B \subseteq \mathcal{N} \\ i \in B, |B| = k}} \sum_{j=1}^{k} \sum_{r=0}^{n-k} \left(\prod_{\ell \in B} \mu_\ell\right)\left(\prod_{\ell \notin B}(1-\mu_\ell)\right) S_r(\mathbf{q}_{\mathcal{N}\setminus B}) S_j(\mathbf{q}_B) \frac{q_i^2 S_{j-1}^2(\mathbf{q}_{B\setminus i})}{S_j^2(\mathbf{q}_B)} \frac{1+r}{j(1+(j+r))}.
\end{aligned}$$

This completes the proof of Proposition 4.1. ■

Proof of Lemma 4.1

By using Lemma 4.4, for any mechanism $\mathcal{M}$, we have

$$\mathcal{I}(\sum_{i=1}^{n} X_i \mid \mathcal{M}) = \mathbb{E}\left[\left(\mathbb{E}\left[\sum_{i=1}^{n} X_i \mid \mathcal{M}\right]\right)^2\right], \quad \mathcal{I}(\theta \mid \mathcal{M}) = \mathbb{E}\left[\left(\mathbb{E}\left[\theta \mid \mathcal{M}\right]\right)^2\right],$$

and

$$\mathcal{I}(Z_i \mid \mathcal{M}) = \mathbb{E}\left[\left(\mathbb{E}\left[Z_i \mid \mathcal{M}\right]\right)^2\right] \text{ for all } i.$$

We next evaluate each term of the above expectations. We can write

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n} X_i \mid \mathcal{M}\right]^2 =& \mathbb{E}\left[n\theta + \sum_{i=1}^{n} Z_i \mid \mathcal{M}\right]^2\\ &= \left(n\mathbb{E}\left[\theta \mid \mathcal{M}\right] + \sum_{i=1}^{n} \mathbb{E}\left[Z_i \mid \mathcal{M}\right]\right)^2\\ &\overset{(a)}{\leq} \left(\mathbb{E}\left[\theta \mid \mathcal{M}\right]^2 + \sum_{i=1}^{n} \mathbb{E}\left[Z_i \mid \mathcal{M}\right]^2\right)(n^2 + \sum_{i=1}^{n} 1)\end{aligned}$$

where (a) follows from Cauchy-Schwarz inequality. Taking expectation over the randomness in $\mathcal{M}$ gives us the desired bound. We next prove that equality holds when

$$\frac{1}{n}\mathbb{E}\left[\theta \mid \mathcal{M}\right] = \mathbb{E}\left[Z_i \mid \mathcal{M}\right] \text{ for all } i,$$

which is the case for $\mathcal{M} = ((q, \ldots, q), 1)$ for any q. To see this, we show that when the mechanism returns k shuffled datapoints $X_1, \cdots, X_k$, we have

$$\mathbb{E}[\theta \mid (X_\ell)_{\ell=1}^{k}] = \frac{1}{k+1}\sum_{\ell=1}^{k} X_\ell \text{ and} \tag{4.24}$$

$$\mathbb{E}[Z_i \mid (X_\ell)_{\ell=1}^{k}] = \frac{1}{n(k+1)}\sum_{\ell=1}^{k} X_\ell. \tag{4.25}$$

Let us prove (4.24) as (4.25) can be established similarly. To see why (4.24) holds, note that θ and $(X_\ell)_{\ell=1}^{k}$ are jointly Gaussian, where the mean of their joint distribution is 0, and the covariance matrix of their joint distribution is given by

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 2 \end{bmatrix}. \tag{4.26}$$

Hence, the distribution of θ given $(X_\ell)_{\ell=1}^k$ is Gaussian, and its mean is given by

$$\mathbb{E}[\theta \mid (X_\ell)_{\ell=1}^k] = [1 \cdots 1] \begin{bmatrix} 2 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 2 \end{bmatrix}^{-1} [X_1, \cdots, X_k]^\top. \tag{4.27}$$

Using the Sherman–Morrison formula for the inverse of rank-1 perturbation of a matrix, we can write

$$\begin{bmatrix} 1+1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1+1 \end{bmatrix}^{-1} = \begin{bmatrix} 1-\frac{1}{k+1} & \cdots & -\frac{1}{k+1} \\ \vdots & \ddots & \vdots \\ -\frac{1}{k+1} & \cdots & 1-\frac{1}{k+1} \end{bmatrix}. \tag{4.28}$$

Plugging this into (4.27), yields

$$\mathbb{E}[\theta \mid (X_\ell)_{\ell=1}^k] = \left(1 - \frac{k}{k+1}\right)[1 \cdots 1][X_1, \cdots, X_k]^\top = \frac{1}{k+1}\sum_{\ell=1}^k X_\ell. \tag{4.29}$$

This completes the proof. ■

Proof of Lemma 4.2

By using Lemma 4.4, for any mechanism $\mathcal{M}$ that has access to some random variable Y which is a function of $X_1, \ldots, X_n$, we have

$$\begin{aligned} \mathcal{I}(\theta \mid \mathcal{M}) = \mathbb{E}\left[(\mathbb{E}[\theta \mid Y])^2\right] &= \int_y f_y(y)\left(\int_\theta \theta f_{\theta|y}(\theta \mid y)d\theta\right)^2 dy \\ &\overset{(a)}{=} \int_y \frac{1}{f_y(y)}\left(\int_\theta \theta f_{\theta,y}(\theta, y)d\theta\right)^2 dy. \end{aligned} \tag{4.30}$$

where (a) follows from Bayes' rule. Similarly, we can write

$$\mathcal{I}(\sum_{i=1}^n X_i \mid \mathcal{M}) = \int_y \frac{1}{f_y(y)}\left(\int_{x_{1:n}}\left(\sum_{i=1}^n x_i\right) f_{x_{1:n},y}(x_{1:n}, y)dx_{1:n}\right)^2 dy. \tag{4.31}$$

We next compare each term of the above expressions and in particular prove that for any y,

$$\int_\theta \theta f_{\theta,y}(\theta, y) d\theta = \frac{1}{n+1} \int_{x_{1:n}} \left(\sum_{i=1}^{n} x_i \right) f_{x_{1:n},y}(x_{1:n}, y) d\theta$$

that together with equations (4.30) and (4.31), completes the proof. We can write

$$\begin{aligned}
\int_\theta \theta f_{\theta,y}(\theta, y) d\theta &\overset{(a)}{=} \int_\theta \theta \int_{x_{1:n}} f_{x_{1:n},\theta,y}(x_{1:n}, \theta, y) dx_{1:n} d\theta \\
&\overset{(b)}{=} \int_\theta \theta \int_{x_{1:n}} f_\theta(\theta) f_{x_{1:n}|\theta}(x_{1:n} \mid \theta) f_{y|x_{1:n},\theta}(y \mid x_{1:n}, \theta) dx_{1:n} d\theta \\
&\overset{(c)}{=} \int_\theta \theta \int_{x_{1:n}} f_{x_{1:n}}(x_{1:n}) f_{\theta|x_{1:n}}(\theta \mid x_{1:n}) f_{y|x_{1:n}}(y \mid x_{1:n}) dx_{1:n} d\theta \\
&\overset{(d)}{=} \int_{x_{1:n}} \left(\int_\theta \theta f_{\theta|x_{1:n}}(\theta \mid x_{1:n}) d\theta \right) f_{x_{1:n}}(x_{1:n}) f_{y|x_{1:n}}(y \mid x_{1:n}) dx_{1:n} \\
&= \int_{x_{1:n}} \mathbb{E}\left[\theta \mid x_{1:n}\right] f_{x_{1:n},y}(x_{1:n}, y) dx_{1:n} \\
&\overset{(e)}{=} \int_{x_{1:n}} \frac{\sum_{i=1}^n x_i}{n+1} f_{x_{1:n},y}(x_{1:n}, y) dx_{1:n},
\end{aligned}$$

where (a) follows from the law of total probability, (b) follows from Bayes' rule, (c) follows from the fact that the mechanism has access to $X_1, \ldots X_n$ and not θ and, therefore, conditional on $X_1, \ldots, X_n$, Y and θ are independent, and (e) follows from (4.27) established in the proof of Lemma 4.1. It is also worth mentioning that we do a change of integration in (d). To see why we are allowed to do so, note that

$$\mathbb{E}\left[\mathbb{E}[|\theta| \mid Y]\right] \leq \mathbb{E}[|\theta|] < \infty,$$

and hence $\mathbb{E}[|\theta| \mid Y]$ is almost surely bounded, i.e.,

$$\mathbb{E}[|\theta| \mid Y = y] = \int_\theta \int_{x_{1:n}} |\theta| \; f_{x_{1:n},\theta|y}(x_{1:n}, \theta \mid y) dx_{1:n} d\theta < \infty \text{ a.s.}$$

Therefore,

$$\int_\theta \int_{x_{1:n}} |\theta| \; f_{x_{1:n},\theta,y}(x_{1:n}, \theta, y) dx_{1:n} d\theta < \infty \text{ a.s.}$$

Thus, by Fubini's theorem, we are allowed to change the order of integrals. This completes the proof. ■

Proof of Lemma 4.3

Among all estimators, we know that the estimator that achieves the minimum mean-squared error is the conditional expectation $\mathbb{E}[\theta \mid X_1, \cdots, X_n]$. Furthermore, the error of this estimator is equal to the error mask-shuffle mechanism $((1, \cdots, 1), 1)$, and hence the proof is complete. ■

Proof of Proposition 4.2

To find the symmetric the user equilibrium, for a fixed μ, suppose user 1 plays q_1 and users $2, \cdots, n$ play q, i.e., $\mathbf{q} = (q_1, q, \cdots, q)$. The symmetric user equilibrium must be such that the maximum of user 1's utility $\mathcal{U}_1(\mathbf{q}, \mu)$ as a function of q_1 is attained for $q_1 = q$. To find such q, we find the maximizer of $\mathcal{U}_1(\mathbf{q}, \mu)$ as a function of q_1 by using first order condition and then finding q such that the maximizer is $q_1 = q$. We also check the boundary cases $q = 0$ and $q = 1$ at the end.

Characterizing $\left.\frac{d\mathcal{I}(\theta \mid \mathbf{q}, \mu)}{d\ q_1}\right|_{q_1=q}$ **:** With action profile $\mathbf{q} = (q_1, q, \cdots, q)$, we have

$$S_j(\mathbf{q}) = q_1 \binom{n-1}{j-1} q^{j-1}(1-q)^{n-j} + (1-q_1)\binom{n-1}{j} q^j (1-q)^{n-1-j}. \tag{4.32}$$

Therefore, using Proposition 4.1, we have

$$\mathcal{I}(\theta \mid \mathbf{q}, \mu) = \sum_{j=1}^{n} \frac{j}{1+j} \left(q_1 \binom{n-1}{j-1} q^{j-1}(1-q)^{n-j} + (1-q_1)\binom{n-1}{j} q^j (1-q)^{n-1-j} \right). \tag{4.33}$$

Hence, we have

$$
\begin{aligned}
\frac{d\mathcal{I}(\theta \mid \mathbf{q}, \mu)}{d\ q_1} &= \sum_{j=1}^{n} \frac{j}{1+j}\binom{n-1}{j-1} q^{j-1}(1-q)^{n-j} - \sum_{j=1}^{n-1} \frac{j}{1+j}\binom{n-1}{j} q^{j}(1-q)^{n-1-j} \\
&= \sum_{j=0}^{n-1} \frac{j+1}{j+2}\binom{n-1}{j} q^{j}(1-q)^{n-1-j} - \sum_{j=0}^{n-1} \frac{j}{1+j}\binom{n-1}{j} q^{j}(1-q)^{n-1-j} \\
&= 1 - \mathbb{E}_{\ j\sim\mathrm{Bin}(n-1,q)}\left[\frac{1}{j+2}\right] - 1 + \mathbb{E}_{\ j\sim\mathrm{Bin}(n-1,q)}\left[\frac{1}{1+j}\right] \\
&= \mathbb{E}_{\ j\sim\mathrm{Bin}(n-1,q)}\left[\frac{1}{1+j} - \frac{1}{j+2}\right] \\
&= \mathbb{E}_{\ j\sim\mathrm{Bin}(n-1,q)}\left[\frac{1}{(j+2)(j+1)}\right]. \qquad (4.34)
\end{aligned}
$$

The above expression becomes (Chao and Strawderman [1972])

$$
\frac{d\mathcal{I}(\theta \mid \mathbf{q}, \mu)}{d\ q_1} = \frac{1-(1+nq)(1-q)^n}{n(n+1)q^2}. \qquad (4.35)
$$

Characterizing $\left.\frac{d\mathcal{I}(Z_1 \mid \mathbf{q},\mu)}{d\ q_1}\right|_{q_1=q}$ **:** Next, we consider the revealed information of Z_1 given this action profile. By Proposition 4.1, we have

$$
\mathcal{I}(Z_1 \mid \mathbf{q}, \mu) = A_1 + A_2, \qquad (4.36)
$$

where

$$
\begin{aligned}
A_1 =&(1-\mu)q_1 \sum_{k=1}^{n} S_{k-1}(\mathbf{q}_{-i})\left(1 - \frac{1}{1+k}\right), \\
A_2 =&\sum_{k=1}^{n} \sum_{\substack{B\subseteq\mathcal{N} \\ 1\in B, |B|=k}} \sum_{j=1}^{k} \sum_{r=0}^{n-k} \mu^k (1-\mu)^{n-k} S_r(\mathbf{q}_{\mathcal{N}\setminus B}) S_j(\mathbf{q}_B) \frac{q_1^2 S_{j-1}^2(\mathbf{q}_{B\setminus 1})}{S_j^2(\mathbf{q}_B)} \frac{1+r}{j(1+(j+r))}.
\end{aligned}
$$

We next evaluate A_1 and A_2. Note that $S_{k-1}(\mathbf{q}_{-1})$ is given by

$$
S_{k-1}(\mathbf{q}_{-1}) = \binom{n-1}{k-1} q^{k-1}(1-q)^{n-k}.
$$

Thus, by using (4.32), (4.38), and $\sum_{k=1}^{n} S_{k-1}(\mathbf{q}_{-1}) = 1$ we can write

$$A_1 = (1-\mu)q_1 \left(1 - \sum_{k=1}^{n} \binom{n-1}{k-1} q^{k-1}(1-q)^{n-k} \frac{1}{1+k}\right). \tag{4.37}$$

The above expression becomes

$$(1-\mu)q_1 \left(1 - \frac{1}{nq}\left(1 - \frac{1-(1-q)^{n+1}}{(n+1)q}\right)\right)$$

whose derivative is

$$\frac{d}{dq_1}A_1 = (1-\mu)\left(1 - \frac{1}{nq}\left(1 - \frac{1-(1-q)^{n+1}}{(n+1)q}\right)\right).$$

We next evaluate derivative of A_2 with respect to q_1 at q. We first upper bound it in general, and then derive its exact form for the special case $\mu = 1$.

Note that $S_{j-1}(\mathbf{q}_{B\setminus 1})$ is given by

$$S_{j-1}(\mathbf{q}_{B\setminus 1}) = \binom{k-1}{j-1} q^{j-1}(1-q)^{k-j}. \tag{4.38}$$

Thus, by using (4.32), (4.38), $\frac{1+r}{1+(j+r)} \le 1$, and $\sum_{r=0}^{n-k} S_r(q_{\mathcal{N}\setminus B}) = 1$ we can write

$$\begin{aligned}
\frac{d}{dq_1}A_2 \le{}& \frac{d}{dq_1}\sum_{k=1}^{n} \sum_{\substack{B\subseteq\mathcal{N} \\ 1\in B, |B|=k}} \mu^k(1-\mu)^{n-k} \sum_{j=1}^{k} \frac{q_1^2 S_{j-1}(\mathbf{q}_{B\setminus 1})^2}{jS_j(\mathbf{q}_B)} \\
={}& \frac{d}{dq_1}\sum_{k=1}^{n} \sum_{\substack{B\subseteq\mathcal{N} \\ 1\in B, |B|=k}} \mu^k(1-\mu)^{n-k} \sum_{j=1}^{k} \frac{q_1^2\left(\binom{k-1}{j-1}q^{j-1}(1-q)^{k-j}\right)^2}{j\left(q_1\binom{k-1}{j-1}q^{j-1}(1-q)^{k-j} + (1-q_1)\binom{k-1}{j}q^{j}(1-q)^{k-1-j}\right)} \\
={}& \frac{d}{dq_1}\sum_{k=1}^{n} \sum_{\substack{B\subseteq\mathcal{N} \\ 1\in B, |B|=k}} \mu^k(1-\mu)^{n-k} \sum_{j=1}^{k} \frac{q_1^2\binom{k-1}{j-1}q^{j-1}(1-q)^{k-j+1}}{(jq_1(1-q) + (k-j)(1-q_1)q)}.
\end{aligned}$$

Therefore, to compute $\frac{d}{dq_1}A_2$ at $q_1 = q$, we need to characterize

$$\frac{d}{dq_1}\ \frac{q_1^2}{jq_1(1-q) + (k-j)(1-q_1)q} \tag{4.39}$$

at q_1 which is given by

$$\begin{aligned}
&\frac{d}{dq_1}\left.\frac{q_1^2}{jq_1(1-q)+(k-j)(1-q_1)q}\right|_{q_1=q}\\
&\quad=\left.\left(\frac{2q_1}{jq_1(1-q)+(k-j)(1-q_1)q}-\frac{(j-kq)q_1^2}{(jq_1(1-q)+(k-j)(1-q_1)q)^2}\right)\right|_{q_1=q}\\
&\quad=\frac{2}{k(1-q)}-\frac{j-kq}{k^2(1-q)^2}=\frac{2k-kq-j}{k^2(1-q)^2}. \qquad (4.40)
\end{aligned}$$

As a consequence, we have

$$\begin{aligned}
\frac{d}{d\ q_1}\left.A_2\right|_{q_1=q} \leq&\sum_{k=1}^{n}\sum_{\substack{B\subseteq\mathcal{N}\\ 1\in B,|B|=k}}\mu^k(1-\mu)^{n-k}\sum_{j=1}^{k}\binom{k-1}{j-1}q^{j-1}(1-q)^{k-j}\frac{2k-kq-j}{k^2(1-q)}\\
=&\sum_{k=1}^{n}\sum_{\substack{B\subseteq\mathcal{N}\\ 1\in B,|B|=k}}\mu^k(1-\mu)^{n-k}\frac{2k-1}{k^2}\\
=&\sum_{k=1}^{n}\binom{n-1}{k-1}\mu^k(1-\mu)^{n-k}\frac{2k-1}{k^2}\\
=&\frac{1}{n}\sum_{k=1}^{n}\binom{n}{k}\mu^k(1-\mu)^{n-k}\frac{2k-1}{k}=\mathcal{O}(\frac{1}{n}).
\end{aligned}$$

For the special case $\mu=1$, we have

$$\begin{aligned}
\frac{d}{d\ q_1}\ A_2 &= \sum_{i=1}^{n}\binom{n-1}{j-1}q^{j-1}(1-q)^{n-j}\frac{2n-nq-j}{n^2(1+j)}\\
&=\frac{2n^2q+2nq-2n-1+(1-q)^n(2n+1-nq)}{n^3(n+1)q^2}. \qquad (4.41)
\end{aligned}$$

Having these characterizations, we next derive the user equilibrium.

Case $\mu<1$: In this case, we first argue qn is bounded. Let define $x:=qn$. Setting the derivative of $\mathcal{U}_1(q_1,\mathbf{q}_{-1})$ evaluated at $q_1=q$ equal to zero implies

$$\left|\alpha\left(1-(1+x)(1-\frac{x}{n})^n\right)-\beta(1-\mu)\left(x(x-1)+1-(1-\frac{x}{n})^n\right)\right|\leq\frac{\kappa x^2}{n}, \qquad (4.42)$$

for some constant κ. Note that the left-hand side grows as a quadratic function with a leading coefficient $\beta(1-\mu)$ while the right-hand is a quadratic with a leading coefficient κ/n. Hence, and since $\mu < 1$, for sufficiently large n, x is bounded. Therefore, we can cast q as c/n. In this case, (4.35) is equal to

$$\frac{1-(c+1)e^{-c}}{c^2} + \mathcal{O}(\frac{1}{n}).$$

Also, derivative of A_1 is equal to

$$\frac{d}{dq_1}A_1 = (1-\mu)\left(1-\frac{1}{c}\left(1-\frac{1-e^{-c}}{c}\right)\right) + \mathcal{O}(\frac{1}{n}).$$

Therefore, the derivative of $\mathcal{U}_1(q_1, \mathbf{q}_{-1})$ evaluated at $q_1 = q = \frac{c}{n}$ becomes

$$\alpha\frac{1-(c+1)e^{-c}}{c^2} = \beta(1-\mu)\left(1-\frac{1}{c}\left(1-\frac{1-e^{-c}}{c}\right)\right) + \mathcal{O}(\frac{1}{n}).$$

We next show that without the $\mathcal{O}(\frac{1}{n})$ term there exists a unique c^* that satisfies the above equation and that the derivative of the difference between the left-hand side and the right-hand side at c^* is away from zero, proving that the fixed point of the above equation is $c^* + \mathcal{O}(\frac{1}{n})$, proving that the symmetric equilibrium is given by $q = \frac{c^*}{n} + \mathcal{O}(\frac{1}{n^2})$.

Case $\mu < 1$; Proof of uniqueness of c: notice that the function

$$\alpha\frac{1-(c+1)e^{-c}}{c^2} - \beta(1-\mu)\left(1-\frac{1}{c}\left(1-\frac{1-e^{-c}}{c}\right)\right)$$

is decreasing in c. Moreover, for $c = 0$, it becomes

$$\alpha\frac{1}{2} - \beta(1-\mu)\frac{1}{2} > 0,$$

where the inequality follows from Assumption 4.1, implying that $\alpha \geq \beta$. For $c \to \infty$, it becomes

$$-\beta(1-\mu) < 0,$$

and thus, for $\mu < 1$, for sufficiently large n, this equation has a unique solution c^*.

Proof of boundedness of the derivative: the derivative of

$$\alpha \frac{1-(c+1)e^{-c}}{c^2} - \beta(1-\mu)\left(1 - \frac{1}{c}\left(1 - \frac{1-e^{-c}}{c}\right)\right)$$

is

$$\alpha \frac{e^{-c}(2+c(c+2)-2e^c)}{c^3} - \beta(1-\mu)\frac{e^{-c}(2+c+(c-2)e^c)}{c^3}.$$

Evaluating the above expression at $c = c^*$ results in

$$\frac{\alpha e^{-c}}{c^3}\left((2+c(c+2)-2e^c) - (2+c+(c-2)e^c)\frac{1-(c+1)e^{-c}}{c^2-c+1-e^{-c}}\right)$$

which is strictly positive for any $c > 0$. Finally, notice that c^* is strictly positive because $\alpha\frac{1}{2} - \beta(1-\mu)\frac{1}{2} > 0$ and therefore $c = 0$ cannot be a solution.

Case $\mu = 1$: In this case, using (4.41), we can write the first order condition as

$$\alpha\left(1-(1+nq)(1-q)^n\right) = \beta\left(2q + \frac{2q-2}{n} - \frac{1}{n^2} + (1-q)^n\frac{2n+1-nq}{n^2}\right). \tag{4.43}$$

If $\alpha > 2\beta$, then one can verify that, for sufficiently large n, this equation has no solution. On the other hand, for $\alpha \le 2\beta$, its solution is in the form of $\frac{\alpha}{2\beta} + \mathcal{O}(\log(n)/n)$.

Boundary cases $q = 0$ and $q = 1$: Finally, we investigate when $(0, 0, \cdots, 0)$ and $(1, 1, \cdots, 1)$ are equilibria.

- First, suppose $q = 0$, and the question is when $q_1 = 0$ is the best response of user one. In this case, we have

  $$\mathcal{I}(\theta \mid \mathbf{q}, \mu) = \frac{q_1}{2}, \quad \mathcal{I}(Z_1 \mid \mathbf{q}, \mu) = \frac{q_1}{2}. \tag{4.44}$$

  Hence, $(0, 0, \cdots, 0)$ is an equilibrium if and only if $\alpha \le \beta$ which is ruled out by Assumption 4.1.

- Now, suppose suppose $q = 1$, and the question is when $q_1 = 1$ is the best response of user one. In this case, we have

$$\mathcal{I}(\theta \mid \mathbf{q}, \mu) = \frac{n-1}{n} + \frac{q_1}{n(n+1)},$$
$$\mathcal{I}(Z_1 \mid \mathbf{q}, \mu) = q_1 \left(\frac{n}{n+1} + \frac{1-(1-\mu)^n}{n} - \mu \right).$$

Thus, one could verify that $(1, 1, \cdots, 1)$ is an equilibrium if and only if

$$\frac{\alpha}{\beta} \geq 1 + (1-\mu)(n^2 - 1), \tag{4.45}$$

and so, for $\mu < 1$ we can choose $N(\mu)$ such that this equilibrium is ruled out. For $\mu = 1$, however, $(1, 1, \cdots, 1)$ is an equilibrium. ■

Proof of Theorem 4.2

The proof simply follows from the fact that the platform's utility is a continuous function and that the set of platform's actions is the interval $[0, 1]$. ■

Proof of Theorem 4.3

We make use of the following two lemmas.

Lemma 4.6. *Suppose Assumption 4.1 holds. Then, for any n and any $\mu < 1$, any intermediary symmetric user equilibrium $\mathbf{q} = (q, \ldots, q)$ satisfies*

$$q \leq \frac{1}{n} \left(\sqrt{\frac{\alpha}{\beta(1-\mu)}} + 1 \right).$$

Proof of Lemma 4.6: Recall from the proof of Proposition 4.2 that any intermediary equilibrium $\mathbf{q} = (q, \cdots, q)$ satisfies

$$\alpha \frac{1-(1+nq)(1-q)^n}{n(n+1)q^2} = \beta \left((1-\mu) \left(1 - \frac{1}{nq} \left(1 - \frac{1-(1-q)^{n+1}}{(n+1)q} \right) \right) + \frac{d}{dq_1} A_2 \right), \tag{4.46}$$

where A_2 is given in the proof of Proposition 4.2. It is straightforward to verify $\frac{d}{dq_1}A_2 \geq 0$, and hence, we have

$$\alpha\frac{1-(1+nq)(1-q)^n}{n(n+1)q^2} \geq \beta(1-\mu)\left(1-\frac{1}{nq}\left(1-\frac{1-(1-q)^{n+1}}{(n+1)q}\right)\right).$$

Simplifying both sides and using the bound $1 \geq 1-(1+nq)(1-q)^n$ yields

$$\frac{\alpha}{\beta(1-\mu)} \geq n(n+1)q^2 - nq + 1 - q - (1-q)^{n+1} \geq (nq)^2 - nq.$$

If $nq \leq 1$, Lemma 4.6 trivially holds. Otherwise, we can lower bound the right-hand side by $(nq)^2 - 2(nq) + 1$ to obtain the desired bound. ■

We next provide an explicit expression for revealed information about the underlying common state θ and private types Z_i's under a symmetric action profile by users.

Lemma 4.7. *For any symmetric action profile* $\mathbf{q} = (q, \cdots, q)$*, we have*

$$\mathcal{I}(\theta \mid \mathbf{q}, \mu) \leq 1 - \frac{1}{n+1}, \tag{4.47}$$

$$\mathcal{I}(Z_i \mid \mathbf{q}, \mu) \leq (1-\mu)q + \frac{1}{n(n+1)} + \frac{1-\mu}{n}. \tag{4.48}$$

Furthermore, by setting $q = c/n$*, we have*

$$\mathcal{I}(\theta \mid \mathbf{q}, \mu) = 1 - \frac{1-e^{-c}}{c} + \mathcal{O}\left(\frac{1}{n}\right), \tag{4.49}$$

$$\mathcal{I}(Z_i \mid \mathbf{q}, \mu) = \frac{(1-\mu)c}{n}\left(1 - \frac{e^{-c}+c-1}{c^2}\right) + \mathcal{O}(\frac{1}{n^2}). \tag{4.50}$$

Proof of Lemma 4.7: To show (4.47) and (4.49), note that, for action profile $\mathbf{q} = (q, \cdots, q)$, we have

$$S_j(\mathbf{q}) = \binom{n}{j} q^j (1-q)^{n-j}. \tag{4.51}$$

Thus, using Proposition 4.1, we have

$$\begin{aligned}\mathcal{I}(\theta \mid \mathbf{q}, \mu) &= \sum_{j=1}^{n} \frac{j}{1+j} \binom{n}{j} q^j (1-q)^{n-j} \\ &= 1 - \sum_{j=1}^{n} \frac{1}{1+j} \binom{n}{j} q^j (1-q)^{n-j} \\ &= 1 - \mathbb{E}_{\ j \sim \mathrm{Bin}(n,q)} \left[\frac{1}{1+j} \right] = 1 - \frac{1-(1-q)^{n+1}}{(n+1)q}\end{aligned}$$

where the last equation follows from derivation of negative moments of binomial distribution (see Chao and Strawderman [1972] for the proof). Also, note that $\frac{1}{1+j}$ is decreasing in j, and hence, $\mathbb{E}_{\ j \sim \mathrm{Bin}(n,q)} \left[\frac{1}{1+j} \right]$ is decreasing in q. Thus, $\mathcal{I}(\theta \mid \mathbf{q}, \mu)$ is increasing in q. Hence, setting $q = 1$ gives us (4.47). Also, setting $q = c/n$ and using the fact that $(1-c/n)^n = e^c + \mathcal{O}(1/n)$ gives us (4.49).

To establish (4.48) and (4.50), it suffices to put $q_1 = q$ in (4.36). More precisely, we have

$$\mathcal{I}(Z_1 \mid \mathbf{q}, \mu) = A_1 + A_2, \tag{4.52}$$

where

$$\begin{aligned}A_1 =& (1-\mu) q \sum_{k=1}^{n} S_{k-1}(\mathbf{q}_{-i}) \left(1 - \frac{1}{1+k} \right), \\ A_2 =& \sum_{k=1}^{n} \sum_{\substack{B \subseteq \mathcal{N} \\ 1 \in B, |B|=k}} \sum_{j=1}^{k} \sum_{r=0}^{n-k} \mu^k (1-\mu)^{n-k} S_r(\mathbf{q}_{\mathcal{N} \setminus B}) S_j(\mathbf{q}_B) \frac{q^2 S_{j-1}^2(\mathbf{q}_{B \setminus 1})}{S_j^2(\mathbf{q}_B)} \frac{1+r}{j(1+(j+r))}.\end{aligned}$$

Using (4.37), with $q_1 = q$, we can characterize A_1 as

$$(1-\mu) q \left(1 - \mathbb{E}_{\ k \sim \mathrm{Bin}(n-1,q)} \left[\frac{1}{k+2} \right] \right) \tag{4.53}$$

$$= (1-\mu) q \left(1 - \frac{1}{nq} \left(1 - \frac{1-(1-q)^{n+1}}{(n+1)q} \right) \right), \tag{4.54}$$

which is bounded by $(1-\mu)q$. Also, plugging $q = c/n$, we obtain

$$A_1 = \frac{(1-\mu)c}{n}\left(1 - \frac{e^{-c}+c-1}{c^2}\right) + \mathcal{O}(\frac{1}{n^2}). \tag{4.55}$$

Therefore, it remains to bound A_2:

Deriving (4.48)**:** Note that A_2/μ is the revealed information regarding Z_1, condition that data of user one has been shuffled. From the definition of revealed information, it is immediate that this term is increasing in q. Hence, we derive an upper bound for A_2 by setting $q = 1$. To do so, note that by simplifying A_2 we have:

$$A_2 = \sum_{k=1}^{n}\sum_{j=1}^{k}\sum_{r=0}^{n-k}\binom{n-1}{k-1}\binom{k-1}{j-1}\binom{n-k}{r}\mu^k(1-\mu)^{n-k}q^{j+r}(1-q)^{n-j-r}\frac{1+r}{k(1+j+r)}. \tag{4.56}$$

Setting $q = 1$, only the terms with $j + r = n$ will be nonzero. This corresponds to $r = n-k$ and $j = k$. Hence, we have

$$\begin{aligned} A_2 &\leq \sum_{k=1}^{n}\binom{n-1}{k-1}\mu^k(1-\mu)^{n-k}\frac{n-k+1}{(n+1)k} \\ &= \mu\left(\mathbb{E}_{k\sim\text{Bin}(n-1,\mu)}\left[\frac{1}{k+1}\right] - \frac{1}{n+1}\right) \\ &= \frac{1-(1-\mu)^n}{n} - \frac{\mu}{n+1} \\ &\leq \frac{1}{n} - \frac{1}{n+1} + \frac{1-\mu}{n+1} \\ &\leq \frac{1}{n(n+1)} + \frac{1-\mu}{n}, \end{aligned} \tag{4.57}$$

which completes the proof of (4.48). It is worth noting that (4.57) follows from the fact that (see Chao and Strawderman [1972])

$$\mathbb{E}_{k\sim\text{Bin}(n-1,\mu)}\left[\frac{1}{k+1}\right] = \frac{1-(1-\mu)^n}{n\mu}.$$

Deriving (4.50)**:** To do so, we bound the term $\frac{1+r}{1+(j+r)} \leq 1$ in A_2 and using $\sum_{r=0}^{n-k} S_r(q_{\mathcal{N}\setminus B}) = 1$ to write

$$A_2 \leq \sum_{k=1}^{n} \sum_{\substack{B\subseteq\mathcal{N} \\ 1\in B, |B|=k}} \sum_{j=1}^{k} \mu^k (1-\mu)^{n-k} S_j(\mathbf{q}_B) \frac{q^2 S_{j-1}^2(\mathbf{q}_{B\setminus 1})}{j\ S_j^2(\mathbf{q}_B)}. \tag{4.58}$$

Next, using (4.51), we simplify the second term on the right hand side:

$$\begin{aligned} A_2 &\leq \sum_{k=1}^{n} \sum_{j=1}^{k} \binom{n-1}{k-1} \binom{k-1}{j-1} \mu^k (1-\mu)^{n-k} q^j (1-q)^{k-j} \frac{1}{k} \\ &= \sum_{k=1}^{n} \frac{1}{k} \binom{n-1}{k-1} \mu^k (1-\mu)^{n-k} \sum_{j=1}^{k} \binom{k-1}{j-1} q^j (1-q)^{k-j}. \end{aligned} \tag{4.59}$$

Note that, we can write the inner sum as

$$\sum_{j=1}^{k} \binom{k-1}{j-1} q^j (1-q)^{k-j} = q. \tag{4.60}$$

Plugging this into (4.59), we obtain

$$\begin{aligned} A_2 &\leq q \sum_{k=1}^{n} \frac{1}{k} \binom{n-1}{k-1} \mu^k (1-\mu)^{n-k} \\ &= q \sum_{k=0}^{n-1} \frac{1}{k+1} \binom{n-1}{k} \mu^{k+1} (1-\mu)^{n-1-k} \\ &= q\mu\ \mathbb{E}_{k\sim \mathrm{Bin}(n-1,\mu)} \left[\frac{1}{k+1} \right] \\ &= \frac{q\ (1-(1-\mu)^n)}{n}, \end{aligned} \tag{4.61}$$

Plugging (4.61) into (4.52) with $q = c/n$ completes the proof of (4.50). ■

We now proceed with the proof of the theorem. We choose $N^e(\epsilon) > N(\epsilon/2)$, with $N(.)$ defined in Proposition 4.2. Note that, by Proposition 4.2, for any $n \geq N(\epsilon/2)$, and for any

$\mu \leq 1 - \epsilon/2$, user equilibrium is in the form of $(c + \mathcal{O}(1/n))/n$ where c satisfies

$$\alpha \frac{1-(c+1)e^{-c}}{c^2} = \beta(1-\mu)\left(1 - \frac{1}{c}\left(1 - \frac{1-e^{-c}}{c}\right)\right). \tag{4.62}$$

We can rewrite this equation as

$$1 - \mu = \frac{\alpha}{\beta} \frac{1-(c+1)e^{-c}}{c^2 - c + 1 - e^{-c}}. \tag{4.63}$$

Using Lemma 4.7 along with the fact that

$$\frac{1-e^{-c}}{c}$$

is Lipschitz continuous as a function of c, platform's problem can be cast as

$$\max_{\mu} 1 - \frac{1-e^{-c}}{c} + \delta(1-\mu)c\left(1 - \frac{e^{-c}+c-1}{c^2}\right) + \mathcal{O}\left(\frac{1}{n}\right) \tag{4.64a}$$

$$\text{s.t.} \quad 1 - \mu = \frac{\alpha}{\beta}\frac{1-(c+1)e^{-c}}{c^2-c+1-e^{-c}} \tag{4.64b}$$

$$\mu \leq 1 - \epsilon/2. \tag{4.64c}$$

The second constraint (4.64c) follows from the fact that this approximation is valid for $\mu \leq 1-\epsilon/2$. We also bound the platform's utility for the case $\mu \in [1-\epsilon/2, 1]$. Using Lemma 4.7, we can write

$$\begin{aligned}\sup_{\mu\in[1-\epsilon/2,1]} \mathcal{U}_{\text{platform}}(\mathbf{q}^e(\mu), \mu) &\leq \sup_{\mu\in[1-\epsilon/2,1]} \left(1 - \frac{1-\delta}{n+1} + (1-\mu)nq + 1 - \mu\right) \\ &\leq \sup_{\mu\in[1-\epsilon/2,1]} \left(1 + (1-\mu)(\sqrt{\frac{\alpha}{\beta(1-\mu)}} + 2)\right) \qquad (4.65) \\ &\leq 1 + \epsilon + \sqrt{\frac{\alpha\epsilon}{2\beta}}, \qquad (4.66)\end{aligned}$$

where (4.65) follows from Lemma 4.6. Next, note that

$$\frac{1-(c+1)e^{-c}}{c^2-c+1-e^{-c}} \tag{4.67}$$

is a deceasing function of c which varies from 1 to 0 as c goes from 0 to ∞. Hence, we could replace $1-\mu$ in (4.64a) using (4.64b) and replace (4.64b) by the following constraint:

$$\underline{c} \leq c \leq \bar{c}, \tag{4.68}$$

where $\underline{c}$ and $\bar{c}$ are such that

$$\left.\frac{1-(c+1)e^{-c}}{c^2-c+1-e^{-c}}\right|_{c=\underline{c}} = \frac{\beta}{\alpha}, \quad \left.\frac{1-(c+1)e^{-c}}{c^2-c+1-e^{-c}}\right|_{c=\bar{c}} = \frac{\epsilon}{2}\frac{\beta}{\alpha}. \tag{4.69}$$

Using these quantities, we obtain

$$\max_c \; 1 - \frac{1-e^{-c}}{c} + \frac{\alpha\delta}{\beta}.\frac{1-(c+1)e^{-c}}{c} + \mathcal{O}\left(\frac{1}{n}\right) \tag{4.70a}$$

$$\text{s.t.} \quad \underline{c} \leq c \leq \bar{c}, \tag{4.70b}$$

where $\underline{c}$ and $\bar{c}$ correspond to $\mu = 0$ and $\mu = 1-\epsilon/2$, respectively. Next, note that we could choose n large enough such that the solution of (4.70) and the following optimization problem in which we have removed the $\mathcal{O}\left(\frac{1}{n}\right)$ term from the objective function differ at most by $\epsilon/2$.

$$\max_c \; 1 - \frac{1-e^{-c}}{c} + \frac{\alpha\delta}{\beta}.\frac{1-(c+1)e^{-c}}{c} \tag{4.71a}$$

$$\text{s.t.} \quad \underline{c} \leq c \leq \bar{c} \tag{4.71b}$$

Hence, it suffices to show there exists $\underline{\alpha}$ and $\bar{\alpha}$ in $[\beta, \infty)$ such that:

1. If $\alpha \leq \underline{\alpha}$, then $c = \bar{c}$ which corresponds to $\mu = 1-\epsilon/2$ being the solution of (4.71). In this case, we will have $\mu^e \geq 1-\epsilon$.

2. If $\alpha \geq \bar{\alpha}$ then $c = \underline{c}$ which corresponds to $\mu = 0$ being the solution of (4.71) and the platform's utility at $c = \underline{c}$ is greater than (4.66). In this case, we will have $\mu^e \leq \epsilon/2$.

To show (i), notice that the derivative of (4.71a) with respect to c is given by

$$\frac{e^{-c}}{c^2}\left((e^c - c - 1)(1 - \frac{\delta\alpha}{\beta}) + c^2\frac{\delta\alpha}{\beta}\right). \tag{4.72}$$

If $\beta \leq \alpha \leq \frac{\beta}{\delta}$, then this derivative is positive, meaning that $c = \bar{c}$ is the solution of (4.71). Thus, we choose $\underline{\alpha} = \frac{\beta}{\delta}$.

To show (ii), note that for α sufficiently large, we have

$$\underline{c} \geq 2, \quad \left.\frac{e^c - c - 1}{e^c - c^2 - c - 1}\right|_{c=2} \leq \frac{\delta\alpha}{\beta}.$$

In this case, it is easy to verify that the derivative is negative over $[\underline{c}, \infty)$, and hence, $c = \underline{c}$ is the optimal solution of (4.71). Furthermore, notice that the limit of (4.71) when c goes to infinity is one. Therefore, because the platform's utility is decreasing over $[\underline{c}, \infty)$, it must be larger than one at $c = \underline{c}$. Hence, using the bound (4.66), we can see that for sufficiently small ϵ and large n, platform's utility at $c = \underline{c}$ would be greater than the maximum of platform's utility for any $\mu \in [1 - \epsilon/2, 1]$. This completes the proof. ■

Proof of Theorem 4.4

Recall the platform's problem given in (4.70). For any $t \in [0, 1]$, we define $c(t)$ as the solution of

$$f_1(c) := \frac{1 - (c+1)e^{-c}}{c^2 - c + 1 - e^{-c}} = t. \tag{4.73}$$

Using this change of variable, we can cast the platform's problem as

$$\max_t \frac{c(t) + e^{-c(t)} - 1}{c(t)}\left(1 - \frac{\delta\alpha}{\beta}t\right) + \frac{\delta\alpha}{\beta}tc(t) + \mathcal{O}\left(\frac{1}{n}\right) \tag{4.74a}$$

$$\text{s.t.} \quad \frac{\epsilon}{2}.\frac{\beta}{\alpha} \leq t \leq \frac{\beta}{\alpha}, \tag{4.74b}$$

Note that (4.73) is equivalent to

$$e^{c(t)} = \frac{c(t) + 1 - t}{1 + tc(t) - t - tc(t)^2}. \tag{4.75}$$

Using this, we can rewrite the platform's problem (4.74) as

$$\max_t \; g(t, \frac{\delta\alpha}{\beta}) + \mathcal{O}\left(\frac{1}{n}\right) \tag{4.76a}$$

$$\text{s.t.} \quad \frac{\epsilon}{2}\frac{\beta}{\alpha} \leq t \leq \frac{\beta}{\alpha}, \tag{4.76b}$$

where

$$g(t,r) := \frac{c(t)(1-t)}{1+c(t)-t}\,(1-rt) + rtc(t) = c(t)\frac{1-t+rtc(t)}{1-t+c(t)} \tag{4.77}$$

Also, note that, by using Lemma 4.7, the user's utility is given by

$$\alpha\left(1 - \frac{1-e^{-c(t)}}{c(t)}\right) + \mathcal{O}\left(\frac{1}{n}\right) = \frac{\beta}{\delta}\; h(t, \frac{\delta\alpha}{\beta}) + \mathcal{O}\left(\frac{1}{n}\right), \tag{4.78}$$

where

$$h(t,r) := r\frac{c(t)(1-t)}{1-t+c(t)}. \tag{4.79}$$

Claim 1. *For any $r > 1$, the function $g(.,r) : [0,1] \to \mathbb{R}$, defined in (4.77), achieves its maximum over $[0,1]$ at the unique $t^*(r)$ that satisfies*

$$\left.\frac{\partial}{\partial t} g(t,r)\right|_{t=t^*(r)} = 0. \tag{4.80}$$

In addition, $t^*(r)$ is an increasing function of r that satisfies $\lim_{r\to 1^+} t^*(r) = 0$. Moreover, there exists $\bar{r} > 1$ such that $h(t^*(r), r)$ is decreasing in r over $(1, \bar{r})$.

First, let us show how this claim gives us the result. Note that for any $\alpha > \beta/\delta$, $g(t, \frac{\delta\alpha}{\beta})$ achieves its maximum at $t^*(\frac{\delta\alpha}{\beta})$. Also, by taking $\alpha \to \beta/\delta$ from right, $t^*(\frac{\delta\alpha}{\beta}) \to 0$. Hence, we can choose $\alpha_L < \alpha_H$ and ϵ small enough such that:

1. $\frac{\beta}{\delta} < \alpha_L < \alpha_H < \bar{r}.\frac{\beta}{\delta}$ and

2. $t^*(\frac{\delta\alpha}{\beta}) \in [\frac{\epsilon}{2}.\frac{\beta}{\alpha}, \frac{\beta}{\alpha}]$ for any $\alpha \in (\alpha_L, \alpha_H)$.

Also, similar to the argument we provided in the proof of Theorem 4.3, we can choose ϵ small enough such that, for $\alpha \in (\alpha_L, \alpha_H)$, the platform's utility at $t = t^*(\frac{\delta\alpha}{\beta})$ be larger than the bound given by (4.66) in the proof of Theorem 4.3. This ensures that μ^e belongs to the interval $[0, 1-\epsilon/2]$ for $\alpha \in (\alpha_L, \alpha_H)$.

Now suppose $\alpha_1 < \alpha_2 \in (\alpha_L, \alpha_H)$. Note that, since $g(., \frac{\delta\alpha_i}{\beta})$ is increasing before its peak and decreasing after that, we have that for any small enough η, there exists $M(\eta)$, such that for any for $n > M(\eta)$ the solution of (4.76) for $\alpha = \alpha_1$ and $\alpha = \alpha_2$ would be in at most η distance of $t^*(\frac{\delta\alpha_1}{\beta})$ and $t^*(\frac{\delta\alpha_2}{\beta})$, respectively.

Next, note that by the above claim, we have

$$h\left(t^*(\frac{\delta\alpha_1}{\beta}), \frac{\delta\alpha_1}{\beta}\right) > h\left(t^*(\frac{\delta\alpha_2}{\beta}), \frac{\delta\alpha_2}{\beta}\right). \tag{4.81}$$

Notice that the user's utility (4.78) at equilibrium for $\alpha = \alpha_i$ with $i \in \{1, 2\}$, is evaluated at the solution of (4.76) which is η-close to $t^*(\frac{\delta\alpha_i}{\beta})$. Hence, by choosing η small enough and n large enough, and by using (4.81), we can establish that the user's utility at equilibrium is larger with $\alpha = \alpha_1$ compared to $\alpha = \alpha_2$. This gives us the desired result. Therefore, it remains to prove the claim.

Maximum of $g(t, r)$ for $r > 1$: Note that $g(t, r)$ can be rewritten as

$$g(t, r) = g_1(c(t), r) \text{ where } g_1(c, r) = 1 - \frac{1-e^{-c}}{c} + r\frac{1-(c+1)e^{-c}}{c}, \tag{4.82}$$

and hence, we have

$$\frac{\partial}{\partial t} g(t, r) = \frac{\partial}{\partial c} g_1(c, r) c'(t). \tag{4.83}$$

Also, note that, by inverse function theorem, $c'(t)$ is given by

$$c'(t) = \frac{1}{f_1'(c(t))}, \tag{4.84}$$

and therefore, $c'(t)$ is negative over $(0, 1)$. Moreover, $\frac{\partial}{\partial c} g_1(c, r)$ is given by

$$\frac{\partial}{\partial c} g_1(c, r) = \frac{e^{-c}}{c^2}\left((e^c - c - 1)(1-r) + c^2 r\right). \tag{4.85}$$

Setting the derivative of $g_1(c, r)$ with respect to c equal to zero for $r > 1$ gives

$$\frac{e^c - c - 1}{c^2} = \frac{r}{r-1}. \tag{4.86}$$

Notice that the left-hand side is an increasing function that goes from 1/2 to infinity as c goes from zero to infinity. Hence, (4.86) has a solution for any $r > 1$ which we denote it by $c^*(r)$. Note that $f_1(c^*(r)) = t^*(r)$.

The derivative $\frac{\partial}{\partial c} g_1(c, r)$ is positive for $c < c^*(r)$ which means $\frac{\partial}{\partial t} g(t, r)$ is negative for $t > t^*(r)$(because $c'(t)$ is negative). In addition, $\frac{\partial}{\partial c} g_1(c, r)$ is negative for $c > c^*(r)$ which means $\frac{\partial}{\partial t} g(t, r)$ is positive for $t \in (0, t^*(r))$. In other words, $g(t, r)$ is increasing over $(0, t^*(r))$ and decreasing over $(t^*(r), \infty)$, and thus, it achieves its maximum at $t^*(r)$.

Also, by increasing r, $r/(r-1)$ decreases which means $c^*(r)$ also decreases. But since f_1 is a decreasing function, $t^*(r)$ increases. Also, by taking $r \to 1^+$, $c^*(r)$ goes to infinity, which implies $t^*(r) \to 0^*$.

$h(t^*(r), r)$ **is decreasing in** r **over** $(1, \bar{r})$**:** Note that

$$\frac{d}{dr} h(t^*(r), r) = \frac{c(t^*(r))(1 - t^*(r))}{1 - t^*(r) + c(t^*(r))} + r \frac{d}{dr} t^*(r) \left(\frac{d}{dt} \frac{c(t)(1-t)}{1 - t + c(t)} \bigg|_{t = t^*(r)} \right). \tag{4.87}$$

Using the fact that

$$\frac{\partial}{\partial t} g(t, r) \bigg|_{t = t^*(r)} = 0,$$

we obtain

$$\frac{d}{dt} \frac{c(t)(1-t)}{1 - t + c(t)} \bigg|_{t = t^*(r)} = - \frac{r \left(c(t) + t c'^2 c'(t) \right)}{(1-t)\left(1 - t + rtc(t)\right)} . \frac{c(t)(1-t)}{1 - t + c(t)} \bigg|_{t = t^*(r)}. \tag{4.88}$$

Plugging this into (4.87) implies

$$\frac{d}{dr} h(t^*(r), r) = \frac{c(t^*(r))(1 - t^*(r))}{1 - t^*(r) + c(t^*(r))} \left(1 - \frac{r^2 \left(c(t) + t c'^2 c'(t) \right)}{(1-t)\left(1 - t + rtc(t)\right)} \bigg|_{t = t^*(r)} . \frac{d}{dr} t^*(r) \right). \tag{4.89}$$

We want to show this derivative is negative over the interval $(1, \bar{r})$. Note that for r close to one, $\frac{r^2}{1-t^*(r)}$ is close to one, and hence, if we show

$$\left.\frac{c(t) + tc'^2c'(t)}{1 - t + rtc(t)}\right|_{t=t^*(r)} \frac{d}{dr}t^*(r) \tag{4.90}$$

is very large when r is close to one, then we are done. We next show that this term goes to infinity as r goes to one.

Recall that $t^*(r) = f_1(c^*(r))$ and thus

$$\frac{d}{dr}t^*(r) = \frac{d}{dc}f_1(c^*(r))\frac{d}{dr}c^*(r). \tag{4.91}$$

Notice that (4.86) implies

$$r = \kappa(c) := \frac{e^c - c - 1}{e^c - c^2 - c - 1}. \tag{4.92}$$

Consequently, by inverse function theorem, we can rewrite (4.91) as

$$\frac{d}{dr}t^*(r) = \frac{f_1'(c^*(r))}{\kappa'^*(r))}. \tag{4.93}$$

Using this derivation along with the fact that $c'(f_1(c))) = 1/f_1'(c)$, $t = f_1(c)$, and $r = \kappa(c)$, we can rewrite (4.90) as a function of c:

$$\left.\frac{cf_1'(c) + f_1(c) - f_1(c)^2}{\kappa'(c)\left(1 - f_1(c) + c\kappa(c)f_1(c)\right)}\right|_{c=c^*(r)},$$

which is equal to

$$\left.\frac{\left(e^c - 1 - c - c^2\right)^3}{c\left(e^c(c-2) + c + 2\right)\left(e^c(c^2 - c + 1) - 1\right)}\right|_{c=c^*(r)}.$$

Recall that $c^*(r)$ goes to infinity as $r \to 0^+$. Thus, this term goes to infinity as r goes to one. This completes the proof of the claim and hence Theorem 4.4. ■

Proof of Theorem 4.5

We denote user's i data after adding noise by $\tilde{X}_i$, i.e., $\tilde{X}_i = X_i + W_i$, where $W_i \sim \mathcal{N}(0, \sigma^2(k))$ if k users share their data. Suppose user one shares her data with probability q_1 and users $2, \cdots, n$ share their data with probability q. Our goal is to show the optimal choice of q_1 for user one is 1. Note that, the utility of user 1 is given by

$$\begin{aligned}
&\alpha \Bigg(q_1 \sum_{k=0}^{n-1} \binom{n-1}{k} q^k (1-q)^{n-1-k} \, \mathcal{I}(\theta \mid k+1 \text{ given users share data}) \\
&\quad + (1-q_1) \sum_{k=0}^{n-1} \binom{n-1}{k} q^k (1-q)^{n-1-k} \, \mathcal{I}(\theta \mid k \text{ given users share data}) \Bigg) \\
&- \beta \, q_1 \sum_{k=0}^{n-1} \binom{n-1}{k} q^k (1-q)^{n-1-k} \, \mathcal{I}(\theta \mid \text{data of user 1 and } k \text{ other given users is shared}).
\end{aligned}$$

Therefore, to show this term is maximized at $q_1 = 1$, we need to show the following inequality holds for any $k \in \{0, \cdots, n-1\}$:

$$\begin{aligned}
\alpha \, \mathcal{I}(\theta \mid k+1 \text{ given users share data}) &\geq \alpha \, \mathcal{I}(\theta \mid k \text{ given users share data}) \\
&+ \beta \, \mathcal{I}(\theta \mid \text{data of user 1 and } k \text{ other given users is shared}).
\end{aligned}$$

To do so, without loss of generality, it suffices to show

$$\alpha \, \mathcal{I}(\theta \mid (\tilde{X}_i)_{i=1}^{k+1}) \geq \alpha \, \mathcal{I}(\theta \mid (\tilde{X}_i)_{i=1}^{k}) + \beta \, \mathcal{I}(\theta \mid (\tilde{X}_i)_{i=1}^{k+1}). \tag{4.94}$$

Note that θ and $(\tilde{X}_i)_{i=1}^{k}$ are jointly Gaussian, where the mean of their joint distribution is 0 and the covariance matrix of their joint distribution is given by

$$\begin{bmatrix}
1 & 1 & \cdots & 1 \\
1 & 2+\sigma^2(k) & \cdots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \cdots & 2+\sigma^2(k)
\end{bmatrix}. \tag{4.95}$$

Therefore, by using Sherman–Morrison formula, we establish that

$$\begin{aligned}\mathbb{E}[\theta \mid (\tilde{X}_i)_{i=1}^k] &= [1 \cdots 1] \begin{bmatrix} 2+\sigma^2(k) & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 2+\sigma^2(k) \end{bmatrix}^{-1} [X_1, \cdots, X_k]^\top \\ &= \frac{1}{k+1+\sigma^2(k)} \sum_{i=1}^k \tilde{X}_i. \end{aligned} \tag{4.96}$$

As a consequence, we have

$$\mathcal{I}(\theta \mid (\tilde{X}_i)_{i=1}^k) = \mathbb{E}\left[\mathbb{E}\left[\theta \mid (\tilde{X}_i)_{i=1}^k\right]^2\right] = \frac{k}{k+1+\sigma^2(k)}. \tag{4.97}$$

Next, notice that the joint distribution of $Z_1, \tilde{X}_1, \ldots, \tilde{X}_k$ is normal with covariance matrix

$$\begin{bmatrix} 1 & 1 & \cdots & 0 \\ 1 & 2+\sigma^2(k) & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 2+\sigma^2(k) \end{bmatrix}.$$

Similar to the previous calculations, we show

$$\mathbb{E}[Z_1 \mid (\tilde{X}_i)_{i=1}^k] = \frac{1}{1+\sigma^2(k)} \left(\frac{k+\sigma^2(k)}{k+1+\sigma^2(k)} \tilde{X}_1 - \frac{1}{k+1+\sigma^2(k)} \sum_{i=2}^k \tilde{X}_i \right). \tag{4.98}$$

Hence, we have

$$\mathcal{I}(Z_1 \mid (\tilde{X}_i)_{i=1}^k) = \mathbb{E}\left[\mathbb{E}\left[Z_1 \mid (\tilde{X}_i)_{i=1}^k\right]^2\right] = \frac{k+\sigma^2(k)}{(k+1+\sigma^2(k))(1+\sigma^2(k))}. \tag{4.99}$$

Plugging (4.97) and (4.99) into (4.94), we need to show

$$\alpha \frac{k+1}{k+2+\sigma^2(k+1)} - \beta \frac{k+1+\sigma^2(k+1)}{(k+2+\sigma^2(k+1))(1+\sigma^2(k+1))} \geq \alpha \frac{k}{k+1+\sigma^2(k)}. \tag{4.100}$$

Notice that we have

$$\frac{k+1}{k+2+\sigma^2(k+1)} \geq \frac{k+1+\sigma^2(k+1)}{(k+2+\sigma^2(k+1))(1+\sigma^2(k+1))}, \tag{4.101}$$

and thus, to show (4.100), it suffices to show

$$(\alpha-\beta)\frac{k+1}{k+2+\sigma^2(k+1)} \geq \alpha\frac{k}{k+1+\sigma^2(k)}. \tag{4.102}$$

We aim to show a slightly stronger inequality by replacing k on the numerator of the left-hand side by $k+1$. In this case, $k+1$ cancels out from both sides, and we need to show

$$k+1+\sigma^2(k) \geq \frac{\alpha}{\alpha-\beta}(k+2+\sigma^2(k+1)). \tag{4.103}$$

Note that, the condition on $\sigma(.)$ implies that $\sigma^2(k)$ by itself is weakly greater than the left-hand side, completing the proof. ■

Proof of Corollary 4.1

By using (4.97) and (4.99), if everyone shares their data and $\sigma(n)=0$, then

$$\mathcal{I}(\theta \mid \text{all sharing}) = \mathcal{I}(Z_i \mid \text{all sharing}) = \frac{n}{n+1}. \tag{4.104}$$

In this case, platform's utility is given by

$$(n\delta+1)\frac{n}{n+1}.$$

This is the utility corresponding to the case in the mask-shuffle mechanism that all users fully share and the platform offers no shuffling. This is the highest possible utility for the platform, but it never happens under the mask-shuffle mechanism since $(q,\mu)=((1,\cdots,1),0)$ is never an equilibrium under the mask-shuffle mechanism. ■

Proof of Proposition 4.3

The proof follows from Theorem 4.5 and (4.104). ■

Bibliography

J. D. Abernethy, R. Cummings, B. Kumar, S. Taggart, and J. Morgenstern. Learning auctions with robust incentive guarantees. In *NeurIPS*, pages 11587–11597, 2019.

D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar. Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics:Micro*, 2022.

D. Acemoglu, A. Fallah, A. Makhdoumi, A. Malekian, and A. E. Ozdaglar. How good are privacy guarantees? data sharing, privacy preservation, and platform behavior. *preprint*, 2023.

J. Acharya, Z. Sun, and H. Zhang. Differentially private assouad, fano, and le cam. In *Algorithmic Learning Theory*, pages 48–78. PMLR, 2021.

M. Akbarpour and S. Li. Credible auctions: A trilemma. *Econometrica*, 88(2):425–467, 2020.

M. Alaggan, S. Gambs, and A.-M. Kermarrec. Heterogeneous differential privacy. *arXiv preprint arXiv:1504.06998*, 2015.

J. Anunrojwong, K. Iyer, and V. Manshadi. Information design for congested social services: Optimal need-based persuasion. *Available at SSRN 3849746*, 2021.

Apple. Differential privacy overview - apple. https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf. Retrieved May 4, 2023.

I. Ashlagi, M. Braverman, Y. Kanoria, and P. Shi. Clearing matching markets efficiently: informative signals and match recommendations. *Management Science*, 66(5):2163–2193, 2020.

I. Ashlagi, F. Monachou, and A. Nikzad. Optimal dynamic allocation: Simplicity through information design. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 101–102, 2021.

S. Asoodeh, M. Aliakbarpour, and F. P. Calmon. Local differential privacy is equivalent to contraction of an f-divergence. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 545–550, 2021.

S. R. Balseiro, O. Besbes, and F. Castro. Mechanism design under approximate incentive compatibility. *Operations Research*, 2022.

R. F. Barber and J. C. Duchi. Privacy and statistical risk: Formalisms and minimax bounds. *arXiv preprint arXiv:1412.4451*, 2014.

D. Bergemann and A. Bonatti. Selling cookies. *American Economic Journal: Microeconomics*, 7(3):259–94, 2015.

D. Bergemann and A. Bonatti. Markets for information: An introduction. *Annual Review of Economics*, 11:85–107, 2019.

D. Bergemann, A. Bonatti, and T. Gan. The economics of social data. *arXiv preprint arXiv:2004.03107*, 2020.

D. P. Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48 (3):334–334, 1997.

O. Besbes and O. Mouchtaki. How big should your data really be? data-driven newsvendor and the transient of learning. *arXiv preprint arXiv:2107.02742*, 2021.

K. Bimpikis, D. Crapis, and A. Tahbaz-Salehi. Information sale and competition. *Management Science*, 65(6):2646–2664, 2019.

K. Bimpikis, I. Morgenstern, and D. Saban. Data tracking under competition. *Available at SSRN 3808228*, 2021.

A. Bittau, Ú. Erlingsson, P. Maniatis, I. Mironov, A. Raghunathan, D. Lie, M. Rudominer, U. Kode, J. Tinnes, and B. Seefeld. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th symposium on operating systems principles*, pages 441–459, 2017.

M. Bun and T. Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer, 2016.

Y. Cai, C. Daskalakis, and C. Papadimitriou. Optimum statistical estimation with strategic data sources. In *Conference on Learning Theory*, pages 280–296. PMLR, 2015.

O. Candogan and K. Drakopoulos. Optimal signaling of content accuracy: Engagement vs. misinformation. *Operations Research*, 68(2):497–515, 2020.

M.-T. Chao and W. Strawderman. Negative moments of positive random variables. *Journal of the American Statistical Association*, 67(338):429–431, 1972.

X. Chen, S. Miao, and Y. Wang. Differential privacy in personalized pricing with nonparametric demand models. *Available at SSRN 3919807*, 2021a.

X. Chen, D. Simchi-Levi, and Y. Wang. Privacy-preserving dynamic personalized pricing with demand learning. *Management Science*, 2021b.

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

Y. Chen, N. Immorlica, B. Lucier, V. Syrgkanis, and J. Ziani. Optimal data acquisition for statistical estimation. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 27–44, 2018.

A. Cheu. Differential privacy in the shuffle model: A survey of separations. *arXiv preprint arXiv:2107.11839*, 2021.

R. Cummings, K. Ligett, A. Roth, Z. S. Wu, and J. Ziani. Accuracy for sale: Aggregating data with a variance constraint. In *Proceedings of the 2015 conference on innovations in theoretical computer science*, pages 317–324, 2015.

R. Cummings, V. Feldman, A. McMillan, and K. Talwar. Mean estimation with user-level privacy under data heterogeneity. In *NeurIPS 2021 Workshop Privacy in Machine Learning*, 2021.

R. Cummings, H. Elzayn, V. Gkatzelis, E. Pountourakis, and J. Ziani. Optimal data acquisition with privacy-aware agents. *arXiv preprint arXiv:2209.06340*, 2022.

O. Dekel, F. Fischer, and A. D. Procaccia. Incentive compatible regression learning. *Journal of Computer and System Sciences*, 76(8):759–777, 2010.

B. Ding, J. Kulkarni, and S. Yekhanin. Collecting telemetry data privately. *arXiv preprint arXiv:1712.01524*, 2017.

J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438. IEEE, 2013.

R. Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer, 2006a.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006b.

C. Dwork, A. Roth, et al. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

Ú. Erlingsson, V. Pihur, and A. Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067, 2014.

I. P. Fainmesser, A. Galeotti, and R. Momot. Digital privacy. *Management Science*, 2022.

A. Fallah, A. Makhdoumi, A. Malekian, and A. Ozdaglar. Optimal and differentially private data acquisition: Central and local mechanisms. *arXiv preprint arXiv:2201.03968*, 2022a.

A. Fallah, A. Makhdoumi, A. Malekian, and A. E. Ozdaglar. Bridging central and local differential privacy in data acquisition mechanisms. *Available at SSRN 4311351*, 2022b.

D. J. Foster, Z. Li, T. Lykouris, K. Sridharan, and E. Tardos. Learning in games: Robustness of fast convergence. *Advances in Neural Information Processing Systems*, 29:4734–4742, 2016.

X. Fu, N. Chen, P. Gao, and Y. Li. Privacy-preserving personalized recommender systems. *Available at SSRN 4202576*, 2022.

A. Ghosh and A. Roth. Selling privacy at auction. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 199–208, 2011.

A. Ghosh, K. Ligett, A. Roth, and G. Schoenebeck. Buying private data without verification. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 931–948, 2014.

A. Goldfarb and C. Tucker. Online display advertising: Targeting and obtrusiveness. *Marketing Science*, 30(3):389–404, 2011.

Y. Gur, G. Macnamara, and D. Saban. On the disclosure of promotion value in platforms with learning sellers. *arXiv preprint arXiv:1911.09256*, 2019.

A. Y. Ha and S. Tong. Contracting and information sharing under supply chain competition. *Management science*, 54(4):701–715, 2008.

Y. Han, Z. Liang, Y. Wang, and J. Zhang. Generalized linear bandits with local differential privacy. *arXiv preprint arXiv:2106.03365*, 2021.

J. Hörner and A. Skrzypacz. Selling information. *Journal of Political Economy*, 124(6): 1515–1562, 2016.

J. Hsu, M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayan, B. C. Pierce, and A. Roth. Differential privacy: An economic method for choosing epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium*, pages 398–410. IEEE, 2014.

M. Hu, R. Momot, and J. Wang. Privacy management in service systems. *HEC Paris Research Paper No. MOSI-2020-1379*, 2020.

S. Ichihashi. Dynamic privacy choices. *Available at SSRN 3472151*, 2020.

N. Immorlica, Y. Kanoria, and J. Lu. When does competition and costly information acquisition lead to a deadlock? *Available at SSRN 3697165*, 2020.

S. Jagabathula, D. Mitrofanov, and G. Vulcano. Inferring consideration sets from sales transaction data. *NYU Stern School of Business*, 2020.

Z. Jorgensen, T. Yu, and G. Cormode. Conservative or liberal? personalized differential privacy. In *2015 IEEE 31St international conference on data engineering*, pages 1023–1034. IEEE, 2015.

G. Kamath, J. Li, V. Singhal, and J. Ullman. Privately learning high-dimensional distributions. In *Conference on Learning Theory*, pages 1853–1902. PMLR, 2019.

G. Kamath, V. Singhal, and J. Ullman. Private mean estimation of heavy-tailed distributions. In *Conference on Learning Theory*, pages 2204–2235. PMLR, 2020.

V. Karwa and S. Vadhan. Finite sample differentially private confidence intervals. *arXiv preprint arXiv:1711.03908*, 2017.

J. Lee and C. Clifton. How much is enough? choosing ε for differential privacy. In *Information Security: 14th International Conference, ISC 2011, Xi'an, China, October 26-29, 2011. Proceedings 14*, pages 325–340. Springer, 2011.

Y. M. Lei, S. Miao, and R. Momot. Privacy-preserving personalized revenue management. *HEC Paris Research Paper No. MOSI-2020-1391*, 2020.

L. Li. Information sharing in a supply chain with horizontal competition. *Management Science*, 48(9):1196–1212, 2002.

L. Li and H. Zhang. Confidentiality and information sharing in supply chain coordination. *Management science*, 54(8):1467–1481, 2008.

K. Ligett and A. Roth. Take it or leave it: Running a survey when privacy comes at a cost. In *International workshop on internet and network economics*, pages 378–391. Springer, 2012.

Y. Liu and Y. Chen. Learning to incentivize: Eliciting effort via output agreement. *arXiv preprint arXiv:1604.04928*, 2016.

Y. Liu and Y. Chen. Sequential peer prediction: Learning to elicit effort using posted prices. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

I. Lobel and W. Xiao. Optimal long-term supply contracts with asymmetric demand information. *Operations Research*, 65(5):1275–1284, 2017.

O. Lucas, M. Sokalski, and R. Fisher. Corporate data responsibility: Bridging the consumer trust gap. *KPMG Advisory*, 2021.

F. McSherry and K. Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.

L. Mehner, S. N. von Voigt, and F. Tschorsch. Towards explaining epsilon: A worst-case study of differential privacy risks. In *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 328–331. IEEE, 2021.

R. Meir, A. D. Procaccia, and J. S. Rosenschein. Algorithms for strategyproof classification. *Artificial Intelligence*, 186:123–156, 2012.

I. Mironov. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pages 263–275. IEEE, 2017.

R. Montes, W. Sand-Zantman, and T. Valletti. The value of personal information in online markets with endogenous privacy. *Management Science*, 65(3):1342–1362, 2019.

R. B. Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

K. Nissim, C. Orlandi, and R. Smorodinsky. Privacy-aware mechanism design. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 774–789, 2012.

K. Nissim, S. Vadhan, and D. Xiao. Redrawing the boundaries on purchasing data from privacy-sensitive individuals. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 411–422, 2014.

B. Niu, Y. Chen, B. Wang, Z. Wang, F. Li, and J. Cao. Adapdp: Adaptive personalized differential privacy. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, pages 1–10. IEEE, 2021.

M. M. Pai and A. Roth. Privacy and mechanism design. *ACM SIGecom Exchanges*, 12(1): 8–29, 2013.

J. Perote and J. Perote-Pena. The impossibility of strategy-proof clustering. *Economics Bulletin*, 4(23):1–9, 2003.

K. Rosling. Inventory cost rate functions with nonlinear shortage costs. *Operations Research*, 50(6):1007–1017, 2002.

A. Roth and G. Schoenebeck. Conducting truthful surveys, cheaply. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 826–843, 2012.

W. Shang, A. Y. Ha, and S. Tong. Information sharing in a supply chain with a common retailer. *Management Science*, 62(1):245–263, 2015.

Y. H. Wang. On the number of successes in independent trials. *Statistica Sinica*, pages 295–312, 1993.

B. Yu. Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, 1997.