Data: Implications for Markets and for Society

Thesis by
Juba Ziani

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2019
Defended May 29, 2019

Juba Ziani
ORCID: 0000-0002-3324-4349

ACKNOWLEDGEMENTS

First, I would like to thank my advisors, Katrina Ligett and Adam Wierman, who guided me every step along the way. Your guidance and feedback made me a better researcher over the years, and are the main reason I made it where I am today. I am grateful for your patience, your support, your kindness, your faith in me, and your genuine care during the past six years. I feel blessed and incredibly lucky to be your student.

I also would like the faculty members at Columbia who were the first to give me a chance to work on research, and who encouraged me to become a researcher, back in 2012, when I was finishing my master's degree. Thank you, Augustin Chaintreau and Maria Chudnovsky.

None of this would have been possible without the help of my incredibly talented co-authors; I learned an awful lot thanks to all of you. Thank you Rachel Cummings, Xiaoqi Ren, Palma London, Venkat Chandrasekaran, Omer Tamuz, Shai Vardi, Steven Wu, Aaron Roth, Sampath Kannan, Hu Fu, Yang Cai, Federico Echenique, Yiling Chen, Nicole Immorlica, Brendan Lucier, Vasilis Syrgkanis, Elie Tamer, and of course, my advisors Katrina Ligett and Adam Wierman.

Of course, I would not have made it without my friends in Pasadena, who were here for me even during the hardest times, and without whom I would not have retained my sanity (not that I am sane by any reasonable definition of the word, but it could have been much worse). Thank you Akshay Sridhar, Nelson Yanes, and Paul Mazur for being amazing roommates; I will definitely miss our flat councils, smash sessions, in-and-out runs, and 6am rugby watching sessions. Thanks to all of the Caltech friends whom I have made many good memories with, in particular Rachel Cummings a.k.a. "Sister", Noah Olsman, Stephan Zheng, Janis Hesse, Mason McGill, Scarlet Eskew, Donal O'Sullivan, Sam Squires, Nikola Georgiev, Gautam Goel, Anish Sarma, Jialin Song, Sumanth Dathathri, Yu Takahashi, as well as my non-Caltech friends in L.A., especially Axel Kurkjian, Aleen Bedrosian. I should also probably thank Ernie, who kept me well-fed over the years, Philz, who provided me with coffee when most needed, and Lucky Baldwin's, for being my go-to hangout spot when I needed to decompress.

I also owe thanks to all of the friends I made during my travels, and who definitely made my experiences in Berkeley, Cambridge, Jerusalem, and Philadelphia that

much more enjoyable. Thanks in particular to Miriam Manevitz, Yahav Bechavod, Steven Wu, Gautam Kamath, Melih Elibol, Christopher Jung, Jieming Mao, Shahin Jabbari, Hadi Elzayn, Neil Vexler, Bexia Shi, Mark Bun, Brendan Avent, Gian Pietro Farina, Marcel Neunhoeffer, Wanrong Zhang, and Yuliia Lut. I also need to thank all of the lifelong friends I made back in New York and in France, whom I am still in touch with despite the physical distance between us, especially Clément Canonne, Yumi Vielpeau, Rémi Rampin, Laurent Ren, Jérôme Thai, Richard Chen, Lucas Violleau, Feiyun Cheng, Alexandre Gras, Guillaume Rageul, and Vincent Barbaresi.

There only way I could possibly end this section, and this is by acknowledging all the support I received from my parents and my brother who loved, encouraged, and supported me at every single step of my life journey, and to whom I owe most of my achievements. I might not be good at showing it, but I am incredibly appreciative of everything you have done for me. It has been hard at times to be away from you guys, and it always fills my heart to be able to spend time with you, even if it is only once in a while.

ABSTRACT

Every day, massive amounts of data are gathered, exchanged, and used to run statistical computations, train machine learning algorithms, and inform decisions on individuals and populations. The quick rise of data, the need to exchange and process it, to take data privacy concerns into account, and to understand how it affects decision-making, introduce many new and interesting economic, game theoretic, and algorithmic challenges.

The goal of this thesis is to provide theoretical foundations to approach these challenges. The first part of this thesis focuses on the design of mechanisms that purchase then aggregate data from many sources, in order to perform statistical tasks. The second part of this thesis revolves around the societal concerns associated with the use of individuals' data. The first such concern we examine is that of privacy, when using sensitive data about individuals in statistical computations; we focus our attention on how privacy constraints interact with the task of designing mechanisms for acquisition and aggregation of sensitive data. The second concern we focus on is that of fairness in decision-making: we aim to provide tools to society that help prevent discrimination against individuals and populations based on sensitive attributes in their data, when making important decisions about them. Finally, we end this thesis on a study of the interactions between data and strategic behavior. There, we see data as a source of information that informs and affects agents' incentives; we study how information revelation impacts agent behavior in auctions, and in turn how a seller should design auctions that take such information revelation into account.

PUBLISHED CONTENT AND CONTRIBUTIONS

Nicole Immorlica, Katrina Ligett, and Juba Ziani (2019). "Access to Population-Level Signaling As a Source of Inequality". In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 249–258. **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** DOI: 10.1145/3287560.3287579. URL: http://doi.acm.org/10.1145/3287560.3287579.

Yang Cai, Federico Echenique, Hu Fu, Katrina Ligett, Adam Wierman, and Juba Ziani (2018). "Third-Party Data Providers Ruin Simple Mechanisms". In: *arXiv preprint arXiv:1802.07407*, **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** URL: http://arxiv.org/abs/1802.07407.

Yiling Chen, Nicole Immorlica, Brendan Lucier, Vasilis Syrgkanis, and Juba Ziani (2018). "Optimal Data Acquisition for Statistical Estimation". In: *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 27–44. **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** DOI: 10.1145/3219166.3219195. URL: http://doi.acm.org/10.1145/3219166.3219195.

Rachel Cummings, Katrina Ligett, Aaron Roth, Zhiwei Steven Wu, and Juba Ziani (2015). "Accuracy for Sale: Aggregating Data with a Variance Constraint". In: *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pp. 317–324. **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** DOI: 10.1145/2688073.2688106. URL: http://doi.acm.org/10.1145/2688073.2688106.

TABLE OF CONTENTS

Part 1

Motivation and Background

Chapter 1

INTRODUCTION

Nowadays, data is everywhere: with quintillions of bytes produced and treated every day, more than 90 percent of the world's data has been created in the past couple years. In the current age of automation, data is gathered and used extensively in computations that can serve many purposes. This data can be used to compute statistics on populations; the U.S. Census Bureau, for example, aims to collect, analyze, and release useful information about the U.S. population; a public health administration may want to collect data on individuals in a population to output statistics on the general health of said population. This data is also often used to train the plethora of machine learning algorithms that are now part of our daily lives, such as the ones used by online recommendation systems, or more recently by self-driving cars; for instance, Netflix recommends specific movies or TV shows to users, and Yelp recommends restaurants one may like, using machine learning algorithms that are trained and make decisions based on user data and search or watch histories. Finally, this data can be used to better understand individuals' behavior, and how to best respond to this behavior; the ability to understand and rationalize how competitors and advertisees act in ad auctions is key to refining one's own bidding strategies in such auctions.

Some companies, such as Google or Facebook, hold enormous amounts of data from their customers or users. Other companies may be looking to build or complement their data-sets by buying and aggregating data from various sources. This supply and demand of data gives rise to interactions in which data is sold and exchanged. Yet exchange of data differs from exchange of physical goods in many ways. Data may be expensive to produce but is cheap to replicate; data from various sources can be aggregated to obtain higher quality, more informative data; data can inform decisions and influences agent behavior. In this regard, much of the current understanding of the economic and algorithmic aspects of markets for physical goods does not transfer to markets for data. Further, many societal concerns arise when using data about individuals in computations and in decision-making. Data privacy is of particular importance; one may want to perform computations that use sensitive data (think, for example, of medical data), while guaranteeing the privacy of the individuals whose data is used. Another crucial concern is fairness: when attributes

of individuals or populations inform important, life-altering decisions about them, can we guarantee that fair decision-making, that does not unjustly discriminate based on these attributes? Therefore, the rise of data, the need to exchange and process it, and to understand how it affects decision-making, are introducing many new and interesting economic, game theoretic, and algorithmic challenges.

The goal of this thesis is to examine the challenges related to the exchange and use of data, and to provide theoretical foundations to address these challenges. We begin by examining the design of markets for data acquisition; in particular, we provide mechanisms that can be used to purchase and aggregate data from companies or individuals, and give theoretical guarantees on their performance. We then examine the privacy and fairness concerns that arise from the use of data. We first look at privacy and analyze how imposing formal privacy constraints on data use impacts the design of markets for data. Then, we shift our attention to fairness; there, we attempt to identify possible sources of disparities between how different populations or individuals are treated, and aim to understand which interventions provably reduce these inequalities. Finally, we conclude this thesis on a study of the interactions between data and strategic behavior. We consider two aspects of these interactions; on the one hand, we study data as a source of information that affects how agents behave in strategic settings, and how such information impact market design; on the other hand, we briefly investigate how data about agent behavior can be used to understand preferences and utilities of said agents.

1.1 Overview of this Thesis

This thesis is divided in three parts. The first part focuses on markets for data, in particular on mechanism design for data acquisition. The second part studies the societal concerns that arise from the use and exchange of data. The third part examines the interaction between data and strategic behavior.

Markets for data

A fundamental challenge in the study of data markets is to develop *new algorithms to gather and aggregate data from many agents*, in order to learn useful lessons from this data. In this thesis, we will focus on data acquisition from *strategic* agents; i.e., informally, agents that act in their own best interest, and may be looking to "game" the algorithm to obtain their preferred outcomes. For example, consider a data analyst looking to buy a data point from a single data provider; the analyst wants to compensate the provider for his cost for producing and reporting his data point, but

simultaneously is on a budget and is looking to minimize his payment to the provider in exchange for the data point. A simple algorithm to solve this problem is to ask the data provider to report his cost for producing or releasing his data point, and to pay the agent exactly this cost. A strategic agent, however, may game this naive algorithm by over-reporting his cost, so as to increase the payment he receives from the analyst and pocket the surplus. It is crucial to design algorithms for acquiring data that are robust to such strategic behavior, while meeting computational goals and budget constraints that an analyst may have.

The problem of designing algorithms for exchanging goods in the presence of strategic behavior falls under the umbrella of *mechanism design* (Mas-Colell et al., 1995; Nisan et al., 2001; Nisan et al., 2007). However, classical market and mechanism design mostly applies to physical goods, and does not fully capture the complexity of markets for data. In particular, an agent's value for a data point usually depends on additional information available to this agent, and data purchasing decisions must be made while taking into account how this data will be aggregated and what computation(s) will be performed. Further, data usually comes from a variety of sources, and may come in different forms, qualities, and at different costs; as such, one of the main challenges that this thesis seeks to address is that of developing mechanisms that must simultaneously decide i) what type of data to buy from which agents, ii) how to compensate agents for their data, and iii) how to aggregate data from various sources. This is the object of Chapters 3 and 4 of the current document.

In Chapter 3—based on joint work with Rachel Cummings, Katrina Ligett, Aaron Roth, and Steven Zhiwei Wu, published at ITCS 2015 (Cummings et al., 2015b)—we take the point of view of an analyst that must choose which data points to buy, how to aggregate them, and how to compensate data providers for the cost they incur by producing and/or providing their data point. More precisely, the analyst has a choice of what quality (or equivalently, accuracy) of data to purchase from each provider. Each quality level has a cost associated with it, reported by the data provider. The analyst's problem is to choose which quality of data to buy from each provider, so as to combine his purchased data into an aggregate estimator that meets a pre-specified accuracy goal, while minimizing the costs incurred by the providers. We provide a mechanism for this problem that is robust to agents acting strategically and misreporting their costs for producing and releasing data.

The work of Chapter 3 considers situations in which the data point provided by an agent is statistically independent of his cost for producing or revealing this data

point. There are, however, many practical situations in which this is seldom the case. Consider the following: a public health organization wants to understand what fraction of a given population has HIV, and does so by asking individuals in the population to report their HIV status. An individual incurs a privacy cost for revealing his health status to the public health organization. In this setting, agents with HIV have more sensitive data than their healthy counterparts, and as such are generally expected to incur higher privacy costs for revealing their data points. A single-minded public health organization focused on obtaining as many data points for as cheap as possible, and in turn unwilling to pay for costly data points, runs the risk of disproportionately aggregating data from agents who do not have HIV, and vastly under-estimating the prevalence of HIV in the considered population. Another situation of interest could be the following: an analyst asks workers on Amazon Mechanical Turk to label pictures, that will be later be used to train a neural network; the accuracy and quality of the labels provided by a worker is correlated with how much time and effort, and hence cost, he put into labelling the data.

Chapter 4—based on joint work with Yiling Chen, Nicole Immorlica, Brendan Lucier and Vasilis Syrgkanis that appeared at EC 2018 (Chen et al., 2018a)—explicitly takes correlation between data and costs into account, by building on the model of Roth et al. (2012). In this chapter, we consider an analyst who wants to compute an unbiased statistic of interest, by purchasing data from agents, under a budget constraint, when the data and cost of the agents are correlated in an unknown fashion. We design a mechanism that optimizes the worst-case mean-squared error of the estimation, where the worst-case is over the unknown correlation between costs and data. We characterize the form of the optimal mechanism in closed-form, and further extend our results to acquiring data for parameter estimation in regression analysis.

Societal concerns from the use of data

As recent events and developments have shown, when using possibly sensitive data about individuals or populations, *important societal concerns and considerations arise*. This thesis focuses on two such societal concerns: data privacy, and fairness in decision-making.

Privacy A first, crucial concern is *privacy*: when using sensitive, individual data for a given computation, one should ensure that the outcome of said computation does not reveal much information about the input data; a well intentioned analyst may want to understand statistics on HIV in a given population, but the resulting computation

must not allow an adversary to identify whether any particular individual, in fact, has HIV; a platform like Netflix may want to leverage user data to improve their recommendation system, but must make sure that this data cannot be used to recover any single user's real-life identity. The following question then arises: how does one compute useful, informative statistics from agents' data while still providing them with privacy guarantees?

Luckily, this is exactly what *Differential Privacy* strives to achieve. Differential privacy is a widely studied notion of privacy introduced by Dwork et al. (2006); it gives theoretical, worst-case guarantees on how much information can be learned about any single individual in a computation, from observing the output of said computation. More precisely, differential privacy ensures that if a single agent in the computation were to modify his data point, the outcome of the computation would be almost unchanged (in distribution). Differential privacy is not only a strong privacy guarantee that has received a lot of attention in the academic computer science literature (for a survey of differential privacy, see the book of Dwork et al. (2014)), but also a privacy guarantee that the technology industry—in particular, Google (Erlingsson et al., 2014) and Apple (Greenberg, 2016)—as well as government agencies such as the U.S. Census Bureau (Abowd, 2018), are starting to adopt.

While there is a long line of work on differential privacy, its interactions with economic considerations have not been studied as carefully. Privacy is an important concern not only in the context of use of sensitive data, but also in the context of data acquisition. The level of differential privacy provided to an agent affects the privacy cost incurred by said agent to reveal his sensitive data, and an analyst must balance out computational accuracy and payments to agents: too much privacy, and the agents' data becomes less and less informative, making it harder to meet accuracy goals of any desired computation; too little privacy, and the agents' privacy costs become high, leading to an analyst lacking budget to compensate agents for their data. Such issues are complicated by the fact that strategic agents may have an incentive to misreport their privacy costs, should the analyst not appropriately design his data acquisition mechanism.

Understanding trade-offs between accuracy and payments, in the presence of strategic behavior, is the object of Chapter 5. There, we formally define the notion of differential privacy, provide the reader with algorithms that guarantee data is treated in a differentially private manner, and show how the results of Chapter 3—based on joint work with Rachel Cummings, Katrina Ligett, Aaron Roth and Steven Z. Wu

(Cummings et al., 2015b)—extend to data acquisition for sensitive data, when agents are given differential privacy guarantees, and compensated for any remaining privacy costs they incur.

Fairness A second fundamental concern is *fairness*: often, data on agents is used to make important decisions; a bank may use loan applicants' features to decide whether to approve loans; algorithms such as COMPAS are used by the criminal justice system to estimate chances of recidivism. When crucial, life-changing decisions must be made based on agents' features, it is of utmost importance that such decisions are made in a fair, non-discriminatory manner.

However, the current status quo is that many decision-making algorithms are, in fact, discriminatory. This could be due to a variety of reasons: decision-making algorithms may be trained on discriminatory data, whose bias they learn to mimic; a lack of data about the performance of a disadvantaged population in certain jobs, due to the fact that said population has been historically prevented from accessing such jobs, may lead employers to not be willing to hire from said population; a decision-maker himself may be, consciously or unconsciously, biased towards or against certain populations or individual profiles.

Examples of disparities are numerous in everyday life. Loan data is known to be historically biased against minorities (Cohen-Cole, 2011) which in turns leads to disparities in how loan approval decisions are made across populations. COMPAS, a widely used tool to assess a defendant's risk of recidivism, and to inform crucial decisions about which defendants should be released and which should remain in jail, has been shown to discriminate against minorities by over-estimating recidivism risk for minority defendants while under-estimating risk for white defendants (Angwin et al., 2016).

It is therefore essential to provide algorithmic tools to society that aim to prevent various individuals or populations from being treated unfairly based on sensitive attributes such as gender or race, and to correct for disparate treatment of populations. A first step to do so, is to understand what causes disparities in treatment, and how well current algorithms and interventions perform when it comes to alleviating these disparities. Chapter 6—based on joint work with Nicole Immorlica and Katrina Ligett, and published at FAT* 2019 (Immorlica et al., 2019)—focuses on unfairness in university admissions. The first contribution of this chapter is to identify unequal access to strategic signaling as a source of unfairness. We then quantify this

unfairness, and show how classical interventions (for instance, having students take a standardized exam) may have the undesirable outcome of increasing the disparities between students in schools with unequal access to signaling.

Fairness in decision-making must also be sought with long-term considerations taken into account. In many practical settings, decisions made today affect future decision-making, either because they impact the priors and information available to decision-making parties further down the pipeline, or simply because they restrict the pool of agents or data future algorithms will make decisions on. This is the object of joint work with Sampath Kannan and Aaron Roth (Kannan et al., 2019); we study whether and—when possible— how a school can influence decisions from other entities down the job market pipeline, and guarantee that similar students from different populations will be treated equivalently; we only allow the school to control its own admission and grading policies, and study the effect of setting different admission rules (effectively, affirmative action) for different populations on long-term fairness. For length reasons, we omit the details of this work in the current thesis, and refer the reader to the publicly available version of this work.

Data and strategic behavior

A last important challenge that this thesis aims to address is to understand *how data explains and informs strategic behavior*. We divide this discussion in two parts. The first part aims to understand the effect of information revelation on mechanism design and strategic behavior, while the second part focuses on the problem of inferring utilities and preferences from behavioral data.

Effect of information on mechanism design and strategic behavior To begin with, data is a source of information that influences the behavior of strategic agents. In many real-life mechanism design settings and auctions in which a seller aims to sell goods to a set of agents (or, in the current context, "bidders"), said bidders may have incomplete information on how much they even value these goods in the first place. Consider, for example, an auction for a piece of land; a bidder may have incomplete information about how fertile the land is, or how much oil can be extracted, and hence about how much he can expect to benefit from said land. In online ad auctions, how much an advertiser can hope to profit from an ad depends on whether the ad reaches a relevant target audience; an advertiser participating in an ad auction for a specific advertising slot on, for example, Google, may be uncertain about what users will see the ad, and therefore may have an incomplete understanding of his

own value for said advertising slot.

However, in many such settings, auxiliary information about the good for auction may be available to the agents, outside of the control of the auctioneer. Geological information about a given piece of land might be publicly available; an advertiser in an ad auction may be able to identify characteristics of the population of advertisees via cookie data. As such, it is important to understand how third-party data influences how agents behave in auctions and mechanism design settings, and how this affects the seller's auction design problem.

We do so in Chapter 7, based a manuscript with Yang Cai, Federico Echenique, Hu Fu, Katrina Ligett, and Adam Wierman (Cai et al., 2018). This chapter builds on a mechanism design setting introduced by Daskalakis et al. (2016), where bidders have incomplete information about the item for sale, that could be of many possible "types". We assume that, in addition to the bidders and the seller found in the work of Daskalakis et al. (2016), there is a third-party source of information that has the ability to reveal information to the bidders about the identity of the item for sale. We aim to understand how the revenue guarantees of a widely used and studied class of simple mechanisms, which are known to achieve a constant-approximation of the revenue in the absence of auxiliary information, translate to mechanism design settings with auxiliary information. Our main contribution is to show that, in fact, this revenue degrades significantly: when using simple mechanisms, a seller can only guarantee a logarithmic (in the number of possible types) fraction of the optimal achievable revenue in the presence of third-party information. Our results are shown to hold in the worst-case over the auxiliary information, but also when the data provider is strategic and trying to optimize some objective (for example, maximizing his own utility from selling their information to bidders).

Inferring strategic incentives from behavioral data Finally, data on strategic agents' behavior reveals information about their utilities and preferences, and one may be interested in understanding just how much knowledge can be acquired in this manner and how it can be used to predict future behavior. There are several reasons for doing so. First, understanding how much information can be learned from observing agent behavior is key to preventing this inferred knowledge from being misused. Second, the ability to compute counterfactuals of how agents would react under new conditions is useful in predicting future agent behavior, and can be leveraged to incentivize self-interested agents to act in a desired (for instance,

socially beneficial) manner.

Inference from observed behavior is the object of joint work with Venkat Chandrasekaran and Katrina Ligett (Chandrasekaran et al., 2016). There, we study the problem of using behavioral data to understand agents' preferences and payoffs, in a game of complete information (i.e., agents perfectly understand their own valuations). We assume an analyst observes agents' actions but does not know the game the agents are playing and their payoff from each action. Similarly, in joint work with Vasilis Syrgkanis and Elie Tamer (Syrgkanis et al., 2017), we provide algorithmic tools to recover the distribution of the common value in a common value auction with signaling from observing the bidding strategies of the agents, under no assumption on the information structure on the common-value available to the bidders. Both works assume that agents act according to variants of the widely studied concept of *Nash equilibrium* (Nash et al., 1950), and leverage linearity properties of these equilibrium concepts to provide convex programming-based, computationally efficient frameworks to recover the parameters of interest. We omit both Chandrasekaran et al. (2016) and Syrgkanis et al. (2017) in the current thesis, and refer the reader to their publicly available versions.

Chapter 2

PRELIMINARIES: GAME THEORY AND MECHANISM DESIGN

For simplicity of notations, in this chapter, we let v_i denote the i-th entry, and v_{-i} denote the vector of all entries except the i-th entry, for any vector v.

2.1 Strategic Behavior, Game Theory, and Equilibrium Play

Much of this thesis focuses on economic and algorithmic challenges in exchanging and using data, in the presence of *strategic* agents; i.e., agents who act in their own, best interest. In this section, we provide a formal, theoretical basis to study the behavior of such agents.

To do so, we will first introduce the concept of *utility* of an agent. A *utility function* is a function that measures an agent's preferences over a set of alternatives, by assigning a numerical value to each alternative. The numerical value associated to a given alternative quantifies how much an agent likes said alternative; the higher the utility, the more appealing the corresponding alternative is. A *strategic* agent is an agent who aims to make decisions that maximize his utility.

Consider, for example, an individual who wants to buy a new phone. Several phones may be available to the individual, and he must decide which one to buy. The individual may make his purchasing decision based on a combination of several considerations, such as how much he values different features of different phones, or how much he has to pay for a given phone. This combination of considerations can be encoded in a utility for each phone; the higher the utility, the better the trade-off between features and price is from the individual's point of view. A strategic agent buys the phone that maximizes his utility among all alternatives.

In the above example, we considered a setting with a single agent that must pick his preferred alternative, or outcome. It is often the case, however, that the alternative or outcome obtained by an agent, is a combination of his own actions and decisions, but also of the actions and decisions of several agents. Consider, for example, what is commonly referred to as an "entry game": several firms must decide whether to enter a market; the revenue, hence utility, that each firm expects to make on this market is a function of how many competitors this firm will face, and thus of how many other firms also decide to enter the market.

Such situations, in which agents aim to make decisions that are in their best interest, but where the outcome they obtain is a function of decisions made by other agents, are called *games*. Formally, a game can be defined as follows:

Definition 2.1.1 ((Normal-Form) Game)**.** *A game $G = (\mathcal{P}, \mathcal{A}, \mathcal{U})$ consists of:*

- *A set of players $\mathcal{P}$, of cardinality n.*
- *A collection of action sets $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_n)$. Each player $i \in \mathcal{P}$ has an action set $\mathcal{A}_i$, and can play an action $a_i \in \mathcal{A}_i$.*
- *A collection of utility functions $\mathcal{U} = (u_1, \ldots, u_n)$. Each player $i \in \mathcal{P}$ has utility function $u_i : \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_n \to \mathbb{R}$. A player's utility $u_i(a_i, a_{-i})$ depends not only on his own action a_i, but also on the actions a_{-i} of the other players.*

In the above entry game example, a player or agent is a firm, an action is to either enter the market, or not enter it, and an agent's utility is a measure of the profit he expects to make by entering (or not entering) the market.

Agent i participates in the game by choosing a strategy ζ_i, which is a probability distribution over actions in $\mathcal{A}_i$, i.e. agents are allowed to randomize over their action sets. A strategy profile $\zeta = (\zeta_1, \ldots, \zeta_n)$ is a n-tuple of strategies, one for each player. In this thesis, we model strategic agents as choosing their strategies so as to form a Nash equilibrium of the game they are playing:

Definition 2.1.2 (Nash Equilibrium)**.** *A strategy profile ζ is a Nash equilibrium of game $G = (\mathcal{P}, \mathcal{A}, \mathcal{U})$ if and only if, for every player i, and for all $\hat{a}_i \in \mathcal{A}_i$,*

$$\mathbb{E}_{(a_i, a_{-i}) \sim \zeta} \left[u_i \left(a_i, a_{-i} \right) \right] \geq \mathbb{E}_{a_{-i} \sim \zeta_{-i}} \left[u_i \left(\hat{a}_i, a_{-i} \right) \right].$$

At a Nash equilibrium, every agent plays a strategy that maximizes his own utility, given the strategies chosen by the remaining agents: i.e., ζ_i is the utility-maximizing strategy for agent i when the other agents follow strategies ζ_{-i}. Informally, a Nash equilibrium is an equilibrium in which *every player's strategy is a best response to the strategies of the remaining players*.

To compute a Nash equilibrium of a (normal-form) game, we have to make the implicit assumption that the utility function u_i of every player is publicly known. This is often not the case. In an auction, for example, how much a given agent values an item for sale may only be known to this agent; in data acquisition settings, the

privacy cost that an individual incurs for revealing his data to an analyst is generally not known to other agents, nor to the analyst. To model such situations, we assume that an agent i has a private type $t_i \in T$, such that his utility function depends on this type. In such cases, an agent's strategy is allowed to depend on his type, and we denote by $\zeta_i(t_i)$ the strategy of player i when his type is t_i.

Such incomplete information settings can typically be dealt with in several manners. One possibility is to assume agents behave according to stronger equilibrium concepts, such as *dominant-strategy equilibria*. A dominant strategy equilibrium is an equilibrium in which each agent's strategy is a best response to *any possible behavior of the remaining agents*.

Definition 2.1.3 (Dominant-Strategy Equilibrium)**.** *A strategy profile ζ is a dominant strategy equilibrium if and only if, for every player i, for all $\zeta_{-i}(t_{-i})$ and for all $\hat{a}_i \in \mathcal{A}_i$,*

$$\mathbb{E}_{(a_i,a_{-i})\sim\zeta(t)}\left[u_i\left(a_i, a_{-i}\right)\right] \geq \mathbb{E}_{a_{-i}\sim\zeta_{-i}(t_{-i})}\left[u_i\left(\hat{a}_i, a_{-i}\right)\right].$$

In Chapter 3, we consider a setting where the agents' types are private; there, the mechanism designer assumes agents participate in his mechanism and decide how to report their types according to a dominant-strategy equilibrium.

Another option is to make *distributional assumptions on agent's types*. It is common to make the assumption that the vector of agents' types t is taken from a known (joint) distribution $\mathcal{F}$, also called "prior". In Chapters 4 and 7, we allow the mechanism designer to make distributional assumptions on the agents' types, and to use such priors to inform the design of his mechanism. In such settings, one can consider agents acting according to the concept of *Bayes-Nash equilibrium*. Informally, a Bayes-Nash equilibrium is an equilibrium in which each agent i's strategy is a best response to every other agent's strategy, *in expectation over the other agents' types*.

Definition 2.1.4 (Bayes-Nash Equilibrium)**.** *A strategy profile ζ is a Bayes-Nash equilibrium if and only if, for every player i, and for all $\hat{a}_i \in \mathcal{A}_i$,*

$$\mathbb{E}_{(a_i,a_{-i})\sim\zeta,t\sim\mathcal{F}}\left[u_i\left(a_i, a_{-i}\right)|t_i\right] \geq \mathbb{E}_{a_{-i}\sim\zeta_{-i},t\sim\mathcal{F}}\left[u_i\left(\hat{a}_i, a_{-i}\right)|t_i\right].$$

Remark 2.1.5 (Dominant-strategy vs Bayes-Nash)**.** *Note that every dominant-strategy truthful equilibrium of a given game is also a Bayes-Nash equilibrium of the same game. Generally, the converse is not true; however, in settings where every agent's utility is a function of only his own type and action (as in Chapter 4), or in*

settings where there is a single agent participating in the game (as in Chapter 7), every Bayes-Nash equilibrium is a dominant-strategy truthful equilibrium of the considered game.

2.2 Mechanism Design

Mechanism design takes an engineering approach to game theory, by influencing and designing incentives and utility functions of strategic agents, so as that it is in their best interest to work towards the objective the designer wants to achieve. While game theory aims to understand how players act within the rules of a game, mechanism design strives to design the rules of the game to incentivize certain behaviors. A mechanism is defined as follows:

Definition 2.2.1 (Mechanism)**.** *A mechanism is a (randomized) mapping from players' actions to outcomes that belong to a set* Ω*:*

$$\mathcal{M} : \mathcal{A}_1 \times \ldots \times \mathcal{A}_n \to \Omega.$$

In this thesis, we restrict our attention to *direct revelation mechanisms*, which are mechanisms in which the available actions are restricted to reporting a (any) type; i.e. the action sets are given by $\mathcal{A}_i = T$. This can be shown to be without loss of generality, i.e. any mechanism can be implemented as a direct-revelation mechanism, by the *Revelation Principle* (Myerson, 1981). A direct revelation mechanism can also often be implemented as a menu of options, such that the mechanism designer lets agents pick their preferred option from said menu; we will do so in Chapters 4 and 7.

For instance, consider an analyst who wants to purchase data from individuals or organizations, given a limited budget, in order to compute a population statistic. To do so, the analyst can run a direct revelation mechanism. The mechanism first asks each agent to report his cost for providing his data point to the analyst. Then, the mechanism decides whether to buy data from each agent according to an *allocation rule*, how to compensate each agent for his data according to a *payment rule*, and how to aggregate the data acquired from the agents. Both the allocation and the payment rules are functions of the collection of agents' reported costs, and the design of these rules controls the utilities the agents obtain when reporting costs to the mechanism. In turn, the reported costs affect which and how many data points a budget-constrained analyst is going to acquire, hence the outcome of the analyst's computation. The analyst may want to optimize some function of this outcome (for

example, its accuracy), under his budget constraint. In Part 2, we focus on the design of such mechanisms for data acquisition and aggregation.

Typically, several desirata are expected to hold for direct-revelation mechanisms: it should be in each agent's best interest i) to participate in the mechanism, and ii) to report his true type to the mechanism, should he decide to participate in the mechanism. This first property is called *individual rationality*, while the second property is referred to as *incentive compatibility*, or *truthfulness*. These properties ensure that a mechanism designer can properly understand the agents' incentives, and anticipate how they will act in the mechanism, so that the outcome of the mechanism aligns with the designer's objective. We give the definitions for the dominant-strategy versions (i.e., that protect an agent against *any* possible reports by the other agents) of these properties below:

Definition 2.2.2 (Dominant-Strategy Individual Rationality (IR))**.** *A mechanism $\mathcal{M}$ is individually rational if and only if, for all players i, and for all reports of the other agents t_{-i},*

$$\mathbb{E}_{\mathcal{M}}\left[u_i(t)\right] \geq 0,$$

where u_i is the (possibly randomized) utility of agent i for participating in (possibly randomized) mechanism $\mathcal{M}$.

Definition 2.2.3 (Dominant-Strategy Incentive Compatibility (IC))**.** *A mechanism $\mathcal{M}$ is incentive-compatible if and only if, for every player i, for any true type $t_i \in T$ for agent i, for every possible misreport $\hat{t}_i \in T$ for agent i, and every possible vector of reports $t_{-i} \in T^{n-1}$ for the remaining agents,*

$$\mathbb{E}_{\mathcal{M}}\left[u_i(t)\right] \geq \mathbb{E}_{\mathcal{M},\, t_{-i}\sim\mathcal{F}}\left[u_i(\hat{t}_i, t_{-i})\right].$$

In Chapter 3, we make no assumption on the agents' types, and therefore immediately aim to design a dominant-strategy IC and IR mechanism for data acquisition. In Chapter 4, our mechanism is *dominant strategy* IC and IR, despite the availability of distributional priors on the agent's types: this is by virtue of an agent's utility from participating in the mechanism being a function only of his own actions, and not of the other agents' actions, as per Remark 2.1.5. In Chapter 7, our mechanism is also dominant strategy IC and IR, due to the presence of a single bidder in the mechanism, once again as per Remark 2.1.5.

Part 2

Markets for Data

Chapter 3

ACQUISITION OF COSTLY DATA FOR STATISTICAL ESTIMATION

Rachel Cummings, Katrina Ligett, Aaron Roth, Zhiwei Steven Wu, and Juba Ziani (2015). "Accuracy for Sale: Aggregating Data with a Variance Constraint". In: *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pp. 317–324. **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** DOI: 10.1145/2688073.2688106. URL: http://doi.acm.org/10.1145/2688073.2688106.

3.1 Introduction

In this chapter, we consider a *data analyst* who wishes to compute an unbiased estimate of some underlying population statistic, by buying and aggregating data from multiple strategic data providers. Each data provider may experience different costs for different levels of data accuracy (variance), and may strategically price access to his data if doing so would benefit him. The analyst must design a mechanism for choosing which level of accuracy to purchase from each provider, and for combining the purchased data into a single aggregate quantity that forms an unbiased estimator of the statistic of interest. Her goal is to do so at minimum cost, given some target level of overall accuracy.

An example of scenario captured by this model is the following: each data provider might be an organization (such as a university) that has the ability to collect a random sample of varying size from a sub-population that it controls (e.g., students, professors, etc). Under the assumption that the individuals in the data provider's populations are sampled i.i.d. from some underlying distribution, the variance of the estimate that they offer is inversely proportional to the number of individuals that they sample. Here, the costs for different levels of variance correspond to the costs required to recruit different numbers of participants to a study. These costs may differ between organizations, and behave in complicated ways: for example, the marginal cost for each additional sample might be decreasing (if there are economies of scale—for example by advertising on a campus TV station), or might be increasing (for example, after exhausting the undergraduate population at a university, obtaining additional samples may require recruiting faculty, which is more difficult). Again,

because we allow data providers to report arbitrary cost schedules corresponding to different variance levels, we need make no assumptions about the form that these costs take.

Summary of contributions

We model the data analyst's problem as a combinatorial optimization problem: From each of the data providers, the analyst buys an unbiased estimator of the population statistic of interest, for which she must choose a variance from a fixed, finite menu of options. Importantly, by assuming the purchased estimators are unbiased no matter what variance level is chosen, we are making the implicit assumption that *the sensitive data held by individuals is independent of their costs*. We will do away with this assumption in Chapter 4.

Given these purchased estimators, the data analyst may then take any convex combination to obtain her final unbiased estimator of the underlying population statistic. The choices made by the data analyst affect both the variance of the final estimator that she derives, as well as the total payment that she must make. This chapter considers the problem of finding the *cheapest* way of constructing an estimator that has variance below some fixed desired level, specified in advance by the data analyst.

Our main tool in solving this problem is linear programming. However, the solution is not straightforward. First, our problem actually consists of two nested optimization problems: we must choose a variance level for each of the estimators, and then we must find the optimal weighted linear combination of these estimators. Rather than solving these problems separately, we use the KKT conditions to derive a closed form for the optimal weights to use in the linear combination of each of the estimators *as a function of their variance*. This allows us to express the problem as a one-shot optimization problem, with decision variables only for the choices of variance for each estimator. Unfortunately, the natural fractional relaxation of this optimization problem (in which the data analyst may fractionally choose different variance levels) is non-convex. Instead, we consider a further (linear) relaxation of the constraint in our problem, which matches the original constraint only for integer solutions. We show that all optimal extreme points of the linear program that result from this relaxation do in fact yield integer choices for all but *at most one* data provider, and then show that if the number of data providers is sufficiently large, then rounding the one fractional assignment to an integer assignment only marginally violates our

target variance constraint.

We note that our algorithm chooses the *minimum expected cost* lottery over purchase decisions from among a pre-specified feasible set of lotteries, and hence is *maximal-in-distributional-range* (i.e. it outputs a lottery that maximizes expected welfare). This means that when paired with VCG payments, truthful reporting of costs is a dominant strategy for each of the data providers. (We recall that although we allow data providers to misreport their costs, they cannot lie about their data or its accuracy.)

In summary, we show the following theorem:

Theorem 3.1.1 (Informal)**.** *Given any finite menu of variance levels, and any feasible aggregate variance level for the data analyst, there exists a dominant strategy truthful mechanism that selects the minimum cost assignment of variance levels to providers, and generates an unbiased linear estimator that satisfies the analyst's variance constraint up to an additive term that tends to 0 as the number of data providers grow large.*

Finally, we observe that VCG payments (although always truthful) do *not* guarantee individual rationality in our setting, because these payments may fail to compensate players for their cost for providing data. We prove an upper bound on the degree to which individual rationality can be violated for any player, and hence can add a fixed amount to the payment given to each player, to guarantee individual rationality for all providers with sufficiently low minimum costs.

Related Work

A growing amount of attention has been placed on understanding interactions between the strategic nature of data holders and the statistical inference and learning tasks that use data collected from these holders.

The line of work considered in this thesis sees individuals as data holders who incur a cost for providing their data point (this could be a cost for producing a data point, or a privacy cost for revealing a sensitive data point), and may strategically misreport these costs to the mechanism designer, but the data itself is *verifiable* (Ghosh et al., 2015; Roth et al., 2012; Fleischer et al., 2012; Ligett et al., 2012; Nissim et al., 2014; Cai et al., 2015; Abernethy et al., 2015; Chen et al., 2018c), i.e., agents do not lie about their data when reporting it. The mechanism designer aims to optimize the trade-off between costs incurred by the agents or payments to the agents, and

accuracy of the desired computation. Chapters 3 and 4 of this thesis fall under this line of work.

An important, related line of work looks at acquisition of unverifiable data points, where agents may lie about their data. This line of work can be divided in two: one part of this line of work does not use monetary payments, and leverages the fact that the agents' utilities, hence the way they manipulate their data, depend on the outcome of the mechanism (Perote-Peña et al., 2003; Dekel et al., 2010; Meir et al., 2011; Meir et al., 2012; Hardt et al., 2016a; Caragiannis et al., 2016; Dong et al., 2018; Chen et al., 2018b); the other part of this line of work incentivizes truthful reporting via payments (Ghosh et al., 2014; Cummings et al., 2015a; Kong et al., 2016b; Kong et al., 2016a; Kong et al., 2018; Liu et al., 2016; Liu et al., 2017; Liu et al., 2018).

3.2 Model and Preliminaries

We consider an analyst who wishes to estimate the expected value μ of some statistic on the underlying population. She has access to a set of n data providers, each of which is capable of providing a data point, that is an unbiased estimate μ_i of the statistic of interest, with different levels of variance $\mathbb{E}\left[(\mu_i - \mu)^2\right]$. The provider may also experience some cost for computing the estimate at each variance level. The analyst's goal is to obtain an accurate unbiased estimate for μ, using the estimates from the providers, while minimizing the social cost for computing such data.

We equip the analyst with a mechanism that offers a menu specifying a discrete, finite range of possible variance levels $0 < v_1 < v_2 < \ldots < v_m < \infty$, and asks each provider i to report back a set of costs $\{c_{ij}\}_{j=1}^{m}$ for computing the estimates at all levels. The mechanism then selects a variance level to purchase from each provider, and generates an estimate for μ that is a weighted sum of the providers' reported estimates μ_i's: $\hat{\mu} = \sum_i w_i \mu_i$. Note that the expectation $\mathbb{E}[\hat{\mu}] = \sum_i w_i \mathbb{E}[\mu_i] = \sum_i w_i \mu$, so $\hat{\mu}$ will be an unbiased estimate as long as $\sum_i w_i = 1$. The following proposition, often called the Bienaymé formula, allows us to express the variance of $\hat{\mu}$ as a linear combination of the variances of μ_i.

Proposition 3.2.1. *Let $Y_1, \ldots, Y_n$ be uncorrelated real-valued random variables, and $w_1, \ldots, w_n$ be any real numbers, then*

$$Var\left(\sum_i w_i Y_i\right) = \sum_i w_i^2 Var(Y_i).$$

The goal of the analyst is to minimize the total cost among all providers, while maintaining a guarantee that the variance of $\hat{\mu}$ is below some threshold α. This can

be expressed in the following program, where each A_{ij} indicates whether we assign provider i to variance level j:

$$\min_{A_{ij}, w_i} \sum_{i,j} A_{ij} c_{ij} \tag{3.1}$$

$$\text{s.t.} \sum_i w_i^2 \left(\sum_j A_{ij} v_j \right) \leq \alpha \tag{3.2}$$

$$\sum_j A_{ij} = 1 \text{ for all } i \tag{3.3}$$

$$A_{ij} \in \{0, 1\} \text{ for all } (i, j) \tag{3.4}$$

$$\sum_i w_i = 1 \text{ and for all } i, w_i \geq 0. \tag{3.5}$$

Mechanism Design for Prior-Free Data Acquisition

The problem we study here is a mechanism design problem, with n players and a set Ω of possible outcomes, as defined in Chapter 2. In the current framework, the analyst wishes to determine a variance level at which to purchase data from each player, so this set Ω corresponds to the set of possible assignments of players to variance levels. Each player has a type, which is given by a cost function $c_i : \Omega \to \mathbb{R}$, where $c_i(\omega)$ is the cost player i incurs when the outcome is ω. Let $c = (c_1, \ldots, c_n)$ denote the profile of cost functions for all players. We want to minimize total cost, so our objective is $\sum_{i=1}^n c_i(\omega)$. We will use Ω_{-i} to denote the set of possible assignments of all players other than i to variance levels, and c_{-i} to denote the vector of reported costs by all players other then i. We assume the players' costs are private, and no prior on these costs is available to the analyst.

In this chapter, a *(direct revelation) mechanism* $\mathcal{M}$ consists of an *allocation rule* A, a function mapping reported cost profiles to outcomes, and a *payment rule* P, a function mapping cost profiles to a payments to each player. Such a mechanism takes as input reported cost functions from the players, and outputs (possibly randomly) an allocation ω and payments to all the players. As discussed in Chapter 2, two important desiderata in mechanism design are *truthfulness* (equivalently, incentive-compatibility) and *individual rationality*. We remind the reader of these desirata, and write them using the notations of this chapter:

Definition 3.2.2 (Dominant-strategy truthfulness-in-expectation)**.** *A mechanism* $\mathcal{M} = (A, P)$ *is* (dominant-strategy) truthful-in-expectation *if for all* $i \in [n]$*, for any*

reported cost profile c_{-i} of other players, and any misreport $\hat{c}_i$ by player i:

$$\mathbb{E}_{\mathcal{M}}\left[P_i(c_i, c_{-i}) - c_i(A(c_i, c_{-i}))\right] \geq \mathbb{E}_{\mathcal{M}}\left[P_i(\hat{c}_i, c_{-i}) - c_i\left(A(\hat{c}_i, c_{-i})\right)\right].$$

Definition 3.2.3 (Individual rationality)**.** *A mechanism $\mathcal{M} = (A, p)$ is* individually rational *(I.R.) if for any reported cost profile c and for all $i \in [n]$:*

$$\mathbb{E}_{\mathcal{M}}\left[P_i(c) - c_i\left(A(c)\right)\right] \geq 0.$$

We will use *VCG-based mechanisms* to minimize total cost while achieving truthfulness. A *VCG mechanism* is defined by the allocation rule that selects the cost-minimizing outcome $\omega^* \in \arg\min_{\omega\in\Omega} \sum_i c_i(\omega)$ for any reported cost functions, and the payment rule P that rewards each player his "externality":

$$P_i(c) = \min_{\omega\in\Omega_{-i}} \sum_{i'\neq i} c_{i'}(\omega) - \sum_{i'\neq i} c_{i'}(\omega^*). \tag{3.6}$$

Let $dist(\Omega)$ be the set of all probability distributions over the set of outcomes Ω, and let $\mathcal{R} \subseteq dist(\Omega)$ be a compact subset. Then a *maximal-in-distributional-range* (MIDR) allocation rule is defined as sampling an outcome ω from distribution $D^* \in \mathcal{R}$, where D^* minimizes the expected total cost $\mathbb{E}_{\omega\sim D^*}[\sum_i c_i(\omega)]$ over all distributions in $\mathcal{R}$. A VCG payment rule can be defined accordingly, where $\mathcal{R}_{-i}$ is the corresponding compact subset of Ω_{-i}:

$$P_i(c) = \min_{D'\in\mathcal{R}_{-i}} \mathbb{E}_{\omega\sim D'}\left[\sum_{i'\neq i} c_{i'}(\omega)\right] - \mathbb{E}_{\omega\sim D^*}\left[\sum_{i'\neq i} c_{i'}(\omega)\right].$$

It is known from (Dobzinski et al., 2009) that when an MIDR allocation rule is paired with a VCG payment rule, the resulting mechanism is truthful-in-expectation.

To guarantee individual rationality, we pay each player some entrance reward R before running the MIDR mechanism so that $R + \mathbb{E}\left[P_i(c) - c_i(A(c))\right] \geq 0$ for all players. It suffices to set $R \geq \max_i \mathbb{E}\left[P_i(c) - c_i(A(c))\right]$, and in Section 3.4 we derive a more refined bound for R to achieve individual rationality.

3.3 Reformulating the Problem

The optimization problem introduced in Section 3.2 is non-convex because the variance constraint (3.2) contains the product of decision variables A_{ij} and w_i. To achieve convexity, we will transform the program in three steps:

1. First, we will eliminate the decision variables w_i by deriving a closed form solution for the weights w_i that minimize variance, once the variables A_{ij} are fixed. However, this will still leave us with a non-convex optimization problem.

2. Next, we will replace the non-convex constraint derived above with a linear constraint, that is identical whenever the A_{ij} variables take on integral values.

3. Finally in Section 3.4, we relax the integrality constraint. Because our linear variance constraint is no longer identical to the original "correct" non-convex variance constraint, we must in the end argue that a rounded solution does not substantially violate the original constraint.

First, to simplify notation, for any assignment $\{A_{ij}\}$, let $\hat{v}_i$ denote the variance level assigned to provider i. We want to write each w_i as a function of the $\hat{v}_k$'s. In particular, given the variance assignments, we want to choose the weights w_i so that the variance of the aggregate statistic $\hat{\mu}$ is minimized.

Lemma 3.3.1. *Given a variance level assignment $\{\hat{v}_i\}$, the weight vector w^* that minimizes the variance of $\hat{\mu} = \sum_i w_i \mu_i$ satisfies*

$$w_i^* = \frac{1/\hat{v}_i}{\sum_k 1/\hat{v}_k} \qquad \textit{for all } i.$$

Proof. The problem can be written as a convex program

$$\min \sum_i w_i^2 \hat{v}_i \qquad \text{subject to } \sum_i w_i = 1 \text{ and } w_i \geq 0 \text{ for all } i.$$

We know that strong duality holds because the program satisfies Slater's condition, and the Lagrangian is given by

$$\begin{aligned}\mathcal{L}(w, \lambda) &= \sum_i \hat{v}_i \cdot w_i^2 - \lambda \left(1 - \sum_i w_i\right) \\ &= w^T V w - \lambda(1 - \mathbb{1}^T w),\end{aligned}$$

where $V = diag(\hat{v_1}, \ldots, \hat{v_n})$. Note that $\nabla_w \mathcal{L}(w, \lambda)^T = 2Vw + \lambda \mathbb{1}$. By KKT conditions, $\nabla_w \mathcal{L}(w^*, \lambda)^T = 0$, and so $w^* = -\frac{\lambda}{2} V^{-1} \mathbb{1}$, which gives $\min_w \mathcal{L}(w, \lambda)^T = -\frac{\lambda^2}{4} \sum_i 1/\hat{v}_i - \lambda$. Now the dual problem can be written as

$$\max_{\lambda} \min_{w \geq 0} \mathcal{L}(w, \lambda) = \max_{\lambda} \left[-\frac{\lambda^2}{4} \sum_i 1/\hat{v}_i - \lambda \right]$$
$$= \max_{\lambda} \left[-\left(\sum_i 1/\hat{v}_i \right) \left(\lambda/2 + \frac{1}{\sum_i 1/\hat{v}_i} \right)^2 + \frac{1}{\sum_i 1/\hat{v}_i} \right].$$

It is easy to see that the maximum is reached at $\lambda^* = \frac{-2}{\sum_i 1/\hat{v}_i}$. It follows that

$$w^* = \frac{-\lambda^*}{2} V^{-1} \mathbb{1} = \frac{V^{-1}\mathbb{1}}{\sum_i 1/\hat{v}_i},$$

and so,

$$w_i^* = \frac{1/\hat{v}_i}{\sum_k 1/\hat{v}_k} \text{ for all } i$$

as suggested by the lemma. □

Section 3.3.1 shows that we can rewrite the variance constraint of $\hat{\mu}$ as

$$\sum_i \left(\frac{1/\hat{v}_i}{\sum_k 1/\hat{v}_k} \right)^2 \hat{v}_i = \sum_i \frac{1/\hat{v}_i}{(\sum_k 1/\hat{v}_k)^2} = \frac{1}{\sum_k 1/\hat{v}_k} \leq \alpha.$$

Changing indices back to i, plugging in $\hat{v}_i = \sum_j A_{ij} v_j$, and taking the inverse on both sides, constraint (3.2) becomes

$$1/\alpha \leq \sum_i \frac{1}{\sum_j A_{ij} v_j}. \tag{3.7}$$

Note that the constraints are not linear, but since each $A_{ij} \in \{0, 1\}$, and only one $A_{ij} = 1$ for each i, we have $1/\sum_j A_{ij} v_j = \sum_j A_{ij}/v_j$. Thus, we can write our whole program as the following ILP.

$$\min_{A_{ij}} \sum_{i,j} A_{ij} c_{ij} \tag{3.8}$$
$$\text{s.t.} 1/\alpha \leq \sum_i \sum_j A_{ij}/v_j \tag{3.9}$$
$$\sum_j A_{ij} = 1 \text{ for all } i \tag{3.10}$$
$$A_{ij} \in \{0, 1\} \text{ for all } (i, j). \tag{3.11}$$

Remark 3.3.2. *Note that our problem is only interesting if the target variance α is in the range of $[v_1/n, v_m/n]$. This is due to the following observation based on constraint* (3.9)*: if $1/\alpha < n/v_m$, then the problem is trivial since the variance constraint is satisfied by any assignment; if $1/\alpha > n/v_1$, then the problem is infeasible, i.e. even if we assign the lowest variance level to all providers, the variance constraint is still violated.*

3.4 An MIDR Mechanism via a Linear Programming Relaxation

In order to obtain a computationally efficient mechanism, we consider the LP relaxation of the integer linear program we derived in the previous section, by replacing constraint (3.11) with $A_{ij} \geq 0$ for all (i, j). We interpret a fractional solution $A_i = (A_{i1}, \ldots, A_{im})$ as a lottery over assignments for player i, i.e., the probabilities of getting assigned to different variance levels. Since the objective is to minimize the total cost, the LP gives a maximal-in-distributional-range allocation rule, where the restricted distributional range is,

$$S_\alpha = \{A \geq 0 \mid \sum_j A_{ij} = 1 \text{ for all } i, \text{ and } \sum_{ij} A_{ij}/v_j \geq 1/\alpha\}.$$

Similarly, the restricted distributional range for $n-1$ players, used to compute VCG payments is,

$$(S_\alpha)_{-i} = \{A \geq 0 \mid \sum_j A_{i'j} = 1 \text{ for all } i' \neq i, \text{ and } \sum_{i' \neq i,\ j} A_{i'j}/v_j \geq 1/\alpha\}.$$

Given a collection of reported costs, our mechanism first computes a distribution A over assignments, based on the MIDR allocation rule defined by the LP. We then pay each provider based on the VCG payment rule, in addition to some entrance reward R. Given the realized variance assignment sampled from A, we ask the providers to compute their estimates μ_i at the corresponding variance levels. Finally, we re-weight the estimates to obtain the linear combination estimator $\hat{\mu}$ with the minimum variance based on the optimal re-weighting rule in Lemma 3.3.1. The formal description of our mechanism is presented in Algorithm 1.

Algorithm 1 MIDR Mechanism for Buying Estimates

Input: Data providers' reported costs $\{c_{ij}\}$ for different variance levels $\{v_1, \ldots, v_m\}$, target variance α, initial payment R

Compute assignment and payment based on MIDR allocation rule and VCG payment rule:

$$A^* \in \arg\min_{A \in S_\alpha} \sum_i c_{ij} A_{ij} \qquad P_i = \min_{A_{-i} \in (S_\alpha)_{-i}} \left[\sum_{i' \neq i} c_{i'}(A) \right] - \sum_{i' \neq i} c_{i'}(A^*) + R$$

Let $\hat{v} = (\hat{v}_1, \ldots, \hat{v_n})$ be the realized variance assignments sampled from A^* and

$$w_i = \frac{1/\hat{v}_i}{\sum_k 1/\hat{v}_k} \qquad \text{for all } i.$$

Collect the estimates from providers $\{\mu_i\}$ based on $\hat{v}$

Output: $\sum_i w_i \mu_i$ as our estimate $\hat{\mu}$

Theorem 3.4.1. *Given n data providers with reported costs $\{c_{ij}\}$ for variance levels $\{v_j\}$ and a feasible target variance level α, Algorithm 1 is a truthful-in-expectation mechanism that selects a minimum expected cost assignment, and*

1. *for any $\varepsilon > 0$, computes an estimate $\hat{\mu}$ with variance $Var(\hat{\mu}) \leq (1 + \varepsilon)\alpha$ as long as*
$$n \geq \left(\frac{v_m}{v_1} - 1 \right) \left(\frac{1}{\varepsilon} + 1 \right),$$
2. *The mechanism is I.R. if the entrance reward $R \geq \max_i \min_j c_{ij}$.*

The properties of cost minimization and truthfulness follow from the MIDR allocation rule and VCG payments. We show the other two properties in the following subsections.

Remark 3.4.2. *To achieve a 2-approximation for the variance (i.e. $\varepsilon = 1$), it will suffice to have $n = 2v_m/v_1$ providers. Plugging in the bound in Remark 1, the meaningful range of target variance should be $v_1^2/2v_m \leq \alpha \leq v_1/2$. Note that $v_1/v_m < 1$, so this range is always non-empty.*

Variance Violation

The fractional solution we obtain could violate the variance constraint (3.7), as could the final assignment sampled from the fractional solution. Let A be an optimal

solution to the LP. Then A violates the variance constraint (3.7) by at most

$$\Delta(A) = \sum_i \sum_j A_{ij}/v_j - \sum_i 1/\sum_j A_{ij}v_j = \sum_i \left(\sum_j A_{ij}/v_j - 1/\sum_j A_{ij}v_j \right).$$

The quantity $\Delta(A)$ represents the distance between the "real" desired variance constraint and our linear relaxation. Note that for any agent who happens to receive an integral allocation, the corresponding terms in the two constraints are equal, but they may diverge for agents who have fractional allocations. To simplify and bound this quantity, we show that at any optimal fractional solution, all but at most one agent receives an integral allocation:

Lemma 3.4.3. *At any extreme point A^* of the feasible region for the LP, there are at least $n-1$ indices i such that $A_{ij} \in \{0,1\}$ for all j.*

Proof. Suppose not. Then let A be a point in the feasible set S_α such that at least two players (without loss of generality, players 1 and 2) are assigned to lotteries. In other words, each of these two players are assigned nonzero weight on at least two different variance levels. Let $a < b, k < l$ be the indices such that $A_{1a}, A_{1b}, A_{2k}, A_{2l} \notin \{0,1\}$. Let $\varepsilon > 0$ be a small enough number such that

$$A_{1a} \pm \varepsilon, A_{1b} \pm \varepsilon, A_{2k} \pm \varepsilon, A_{2l} \pm \varepsilon \in [0,1]$$

and

$$A_{1a} \pm \varepsilon', A_{1b} \pm \varepsilon', A_{2k} \pm \varepsilon', A_{2l} \pm \varepsilon' \in [0,1],$$

where $\varepsilon' = \varepsilon \left(\frac{1/v_a - 1/v_b}{1/v_k - 1/v_l} \right)$. Now consider the following two points that differ from A only in four coordinates:

$$\begin{aligned} x: &\ x_{1a} = A_{1a} + \varepsilon,\ x_{1b} = A_{1b} - \varepsilon,\ x_{2k} = A_{2k} - \varepsilon',\ \text{and } x_{2l} = A_{2l} + \varepsilon' \\ x': &\ x'_{1a} = A_{1a} - \varepsilon,\ x'_{1b} = A_{1b} + \varepsilon,\ x'_{2k} = A_{2k} + \varepsilon',\ \text{and } x'_{2l} = A_{2l} - \varepsilon'. \end{aligned}$$

Note that $A = \frac{1}{2}(x + x')$, and recall that $1/\alpha \leq \sum_i \sum_j A_{ij}/v_j$ because $A \in S_\alpha$. Furthermore,

$$\begin{aligned} \sum_i \sum_j x_{ij}/v_j &= \sum_i \sum_j A_{ij}/v_j + \varepsilon(1/v_a - 1/v_b) + \varepsilon'(1/v_l - 1/v_k) \\ &= \sum_i \sum_j A_{ij}/v_j + \varepsilon \left[1/v_a - 1/v_b + (1/v_l - 1/v_k) \frac{1/v_a - 1/v_b}{1/v_k - 1/v_l} \right] \\ &= \sum_i \sum_j A_{ij}/v_j \geq 1/\alpha. \end{aligned}$$

Similarly, $\sum_{i,j} x'_{ij}/v_j = \sum_{i,j} A_{ij}/v_j \geq 1/\alpha$, so both x and x' are in the feasible region S_α. Since A is a convex combination of x and x' that are both in S_α, we know that A cannot be an extreme point of the feasible region. □

Lemma 3.4.3 says that at any extreme point A, at least $n-1$ players have an integral assignment in A. To use this property, we will compute the solution using an (ellipsoid-based) polynomial-time LP solver from (Nemhauser et al., 1988) that always returns an optimal extreme point solution.[1] Now we can bound the variance of our aggregate estimate $\hat{\mu}$.

Lemma 3.4.4. *For any $\varepsilon > 0$, the variance of our estimate $Var(\hat{\mu}) \leq (1+\varepsilon)\alpha$, as long as*

$$n \geq \left(\frac{v_m}{v_1} - 1\right)\left(\frac{1}{\varepsilon} + 1\right).$$

Proof. Suppose that n satisfies the bound above. If the solution A is fully integral, then the variance is no more than α. Otherwise let k be the data provider receiving a lottery in A. Since for every player i with an integral assignment $\sum_j A_{ij}/v_j = \sum_j 1/\sum_j A_{ij}v_j$, we can further simplify,

$$\Delta(A) = \sum_j A_{kj}/v_j - 1/\sum_j A_{kj}v_j.$$

Then we can bound the violation of (3.7) by the final assignment $\hat{v}$:

$$\sum_j A_{kj}/v_j - 1/v_m \leq 1/v_1 - 1/v_m.$$

In other words, the resulting variance $\mathrm{Var}(\hat{\mu})$ satisfies

$$\frac{1}{\mathrm{Var}(\hat{\mu})} \geq \frac{1}{\alpha} - (\frac{1}{v_1} - \frac{1}{v_m}).$$

Since we assume $n > v_m/v_1 - 1$, we have $n/v_m - (1/v_1 - 1/v_m) > 0$. As stated earlier in Remark 1, the only interesting range of α is $v_1/n \leq \alpha \leq v_m/n$. (Recall that if $\alpha < v_1/n$, then the problem is infeasible; if $\alpha > v_m/n$, then the problem is trivial.) For the remainder of the proof, we assume $\alpha \in [v_1/n, v_m/n]$. By this assumption,

[1] The algorithm consists of two steps: first compute a sufficiently near optimal solution $\hat{A}$ using the ellipsoid algorithm; then round the solution $\hat{A}$ to an optimal extreme point solution A^* using the method of continued fractions. For more details, see (Nemhauser et al., 1988).

$1/\alpha - (1/v_1 - 1/v_m) > 0$, and so,

$$\begin{aligned}\mathrm{Var}(\hat{\mu}) \leq \frac{1}{\frac{1}{\alpha} - \frac{1}{v_1} + \frac{1}{v_m}} &= \alpha \left(\frac{1}{1 - \frac{\alpha}{v_1 v_m}(v_m - v_1)} \right) \\ &\leq \alpha \left(\frac{n}{n - (\frac{v_m}{v_1} - 1)} \right) \\ &\leq (1 + \varepsilon)\alpha,\end{aligned}$$

which recovers our lemma. □

We give an example in Section 3.5 showing that this analysis cannot be improved, and we do need $n = \Omega(v_m/v_1)$ to approximately satisfy the target variance constraint.

Individual Rationality

In order to ensure individual rationality, we need to set the entrance reward R large enough, so that for each player i, $R + P_i - c_i \geq 0$, where c_i denotes the cost for player i to provide its assigned estimate. To reason about the payment player i gets, we need to compute the following two costs C_1 and C_2, for all players except i. Let A^* be the optimal (fractional) solution for our LP, and $\hat{v}_i$ be the expected variance level assigned to player i: $\hat{v}_i = \sum_j A^*_{ij} v_j$. Let OPT denote the optimal min-cost value of the LP, and C_1 denote the total cost for all players except i in A^*:

$$\begin{aligned} C_1 = \min & \sum_{k \neq i, j} A_{kj} c_{kj} \\ \text{s.t.} & \sum_{k \neq i, j} A_{kj}/v_j \geq 1/\alpha - 1/\hat{v}_i \\ & \sum_j A_{kj} = 1 \text{ for all } k \\ & A_{kj} \geq 0 \text{ for all } (k, j). \end{aligned}$$

Let C_2 be the minimum cost had we removed agent i from the input:

$$\begin{aligned} C_2 = \min & \sum_{k \neq i, j} A_{kj} c_{kj} \\ \text{s.t.} & \sum_{k \neq i, j} A_{kj}/v_j \geq 1/\alpha \\ & \sum_j A_{kj} = 1 \text{ for all } k \\ & A_{kj} \geq 0 \text{ for all } (k, j). \end{aligned}$$

The VCG payment given to player i in Algorithm 1 is $P_i = C_2 - C_1$. Note that since the second LP is more constrained than the first, we know $C_2 \geq C_1$ and the payment is always non-negative. We can write down the expected utility of player i:

$$R + P_i - c_i = R + C_2 - C_1 - c_i = R + C_2 - \text{OPT}\,.$$

Lemma 3.4.5. *The mechanism in Algorithm 1 is individually rational if the entrance reward satisfies*

$$R \geq \max_i \min_j c_{ij}.$$

Proof. Let A_{-i} be the optimal assignment for the second program (with optimal objective value at C_2). Now let's add back player i to the problem, and construct an assignment A such that $A = (A_i, A_{-i})$, where A_i assigns player i to the assignment with minimum cost ($\min_j c_{ij}$).

Note that A is a feasible solution to our original problem since A_{-i} already satisfies the variance constraint. It follows that the cost given by A is at least as large as OPT, the optimal solution: $\text{OPT} \leq C_2 + \min_j c_{ij}$.

Therefore, as long as $R \geq \min_j c_{ij}$ for each player i, we have individual rationality. □

We give an example in Section 3.5 to show that this bound is tight. In particular, our example shows that without an entrance reward, the individual rationality constraint could be violated by up to $\min_j c_{ij}$ for each player i.

Remark 3.4.6. *Let $c_{min} = \max_i \min_j c_{ij}$. If costs are drawn from a known distribution, the analyst can set R to ensure that with high probability, all players have $c_{min} \leq R$. If c_{min} is unbounded, it is clear that no Groves mechanism[2] can be individually rational for all players in this setting. The Green-Laffont-Holmström theorem (Jerry R. Green et al., 1977; Holmström, 1979) shows that under certain technical conditions, any mechanism which is dominant strategy incentive compatible and maximizes welfare must be a Groves mechanism. Thus without additional information on the players' costs, we should not hope to satisfy individual rationality for all players while still achieving our other desiderata.*

[2] A *Groves mechanism* is one which selects the welfare maximizing outcome, and each player's payment is his externality plus an amount that is independent of his report. In particular, the payments induced by any Groves mechanism to a player i are shifts of the payments induced by our mechanism, by an amount that is independent of player i's report. Therefore, by reporting a large enough value of c_{min}, individual rationality can always be violated by a Groves mechanism.

3.5 Tightness of Our Bounds

Example for Variance Violation Bound

Consider an example where there are only two options of variance levels, v_1 and v_2, and we set the target variance $\alpha = \frac{v_1 v_2}{n v_1 + \delta(v_2 - v_1)}$. Suppose the reported costs $c_{i1} = t_1$ and $c_{i2} = t_2$ for each player $i \in [n-1]$, and $c_{n1} < t_1$ and $c_{n2} = t_2$ for player n. We also assume that $t_2 < t_1$. Let A denote the assignment such that $A_{i1} = 0$ and $A_{i2} = 1$ for each $i \in [n-1]$, and $A_{n1} = \delta \in (0,1)$ and $A_{n2} = 1 - \delta$. That is, the assignment gives v_2 to the first $(n-1)$ players, and give a lottery between the two levels to player n. Note that

$$\frac{1}{\alpha} = \frac{n-\delta}{v_2} + \frac{\delta}{v_1}.$$

We know that the fractional solution A exactly satisfies the variance constraint (3.7), and is also the optimal min-cost solution. Therefore, with probability $(1-\delta)$, the realized variance satisfies,

$$\frac{1}{\mathrm{Var}(\hat{\mu})} = \frac{n}{v_2} = \frac{n-\delta}{v_2} + \frac{\delta}{v_1} + \frac{\delta}{v_2} - \frac{\delta}{v_1} = \frac{1}{\alpha} - \delta\left(\frac{1}{v_1} - \frac{1}{v_2}\right) > 0.$$

It follows that

$$\mathrm{Var}(\hat{\mu}) = \frac{\alpha}{1 - \alpha\delta(\frac{1}{v_1} - \frac{1}{v_2})} = \alpha\left(1 - \frac{\delta\left(\frac{1}{v_1} - \frac{1}{v_2}\right)}{\frac{n}{v_2} + \delta\left(\frac{1}{v_1} - \frac{1}{v_2}\right)}\right)^{-1} = \alpha\left(1 + \frac{\delta\left(\frac{v_2}{v_1} - 1\right)}{n}\right).$$

If we want $\frac{\delta(v_2/v_1 - 1)}{n} \le \varepsilon$, we would need to have the number of providers

$$n \ge \left(\frac{v_2}{v_1} - 1\right)\frac{\delta}{\varepsilon}.$$

For δ close to 1 and constant ε, the number of providers we need does scale with v_2/v_1, which shows that the $\Omega(v_m/v_1)$ for n is essentially tight.

Example for Entrance Reward Bound

Consider an example with two providers, two possible variance levels v_1, v_2 such that $v_2 = 2v_1$, and target variance $\alpha = v_1$. Suppose the costs satisfy $c_{11} = c_{21} = t$ and $c_{12} = c_{22} = t - \varepsilon$ for some $\varepsilon > 0$. Since we need to an estimate from each provider, the optimal solution is to assign v_2 to both players, which yields cost $\mathrm{OPT} = 2t - 2\varepsilon$. Now suppose we remove any provider from the mechanism. Then we would assign the remaining provider to v_1, which yield cost $C_2 = t$. Therefore, the utility for each

provider is

$$\begin{aligned} R + C_2 - \mathrm{OPT} &= R + t - 2(t + \varepsilon) \\ &= R + 2\varepsilon - t \\ &= R + \varepsilon - t. \end{aligned}$$

In order to ensure non-negative utility, we need $R \geq t - \varepsilon$. Note that the right hand side tends to $\max_i \min_j c_{ij}$ when ε tends to 0. Therefore, the bound in Lemma 3.4.5 is tight.

Chapter 4

ACQUISITION OF COSTLY DATA FOR STATISTICAL ESTIMATION, REVISITED

Yiling Chen, Nicole Immorlica, Brendan Lucier, Vasilis Syrgkanis, and Juba Ziani (2018). "Optimal Data Acquisition for Statistical Estimation". In: *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 27–44. **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** DOI: 10.1145/3219166.3219195. URL: http://doi.acm.org/10.1145/3219166.3219195.

4.1 Introduction

In this chapter, we further the study of mechanism design for buying verifiable data from a population in order to estimate a statistic of interest, such as the expected value of some function of the underlying data. We assume each individual has a private cost, or dis-utility, for revealing his or her sensitive data to the analyst. Importantly, *this cost may be correlated with the private data*. For example, individuals with HIV are expected to have a higher cost of revealing their data than people of a healthy weight than their healthy counterparts. This is a departure from Chapter 3, where we implicitely make the assumption that data points and independent of costs (in particular, in Chapter 3, an agent's data point is always unbiased, independently of the agent's cost).

The analyst has a fixed budget for buying data. The analyst does not know the distribution of the data: properties of the distribution is what she is trying to learn from the data samples, therefore it is important that she uses the data she collects to learn it rather than using an inaccurate prior distribution (for example, the analyst may have a prior on weight distribution within a population from DMV records or previous surveys, but such a prior may be erroneous if people do not accurately report their weights). However, we do assume the analyst *has a prior for the marginal distribution of costs*, and that she estimates how much a survey may cost her as a function of said prior.[1]

[1]This prior could come from similar past exercises. Alternatively, when no prior is known, the analyst can allocate a fraction of his budget to buying data for the sake of learning this distribution of costs. In this chapter, we follow prior work (Roth et al., 2012) and assume that a prior distribution is known, instead of focusing on how one might learn the distribution of costs. Chen et al. (2018c)

The analyst would like to buy data subject to her budget, then use that data to obtain an unbiased estimator for the statistic of interest. To this end, the analyst posts a menu of probability-price pairs. Each individual i with cost c selects a pair (A, P) from the menu, at which point the analyst buys the data with probability A at price P. The expected utility of the individual is thus $(P - c)A$.[2] To form an estimate based on this collected data, we assume the analyst uses *inverse propensity scoring*, pioneered by (Horvitz et al., 1952). This is the unique unbiased linear estimator; it works by upweighting the data from individual i by the inverse of his/her selection probability, $1/A$.

The Horvitz-Thompson estimator always generates an unbiased estimate of the statistic being measured, regardless of the price menu. However, the precision of the estimator, as measured by the variance or mean-squared error of the estimate, depends on the menu of probability-price pairs offered to each individual. For example, offering a high price would generate data samples with low bias (since many individuals would accept such an offer), but the budget would limit the number of samples. Offering low prices allows the mechanism to collect more samples, but these would be more heavily biased, requiring more aggressive correction which introduces additional noise. The goal of the analyst is to strike a balance between these forces and post a menu that minimizes the variance of her estimate in the worst-case over all possible joint distributions of the data and cost consistent with the cost prior. We note that this problem setting was first studied by Roth et al. (2012), who characterized an approximately optimal mechanism for moment estimation. The current chapter aims to generalize the results of the work of Roth et al. (2012), by giving an exactly optimal mechanism to this problem, exhibiting new structure in the optimal solution, and extending mechanism design for data acquisition to parameter estimation via regression.

Note that unlike Chapter 3 where we aim to optimize the budget under a variance constraint, we formulate the problem in this chapter as minimizing variance under a budget constraint, for simplicity of exposition; such formulations are equivalent, in the sense that they lead to the same trade-off curve between variance and budget.

---

consider an extension of this chapter that does away with this distributional prior on the costs.

[2] As we show, this menu-based formulation is fully general and captures arbitrary data-collection mechanisms.

Summary of contributions

Our main contribution comes in the form of an exact solution for the optimal menu, as discussed in Section 4.3. As one would expect, if the budget is large, the optimal menu offers to buy, with probability 1, all data at a cost equal to the maximum cost in the population. If the budget is small, the optimal menu buys data from an individual with probability inversely proportional to the square root of their cost.[3] Interestingly, in intermediate regimes, we show the optimal menu employs pooling: for all individuals with sufficiently low private cost, it buys their data with equal probability; for the remaining high cost agents, it buys their data with probability inversely proportional to the square root of their costs. Revisiting the example of estimating the weight of a population of size n, our scheme suggests the following solution. Imagine the costs are $0, 4, 8$ with probability $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$, and the total budget of the analyst is $B = 7n$. The analyst brings a scale to a public location and posts the following menu of pairs of allocation probability and price: $\{(1, \frac{36}{5}), (\frac{4}{5}, 8)\}$. A simple calculation shows that individuals with cost 0 or 4 will pick the first menu option: stepping on the scale and having their weight recorded with probability 1, and receiving a payment of $\frac{36}{5}$ dollars. Individuals with cost 8 will pick the second menu option; if they are selected to step on the scale, which happens with probability $\frac{4}{5}$, the analyst records their weight scaled by a factor of $\frac{5}{4}$. This scaling is precisely the upweighting from inverse propensity scoring. In expectation over the population, the analyst spends exactly his budget $7n$. The estimate is the average of the scaled weights.

We show how to extend our approach beyond moment estimation to the common task of (multi-dimensional) linear regression, in Section 4.4. In this regression problem, an individual's data includes both features (which are assumed to be insensitive or publicly available) and outcomes (which may be sensitive). The analyst's goal is to estimate the linear regression coefficients that relate the outcomes to the features. We make the assumption that an individual's cost is independent of her features, but may be arbitrarily correlated with her outcome. For example, the goal might be to regress a health outcome (such as severity of a disease) on demographic information. In this case, we might imagine that an agent incurs no cost for reporting his age, height, or gender, but his cost might be highly correlated with his realized health outcome.

[3] Of course, the individual is him/herself selecting the menu option and so the use of an active verb in this context is perhaps a bit misleading. What we mean here is that, given his/her incentives based on his/her private cost, the choice the individual selects is one that buys his/her data with probability inversely proportional to the square root of his/her cost.

In such a setting, we show that the asymptotically optimal allocation rule, given a fixed average budget per agent as the number of agent grows large, can be calculated efficiently and exhibits a pooling region as before. However, unlike for moment estimation, agents with intermediate costs can also be pooled together.

Our techniques rely on i) reducing the mechanism design problem to an optimization problem through the classical notion of virtual costs, then ii) reducing the problem of optimizing the worst-case variance to that of finding an equilibrium of a zero-sum game between the analyst and an adversary. The adversary's goal is to pick a distribution of data, conditional on agents' costs, that maximizes the variance of the analyst's estimator. We then characterize such an equilibrium through the optimality conditions for convex optimization described in (Boyd et al., 2004).

Related work

This chapter also belongs to the line of work on data acquisition where agents incur a cost for producing or revealing their data point, and a mechanism designer needs to balance budget and accuracy of the final estimate. Prior papers by Roth et al. (2012) and Abernethy et al. (2015) are closest to the setting of the current chapter. Similarly to our work, both Roth et al. (2012) and Abernethy et al. (2015) consider an analyst's problem of purchasing data from individuals with private costs subject to a budget constraint, allow the cost to be correlated with the value of data, and assume that individuals cannot fabricate their data. Roth et al. (2012) aim at obtaining an optimal unbiased estimator with minimum worst-case variance for population mean, while their mechanism achieves optimality only approximately: instead of the actual worst-case variance, a bound on the worst-case variance is minimized. While our setting is identical to that of Roth et al. (2012), our work precisely minimizes worst-case variance (under a regularity assumption on the cost distribution), and our main contribution is to exhibit the structure of the optimal mechanism, as well as to extend our results to broader classes of statistical inference, moment estimation and linear regression. In particular, compared to (Roth et al., 2012), our solution exhibits new structure in the form of a pooling region for low cost agents; i.e., the optimal mechanism pools agents with the lowest costs together and treats them identically. Such structure does not arise in (Roth et al., 2012) under a regularity assumption on the cost distribution. Abernethy et al. (2015) consider general supervised learning. They do not seek to achieve a notion of optimality; instead, they take a learning-theoretic approach and design mechanisms to obtain learning guarantees (risk bounds).

4.2 Model and Preliminaries

Survey Mechanisms: There is a population of n agents. Each agent i has a private pair (z_i, c_i) (the agent's type), where $z_i \in \mathcal{Z}$ is a data point and $c_i > 0$ is a cost. We think of c_i as the cost or dis-utility agent i incurs by releasing her data z_i. The pair is drawn from a distribution $\mathcal{D}$, unknown to the mechanism designer. We denote with $\mathcal{F}$ the CDF of the marginal distribution of costs,[4] supported on a set C. We assume that $\mathcal{F}$ and the support of the data points, $\mathcal{Z}$, are known. However, the joint distribution $\mathcal{D}$ of data and costs is unknown.

A *survey mechanism* is defined by an allocation rule $A : C \to [0, 1]$ and a payment rule $P : C \to \mathbb{R}$, and works as follows. Each agent i arrives at the mechanism in sequence and reports a cost $\hat{c}_i$. The mechanism chooses to buy the agent's data with probability $A(\hat{c}_i)$. If the mechanism buys the data, then it learns the value of z_i (i.e., agents cannot misreport their data) and pays the agent $P(\hat{c}_i)$. Otherwise the data point is not learned and no payment is made.

We assume agents have quasi-linear utilities, so that the utility enjoyed by agent i when reporting $\hat{c}_i$ is

$$u(\hat{c}_i; c_i) = (P(\hat{c}_i) - c_i) \cdot A(\hat{c}_i). \tag{4.1}$$

We will restrict attention to survey mechanisms that are truthful and individually rational, as defined in Chapter 2. In the context of the current chapter, these definitions can be rewritten as follows:

Definition 4.2.1 (Truthful and Individually Rational - TIR)**.** *A survey mechanism is truthful if for any cost c it is in the agent's best interest to report their true cost, i.e. for any report $\hat{c}$:*

$$u(c; c) \geq u(\hat{c}; c). \tag{4.2}$$

It is individually rational if, e. for any cost $c \in C$, $P(c) \geq c$.

We assume that the mechanism is constrained in the amount of payment it can make to the agents. We will formally define this as an expected budget constraint for the survey mechanism.

Definition 4.2.2 (Expected Budget Constraint)**.** *A mechanism respects a budget constraint B if:*

$$n \cdot \mathbb{E}_{c \sim \mathcal{F}}[P(c) \cdot A(c)] \leq B. \tag{4.3}$$

[4] Throughout the text we will use the CDF to refer to the distribution itself.

Estimators: The designer (or *data analyst*) wishes to use the survey mechanism to estimate some parameter $\theta \in \mathbb{R}$ of the marginal distribution of data points. For example, it might be that $\mathcal{Z} = \mathbb{R}$ and θ is the mean of the distribution over data points in the population. To this end, the designer will apply an *estimator* to the collection of data points S elicited by the survey mechanism. We will write $\hat{\theta}_S$ for the estimator used. Note that the value of the estimator $\hat{\theta}_S$ depends on the sample S, but might also depend on the distribution of costs $\mathcal{F}$ and the survey mechanism. Due to the randomness inherent in the survey mechanism (both in the choice of data points sampled and the values of those samples), we think of $\hat{\theta}_S$ as a random variable, drawn from a distribution $\mathcal{T}(\mathcal{D}, A)$. We will focus exclusively on *unbiased* estimators.

Definition 4.2.3 (Unbiased Estimator)**.** *Given an allocation function A, an estimator $\hat{\theta}_S$ for θ is* unbiased *if for any instantiation of the true distribution $\mathcal{D}$ its expected value is equal to θ:*

$$\mathbb{E}_{\hat{\theta}_S \sim \mathcal{T}(\mathcal{D},A)} \left[\hat{\theta}_S\right] = \theta. \tag{4.4}$$

Given a fixed choice of estimator, the mechanism designer wants to construct the survey mechanism to minimize the variance (finite sample or asymptotic as the population grows) of that estimator. Since the designer does not know the distribution $\mathcal{D}$, we will work with the worst-case variance over all instantiations of $\mathcal{D}$ that are consistent with the cost marginal $\mathcal{F}$.

Definition 4.2.4 (Worst-Case Variance)**.** *Given an allocation function A and an instance of the true distribution $\mathcal{D}$, the* variance *of an estimator $\hat{\theta}_S$ is defined as:*

$$Var(\hat{\theta}_S; \mathcal{D}, A) = \mathbb{E}_{\hat{\theta}_S \sim \mathcal{T}(\mathcal{D},A)} \left[\left(\hat{\theta}_S - \mathbb{E}\left[\hat{\theta}_S\right]\right)^2\right]. \tag{4.5}$$

The worst-case variance *of $\hat{\theta}_S$ is*

$$Var^*(\hat{\theta}_S; \mathcal{F}, A) = \sup_{\mathcal{D} \text{ consistent with } \mathcal{F}} Var(\hat{\theta}_S; \mathcal{D}, A). \tag{4.6}$$

We are now ready to formally define the mechanism design problem faced by the data analyst.

Definition 4.2.5 (Analyst's Mechanism Design Problem)**.** *Given an estimator $\hat{\theta}_S$ and cost distribution $\mathcal{F}$, the goal of the designer is to design an allocation rule A and payment rule P so as to minimize worst-case variance subject to the truthfulness,*

individual rationality and budget constraints:

$$\begin{aligned} &\inf_{A,P} \ Var^*(\hat{\theta}_S; \mathcal{F}, A) \\ &\text{s.t. } n \cdot \mathbb{E}_{c\sim\mathcal{F}}[P(c) \cdot A(c)] \leq B \\ &\qquad A, P \text{ define a TIR mechanism.} \end{aligned} \tag{4.7}$$

Remark 4.2.6 (Implementing Surveys as Posted Menus.)**.** *The formulation above describes surveys as direct-revelation mechanisms, where agents report costs. We note that an equivalent indirect implementation might be more natural: a* posted menu survey *offers each agent a menu of (price, probability) pairs* $(P_1, A_1), \ldots, (P_k, A_k)$*. If the agent chooses* (P_m, A_m) *then their data is elicited with probability* A_m*, in which case they are paid* P_m*. Each agent can choose the item that maximizes their expected utility, i.e.,* $\text{argmax}_{m\in[k]} \ (P_m - c) \cdot A_m$*. By the well-known* taxation principle*, any survey mechanism can be implemented as a posted menu survey, and the number of menu items required is at most the size of the support of the cost distribution.*

Reducing Mechanism Design to Optimization

We begin by reducing the mechanism design problem to a simpler full-information optimization problem where the designer knows the private cost of each player and can acquire their data by paying them exactly that cost. However, the designer is constrained to using *monotone* allocation rules, in which players with higher costs have weakly lower probability of being chosen.

Definition 4.2.7 (Analyst's Optimization Problem)**.** *Given an estimator* $\hat{\theta}_S$ *and cost distribution* $\mathcal{F}$*, the optimization version of the designer's problem is to find a non-increasing allocation rule A that minimizes worst-case variance subject to the budget constraint, assuming agents are paid their cost:*

$$\begin{aligned} &\inf_{A} \ Var^*(\hat{\theta}_S; \mathcal{F}, A) \\ &\text{s.t. } n \cdot \mathbb{E}_{c\sim\mathcal{F}}[c \cdot A(c)] \leq B \\ &\qquad A \text{ is monotone non-increasing.} \end{aligned} \tag{4.8}$$

The mechanism design problem in Definition 4.2.5 reduces to the optimization problem given by Definition 4.2.7, albeit with a transformation of costs to *virtual cost*.

Definition 4.2.8 (Virtual Costs and Regular Distributions)**.** *If* $\mathcal{F}$ *is continuous and admits a density* f *then define the virtual cost function as* $\phi(c) = c + \frac{\mathcal{F}(c)}{f(c)}$*. For*

discrete $\mathcal{F}$ with support $C = \{c_1, \ldots, c_{|C|}\}$ and PDF f, we define the virtual cost function as: $\phi(c_t) = c_t + \frac{c_t - c_{t-1}}{f(c_t)}\mathcal{F}(c_{t-1})$, with $c_0 = 0$. We also denote with $\phi(\mathcal{F})$ the distribution of virtual costs; i.e., the distribution created by first drawing c from $\mathcal{F}$ and then mapping it to $\phi(c)$. A distribution $\mathcal{F}$ is regular if the virtual cost function is increasing.

When $\mathcal{F}$ is twice-continuously differentiable, $\mathcal{F}$ is regular if and only if $\mathcal{F}(c)f'(c) < 2f(c)^2$ for all $c \in C$. Importantly, in this case, the allocation rule described by Roth et al. (2012) is monotone strictly decreasing in c and does not exhibit a pooling region at low-cost as our solution does. The following is an analogue of Myerson (1981)'s reduction of mechanism design to virtual welfare maximization, adapted to the survey design setting.

Lemma 4.2.9. *If the distribution of costs $\mathcal{F}$ is regular, then solving the Analyst's Mechanism Design Problem reduces to solving the Analyst's Optimization Problem for distribution of costs $\phi(\mathcal{F})$.*

Proof. The proof is given in Section 4.5. □

Unbiased Estimation and Inverse Propensity Scoring

We now describe a class of estimators $\hat{\theta}_S$ that we will focus on for the remainder of the chapter. Note that simply calculating the quantity of interest, θ, on the sampled data points can lead to bias, due to the potential correlation between costs and data. For instance, suppose that $z \in \mathbb{R}$ and the goal is to estimate the mean of the distribution of z. A natural estimator is the average of the collected data: $\hat{\theta}_S = \frac{1}{|S|}\sum_{i \in S} z_i$. However, if players with lower z tend to have lower cost, and are therefore selected with higher probability by the analyst, then this estimator will consistently underestimate the true mean.

This problem can be addressed using *inverse propensity scoring* (IPS), pioneered by Horvitz et al. (1952). The idea is to recover unbiasedness by weighting each data point by the inverse of the probability of observing it.

This IPS approach can be applied to any parameter estimation problem where the parameter of interest is the expected value of an arbitrary *moment function* $m : \mathcal{Z} \to \mathbb{R}$.

Definition 4.2.10 (Horvitz-Thompson Estimator)**.** *The Horvitz-Thompson estimator for the case when the parameter of interest is the expected value of a (moment)*

function $m : \mathcal{Z} \to \mathbb{R}$ is defined as:

$$\hat{\theta}_S = \frac{1}{n} \sum_{i \in [n]} \frac{m(z_i) \cdot 1\{i \in S\}}{A(c_i)}. \tag{4.9}$$

The Horvitz-Thompson estimator is the unique unbiased estimator that is a linear function of the observations $m(z_i)$ (Roth et al., 2012). It is therefore without loss of generality to focus on this estimator if one restricts to unbiased linear estimators.[5]

IPS beyond moment estimation: We defined the Horvitz-Thompson estimator with respect to moment estimation problems, $\theta = \mathbb{E}[m(z)]$. As it turns out, this approach to unbiased estimation extends even beyond the moment estimation problem to parameter estimation problems defined as the solution to a system of moment equations $\mathbb{E}[m(z;\theta)] = 0$ or parameters defined as the minima of a moment function $\operatorname{argmin}_\theta \mathbb{E}[m(z;\theta)]$. We defer this discussion to Section 4.4.

4.3 Estimating Moments of a Data Distribution

In this section we consider the case where the analyst's goal is to estimate the (one-dimensional) mean of a given moment function of the distribution. That is, there is some function $m : C \to [0, 1]$ such that both 0 and 1 are in the support of random variable $m(z)$, and the goal of the analyst is to estimate $\theta = \mathbb{E}[m(z)]$.[6] We assume that $\hat{\theta}_S$, the estimator being applied, is the Horvitz-Thompson estimator given in Definition 4.2.10. For multi-dimensional moment estimation, please refer to the full version this work (Chen et al., 2018a).

For convenience we will assume that the agents' types/cost distribution $\mathcal{F}$ have finite support, say $C = \{c_1, \ldots, c_{|C|}\}$ with $c_1 < \ldots < c_{|C|}$; we abuse notations and let c_t be the cost of an agent of type t. Note that we relax this finite support assumption in the full version (Chen et al., 2018a). Write $\pi_t = f(c_t)$ for the probability of cost c_t in $\mathcal{F}$. Also, for a given allocation rule A, we will write $A_t = A(c_t)$ for convenience. That is, we can interpret an allocation rule A as a vector of $|C|$ values $A_1, \ldots, A_{|C|}$. For further convenience, we will write $q_t = \Pr[m(z) = 1 | c_t]$. This is the probability

[5] We note that we have assumed, for convenience, that $A(c_i) > 0$ for all $i \in [n]$ in the expression of this estimator, for it to be unbiased and well-defined. It is easy to see from the expression for the variance given in Section 4.3 that the variance-minimizing allocation rule will indeed be non-zero for each cost.

[6] Observe that it is easy to deal with the more general case of $m(z) \in [a, b]$ by a simple linear translation, i.e., estimate $\tilde{m}(z) = \frac{m(z)-a}{b-a}$ instead, which is in $[0, 1]$ and then translate the estimator back to recover $m(z)$.

that the moment takes on its maximum value when the cost is c_t. Finally, we will assume that the distribution of costs is regular.

Our goal is to address the analyst's mechanism design problem for this restricted setting. By Lemma 4.2.9 it suffices to solve the analyst's optimization problem. We start by characterizing the worst-case variance for this setting.

Lemma 4.3.1. *The worst-case variance of the Horvitz-Thompson estimator of a moment $m : C \to [0,1]$, given cost distribution $\mathcal{F}$ and allocation rule A, is:*

$$n \cdot Var^*(\hat{\theta}_S; \mathcal{F}, A) = \sup_{q \in [0,1]^{|C|}} \sum_{t=1}^{|C|} \pi_t \cdot \frac{q_t}{A_t} - \left(\sum_{t=1}^{|C|} \pi_t \cdot q_t \right)^2. \tag{4.10}$$

Proof. For any distribution $\mathcal{D}$, observe that the Horvitz-Thompson estimator can be written as the sum of n i.i.d. random variables each with a variance:

$$\begin{aligned} \text{Var} &= \mathbb{E}\left[\left(\frac{m(z_i) \cdot 1\{i \in S\}}{A(c_i)} \right)^2 \right] - \mathbb{E}\left[\frac{m(z_i) \cdot 1\{i \in S\}}{A(c_i)} \right]^2 \\ &= \sum_{t=1}^{|C|} \pi_t \cdot \frac{\mathbb{E}\left[m(z)^2 | c_t \right]}{A_t} - \mathbb{E}[m(z)]^2. \end{aligned}$$

Therefore, the variance of the estimator is $\frac{\text{Var}}{n}$. Observe that conditional on any value c, the worst-case distribution $\mathcal{D}$, will assign positive mass only to values $z \in \mathcal{Z}$ such that $m(z) \in \{0,1\}$. This is because any other conditional distribution can be altered by a mean-preserving spread, pushing all the mass on these values, while preserving the conditional mean $\mathbb{E}[m(z)|c]$. This would strictly increase the latter variance. Thus we can assume without loss of generality that $m(z) \in \{0,1\}$, in which case $m(z)^2 = m(z)$ and $\mathbb{E}[m(z)|c] = \Pr[m(z) = 1|c]$. Recall that $q_t = \Pr[m(z) = 1|c_t]$. Then we can simplify the variance as:

$$n \cdot \text{Var}(\hat{\theta}_S; \mathcal{D}, A) = \sum_{t=1}^{|C|} \pi_t \cdot \frac{\mathbb{E}[m(z)|c_t]}{A_t} - \mathbb{E}[m(z)]^2 = \sum_{t=1}^{|C|} \pi_t \cdot \frac{q_t}{A_t} - \left(\sum_{t=1}^{|C|} \pi_t \cdot q_t \right)^2.$$

The theorem follows since the worst-case variance is a supremum over all possible consistent distributions, and equivalently a supremum over conditional probabilities $q : [0,1]^{|C|}$. □

Given the above characterization of the variance of the estimator, we can greatly simplify the analyst's optimization problem for this setting. Indeed, it suffices to find the allocation rule $A \in (0,1]^{|C|}$ that minimizes (4.10), subject to A being monotone non-decreasing and satisfying the expected budget constraint.

Characterization of the Optimal Allocation Rule

We are now ready to solve the analyst's optimization problem for moment estimation. In this and all following sections, we denote $\bar{B} = \frac{B}{n}$ for simplicity of notations, and refer to $\bar{B}$ as the "average budget per agent". Note that different agents with different costs may be allocated different fractions of the total budget B that in general do not coincide with $\bar{B}$. We remark that if $\bar{B}$ is larger than the expected cost of an agent, then it is feasible (and hence optimal) for the analyst to set the allocation rule to pick any type with probability 1. We therefore assume without loss of generality that $\mathbb{E}[c] > \bar{B}$.

Our analysis is based on an equilibrium characterization, where we view the analyst choosing A and the adversary choosing z as playing a zero-sum game and solve for its equilibria. We first present the characterization and some qualitative implications and then present an outline of our proof. We defer the full details of the proof to Section 4.6.

Theorem 4.3.2 (Optimal Allocation for Moment Estimation)**.** *The optimal allocation rule A is determined by two constants $\bar{A}$ and $t^* \in \{0, \dots, |C|\}$ such that:*

$$A_t = \begin{cases} \bar{A} & \text{if } t \le t^* \\ \frac{\alpha}{\sqrt{c_t}} & \text{o.w.} \end{cases} \tag{4.11}$$

with α uniquely determined such that the budget constraint is binding.[7] *Moreover, the parameters $\bar{A}$ and t^* can be computed in time $O(\log(|C|))$.*

The parameters $\bar{A}$ and t^* in Theorem 4.3.2 are explicitly derived in closed form in Section 4.6. For instance, when $\bar{B} \ge \frac{c_{|C|}}{2}$, then $t^* = |C|$ and $A_t = \bar{A} = \min\left\{1, \frac{\bar{B}}{\mathbb{E}[c]}\right\}$ for all t. When $\bar{B} \le \frac{\sqrt{c_1}\mathbb{E}[\sqrt{c}]}{2}$ then $t^* = 0$ and $A_t = \frac{\bar{B}}{\sqrt{c_t}\mathbb{E}[\sqrt{c}]}$. In fact, it can be shown (see full proof) that in this latter case, the worst-case distribution is given by $q = 1$. In particular, in this restricted case, the approximation of (Roth et al., 2012) is in fact optimal, and indeed our allocation rule is expressing the solution of (Roth et al., 2012) as a posted menu for a discrete, regular distribution of costs. In every other case, $q \neq 1$ and our solution differs from that of (Roth et al., 2012), exhibiting a pooling region for low-cost agents. More generally, the computational part of Theorem 4.3.2 follows by performing binary search over the support of $\mathcal{F}$, which can be done in $O(\log(|C|))$ time.

[7] The explicit form of this is $\alpha = \frac{\bar{B} - \bar{A}\mathbb{E}[c \cdot 1\{c \le c_{t^*}\}]}{\mathbb{E}[\sqrt{c} \cdot 1\{c > c_{t^*}\}]}$.

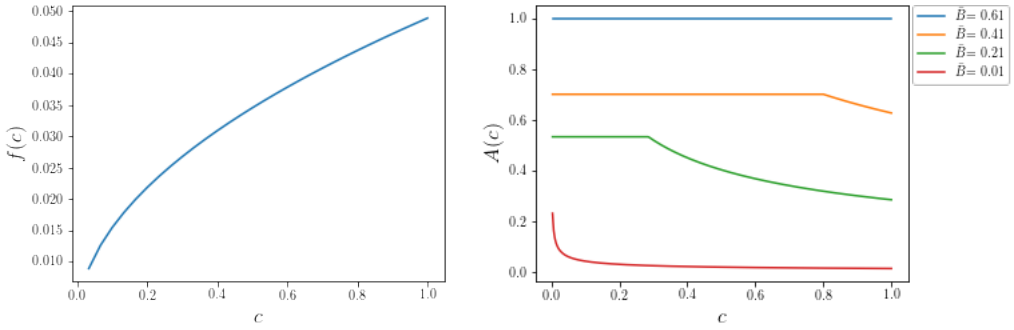

Figure 4.1: The pdf (left) of a distribution of costs and the corresponding optimal allocation rule for varying levels of per-agent budget (right). Note that for sufficiently large budgets, a flat pooling region forms for agents with low costs.

We note that the optimal rule essentially allocates to each agent inversely proportionally to the square root of their cost, but may also "pool" the allocation probability for agents at the lower end of the cost distribution. See Figure 4.1 for examples of optimal solutions.

The proof of Theorem 4.3.2 appears in Section 4.6. The main idea is to view the optimization problem as a zero-sum game between the analyst who designs the allocation rule A, and an adversary who designs q so as to maximize the variance of the estimate. We show how to compute an equilibrium (A^*, z^*) of this zero-sum game via Lagrangian and KKT conditions, and then note that the obtained A^* must in fact be an optimal allocation rule for worst-case variance.

The analysis above applied to a discrete cost distribution over a finite support of possible costs. We show how to extend this analysis to a continuous distribution below, noting that the continuous variant of the Optimization Problem for Moment Estimation can be derived by taking the limit over finer and finer discrete approximations of the cost distribution, in the full version of this work.

4.4 Multi-dimensional Parameter Estimation via Linear Regression

In this section, we extend beyond moment estimation to a multi-dimensional linear regression task (we discuss the non-linear case in the full version of (Chen et al., 2018a)). For this setting we will impose additional structure on the data held by each agent. Each agent's private information consists of a feature vector $x_i \in \mathbb{R}$, an outcome value $y_i \in \mathbb{R}$, and a residual value $\varepsilon_i \in \mathbb{R}$, that are i.i.d among agents. Each agent also has a cost c_i. The data is generated in the following way: first, x_i is drawn from an unknown distribution $\mathcal{X}$. Then, independently from x_i, the pair (c_i, ε_i) is

drawn from a joint distribution $\mathcal{D}$ over $\mathbb{R}^2$. The marginal distribution over costs, $\mathcal{D}_c$, is known to the designer, but not the full joint distribution $\mathcal{D}$. Then y_i is defined to be

$$y_i = x_i^\top \theta^* + \varepsilon_i, \tag{4.12}$$

where $\theta^* \in \Theta$ with Θ a compact subset of $\mathbb{R}^d$. We further require that θ^* is in the interior of Θ. We write $\mathcal{D}_\varepsilon$ for the marginal distribution over ε_i, which is supported on some bounded range $[L, U]$ and has mean 0. (In particular, $L \leq 0 \leq U$.) We remark that it may be the case, however, that $\mathbb{E}[\varepsilon_i | c_i] \neq 0$.

When a survey mechanism buys data from agent i, the pair (x_i, y_i) is revealed. Crucially, the value of ε_i is not revealed to the survey mechanism. The goal of the designer is to estimate the parameter vector θ^*.

Note that the single-dimensional moment estimation problem from Section 4.3 is a special case of linear regression. Indeed, consider setting $d = 1$, $\varepsilon_i = m(z_i) - \mathbb{E}[m(z_i)]$ for each i, $\theta^* = \mathbb{E}[m(z_i)]$, and x_i to be the constant -1. Then, when the survey mechanism purchases data from agent i, it learns $y_i = m(z_i)$, and estimating θ^* is equivalent to estimating the expected value of $m(z_i)$.

More generally, one can interpret x_i as a vector of publicly-verifiable information about agent i, which might influence a (possibly sensitive) outcome y_i. For example, x_i might consist of demographic information, and y_i might indicate the severity of a medical condition. The coefficient vector θ^* describes the average effect of each feature on the outcome, over the entire population. Under this interpretation, ε_i is the residual agent-specific component of the outcome, beyond what can be accounted for by the agent's features. We can interpret the independence of x_i from (c_i, ε_i) as meaning that each agent's cost to reveal information is potentially correlated with their (private) residual data, but is independent of the agent's features.

As in Section 4.3, the analyst wants to design a survey mechanism to buy from the agents, obtain data from the set S of elicited agents, then compute an estimate $\hat{\theta}_S$ of θ. The expected average payment to each of the n agents should be no more than $\bar{B}$. As in Section 4.2, we note that the problem of designing a survey mechanism in fact reduces to that of designing an allocation rule A that minimizes said variance and satisfies a budget constraint in which the prices are replaced by known virtual costs. To this end, the analyst designs an allocation rule A and a pricing rule P so as to minimize the $\sqrt{n}$-normalized worst-case asymptotic mean-squared error of $\hat{\theta}_S$ as the population size goes to infinity. Our mechanism will essentially be optimizing

the coefficient in front of the leading $1/n$ term in the mean squared error, ignoring potential finite sample deviations that decay at a faster rate than $1/n$. Note that we will design allocation and pricing rules to be independent of the population size n; hence, the analyst can use the designed mechanism even if the exact population size in unknown.

Estimators for Regression

Let S be the set of data points elicited by a survey mechanism. The analyst's estimate will then be the value $\hat{\theta}_S$ that minimizes the Horvitz-Thompson mean-squared error $\mathbb{E}[(y_i - x_i^\top \theta^*)^2]$, i.e.,

$$\hat{\theta}_S = \operatorname*{argmin}_{\theta \in \Theta} \sum_i \frac{1\{i \in S\}}{A(c_i)} (y_i - x_i^\top \theta)^2. \tag{4.13}$$

Further, we make the following assumptions on the distribution of data points:

Assumption 4.4.1 (Assumption on the distribution of features)**.** $E[x_i x_i^\top]$ *is finite and positive-definite, and hence invertible.*

Finite expectation is a property one may expect real data such has age, height, weight, etc. to exhibit. The second part of the assumption is satisfied by common classes of distributions, such as multivariate normals. We first show that $\hat{\theta}_S$ is a consistent estimator of θ.

Lemma 4.4.2. *Under Assumption 4.4.1, for any allocation rule $A > 0$ that does not depend on n, $\hat{\theta}_S$ is a consistent estimator of θ^*.*

Proof of Lemma 4.4.2. Let $m(\theta; x, y) = (y - x^\top \theta)^2$, and let $\mathbb{1}_i = 1\{i \in S\}$ for simplicity. The following holds:

1. First, we note that θ^* is the unique parameter that minimizes $\mathbb{E}[(y_i - x_i^\top \theta)^2]$; indeed, take any $\theta \neq \theta^*$, we have that

$$\begin{aligned}\mathbb{E}[(y_i - \theta^\top x_i)^2] &= \mathbb{E}\left[\left(y_i - x_i^\top \theta^* + x_i^\top(\theta^* - \theta)\right)^2\right] \\ &= \mathbb{E}\left[\left(y_i - x_i^\top \theta^*\right)^2\right] + \mathbb{E}\left[\left(x_i^\top(\theta^* - \theta)\right)^2\right] + 2\mathbb{E}\left[\varepsilon_i(\theta^* - \theta)^\top x_i\right].\end{aligned}$$

As x and ε are independent, ε has mean 0, this simplifies to

$$\begin{aligned}
&\mathbb{E}[(y_i - \theta^\top x_i)^2] \\
&= \mathbb{E}\left[\left(y_i - \theta^{*\top} x_i\right)^2\right] + (\theta^* - \theta)^\top \mathbb{E}\left[x_i x_i^\top\right](\theta^* - \theta) + 2(\theta^* - \theta)^\top \mathbb{E}\left[\varepsilon_i x_i\right] \\
&= \mathbb{E}\left[\left(y_i - \theta^{*\top} x_i\right)^2\right] + (\theta^* - \theta)^\top \mathbb{E}\left[x_i x_i^\top\right](\theta^* - \theta) \\
&> \mathbb{E}\left[\left(y_i - \theta^{*\top} x_i\right)^2\right].
\end{aligned}$$

where the last step follows from $\mathbb{E}[x_i x_i^\top]$ being positive-definite by Assumption 4.4.1.

2. By definition, Θ is compact.

3. $m(\theta; x, y)$ is continuous in θ, and so is its expectation.

4. $m(.;.)$ is also bounded (lower-bounded by 0, and upper-bounded by either L^2 or U^2), implying that $\theta \to \frac{\mathbb{1}_i}{A(c_i)} m(\theta; x_i, y_i)$ is continuous and bounded. Hence, by the uniform law of large number, remembering that $\frac{\mathbb{1}_i}{A(c_i)} m(\theta; x_i, y_i)$ are i.i.d,

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{1}_i}{A(c_i)} m(\theta; x_i, y_i) - \mathbb{E}\left[\frac{\mathbb{1}_i}{A(c_i)} m(\theta; x_i, y_i)\right] \right| \to 0.$$

Finally, noting that conditional on c_i, $m(\theta; x_i, y_i)$ and $\frac{\mathbb{1}_i}{A(c_i)}$ are independent, we have:

$$\mathbb{E}\left[\frac{\mathbb{1}_i}{A(c_i)} m(\theta; x_i, y_i)\right] = \mathbb{E}\left[\mathbb{E}\left[\left.\frac{\mathbb{1}_i}{A(c_i)}\right| c_i\right] \mathbb{E}\left[\, m(\theta; x_i, y_i) \mid c_i \,\right]\right] = \mathbb{E}\left[m(\theta; x_i, y_i)\right]$$

using $\mathbb{E}\left[\left.\frac{\mathbb{1}_i}{A(c_i)}\right| c_i\right] = 1$.

Therefore, all of the conditions of Theorem 2.1 of Newey et al., 1994 are satisfied, which is enough to prove the result. □

Similarly to the moment estimation problem in Section 4.3, the goal of the analyst is to minimize the worst-case (over the distribution of data and the correlation between c_i's and ε_i's) asymptotic mean-squared error of the estimator $\hat{\theta}_S$. Here "asymptotic" means the worst-case error as $\hat{\theta}_S$ approaches the true parameter θ^*. The following theorem characterizes the asymptotic covariance matrix of $\hat{\theta}_S$. (In fact, it fully characterizes the asymptotic distribution of $\hat{\theta}_S$.)

Lemma 4.4.3. *Under Assumption 4.4.1, for any allocation rule $A > 0$ that does not depend on n, the asymptotic distribution of $\hat{\theta}_S$ is given by*

$$\sqrt{n}(\hat{\theta}_S - \theta^*) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}[x_i x_i^\top]^{-1}\mathbb{E}\left[\varepsilon_i^2 \frac{1\{i \in S\}}{A^2(c_i)}\right]\right),$$

where d denotes convergence in distribution and where randomness in the expectations is taken on the costs c_i, the set of elicited data points S, the features of the data x_i, and the noise ε_i.

Proof of Lemma 4.4.3. For simplicity, let $\mathbb{1}_i = 1\{i \in S\}$ and note that the $\mathbb{1}_i$'s are i.i.d. Let $m(\theta; x_i, y_i) = (y_i - x_i^\top\theta)^2$. First we remark that $\nabla_\theta m(\theta; x_i, y_i) \cdot \frac{\mathbb{1}_i}{A(c_i)} = 2\frac{\mathbb{1}_i}{A(c_i)} x_i(x_i^\top\theta - y_i)$ and $\nabla^2_{\theta\theta} m(\theta; x_i, y_i) \cdot \frac{\mathbb{1}_i}{A(c_i)} = 2\frac{\mathbb{1}_i}{A(c_i)} x_i x_i^\top$. We then note the following:

1. θ^* is in the interior of Θ.
2. $\theta \to m(\theta; x_i, y_i) \cdot \frac{\mathbb{1}_i}{A(c_i)}$ is twice continuously differentiable for all $x_i, y_i, c_i, \mathbb{1}_i$.
3. $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \nabla_\theta\, m(\theta^*; x_i, y_i) \cdot \frac{\mathbb{1}_i}{A(c_i)}\right) \to \mathcal{N}\left(0, 4\mathbb{E}\left[\frac{\mathbb{1}_i}{A^2(c_i)} x_i x_i^\top (x_i^\top\theta^* - y_i)^2\right]\right)$. This follows directly from applying the multivariate central limit theorem, noting that

$$\begin{aligned}\mathbb{E}\left[\nabla_\theta m(\theta^*; x_i, y_i) \cdot \frac{\mathbb{1}_i}{A(c_i)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\, 2x_i\varepsilon_i \mid c_i \,\right] \cdot \mathbb{E}\left[\, \frac{\mathbb{1}_i}{A(c_i)} \,\middle|\, c_i \,\right]\right] \\ &= \mathbb{E}\left[2x_i\varepsilon_i\right] \\ &= 0.\end{aligned}$$

   where the first step follows from conditional indepence on c of x, ε with $A(c), S$, the second step from $\mathbb{E}\left[\, \frac{\mathbb{1}_i}{A(c_i)} \,\middle|\, c_i \,\right] = 1$, and the last equality follows from the fact that x and ε are independent and $\mathbb{E}[\varepsilon_i] = 0$.
4. $\sup_{\theta\in\Theta} \left\| \mathbb{E}\left[\nabla^2_{\theta\theta} m(\theta; x_i, y_i) \cdot \frac{\mathbb{1}_i}{A(c_i)}\right] - \frac{1}{n}\sum_i \nabla^2_{\theta\theta} m(\theta; x_i, y_i) \cdot \frac{\mathbb{1}_i}{A(c_i)} \right\| \to 0$, applying the uniform law of large numbers as $\nabla^2_{\theta\theta} m(\theta; x_i, y_i) \cdot \frac{\mathbb{1}_i}{A(c_i)} = 2\frac{\mathbb{1}_i}{A(c_i)} x_i x_i^\top$ is i) continuous in θ, and ii) constant in θ, thus bounded coordinate-by-coordinate by $2\frac{\mathbb{1}_i}{A(c_i)} x_i x_i^\top$ that is independent of θ and has finite expectation $2\mathbb{E}[x_i x_i^\top]$.
5. $\mathbb{E}\left[\nabla^2_{\theta\theta} m(\theta; x_i, y_i) \cdot \frac{\mathbb{1}_i}{A(c_i)}\right] = 2\mathbb{E}[x_i x_i^\top]$ is invertible as it is positive-definite.

Therefore the sufficient conditions i)-v) in Theorem 3.1 of Newey et al., 1994 hold, proving that the asymptotic distribution is normal with mean 0 and variance

$$\mathbb{E}[2x_i x_i^\top]^{-1}\mathbb{E}\left[4\frac{\mathbb{1}_i}{A^2(c_i)}(x_i^\top\theta - y_i)^2 x_i x_i^\top\right]\mathbb{E}[2x_i x_i^\top]^{-1}.$$

To conclude the proof, we remark that by independence of x_i with c_i and ε_i,

$$\mathbb{E}\left[\frac{\mathbb{1}_i}{A^2(c_i)}(x_i^\top\theta - y_i)^2 x_i x_i^\top\right] = \mathbb{E}[x_i x_i^\top]\mathbb{E}\left[\varepsilon_i^2\frac{\mathbb{1}_i}{A^2(c_i)}\right].$$

□

Lemma 4.4.3 implies that the worst-case asymptotic mean-squared error, under a budget constraint, is given by the worst-case trace of the variance matrix. That is,

$$\mathcal{R}^*(\mathcal{F}, A) \triangleq \sup_{\mathcal{X}} \sup_{\mathcal{D}_\varepsilon} \mathbb{E}\left[\varepsilon_i^2\frac{1\{i \in S\}}{A^2(c_i)}\right] \cdot \sum_{j=1}^{d}\mathbb{E}[x_i x_i^\top]_{jj}^{-1} \quad \text{s.t. } \mathbb{E}[\varepsilon_i] = 0. \tag{4.14}$$

where recall that $\mathcal{D}_\varepsilon$ is the marginal distribution over ε and $\mathcal{X}$ the distribution over x. Importantly, this can be rewritten as

$$\mathcal{R}^*(\mathcal{F}, A) \triangleq \left(\sup_{\mathcal{X}}\sum_{j=1}^{d}\mathbb{E}[x_i x_i^\top]_{jj}^{-1}\right) \cdot \sup_{\mathcal{D}_\varepsilon} \mathbb{E}\left[\varepsilon_i^2\frac{1\{i \in S\}}{A^2(c_i)}\right] \quad \text{s.t. } \mathbb{E}[\varepsilon_i] = 0. \tag{4.15}$$

Therefore, the analyst's decision solely depend on the worst-case correlation between costs c_i and noise ε_i, and not on the worst-case distribution $\mathcal{X}$. In turn, the analyst's allocation is completely independent of and robust in $\mathcal{X}$.

Characterizing the Optimal Allocation Rule for Regression

As in Section 4.3, we assume costs are drawn from a discrete set, say $C = \{c_1, \ldots, c_{|C|}\}$. We will then write A_t for an allocation rule conditional on the cost being c_t, and π_t the probability of the cost of an agent being c_t. We will assume that $\bar{B} < \sum_{t=1}^{|C|}\pi_t c_t$, meaning that it is not feasible to accept all data points, since otherwise it is trivially optimal to set $A_t = 1$ for all t.

The following lemma describes the optimization problem faced by an analyst wanting to design an optimal survey mechanism. Recall that residual values lie in the interval $[L, H]$.

Lemma 4.4.4 (Optimization Problem for Parameter Estimation)**.** *The optimization*

program for the analyst is given by:

$$\begin{aligned} \inf_{A\in[0,1]^{|C|}} \quad & \sup_{q\in[0,1]^{|C|}} \sum_{t=1}^{l} \frac{\pi_t}{A_t}\left((1-q_t)\cdot L^2 + q_t\cdot U^2\right) \\ \text{s.t.} \quad & \sum_{t=1}^{|C|} \pi_t\left((1-q_t)\cdot L + q_t\cdot U\right) = 0 \\ & \sum_{t=1}^{|C|} \pi_t c_t A_t \le \bar{B} \\ & A \text{ is monotone non-increasing} \end{aligned} \tag{4.16}$$

Proof of Lemma 4.4.4. First we note that

$$\begin{aligned} \mathcal{R}^*(\mathcal{F}, A) \triangleq \left(\sup_{X} \sum_{j=1}^{d} \mathbb{E}[x_i x_i^\top]^{-1}_{jj}\right)\cdot \sup_{\mathcal{D}_\varepsilon} \mathbb{E}\left[\varepsilon_i^2 \frac{1\{i\in S\}}{A^2(c_i)}\right] \\ \text{s.t. } \mathbb{E}[\varepsilon_i] = 0. \end{aligned} \tag{4.17}$$

We can therefore renormalize the worst-case variance by $\sup_X \sum_{i=1}^{d} \mathbb{E}[x_i x_i^\top]^{-1}_{ii}$, as it does not depend on any other parameter of the problem. The analyst's objective is now given by

$$\begin{aligned} \sup_{\mathcal{D}_\varepsilon} \sum_{t=1}^{|C|} \pi_t \frac{\mathbb{E}[\varepsilon_i^2|c_t]}{A_t} \\ \text{s.t.} \mathbb{E}[\varepsilon_i] = 0. \end{aligned} \tag{4.18}$$

The worst case distribution is reached when $\varepsilon_i|c_t$ is binomial between L and U (and such a distribution is feasible for $\varepsilon_i|c_t$), therefore letting $q_t = P[\varepsilon_i = U \mid c_t]$, we obtain the lemma. □

We can now characterize the form of the optimal survey mechanism. For simplicity, we will assume that $U^2 \ge L^2$. This is without loss of generality, since the optimization program is symmetric in L and U; if $L^2 > U^2$, the analyst can set $q_t = 1 - q_t$, $L = U$ and $U = L$ to obtain Program (4.16) with $U^2 > L^2$.

Theorem 4.4.5. *Under the assumptions described above, an optimal allocation rule A has the form*

1. $A_t = \min\left(1, \alpha\frac{|L|}{\sqrt{c_t}}\right)$ *for* $t < t^-$
2. $A_t = \bar{A}$ *for all* $t \in \{t^-, \ldots, t^+\}$

3. $A_t = \min\left(1, \alpha\frac{U}{\sqrt{c_t}}\right)$ *for* $t > t^+$

for $\bar{A}$ *and* α *positive constants that do not depend on n, and* t^- *and* t^+ *integers with* $t^- \leq t^+$. *Further,* $\bar{A}$ *and* α *can be computed efficiently given knowledge of* t^-, t^+.

We remark that the allocation rule that we designed is strictly positive and independent of n (as the optimization program itself does not depend on n), so Lemmas 4.4.2 and 4.4.3 apply. Theorem 4.4.5 immediately implies that an optimal allocation rule can be obtained by simply searching over the space of parameters (t^-, t^+), which can be done in at most $|C|^2$ steps. For each pairs of parameters (t^-, t^+), A can be computed efficiently as stated in the Theorem. Then the analyst only needs to pick the allocation rule that minimizes the objective value among the obtained allocation rules that are feasible for Program (4.16). Further, we remark that the solution for the linear regression case exhibits a structure that is similar to the structure of the optimal allocation rule for moment estimation (see Theorem 4.3.2): it exhibits a pooling region in which all cost types are treated the same way, and changes in the inverse of the square root of the cost outside said pooling region. However, we note that we may now choose to pool agents together in an intermediate range of costs, instead of pooling together agents whose costs are below a given threshold.

Proof sketch. We first compute the best response q^* of the adversary; we note that this best response is in fact the solution to a knapsack problem that is independent of the value taken by the allocation rule A. We can therefore plug the adversary's best response into the optimization problem, and reduce the minimax problem above in a simple minimization problem on A. We then characterize the solution as a function of the parameters $(t^-, t^+) \in [|C|]^2$ through KKT conditions. The full proof is given in Section 4.7. □

4.5 Proofs: Reduction from Mechanism Design to Optimization

We give the proof when the cost support is discrete. In the whole proof, we let $C \triangleq \{c_1, \ldots, c_{|C|}\}$ with $c_1 < \ldots < c_{|C|}$, $\pi_t = f(c_t)$, and $A_t = A(c_t)$. We first show that given a fixed monotone non-increasing allocation rule A with $A_1 \geq \ldots \geq A_{|C|}$, there exists optimal prices $P^*(A)$ such that the payments of any individually rational and truthful mechanism are lower-bounded by $P^*(A)$, and such that mechanism with allocation rule A and prices $P^*(A)$ is individually rational and truthful:

Claim 4.5.1. *Let* $P^*_{|C|}(A) = c_{|C|}$, *and* $P^*_t(A) = c_t + \sum_{j=t+1}^{|C|} \frac{A_j}{A_t}(c_j - c_{j-1})$ *for all* $t < |C|$. *Then for every IC and IR mechanism with monotone non-increasing allocation rule*

A, the pricing rule P must satisfy $P_t \geq P_t^*(A)$ for all t. Further, the mechanism with allocation rule A and pricing rule $P^*(A)$ is IC and IR.

Note that this directly implies that if there exists a variance-minimizing mechanism, then there exists an IC and IR variance-minimizing mechanism with pricing rule $P^*(A)$ given allocation rule A. Therefore, we can reduce our attention to such mechanisms.

Proof. We show the first part of the lemma by induction: clearly, it must be the case that $P_{|C|} \geq c_{|C|}$ for the mechanism to be IR when using pricing rule P. Now, suppose by induction that for any IC and IR mechanism, $P_{t+1} \geq P_{t+1}^*(A)$. We require by IC constraint and induction hypothesis that for $t < |C| - 1$,

$$\begin{aligned} P_t &\geq c_t + \frac{A_{t+1}}{A_t}(P_{t+1}^*(A) - c_t) \\ &= c_t + \frac{A_{t+1}}{A_t}\left(\sum_{j=t+2}^{|C|} \frac{A_j}{A_{t+1}}(c_j - c_{j-1}) + c_{t+1} - c_t\right) \\ &= P_t^*(A) \end{aligned}$$

and for $t = |C| - 1$,

$$\begin{aligned} P_{|C|-1} &\geq c_{|C|-1} + \frac{A_{|C|}}{A_{|C|-1}}(P_{|C|}^*(A) - c_{|C|-1}) \\ &= c_{|C|-1} + \frac{A_{|C|}}{A_{|C|-1}}(c_{|C|} - c_{|C|-1}) \\ &= P_{|C|-1}^*(A). \end{aligned}$$

This proves the first part of the claim. Now, we note that the mechanism with prices $P^*(A)$ is IR as clearly $P_t^*(A) \geq c_t$. It remains to show the mechanism is IC to complete the proof of the claim. Take $t \neq t'$, we have:

$$\begin{aligned} &A_t(P_t^* - c_t) - A_{t'}(P_{t'}^* - c_t) \\ &= \sum_{j=t+1}^{|C|} A_j(c_j - c_{j-1}) - \sum_{j=t'+1}^{|C|} A_j(c_j - c_{j-1}) + A_{t'}(c_t - c_{t'}). \end{aligned}$$

If $t > t'$, we have

$$A_t(P_t^* - c_t) - A_{t'}(P_{t'}^* - c_{t'}) \geq A_{t'}(c_t - c_{t'}) - A_{t'} \sum_{j=t'+1}^{t} (c_j - c_{j-1}) = 0$$

as $A_j \leq A_{t'}$ for all $j \geq t'$, while if $t < t'$, we have

$$A_t(P_t^* - c_t) - A_{t'}(P_{t'}^* - c_t) \geq A_{t'} \sum_{j=t+1}^{t'} (c_j - c_{j-1}) - A_{t'}(c_{t'} - c_t) = 0.$$

as $A_j \geq A_{t'}$ for all $j \leq t'$. This concludes the proof. □

We conclude the proof by showing that the budget constraint can be rewritten in the desired form, i.e. such that the true costs are replaced by the virtual costs in the budget expression:

Lemma 4.5.2. *The expected budget used by a mechanism with allocation rule A and payment rule $P^*(A)$ can be written*

$$\sum_{t=1}^{n} \sum_{t=1}^{|C|} \pi_t \phi(c_t) A_t.$$

Proof. The expected budget spent on an agent can be written, using the previous claim:

$$\begin{aligned}
\sum_{t=1}^{|C|} \pi_t P_t^*(A) A_t &= \sum_{t=1}^{|C|} \pi_t c_t A_t + \sum_{t=1}^{|C|} \pi_t \sum_{j=t+1}^{|C|} A_j (c_j - c_{j-1}) \\
&= \sum_{t=1}^{|C|} \pi_t c_t A_t + \sum_{j=2}^{|C|} \sum_{t=1}^{j-1} \pi_t A_j (c_j - c_{j-1}) \\
&= \pi_1 c_1 A_1 + \sum_{j=2}^{|C|} \left(\pi_j c_j + (c_j - c_{j-1}) \sum_{t=1}^{j-1} \pi_t \right) A_j \\
&= \sum_{j=1}^{|C|} \pi_j \phi(c_j) A_j.
\end{aligned}$$

□

4.6 Proofs: Theorem 4.3.2

A More Structural Characterization

We begin by giving a strengthening of our main Theorem 4.3.2 that exactly pin-points the optimal allocation rule in a closed form. We then give a proof of this stronger theorem.

Theorem 4.6.1 (Closed Form for Optimal Allocation for Moment Estimation)**.** *The optimal allocation rule A is determined by two constants $\bar{A}$ and $t^* \in \{0, \ldots, |C|\}$*

such that:

$$A_t = \begin{cases} \bar{A} & \text{if } t \le t^* \\ \frac{1}{\sqrt{c_t}} \cdot \frac{\bar{B} - \bar{A}\mathbb{E}[c \cdot 1\{c \le c_{t^*}\}]}{\mathbb{E}[\sqrt{c} \cdot 1\{c > c_{t^*}\}} & o.w. \end{cases} \tag{4.19}$$

The parameters $\bar{A}$ and t^* are determined as follows. For $k \in \{0, 1, \ldots, |C| + 1\}$ and $x \in [0, 1]$ let (for $c_0 = 0$ and $c_{|C|+1} = \infty$):

$$Q(k, x) = \sum_{t=1}^{k} \pi_t c_t + \sum_{t=k+1}^{|C|} \pi_t \sqrt{\frac{c_t \cdot c_k}{x}}. \tag{4.20}$$

$$R(k, x) = 2\left(\sum_{t=1}^{k} \pi_t \frac{c_t \cdot x}{c_k} + \sum_{t=k+1}^{|C|} \pi_t\right). \tag{4.21}$$

$$B(k, x) = \frac{Q(k, x)}{R(k, x)}. \tag{4.22}$$

Let k^* be the unique k s.t. $B(k, 1) \le \bar{B} < B(k + 1, 1)$. If $k^* = 0$ then $t^* = 0$. Otherwise let $x^* \in [0, 1]$ be the unique solution to: $\bar{B} = B(k^*, x^*)$.[8] *If $R(k^*, x^*) \ge 1$ then $t^* = k^*$ and $\bar{A} = \frac{1}{R(k^*, x^*)}$. If $R(k^*, x^*) < 1$ then $t^* = \max\{k : \bar{B} > Q(k, 1)\}$ and $\bar{A} = 1$.*

Proof of Theorem 4.6.1: Optimal Survey for Moment Estimation

In all that follows, we will drop the monotonicity constraints on A. We write

$$\begin{aligned} P = \inf_{A \in (0,1]^{|C|}} \sup_{q \in [0,1]^{|C|}} & \sum_{t=1}^{|C|} \pi_t \cdot \frac{q_t}{A_t} - \left(\sum_{t=1}^{|C|} \pi_t \cdot q_t\right)^2 \\ & \text{s.t.} \sum_{t=1}^{|C|} \pi_t \cdot c_t \cdot A_t \le \bar{B}. \end{aligned} \tag{4.23}$$

We will show that nevertheless, the solution to this optimization program satisfies the monotonicity constraint in A, hence dropping the constraint can be done without loss of generality. For further simplification of notation, for any two vectors x, y we let $x \circ y$ be their component-wise product vector, $x./y$ their component-wise division and $\langle x, y \rangle$ their inner product. Finally we denote with $x_{i:k}$ the sub-vector $(x_i, x_{i+1}, \ldots, x_k)$. Thus we can write the objective function inside the minimax problem as:

$$V(A, q) = \langle \pi, q./A \rangle - \langle \pi, q \rangle^2, \tag{4.24}$$

[8] The latter amounts to solving a simple cubic equation of the form $\frac{A}{\sqrt{x}} + B = xC + D \Leftrightarrow \sqrt{x}^3 C + (D - B)\sqrt{x} - A = 0$, which admits a closed form solution.

where π is the pdf vector and we can write the budget constraint as $\langle \pi \circ c, A \rangle \leq \bar{B}$.

Rather than solving for simply the optimal solution for A in the latter minimax problem, we will instead address the harder problem of finding an equilibrium of the zero-sum game $\mathcal{G}$ associated with this minimax, i.e. a game where the minimizer player is choosing A and the maximizing player is choosing q and the utility is $V(A, q)$. An equilibrium of this game is then defined as:

Definition 4.6.2 (Equilibrium Pair)**.** *A pair of solutions* (A^*, q^*) *is an equilibrium if:*

$$V(A^*, q^*) = \inf_{A \in [0,1]^{|C|} : \langle \pi \circ c, A \rangle \leq \bar{B}} V(A, q^*) = \sup_{q \in [0,1]^{|C|}} V(A^*, q). \tag{4.25}$$

Observe that the function $V(A, q)$ is convex in A and concave in q, hence it defines a convex-concave zero-sum game. From standard results on zero-sum games, if (A^*, q^*) is an equilibrium solution, then A^* is a solution to the minimax problem P that we are interested in (see e.g. Freund et al., 1999), since:

$$\inf_A V(A, q^*) \leq \inf_A \sup_q V(A, q) \leq \sup_q V(A^*, q) = V(A^*, q^*) = \inf_A V(A, q^*),$$

directly implying $\sup_z V(A^*, z) = \inf_A \sup_q V(A, q)$.

Characterizing the best responses of the minimizing and maximizing player: In this paragraph we characterize the best-responses of the minimizing and maximizing players in the zero-sum game formulation of our problem.

Lemma 4.6.3 (Best Response of Min Player)**.** *Fix* q *such that* $q_t > 0$ *for all* t*. Let* $\lambda^* > 0$ *be such that*

$$\sum_{t=1}^{|C|} \pi_t \cdot c_t \cdot \min\left(1, \sqrt{\frac{q_t}{\lambda^* c_t}}\right) = \bar{B}. \tag{4.26}$$

Then the allocation rule A *such that* $A_t = \min\left(1, \sqrt{\frac{q_t}{\lambda^* c_t}}\right)$ *is a best-response to* q *in game* $\mathcal{G}$*.*

Proof. Fix q . Note that when A_t goes to 0 for any t, the objective values tends to infinity, but the optimal solution is clearly finite (simply splitting the budget evenly among cost types c_t is feasible and leads to a finite objective value). This implies that a best-response exists for the analyst: indeed, the objective is convex and continuous on $(0, 1]^{|C|}$, and the feasible set can be restricted without loss of generality to be

compact and convex by adding the constraints $A_t \geq \gamma$ for a small enough $\gamma > 0$. The Lagrangian (for more on Lagrangians and KKT conditions, see Boyd et al., 2004) of the minimization problem solved by the minimizing player is therefore given by:

$$\mathcal{L}(A, \lambda, \lambda_t^1) = V(A, q) + \lambda \sum_t (\pi_t c_t A_t - \bar{B}) + \sum_{t=1}^{|C|} \lambda_t^1 (A_t - 1)$$

where $\lambda, \lambda_t^1 \geq 0$. Let $(\lambda^*, \lambda_t^{1*})$ denote optimal dual variables and A an optimal primal variable, we have that $\frac{\partial \mathcal{L}(A, \lambda^*, \lambda_t^{1*})}{\partial A_t} = 0$ for all t by the KKT conditions, implying

$$-\pi_t \frac{q_t}{A_t^2} + \lambda^* \pi_t c_t + \lambda_t^{1*} = 0 \Rightarrow A_t = \sqrt{\frac{\pi_t q_t}{\lambda^* \pi_t c_t + \lambda_t^{1*}}}$$

where we note that the denominator is non-zero as we have $c_t > 0$ and $\lambda_t^{1*} \geq 0$. Further, if $A < 1$, we must also have by the KKT conditions that $\lambda_t^{1*} = 0$. This directly implies that at a best response, we must have $A_t = \min\left(1, \sqrt{\frac{q_t}{\lambda^* c_t}}\right)$, for some $\lambda^* \geq 0$. Since the budget constraint will always be binding (as increasing the allocation probability can only help the variance), λ^* must be solving Equation (4.26). The latter concludes the proof of the Lemma. □

This gives the best response of the minimizing player. The best response of the maximizing player can be obtained by similar techniques:

Lemma 4.6.4 (Best-Response of Max Player)**.** *Fix A such that $A_j > 0$ for all j. Take any $q \in [0, 1]^{|C|}$ such that for every j, at least one of the following holds:*

1. $q_j = 0$ *and* $\frac{1}{A_j} < 2 \langle \pi, q \rangle$
2. $q_j = 1$ *and* $\frac{1}{A_j} > 2 \langle \pi, q \rangle$
3. $0 \leq q_j \leq 1$ *and* $\frac{1}{A_j} = 2 \langle \pi, q \rangle$

Then q is a best response to A in game $\mathcal{G}$.

Proof. Fix A. The Lagrangian of the maximization problem solved by the maximizing player is given by:

$$\mathcal{L}(q, \lambda_t^0, \lambda_t^1) = V(A, q) + \sum_{t=1}^{|C|} \lambda_t^1 (1 - q_t) + \sum_{t=1}^{|C|} \lambda_t^0 q_t$$

For optimal primal and dual variables, the KKT conditions are given by the following: for all $t \in \{1, \ldots, |C|\}$

- *First order.* $\frac{\partial \mathcal{L}(q,\lambda_t^0,\lambda_t^1)}{\partial q_t} = 0$, which can be rewritten as: $\frac{\pi_t}{A_t} - 2\pi_t \langle \pi, q \rangle - \lambda_t^1 + \lambda_t^0 = 0$.
- *Feasibility.* The variables $q_t, \lambda_t^1, \lambda_t^0$ are feasible, i.e. $q_t \in [0, 1]$ and $\lambda_t^0, \lambda_t^1 \geq 0$
- *Complementarity.* Either $\lambda_t^0 = 0$ or $q_t = 0$. Either $\lambda_t^1 = 0$ or $q_t = 1$

Because the objective value is convex and differentiable in q, the KKT conditions are necessary and sufficient for optimality of the primal and dual variables (see Boyd et al., 2004). It is therefore enough to show that $(q_t, \lambda_t^0, \lambda_t^1)$ satisfies the KKT conditions for some well-chosen $q_t, \lambda_t^0, \lambda_t^1$. If $\frac{1}{A_t} < 2 \langle \pi, q \rangle$, then $q_t = 0$, $\lambda_t^1 = 0$ and $\lambda_t^0 = -\frac{\pi_t}{A_t} + 2\pi_t \langle \pi, q \rangle \geq 0$ is a solution. If $\frac{1}{A_t} > 2 \langle \pi, q \rangle$ then $q_t = 1$, $\lambda_j^0 = 0$ and $\lambda_j^1 = \frac{\pi_j}{A_j} - 2\pi_j \langle \pi, q \rangle \geq 0$ is a solution. Otherwise if $\frac{1}{A_t} = 2 \langle \pi, q \rangle$ then $\lambda_t^0 = \lambda_t^1 = 0$ and any feasible value for q_t that respects the equality $\frac{1}{A_t} = 2 \langle \pi, q \rangle$ is a solution. □

Properties of functions $Q(k, x)$***,*** $R(k, x)$ ***and*** $B(k, x)$***:*** Before proceeding to the main proof of the Theorem, we show some useful properties of $Q(k, x)$, $R(k, x)$ and $B(k, x)$.

Claim 4.6.5. *For all* $k < |C|$, $R(k, 1) \geq R(k + 1, 1)$, $Q(k, 1) \leq Q(k + 1, 1)$ *and* $B(k, 1) \leq B(k + 1, 1)$. *Therefore, the exists a unique* k *such that* $B(k, 1) \leq \bar{B} < B(k + 1, 1)$.

Proof. By expanding:

$$\begin{aligned} R(k,1) - R(k+1,1) &= \sum_{t=1}^{k} \pi_t \frac{c_t}{c_k} + \sum_{i=k+1}^{|C|} \pi_t - \sum_{t=1}^{k+1} \pi_t \frac{c_t}{c_{k+1}} - \sum_{t=k+2}^{|C|} \pi_t \\ &= \left(\frac{1}{c_k} - \frac{1}{c_{k+1}} \right) \sum_{t=1}^{k} \pi_t c_t \geq 0. \end{aligned}$$

Additionally,

$$\begin{aligned} Q(k+1,1) - Q(k,1) &= \sum_{t=1}^{k+1} \pi_t c_t - \sum_{t=1}^{k} \pi_t c_t + \sum_{t=k+2}^{|C|} \pi_t \sqrt{c_{k+1} c_t} - \sum_{t=k+1}^{|C|} \pi_t \sqrt{c_k c_t} \\ &= (\sqrt{c_{k+1}} - \sqrt{c_k}) \sum_{t=k+1}^{|C|} \pi_t \sqrt{c_t} \geq 0. \end{aligned}$$

This concludes the proof. □

Claim 4.6.6. *There is a unique* $x^* \in \left[\frac{c_{k^*}}{c_{k^*+1}}, 1\right]$ *such that* $\bar{B} = B(k^*, x^*)$ *for* $B(1,1) \leq \bar{B} < B(|C|, 1)$.

Proof. Observe that $B(k^*, x)$ is continuous decreasing in x, proving uniqueness of a solution if it exists. Existence within $\left[\frac{c_{k^*}}{c_{k^*+1}}, 1\right]$ follows by noting that $B(k^*, x)$ is continuous decreasing in x, and that trivially $B(k^*, 1) \leq \bar{B} < B(k^* + 1, 1) = B\left(k^*, \frac{c_{k^*}}{c_{k^*+1}}\right)$, proving the result for $B(1,1) \leq \bar{B} < B(|C|, 1)$. □

Proof of the Theorem: We split the proof of the theorem in the two corresponding cases when $B \geq B(1,1)$ i.e. $t^* \geq 1$, then deal with the corner case when $B < B(1,1)$.

Lemma 4.6.7 (Case 1: $R(k^*, x^*) \geq 1$, $B \geq B(1,1)$)**.** *For this case,* $t^* = k^*$ *and* $\bar{A} = \frac{1}{R(k^*,x^*)} = \frac{\bar{B}}{Q(k^*,x^*)}$.

Proof. Let q be such that $q_t = \frac{c_t}{c_{k^*}} x^*$ for $t \leq k^*$, $q_{k^*+1} = \ldots = q_{|C|} = 1$. We show that A defined by the parameters $\bar{A}$ and t^* given in the lemma, is a best response to q and vice-versa.

First, let us show that q is a best response to A. For $j \leq k^*$, we note that $0 \leq q_j \leq 1$ and we have

$$2\langle \pi, q \rangle = 2\left(\frac{x^*}{c_{k^*}} \sum_{t=1}^{k^*} \pi_t c_t + \sum_{t=k^*+1}^{|C|} \pi_t\right) = R(k^*, x^*) = \frac{1}{A_j}.$$

For $t \geq k^* + 1$ (when such a case exists, i.e. when $k^* < |C|$), $q_j = 1$. Moreover, the allocation takes the form:

$$\begin{aligned} A_t = \frac{1}{\sqrt{c_t}} \cdot \frac{\bar{B} - \bar{A}\mathbb{E}[c \cdot 1\{c \leq c_{t^*}\}]}{\mathbb{E}[\sqrt{c} \cdot 1\{c > c_{t^*}\}]} &= \frac{\bar{B}}{\sqrt{c_t}} \cdot \frac{1 - \frac{1}{Q(k^*,x^*)}\mathbb{E}[c \cdot 1\{c \leq c_{t^*}\}]}{\mathbb{E}[\sqrt{c} \cdot 1\{c > c_{t^*}\}]} \\ &= \frac{\bar{B}}{\sqrt{c_t}} \cdot \frac{\sqrt{\frac{c_{k^*}}{x^*}}}{Q(k^*, x^*)}. \end{aligned}$$

By Claim 4.6.6, $\frac{c_{k^*+1}}{c_{k^*}} \cdot x^* \geq 1$. Thus:

$$A_t \leq \frac{\bar{B}}{\sqrt{c_t}} \cdot \frac{\sqrt{c_{k^*+1}}}{Q(k^*, x^*)} \leq \frac{\bar{B}}{Q(k^*, x^*)} = \frac{1}{R(k^*, x^*)} = \frac{1}{2\langle \pi, q \rangle}. \tag{4.27}$$

By Lemma 4.6.4, q is a best response to A. Note that combined with the costs being non-decreasing, this also proves that A is monotone non-increasing.

It remains to show that A is a best response to q. By Lemma 4.6.3, we only need to check that for all $j \in \{1, \ldots, |C|\}$, $A_j = \min\left(1, \sqrt{\frac{q_j}{\lambda^* c_j}}\right)$, where λ^* is chosen to make

the budget constraint tight. First, we note that the A given by the lemma does make the budget constraint tight, as any allocation of the form given by Equation (4.19) does so by construction. Now, let $\lambda^* = \left(\frac{\sum_t \pi_t \sqrt{c_t q_t}}{\bar{B}}\right)^2$. We have that for $j \leq k^*$:

$$\begin{aligned}
\sqrt{\frac{q_j}{\lambda^* c_j}} &= \sqrt{\frac{x^*}{c_{k^*}}} \frac{\bar{B}}{\sum_t \pi_t \sqrt{c_t q_t}} \\
&= \sqrt{\frac{x^*}{c_{k^*}}} \frac{\bar{B}}{\sqrt{\frac{x^*}{c_{k^*}}} \sum_{t=1}^{k^*} \pi_t c_t + \sum_{t=k^*+1}^{|C|} \pi_t \sqrt{c_t}} \\
&= \frac{\bar{B}}{Q(k^*, x^*)} \\
&= A_j
\end{aligned}$$

and by a similar calculation that for $j \geq k^* + 1$ that

$$\sqrt{\frac{q_j}{\lambda^* c_j}} = \sqrt{\frac{1}{c_j}} \frac{\bar{B}}{\sum_t \pi_t \sqrt{c_t q_t}} = \frac{1}{\sqrt{c_j}} \frac{\bar{B}}{\sqrt{\frac{x^*}{c_{k^*}}} Q(k^*, x^*)} = A_j.$$

Since $A_j \leq 1$ for all j, by the conditions of the Lemma, we therefore have:

$$A_j = \min\left(1, \sqrt{\frac{q_j}{\lambda^* c_j}}\right).$$

□

Lemma 4.6.8 (Case 2: $R(k^*, x^*) < 1$, $\bar{B} \geq B(1, 1)$)**.** *For this case,* $t^* = \max\{k : \bar{B} > Q(k, 1)\}$ *and* $\bar{A} = 1$.

We start with the following claim:

Claim 4.6.9. *If* $R(k^*, x^*) < 1$*, then there exist* $\tilde{B}$*,* $\tilde{k}$ *and* $\tilde{x}$ *such that* $\tilde{k} \leq k^*$ *is the unique* k *such that* $B(k, 1) \leq \tilde{B} < B(k + 1, 1)$*,* $\tilde{x}$ *is the unique solution to* $\tilde{B} = B(\tilde{k}, \tilde{x})$*, and* $R(\tilde{k}, \tilde{x}) = 1$*. Further,* $\tilde{x} \in \left[\frac{c_{\tilde{k}}}{c_{\tilde{k}+1}}, 1\right]$.

Proof. By the fact that $R(k, x)$ is increasing in x and by Claim 4.6.6, we get that: $R(k^* + 1, 1) = R\left(k^*, \frac{c_{k^*}}{c_{k^*+1}}\right) \leq R(k^*, x^*) < 1$. Because $R(0, 1) = 2$ and $R(k, 1)$ is decreasing in k, there exists $\tilde{k} \leq k^*$ such that $R(\tilde{k}, 1) > 1 \geq R(\tilde{k} + 1, 1) = R\left(\tilde{k}, \frac{c_{\tilde{k}}}{c_{\tilde{k}+1}}\right)$. Because $R(\tilde{k}, x)$ is increasing continuous in x, there exists $\tilde{x} \in \left[\frac{c_{\tilde{k}}}{c_{\tilde{k}+1}}, 1\right]$ such that $R(\tilde{k}, \tilde{x}) = 1$. Let $\tilde{B} = B(\tilde{k}, \tilde{x}) = Q(\tilde{k}, \tilde{x})$, we have $B(\tilde{k}, 1) \leq \tilde{B} \leq B\left(\tilde{k}, \frac{c_{\tilde{k}}}{c_{\tilde{k}+1}}\right) = B(\tilde{k} + 1, 1)$ as B is decreasing in x. Thus $(\tilde{B}, \tilde{k}, \tilde{x})$ satisfies the claim. □

Now we show the proof of the lemma:

Proof of Lemma 4.6.8. Let us define $(\tilde{B}, \tilde{k}, \tilde{x})$ as in the statement of Claim 4.6.9. First we show that $t^* \geq \tilde{k}$. This follows from noting that $\bar{B} > Q(k^*, x^*) \geq Q(k^*, 1) \geq Q(\tilde{k}, 1)$ as clearly $Q(k, x)$ is increasing in k and decreasing in x. Now, let $q_t = \frac{c_t}{c_{\tilde{k}}}\tilde{x}$ for $t \leq \tilde{k}$, $q_{\tilde{k}+1} = \ldots = q_{|C|} = 1$.

We prove that q is a best response to A. We have that

$$2\langle \pi, q \rangle = 2\left(\frac{\tilde{x}}{c_{\tilde{k}}} \sum_{t=1}^{\tilde{k}} \pi_t c_t + \sum_{t=\tilde{k}+1}^{|C|} \pi_t\right) = R(\tilde{k}, \tilde{x}) = 1$$

For $j \leq \tilde{k} \leq t^*$, $A_j = 1$ hence $2\langle \pi, q \rangle = \frac{1}{A_j}$. For $j > t^*$, we note that by definition of t^*,

$$\bar{B} \leq Q(t^* + 1, 1) = \sum_{t=1}^{t^*+1} \pi_t c_t + \sum_{t=t^*+2}^{|C|} \pi_t \sqrt{c_t \cdot c_{t^*+1}} = \sum_{t=1}^{t^*} \pi_t c_t + \sqrt{c_{t^*+1}} \sum_{t=t^*+1}^{|C|} \pi_t \sqrt{c_t}$$

$$\Rightarrow \frac{\bar{B} - \sum_{t=1}^{t^*} \pi_t c_t}{\sqrt{c_{t^*+1}} \sum_{t=t^*+1}^{|C|} \pi_t \sqrt{c_t}} \leq 1$$

and this directly implies that $A_{t^*+1} \leq 1$. By monotonicity of c_t, it then holds that $A_j \leq 1$ for any $j > t^*$. Hence $2\langle \pi, q \rangle = 1 \geq \frac{1}{A_j}$. This proves q is a best response to A by Lemma 4.6.4, and simultaneously proves monotonicity of A.

Finally, we prove that A is a best response to q. By Lemma 4.6.3 it suffices to show that $\min\left(1, \sqrt{\frac{q_t}{\lambda^* c_t}}\right) = A_t$ for λ^* that makes the budget constraint binding. First, we note that the A given by the lemma makes the budget constraint tight, as any allocation of the form given by Equation (4.19) does so by construction. Now, let $\lambda^* = \left(\frac{\sum_{t=t^*+1}^{|C|} \pi_t c_t}{\bar{B} - \sum_{t=1}^{t^*} \pi_t c_t}\right)^2$, the following holds:

1. For $j > t^*$:
$$\sqrt{\frac{q_j}{\lambda^* c_j}} = \frac{1}{\sqrt{c_j}} \cdot \frac{\bar{B} - \sum_{t=1}^{t^*} \pi_t c_t}{\sum_{t=t^*+1}^{|C|} \pi_t \sqrt{c_t}} = A_j \leq 1.$$

2. For $j \in \{\tilde{k} + 1, \ldots, t^*\}$ (when this regime exists, i.e. when $\tilde{k} < t^*$), remember that by definition of t^*,
$$\bar{B} > Q(t^*, 1) = \sum_{t=1}^{t^*} \pi_t c_t + \sqrt{c_{t^*}} \sum_{t=t^*+1}^{|C|} \pi_t \sqrt{c_t}$$

which implies that

$$\sqrt{\frac{1}{\lambda^* c_{t^*}}} = \frac{1}{\sqrt{c_{t^*}}} \frac{\bar{B} - \sum_{t=1}^{t^*} \pi_t c_t}{\sum_{t=t^*+1}^{|C|} \pi_t \sqrt{c_t}} > 1.$$

Thus

$$\sqrt{\frac{1}{\lambda^* c_j}} \geq \sqrt{\frac{1}{\lambda^* c_{t^*}}} > 1$$

and $\min\left(1, \sqrt{\frac{1}{\lambda^* c_j}}\right) = 1 = A_j$.

3. For $j \leq \tilde{k}$, by Claim 4.6.9 we have $\tilde{x} \geq \frac{c_{\tilde{k}}}{c_{\tilde{k}+1}}$. Therefore:

$$\sqrt{\frac{q_j}{\lambda^* c_j}} = \sqrt{\frac{\tilde{x}}{\lambda^* c_{\tilde{k}}}} \geq \sqrt{\frac{1}{\lambda^* c_{\tilde{k}+1}}} > 1,$$

which follows from the previous case. Hence, $\min\left(1, \sqrt{\frac{1}{\lambda^* c_j}}\right) = 1 = A_j$.

□

Lemma 4.6.10 (Case 3: $\bar{B} < B(1,1) = \frac{\sqrt{c_1}\mathbb{E}[\sqrt{c}]}{2}$)**.** *For this case, an optimal solution is given by* $t^* = 0$.

Proof. We let $q_t = 1$ for all t. We first show that q_t is a best response to A. This trivially follows by Lemma 4.6.4 by remarking that

$$2\langle \pi, q \rangle = 2 < \frac{\sqrt{c_1}\mathbb{E}[\sqrt{c}]}{\bar{B}} \leq \frac{\sqrt{c_j}\mathbb{E}[\sqrt{c}]}{\bar{B}} = \frac{1}{A_j}$$

Now, we show that A is a monotone best response to q. Monotonicity directly follows from the fact that the costs are non-decreasing. Now, to check that A is a best response let us set $\lambda^* = \left(\frac{\mathbb{E}[\sqrt{c}]}{\bar{B}}\right)^2$. We have $\sqrt{\frac{q_j}{\lambda^* c_j}} = \sqrt{\frac{1}{c_j}} \frac{\bar{B}}{\mathbb{E}[\sqrt{c}]} = A_j \leq \frac{\bar{B}}{\sqrt{c_1}\mathbb{E}[\sqrt{c}]} \leq 1/2$. Hence $A_j = \min\left(1, \sqrt{\frac{q_j}{\lambda^* c_j}}\right)$. By Lemma 4.6.3, A is a best response to q. This concludes the proof. □

4.7 Proof of Theorem 4.4.5

Before starting to prove the Theorem, we note that we can assume without loss of generality that $U > 0$. Otherwise, $U = L = 0$ (as we are assuming $U^2 \geq L^2$), and the objective value of the optimization program is 0 and independent of the allocation rule, thus any feasible allocation rule is optimal. In particular, any monotone allocation rule of the form given in the theorem statement works.

Simplifying the analyst's problem

The following lemma reduces the minimax problem that the analyst needs to solve to a simple convex minimization problem:

Lemma 4.7.1. *The optimization program solved by the analyst can be written as:*

$$\begin{aligned} \inf_{A\in(0,1]^{|C|}} \quad & \sum_{t=1}^{t^*-1} \pi_t \frac{L^2}{A_t} + \pi_{t^*}\frac{R^2}{A_{t^*}} + \sum_{t=t^*+1}^{|C|} \pi_t \frac{U^2}{A_t} \\ \text{s.t.} \quad & \sum_{t=1}^{|C|} \pi_t c_t A_t \le \bar{B} \\ & A \text{ is monotone non-increasing,} \end{aligned} \tag{4.28}$$

where $t^* = \min\{j : \ -\frac{L}{U-L} > \sum_{t=j+1}^{|C|} \pi_t\}$, $q_t^* = \frac{1}{\pi_t^*}\left(= -\frac{L}{U-L} - \sum_{t=t^*+1}^{|C|} \pi_t\right) > 0$ *and* $R^2 = (U^2 - L^2)q_{t^*} + L^2 > 0$.

Proof. We first note that for a given A, the optimization program solved by the adversary can be rewritten as

$$\begin{aligned} \sup_{q\in[0,1]^{|C|}} \quad & \sum_{t=1}^{|C|} \frac{\pi_t}{A_t} q_t (U^2 - L^2) + L^2 \sum_{t=1}^{|C|} \frac{\pi_t}{A_t} \\ \text{s.t.} \quad & \sum_t \pi_t q_t = -\frac{L}{U-L} \end{aligned}$$

As such, the adversary is exactly solving a knapsack problem with capacity $-\frac{L}{U-L} \ge 0$, weights π_t and utilities $\frac{\pi_t}{A_t}$. Therefore, an optimal solution exists and is to put the weight on the t's with the higher values of $\frac{1}{A_t}$, i.e., the lowest values of A_t first. Because A is non-increasing in the costs and therefore in t, an optimal solution is given by:

$$q_1 = \ldots = q_{t^*-1} = 0, \pi_{t^*} q_{t^*} = -\frac{L}{U-L} - \sum_{t=t^*+1}^{|C|} \pi_t, q_{t^*+1} = \ldots = q_l = 1.$$

This holds independently of A as long as A is feasible (hence monotone), proving the result. □

Solving the optimization problem

We start with the following lemma that characterizes the form of the solution:

Lemma 4.7.2. *There exist* $\lambda \ge 0$*, non-negative integers* t^-, t^+ *such that* $t^- \le t^* \le t^+$*, and an optimal allocation rule* A *that satisfy*

1. $A_1 \geq \ldots \geq A_{t^- - 1} \geq A_{t^-} = \ldots = A_{t^+} \geq A_{t^+ + 1} \geq \ldots A_l$
2. $A_t = \min\left(1, \sqrt{\frac{L^2}{\lambda c_t}}\right) \ \forall t < t^-$
3. $A_t = \min\left(1, \sqrt{\frac{U^2}{\lambda c_t}}\right) \ \forall t > t^+,$

and that make the budget constraint tight. In the rest of the proof, we denote $\bar{A} \triangleq A_{t^-} = \ldots = A_{t^+}$.

Proof. First, we show existence of an optimal solution A. Because the optimal value of the program is finite but the objective tends to infinity when any $A_{|C|}$ tends to 0 (as we have $U^2, R^2 > 0$), there must exists $\gamma > 0$ such that the analyst's program is given by

$$\begin{aligned} \inf_{A \in [\gamma, 1]^{|C|}} \ & \sum_{t=1}^{t^*-1} \pi_t \frac{L^2}{A_t} + \pi_{t^*} \frac{R^2}{A_{t^*}} + \sum_{t=t^*+1}^{|C|} \pi_t \frac{U^2}{A_t} \\ \text{s.t.} \quad & \sum_{t=1}^{|C|} \pi_t c_t A_t \leq \bar{B} \\ & A \text{ is monotone non-increasing.} \end{aligned} \tag{4.29}$$

The objective is convex and continuous in A over $[\gamma, 1]^{|C|}$, and the feasible set is convex and compact, therefore the above program admits an optimal solution. Now, consider the following program with partial monotonicity constraints:

$$\begin{aligned} \inf_{A \in (0, 1]^{|C|}} \ & \sum_{t=1}^{t^*-1} \pi_t \frac{L^2}{A_t} + \pi_{t^*} \frac{R^2}{A_{t^*}} + \sum_{t=t^*+1}^{|C|} \pi_t \frac{U^2}{A_t} \\ \text{s.t.} \quad & \sum_{t=1}^{|C|} \pi_t c_t A_t \leq \bar{B} \\ & A_1, \ldots, A_{t^*-1} \geq A_t \geq A_{t*+1}, \ldots, A_l \end{aligned} \tag{4.30}$$

In fact, there exists an optimal solution to this problem that makes the budget constraint tight (one can increase the allocation rule without decreasing the objective value until the budget constraint becomes tight). Considering such a solution, the Lagrangian of the program is given by

$$\begin{aligned} \mathcal{L}(A, \lambda, \lambda_t^1, \lambda_t) = & \sum_{t=1}^{t^*-1} \pi_t \frac{L^2}{A_t} + \sum_{t=t^*+1}^{|C|} \pi_t \frac{U^2}{A_t} + \pi_{t^*} \frac{R^2}{A_{t^*}} + \lambda \sum_t \pi_t c_t A_t \\ & + \sum_t \lambda_t^1 A_t - \lambda B/n - \sum_t \lambda_t^1 + \sum_{t<t^*} \lambda_t (A_{t^*} - A_t) + \sum_{t>t^*} \lambda_t (A_t - A_{t^*}). \end{aligned}$$

It must necessarily be the case at optimal that whenever $A_t < 1$, $\lambda_t^1 = 0$ and whenever $A_t > A_{t^*}$ (for $t < t^*$) or $A_t < A_{t^*}$ (for $t > t^*$), $\lambda_t = 0$ by the KKT conditions, and that A satisfies the first order conditions $\nabla_A \mathcal{L}(A, \lambda, \lambda_t^1, \lambda^k) = 0$. Thus, an optimal solution must necessarily satisfy:

1. $A_t = \min\left(1, \max\left(A_{t^*}, \sqrt{\frac{\pi_t L^2}{\lambda \pi_t c_t}}\right)\right)$ for $t < t^*$
2. $A_t = \min\left(1, A_{t^*}, \sqrt{\frac{\pi_t U^2}{\lambda \pi_t c_t}}\right)$ for $t > t^*$

Note that this implies that A is monotone non-increasing (as the virtual costs are monotone non-decreasing), therefore is an optimal solution to the analyst's problem, and that there must exist t^- and t^+ such that

$$A_1 \geq \ldots \geq A_{t^- - 1} \geq A_{t^-} = \ldots = A_{t^*} = \ldots = A_{t^+} \geq A_{t^+ + 1} \geq \ldots A_l$$

with $A_t = \min\left(1, \sqrt{\frac{L^2}{\lambda c_t}}\right)$ $\forall t < t^-$ and $A_t = \min\left(1, \sqrt{\frac{U^2}{\lambda c_t}}\right)$ $\forall t > t^+$. □

We now proceed onto proving the main statement. Let $\gamma_t = |L|$ (resp. R, U), for $t < t^*$ (resp. $t = t^*$, $t > t^*$). Let t^-, t^+ be such that $t^- \leq t^* \leq t^+$ and $A_{t^-} = \ldots = A_{t^+}$ at optimal. Suppose the analyst has knowledge of t^-, t^+. Then, replacing A as a function of $t^-, t^+, \lambda, \bar{A}$ in the analyst's problem (4.28) reduces to the following problem of two variables λ and $\bar{A}$:

$$\begin{aligned}
\inf_{\lambda \geq 0,\ \bar{A} \in [0,1]} \quad & \sum_{t=1}^{t^- - 1} \pi_t \frac{\gamma_t^2}{\min(1, \frac{\gamma_t}{\sqrt{\lambda c_t}})} + \frac{1}{\bar{A}} \cdot \sum_{t=t^-}^{t^+ - 1} \pi_t \gamma_t + \cdot \sum_{t=t^+}^{|C|} \pi_t \frac{\gamma_t^2}{\min(1, \frac{\gamma_t}{\sqrt{\lambda c_t}})} \\
\text{s.t.} \quad & \sum_{t=1}^{t^- - 1} \pi_t c_t \min(1, \frac{\gamma_t}{\sqrt{\lambda c_t}}) + \bar{A} \sum_{t=t^-}^{t^+} \pi_t c_t + \sum_{t=t^+ + 1}^{|C|} \pi_t c_t \min(1, \frac{\gamma_t}{\sqrt{\lambda c_t}}) = \bar{B} \\
& \min\left(1, \frac{\gamma_t}{\sqrt{\lambda c_t}}\right) \geq \bar{A} \ \ \forall t \in \{t^1, \ldots, t^- - 1\} \\
& \min\left(1, \frac{\gamma_t}{\sqrt{\lambda c_t}}\right) \leq \bar{A} \ \ \forall t \in \{t^+ + 1, \ldots, |C|\}.
\end{aligned} \tag{4.31}$$

Hence there exists an optimal solution of the form given by Lemma 4.7.2 with

$$\bar{A} = A(\lambda) \triangleq \frac{1}{\sum_{t=t^-}^{t^+} \pi_t c_t} \left(\bar{B} - \sum_{t=1}^{t^- - 1} \pi_t c_t \min\left(1, \frac{\gamma_t}{\sqrt{\lambda c_t}}\right) - \sum_{t=t^+ + 1}^{|C|} \pi_t c_t \min\left(1, \frac{\gamma_t}{\sqrt{\lambda c_t}}\right) \right),$$

as $\bar{A}$ as a function of λ is entirely determined by the budget constraint. Plugging this back in the above program, we can rewrite the program as depending only on

the variable λ as follows, also remembering that the (virtual) costs are monotone non-decreasing:

$$\begin{aligned}
\inf_{\lambda \geq 0} \quad & \sum_{t=1}^{t^- - 1} \pi_t \frac{\gamma_t^2}{\min\left(1, \frac{\gamma_t}{\sqrt{\lambda c_t}}\right)} + \frac{1}{A(\lambda)} \cdot \sum_{t=t^-}^{t^+} \pi_t \gamma_t + \cdot \sum_{t=t^+ + 1}^{|C|} \pi_t \frac{\gamma_t^2}{\min\left(1, \frac{\gamma_t}{\sqrt{\lambda c_t}}\right)} \\
\text{s.t.} \quad & \min\left(1, \frac{|L|}{\sqrt{\lambda c_{t^- - 1}}}\right) \geq A(\lambda) \\
& \min\left(1, \frac{U}{\sqrt{\lambda c_{t^+ + 1}}}\right) \leq A(\lambda).
\end{aligned} \tag{4.32}$$

Suppose that the analyst also has knowledge of t^1, the maximum value of $t \in \{0, \ldots, t^- - 1\} \cup \{t^+ + 1, \ldots, |C|\}$ such that $\min(1, \frac{\gamma_t}{\sqrt{\lambda c_t}}) = 1$. Then λ^* must satisfy one of the three following conditions:

1. $\min\left(1, \frac{|L|}{\sqrt{\lambda c_{t^- - 1}}}\right) = A(\lambda)$. Letting $\mu = \frac{1}{\sqrt{\lambda}}$, this is a linear equation in μ and therefore can be solved efficiently given knowledge of t^1. This follows from noting the equation can be written

$$\begin{aligned}
& \min\left(1, \frac{\mu |L|}{\sqrt{c_{t^- - 1}}}\right) \\
& = \frac{1}{\sum_{t=t^-}^{t^+} \pi_t c_t} \left(\bar{B} - \sum_{t=1}^{t^- - 1} \pi_t c_t \min\left(1, \frac{\mu \gamma_t}{\sqrt{c_t}}\right) - \sum_{t=t^+ + 1}^{|C|} \pi_t c_t \min\left(1, \frac{\mu \gamma_t}{\sqrt{c_t}}\right) \right),
\end{aligned}$$

   which is of the form $a\mu + b = 0$ for some constants a, b.

2. $\min(1, \frac{U}{\sqrt{\lambda c_{t+1}}}) = A(\lambda)$. Letting $\mu = \frac{1}{\sqrt{\lambda}}$, this is a linear equation in μ and can be solved efficiently given knowledge of t^1. This follows from the same argument as above.

3. λ minimizes the optimization problem with only the non-negativity constraint $\lambda \geq 0$. Then, letting $\mu = \frac{1}{\sqrt{\lambda}}$, the objective value is the following function of μ:

$$\begin{aligned}
OPT(\mu) = & \sum_{t=1}^{t^- - 1} \pi_t \gamma_t^2 \max\left(1, \frac{\sqrt{c_t}}{\mu \gamma_t}\right) + \sum_{t=t^+}^{|C|} \pi_t \gamma_t^2 \max\left(1, \frac{\sqrt{c_t}}{\mu \gamma_t}\right) \\
& + \frac{\sum_{t=t^-}^{t^+} \pi_t c_t \cdot \sum_{t=t^-}^{t^+ - 1} \pi_t \gamma_t}{\bar{B} - \sum_{t=1}^{t^- - 1} \pi_t c_t \min\left(1, \frac{\mu \gamma_t}{\sqrt{c_t}}\right) - \sum_{t=t^+ + 1}^{|C|} \pi_t c_t \min\left(1, \frac{\mu \gamma_t}{\sqrt{c_t}}\right)},
\end{aligned}$$

   i.e. can be written (with knowledge of t^1) $OPT(\mu) = C + \frac{K}{\mu} + \frac{1}{\gamma - \kappa\mu}$ for some constants C, γ, K, κ. The first order condition is given by $\frac{K}{\mu^2} = \frac{\kappa}{(\gamma - \kappa\mu)^2}$ and

a minimizer can therefore be computed efficiently (either as 0, $+\infty$, or as a solution of the first order conditions—whichever leads to the smallest objective value).

The analyst only needs to pick the value of λ among the three cases above that is feasible and minimizes the objective value of Program (4.32). In practice, the analyst does not know t^1 but can search over the space of possible t^1's, which can be done in at most $|C|$ steps (note that there may be values of t^1 for which the program is infeasible, showing that said value of t^1 is impossible). The analyst can solve the optimization problem absent knowledge of t^- and t^+ by searching over (t^-, t^+) pairs, and the analyst obtains a solution by picking the (t^1, t^-, t^+) tuple for which Program (4.32) is feasible and the corresponding optimal λ that lead to the best objective value over all tuples. This can be done in a most $|C|^3$ steps.

Part 3

Societal Concerns from the Use of Data

Chapter 5

PRIVACY CONCERNS FROM THE USE OF DATA

In this chapter, we consider the question of mechanism design for acquisition of data. Similarly to Chapter 3, we consider a setting in which a data analyst wishes to compute an unbiased estimate of some underlying population statistic, by buying and aggregating data points from multiple strategic data providers. However, unlike Chapter 3, the providers hold sensitive, private data, care about the privacy losses incurred by revealing these data points to the analyst, and require their data to be treated in a formally, differentially private fashion.

There has been significant interest in buying sensitive data from individuals (Ghosh et al., 2015; Ligett et al., 2012; Roth et al., 2012; Fleischer et al., 2012; Ghosh et al., 2013; Nissim et al., 2014; Ghosh et al., 2014; Cummings et al., 2015a). This line of work considers the problem of incentivizing individuals to provide their data to an analyst, when they experience a cost—usually due to privacy loss—from sharing their data. These papers have used differential privacy, defined by Dwork et al. (2006), to combat this privacy loss, but have generally offered only a single privacy level to participants, or have made assumptions about the functional form of this privacy loss in terms of the differential privacy parameter. The work described in the current chapter, on the other hand, allows the analyst to offer each data provider a menu of different levels of differential privacy, and allows the agents to express *arbitrary* costs for each level independently. This requires no assumptions at all about the functional form of agent costs.

5.1 Preliminaries: Differential Privacy

Defining Differential Privacy

Differential privacy was introduced by (Dwork et al., 2006), and provides a formal framework for privacy, that that aims to prevent inference about individuals' data from observing the output of a computational query that involves this data. Differential privacy does so by introducing randomness in the output of a query, and by ensuring that the distribution over outputs cannot be affected much by a change in a data point of a single agent. As such, because a change in an individual's data point (almost) does not affect the outcome of the computation, it becomes to hard to identify *what* was the data that any given individual provided to the query.

We consider the case of an analyst who collects data points, that live in a universe $\mathcal{Z}$, from n individuals. To formally define differential privacy, we introduce the concept of database: a database $Z \in \mathcal{Z}^n$ is simply a vector comprised of the data of these n individuals; we let z_i denote the data point of individual i. This database Z is given as an input to a mechanism $\mathcal{M}$ that runs some query, or equivalently performs a computation, on the data in Z; notation-wise, we let $\mathcal{M}(Z)$ be the outcome of a mechanism $\mathcal{M}$ when using data from Z.

To formalize the idea of change in the data of a single individual, or equivalently in a single entry of a database, we introduce the concept of *neighboring databases*:

Definition 5.1.1 (Neighboring Databases)**.** *Two databases Z and $\hat{Z}$ are neighboring if and only if*

$$\left\| Z - \widehat{Z} \right\|_1 \triangleq \sum_{i=1}^{n} \mathbb{1}\left[z_i \neq \hat{z}_i \right] \leq 1,$$

i.e., Z and $\hat{Z}$ differ in the data point of a single individual.

With these notations and definitions on hand, we can now formally introduce the concept of differential privacy:

Definition 5.1.2 (Differential Privacy)**.** *A randomized mechanism $\mathcal{M} : \mathcal{Z}^n \to \Omega$ is ε-differentially private for some parameter $\varepsilon > 0$ if and only if for every possible subset of outputs $S \subset \Omega$, and for all neighboring databases Z and $\widehat{Z}$,*

$$\Pr\left[\mathcal{M}(Z) \in S\right] \leq \exp(\varepsilon) \cdot \Pr\left[\mathcal{M}(\widehat{Z}) \in S\right], \tag{5.1}$$

when the randomness is taken on the coin flips of mechanism $\mathcal{M}$.

ε is a parameter that allows controlling the level of privacy that is provided to the participants in the mechanism. As ε becomes smaller, Constraint 5.1 forces the distributions of outputs between neighboring databases to be closer to each other, leading in turn to stronger privacy guarantees.

The Laplace Mechanism

Now that we have defined differential privacy, we give a mechanism that is guaranteed to be differentially private for queries with numeric answers. This mechanism is called the *Laplace mechanism*. To do so, we introduce the concept of *sensitivity* of a query, i.e., how much the value of a query can change between two neighboring databases, that only differ in one entry. Formally:

Definition 5.1.3 (Sensitivity of a query)**.** *The sensitivity of a query $q : \mathcal{Z}^n \to \mathbb{R}$ is given by*

$$\Delta q = \max_{\|Z - \widehat{Z}\|_1 \leq 1} \left| q(Z) - q(\widehat{Z}) \right|$$

The Laplace Mechanism is based on the Laplace distribution, defined as follows:

Definition 5.1.4 (Laplace Distribution)**.** *The Laplace Distribution with parameter b is the distribution with probability density function*

$$Lap(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right),$$

and has mean 0 *and variance* $2b^2$.

The Laplace Mechanism for releasing the answer to a query q is then defined as follows:

Lemma 5.1.5 (Laplace Mechanism)**.** *Given any query $q : \mathcal{Z}^n \to \mathbb{R}$, the Laplace Mechanism with parameter ε is defined as*

$$\mathcal{M}(Z, q, \varepsilon) = q(Z) + Y,$$

where Y is drawn from a Laplace distribution with parameter $\frac{\Delta q}{\varepsilon}$. The Laplace Mechanism with parameter ε is ε-differentially private.

Note that as ε grows smaller, the Laplace mechanism provides stronger privacy guarantees, but the variance of the output of the mechanism (given by $2\Delta f^2/\varepsilon^2$) increases, meaning the final estimate becomes less accurate. This trade-off between accuracy and privacy is, in fact, not an artifact of the Laplace mechanism, but an unavoidable property of differential privacy. The proof of the differential privacy guarantee of the Laplace mechanism, as well as a detailed discussion of differential privacy, can be found in (Dwork et al., 2014).

5.2 Model

In this chapter, an analyst wishes to estimate the expected value μ of some statistic on the underlying population. She has access to a set of n data providers, each of which is capable of providing an unbiased estimate μ_i of the statistic of interest. We assume providers protect their data points z_i by using differential privacy—let us say via the Laplace mechanism, for simplicity—and the μ_i's are the unbiased but *noisy*

estimates reported to the analyst. Each provider incurs a cost for revealing his data point, that depends on the level of differential privacy used to protect said data point. The analyst has the ability to require a specific differential privacy level from each provider, that she chooses from a set $\{\varepsilon_1, \ldots, \varepsilon_m\}$ (she may require different privacy levels from different providers); the providers must comply and are not allowed to lie about the level of privacy they provide. Choosing lower values of ε lead the providers to incur lower privacy costs, but makes the reported data points noisier, and decreases the accuracy and variance of the analyst's estimate.

For example, each data provider might in fact be a single individual, who is selling a (possibly perturbed) bit signifying some property of interest to the data analyst (e.g., the cancer or HIV status of the individual). An individual may add noise to his data in order to guarantee a certain level of (differential) privacy, and can potentially offer the data analyst access to his data at a menu of different levels of privacy protection. The cost an individual experiences for a given privacy level is determined by his preferences for privacy; different individuals may value their privacy differently, hence incur different costs.

The goal of the analyst is to minimize the total cost among all providers, while maintaining a guarantee that the variance of her estimate is below some threshold α. To do so, the analyst designs a mechanism to buy and aggregate data from the providers. First, the mechanism ask each strategic provider i to report his privacy cost c_{ij} for each possible privacy level ε_j; a provider may misreport his costs, if he benefits from doing so. The mechanism then selects a privacy level to obtain from each provider, compensate the providers from their data, and generates an estimate for μ that is a weighted sum of the providers' reported estimates μ_i's: $\hat{\mu} = \sum_i w_i \mu_i$. The analyst wants to design a mechanism that is dominant-strategy incentive compatible and individually rational.

We assume here, as in Chapter 3, that an agent's data point is independent of his (privacy) cost for reporting this data point to the analyst. This is motivated by impossibility results of Ghosh et al. (2015) and Nissim et al. (2014), showing that when privacy costs are correlated with data, no mechanism can satisfy individual rationality and estimate the statistic of interest with non-trivial accuracy, while making finite payments.

5.3 An MIDR Mechanism for Private Data Acquisition

This problem can be cast exactly within the framework of Chapter 3. This stems from the following, simple observation: every level of differential privacy ε_j corresponds to a level of variance v_j for that same estimate, controlled by the amount of noise added for privacy. It is a simple task to translate levels of privacy to levels of variance, and we as such obtain the following result, that follows immediately from Chapter 3:

Theorem 5.3.1. *Given n data providers, such that agent i reports cost* $\{c_{ij}\}$ *for variance level* $\{v_j\}$*, and a feasible target variance level* α*, Algorithm 1 in Chapter 3 is a dominant-strategy incentive compatible mechanism that selects a minimum expected privacy cost assignment, and*

1. *for any* $\varepsilon > 0$*, computes an estimate* $\hat{\mu}$ *with variance* $Var(\hat{\mu}) \leq (1 + \varepsilon)\alpha$ *as long as*
$$n \geq \left(\frac{v_m}{v_1} - 1\right)\left(\frac{1}{\varepsilon} + 1\right),$$
2. *The mechanism is I.R. for any entrance reward* $R \geq \max_i \min_j c_{ij}$.

Chapter 6

FAIRNESS IN DECISION-MAKING

Nicole Immorlica, Katrina Ligett, and Juba Ziani (2019). "Access to Population-Level Signaling As a Source of Inequality". In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 249–258. **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** DOI: 10.1145/3287560.3287579. URL: http://doi.acm.org/10.1145/3287560.3287579.

6.1 Introduction

Settings where personal data drive consequential decisions, at large scale, abound—financial data determine loan decisions, personal history affects bail and sentencing, academic records feed into admissions and hiring. Data-driven decision-making is not reserved for major life events, of course; on a minute-by-minute basis, our digital trails are used to determine the news we see, the job ads we are shown, and the behaviors we are nudged towards.

There has been an explosion of interest recently in the ways in which such data-driven decision-making can reinforce and amplify injustices. One goal of the literature has been to identify the points in the decision-making pipeline that can contribute to unfairness. For example, are *data more noisy or less plentiful* for a disadvantaged population than for an advantaged one? Are the available *data less relevant* to the decision-making task with respect to the disadvantaged population? Has the disadvantaged population historically been *prevented or discouraged from acquiring good data profiles* that would lead to favorable decisions? Is the decision-maker simply *making worse decisions* about the disadvantaged population, despite access to data that could prevent it?

In this chapter, we study *access to population-level signaling* as a source of inequity that, to the best of our knowledge, has not received attention in the literature. We consider settings where the data of individuals in a population passes to a *population-level signaler*, and the signaler determines what function of the data is provided as a signal to a decision-maker. The signaler can serve as an advocate for the population by filtering or noising its individuals' data, but cannot outright lie to the decision-maker; whatever function the signaler chooses to map from individuals'

data to signals must be fixed and known to the decision-maker.

Examples of population-level strategic signalers include high schools, who, in order to increase the chances that their students will be admitted to prestigious universities, inflate their grades, refuse to release class rankings (Ostrovsky et al., 2010), and provide glowing recommendation letters for more than just the best students. Likewise, law firms advocate on behalf of their client populations by selectively revealing information or advocating for trial vs. plea bargains. Even the choice of advertisements we see online is based on signals about us sold by exchanges, who wish to make their ad-viewing population seem as valuable as possible.

Our interest in asymmetric information in general and in population level strategic signaling in particular are inspired by the recent wave of interest in these issues in the economics literature (see related work for an overview). In particular, the model we adopt to study these issues in the context of inequity parallels the highly influential work on Bayesian persuasion (Kamenica et al., 2011) and information design (Bergemann et al., 2019).

In order to explore the role that population-level strategic signaling can play in reinforcing inequity, we investigate its impact in a stylized model of university admissions.

We consider a setting in which a high school's information about its students is noisy but unbiased. Throughout, we call this noisy information *grades*, but emphasize that it may incorporate additional sources of information such as observations of personality and effort, that are also indicative of student quality. Importantly, all relevant information about student quality is observed directly by the school alone.

The school then aggregates each student's information into a signal about that student that is transmitted to the university. This aggregation method is called a *signaling scheme*, or informally, a (randomized) mapping from a student's information to a recommendation. A school could, for instance, choose to give the same recommendation for all its students, effectively aggregating the information about all students into one statement about average quality. Or, for example, the school could choose to provide positive recommendations to only those students that it believes, based on its information, to have high ability.

The university makes admission decisions based on these recommendations, with the goal of admitting qualified students and rejecting unqualified ones.[1] A school might

[1]In our simple model, the university does not have a fixed capacity, nor does it consider

make recommendations designed to maximize the number of their students admitted by the university. We call such a school *strategic*. Alternatively, a school might simply report the information it has collected on its students to the university directly. We call such a school *revealing*. As is common in economics, we assume that the university knows the signaling scheme chosen by the school (but does not know the realization of any randomness the school uses in its mapping). One justification typically given for such an assumption is that the university could learn this mapping over time, as it observes student quality from past years.

As expected, we find that strategic schools with accurate information about their students have a significant advantage over revealing schools, and, in the absence of intervention, strategic schools get more of their students (including unqualified ones) admitted by the university.

A common intervention in this setting is the standardized test. The university could require students to take a standardized test before being considered for admission, and use test scores in addition to the school's recommendations in an effort to enable more-informed admissions decisions. Intuitively, the role of the standardized test is that it "adds information back in" that was obfuscated by a strategic school in its recommendations, and so one might naturally expect the test to reduce inequity in the admissions process. While such a standardized test does increase the accuracy of admissions decisions, we show that when the test is a noisy estimate of student quality, it may in fact exacerbate the impact of disparities in signaling between schools.

Summary of contributions

We highlight access to strategic population level signaling, as studied in the economics literature, as a potential source of inequity. We derive the optimal signaling scheme for a school and compute the resulting school utility and false positive and negative rates in Section 6.3. We then show, still in Section 6.3, that disparities in abilities to signal strategically can constitute a non-negligible source of inequity. In Section 6.4, we study the effect of a standardized test that students must take before applying to the university, and highlight its limitations in addressing signaling-based inequity.

Related work

There is a large literature on individual-level signaling in economics, following on the Nobel-prize-winning work of Spence (1973). The general model there is quite

complementarities between students.

different from our population-level signaling model; in the Spence model, *individuals* (not populations) invest in *costly* (in terms of money or effort) signals whose costs correlate with the individual's type. In that model, equilibria can emerge where high-type individuals are more likely to invest in the signal than low-types, which can result in the signal being useful for admissions or hiring.

Closer to our setting, Ostrovsky et al. (2010) study a model in which schools provide noisy information about their students to potential employers. Their focus is on understanding properties of the equilibria of the system; they do not fully characterize the equilibria, they do not consider the role of signaling in compounding inequity, and they do not investigate the impact of interventions like our standardized test. Unlike us, they do not consider the case where the schools have imperfect observations of the students' types. Such work falls into a broader literature on optimal information structures (e.g., (Rayo et al., 2010)).

The impact of information asymmetries is a common theme in economics today, with key early work including Brocas et al. (2007). Our model of signaling is inspired by the influential work on Bayesian Persuasion (Kamenica et al., 2011), where a persuader (played, in our model, by the school) commits to revealing some fixed function of the types of the population it serves; this revelation is used as the basis of a decision that impacts the welfare of both the decider and the persuader (and the persuader's constituents). The Bayesian Persuasion model has been applied to a variety of domains, e.g. (Bergemann et al., 2015; Rabinovich et al., 2015; Xu et al., 2015; "Implementing the "Wisdom of the Crowd"" 2014; Bergemann et al., 2007; Emek et al., 2014; Johnson et al., 2006; Anderson et al., 2006), and generalizations and alternatives to this model have been studied in (Rayo et al., 2010; Arieli et al., 2016; Gentzkow et al., 2017; Alonso et al., 2016; Gentzkow et al., 2014; Kolotilin et al., 2017). Recent work (Dughmi, 2014; Dughmi, 2017; Dughmi et al., 2016; Dughmi et al., 2017; Emek et al., 2014; Guo et al., 2013; Dughmi et al., 2014) has explored algorithmic aspects of persuasion settings. To our knowledge, ours is the first work to consider access to population-level signaling, Bayesian Persuasion, or information design as a source of inequity.

Recent work on fairness has highlighted a number of objectives that one might wish to enforce when allocating resources to or making decisions about large numbers of individuals. At a high level, these objectives tend to focus either on ensuring group-level fairness (Feldman et al., 2015; Kamiran et al., 2012; Hajian et al., 2013; Hardt et al., 2016b; Friedler et al., 2016; Chouldechova, 2017; Kleinberg et al., 2016;

Zafar et al., 2017; Kearns et al., 2018) or individual-level fairness (Dwork et al., 2012; Joseph et al., 2016; Kannan et al., 2017). The metrics we study—expected utility, false positive rates, and false negative rates—are generally considered to be metrics of group fairness, but they also (coarsely) compare the extent to which similar individuals are being treated similarly.

One very interesting recent paper on fairness (Hu et al., 2017) does incorporate Spence-style individual-level signaling; in their model, a worker can choose whether and how much to invest in human capital, and this acts as an imperfect signal on whether the worker is qualified. Although their model and its implications are very different from ours, they similarly investigate the impact of upstream interventions on downstream group-level unfairness. Similar notions of individual-level signaling can also be found in (Foster et al., 1992; Coate et al., 1993).

6.2 Model and Preliminaries

We consider a setting with high schools (henceforth, "schools"), and a single university. A school has a population of students. Each student i has a binary type $t_i \in \{0, 1\}$ that represents the quality of the student. The students' types are drawn i.i.d. from a Bernoulli distribution with mean p; that is, a student has type 1 w.p. p and 0 w.p. $1 - p$. A student's type is private, that is, known to the student but unknown to both the school and the university. The prior p is public and common knowledge to all agents.

A school observes noisy information about the types of each of its students. To formally model this, we assume student i has a grade $g_i \in \{0, 1\}$, which is observed by the school but is unknown to the university.

The grade g_i for student i is drawn as follows: $\Pr[g_i = 0|t_i = 0] = \Pr[g_i = 1|t_i = 1] = q$, for $q \in [1/2, 1]$.[2] That is, the student's type is flipped with some probability $1 - q$. As q increases, the grade g_i becomes a more accurate estimate of the student's type t_i. The grade g_i is known to the school but *not* the university. The distribution q of the grade, however, is public, i.e., common knowledge to all parties.

A school has access to a (possibly trivial or uncountably infinite) set of signals Σ, and commits to a signaling scheme mapping grades g to probability distributions over signals in Σ. For each student i, the university makes an accept/reject decision based on the distribution of the types p, the distribution of the grades q, and the realization

[2]The assumption that $q \geq 1/2$ is without loss of generality; when $q < 1/2$, one can set $q = 1 - q, g_i = 1 - g_i$ and all results carry through by symmetry.

of the signal chosen by the school. The goal of the university is to maximize the quality of the students it accepts.[3] In particular, we model the university as having additive utility over the set of students it accepts, with utility 1 for accepting a student of high type ($t_i = 1$), and utility -1 for a student with low type ($t_i = 0$). We assume that the university has unlimited capacity; therefore, the university accepts exactly those students who induce non-negative expected utility given the common priors and the signal.[4] We measure a school's utility by the expected fraction of its students who are admitted to the university. We note that this choice of utility measures the *access to opportunity* (defined as admittance to university) of the school's students. We refer to a school as *revealing* if it simply transmits the grade to the university as the signal. We refer to a school as *strategic* if it employs the optimal strategic signaling scheme, as examined in Section 6.3. A strategic school thus maximizes its expected utility.

In several places, we will discuss the distribution of students accepted by the university. To do so, it is useful to introduce the notions of *false positive* and *false negative* rates. The *false positive rate* of a school is the (expected) probability that a student with type 0 is accepted by the university. The *false negative rate* of a school is the (expected) probability that a student with type 1 is rejected by the university.

We introduce several assumptions that restrict our attention to settings of interest. First, we assume the expected quality of a student is negative, such that the university would reject students without any signal from the school.

Assumption 6.2.1. *The university's expected utility for accepting any given student, absent any auxiliary information, is negative, i.e.,* $p - (1 - p) < 0$*, and therefore* $p < 1/2$.

Next we assume the university's expected utility of accepting a student with a high (resp. low) grade is positive (resp. negative).

Assumption 6.2.2. *The university has non-negative expected utility for accepting a student with a high grade, and negative expected utility for accepting a student with a low grade:*

$$\Pr[t = 1|g = 1] - \Pr[t = 0|g = 1] \geq 0;$$
$$\Pr[t = 1|g = 0] - \Pr[t = 0|g = 0] < 0.$$

[3] There is no notion here of students "applying" to the university or not; the university considers *all* students for admission.

[4] When indifferent, the university accepts the student.

These can be rewritten as:

$$pq - (1-p)(1-q) \geq 0;$$
$$p(1-q) - (1-p)q < 0.$$

We note that if the expected utility of accepting a student with a high grade were negative, then none of the school's students would be admitted by the university under any signaling scheme. On the other hand, if the expected utility of accepting a student with a low grade were positive, then the university would always accept every student.[5] Thus, this assumption restricts our analysis to the regime in which the utilities of revealing and strategic schools may differ.

The following easy consequence of these assumptions will be useful in our analysis.

Observation 6.2.3. *Under Assumption 6.2.1, Assumption 6.2.2 implies* $q \geq 1 - p$.

We conclude with the following well-known result (see, e.g., (Kamenica et al., 2011)) that an optimal signaling scheme contains, without loss of generality, at most as many signals as there are actions available to the decision-maker. In our setting, this corresponds to restricting $|\Sigma| = 2$ as the university makes an accept/reject decision for each student.

The result, reproduced below for our setting, follows from a revelation-principle type argument. The idea is to replicate the utilities of a signaling scheme with many signals by first producing a signal according to the original scheme and then simply reporting to the university, as a signal in the simplified scheme, the action σ^+ = *accept* or σ^- = *reject* that it would choose to take as a result of seeing the original signal.

Theorem 6.2.4 (Kamenica and Gentzkow (Kamenica et al., 2011))**.** *Suppose* Σ *is a measurable (but potentially uncountable) set with at least two elements. Let* Σ' *be such that* $|\Sigma'| = 2$*. Given any original signaling scheme mapping to* $\Delta(\Sigma)$*, there exists a new signaling scheme mapping to* $\Delta(\Sigma')$ *that induces the same utilities for the school and the university as those induced by the original scheme. Further, one can write* $\Sigma' = \{\sigma^-, \sigma^+\}$ *such that a student with signal* σ^+ *is accepted by the university with probability* 1*, and a student with signal* σ^- *is rejected with probability* 1.

[5] In fact, this condition is already ruled out by Assumption 6.2.1.

When $|\Sigma| = 1$, signals carry no information, making mute the question of access to signaling schemes. Therefore, throughout the chapter, we make the assumption that $|\Sigma| = 2$ and denote its elements by $\Sigma = \{\sigma^+, \sigma^-\}$. This is without loss of generality, by the argument above.

6.3 Impact of Signaling Schemes

The goal of this chapter is to highlight the role of access to strategic signaling in creating unequal access to opportunity and explore the intervention of a standardized test as a way to combat this inequity. In order to do so, we first formulate optimal signaling schemes, and then we study their impact on students and their relationship to noisy grades.

Optimal signaling scheme

We first derive the optimal signaling scheme. The idea is to pack low-quality students together with high quality students by giving both the *accept* signal σ^+. A school is limited in the extent to which it can do so, as it must ensure the university obtains non-negative expected utility by accepting all the students who have signal σ^+. The following theorem provides the right balance.

Theorem 6.3.1. *The optimal signaling scheme for a school is*

$$\Pr\left[\sigma^+ |\ g = 0\right] = \frac{p + q - 1}{q - p}$$
$$\Pr\left[\sigma^+ |\ g = 1\right] = 1.$$

Proof. As per the revelation principle in Theorem 6.2.4, we can let σ^+ be a signal such that all students with that signal are accepted by the university, and σ^- a signal such that all students with that signal are rejected. Conditional on σ^+, we can write the probabilities that a student is of each type as

$$\begin{aligned}
\Pr[t = 1|\sigma^+] =& \frac{\Pr\left[t = 1, \sigma^+\right]}{\Pr\left[\sigma^+\right]} \\
=& \Pr[t = 1] \cdot \frac{\Pr[\sigma^+|t = 1]}{\Pr[\sigma^+]} \\
=& \Pr[t = 1] \cdot \frac{\Pr[\sigma^+|g = 1] \Pr[g = 1|t = 1]}{\Pr[\sigma^+]} \\
&+ \Pr[t = 1] \cdot \frac{\Pr[\sigma^+|g = 0] \Pr[g = 0|t = 1]}{\Pr[\sigma^+]} \\
=& p \cdot \frac{q \Pr[\sigma^+|g = 1] + (1 - q) \Pr[\sigma^+|g = 0]}{\Pr\left[\sigma^+\right]}
\end{aligned}$$

and, similarly,

$$\Pr[t = 0|\sigma^+] = (1-p) \cdot \frac{(1-q)\Pr[\sigma^+|g=1] + q\Pr[\sigma^+|g=0]}{\Pr[\sigma^+]}.$$

The university's expected utility when accepting all those students with signal σ^+ is non-negative if and only if such a student is at least as likely to be of type 1 as of type 0, that is, $\Pr[t = 0|\sigma^+] \leq \Pr[t = 1|\sigma^+]$. Plugging in and rearranging, this gives the constraint

$$\begin{aligned}\Pr[\sigma^+|g=0] \cdot (q(1-p) - p(1-q)) \\ \leq \Pr[\sigma^+|g=1] \cdot (pq - (1-q)(1-p)).\end{aligned}$$

Recall that $q(1-p) - p(1-q) > 0$ by Assumption 6.2.2, and thus the constraint can be rewritten as

$$\begin{aligned}\Pr[\sigma^+|g=0] &\leq \frac{pq-(1-q)(1-p)}{q(1-p)-p(1-q)} \cdot \Pr[\sigma^+|g=1] \\ &= \frac{p+q-1}{q-p} \cdot \Pr[\sigma^+|g=1].\end{aligned}$$

The school's expected utility is

$$\Pr[\sigma^+] = \Pr[\sigma^+|g=0]\Pr[g=0] + \Pr[\sigma^+|g=1]\Pr[g=1].$$

Since $\Pr[\sigma^+|g=1]$ is unconstrained, the school's utility is maximized by setting it to 1. The school's utility is, similarly, maximized by maximizing the value of $\Pr[\sigma^+|g=0]$, which, given the constraint, occurs by setting

$$\Pr[\sigma^+|g=0] = \frac{p+q-1}{q-p} \cdot \Pr[\sigma^+|g=1] = \frac{p+q-1}{q-p}. \square$$

School's utility, false positive and false negative rates

In this section, we calculate the expected utility, false positive, and false negative rate achieved by a school, depending on the accuracy of its grades and whether it uses the optimal strategic signaling scheme when transmitting information about its students to the university. These lemmas will form the basis of our evaluation of the impacts of strategic signaling, later in Section 6.3. Recall that we refer to a school that does not strategically signal and instead transmits its raw grades to the university as *revealing*.

The proofs of the following Lemmas follow by direct calculations. We provide an exposition of the more involved calculations of Lemmas 6.3.3 and 6.3.5 in Section 6.5.

Lemma 6.3.2 (Revealing school's utility)**.** *The expected utility $U_r(p, q)$ of a revealing school is*

$$U_r(p, q) = pq + (1 - p)(1 - q).$$

For the special case of a revealing school with accurate grades (when $q = 1$), we have

$$U_r(p, 1) = p.$$

A revealing school gets exactly the students with high grades accepted, as per Assumption 6.2.2; in particular, a q fraction of high-type students will have a high grade and be accepted, while a $(1 - q)$ fraction of the low-type students will be accepted.

Lemma 6.3.3 (Strategic school's utility)**.** *A school's expected utility $U_s(p, q)$ when it signals strategically is given by*

$$U_s(p, q) = 1 + (p + q - 2pq) \cdot \frac{2p - 1}{q - p}.$$

For the special case of a strategic school with accurate grades (when $q = 1$), we have

$$U_s(p, 1) = 2p.$$

A school that signals strategically gets exactly those students with a signal of σ^+ accepted, as per the revelation principle argument of Theorem 6.2.4; a student with a high grade will be accepted with probability $\Pr[\sigma^+ | \, g = 1]$ and a student with a low grade with probability $\Pr[\sigma^+ | \, g = 0]$, with the probabilities chosen according to Theorem 6.3.1.

Lemma 6.3.4 (Revealing school's FPR/FNR)**.** *When a school is revealing, the false positive rate is given by*

$$FPR_r(p, q) = 1 - q$$

and the false negative rate by

$$FNR_r(p, q) = 1 - q.$$

For the special case of a revealing school with accurate grades (when $q = 1$), we have $FPR_r(p, 1) = FNR_r(p, 1) = 0$.

In the case of a revealing school, a low-type (resp. high-type) student obtains a low (resp. high) grade and gets rejected (resp. accepted) with probability $1 - q$, i.e., if the grade does not match the type.

Lemma 6.3.5 (Strategic school's FPR/FNR)**.** *When a school signals strategically, the false positive rate is given by*

$$FPR_s(p, q) = 1 - q + q \cdot \frac{p + q - 1}{q - p}$$

and the false negative rate by

$$FNR_s(p, q) = (1 - q)\frac{1 - 2p}{q - p}.$$

For the special case of a strategic school with accurate grades (when $q = 1$), we have $FPR_s(p, 1) = \frac{p}{1-p}$ and $FNR_s(p, 1) = 0$.

In the case of a school that signals strategically according to Theorem 6.3.1, a low-type student gets accepted with probability $\Pr[\sigma^+ |\ g = 1] = 1$ if his grade is 1 (which occurs with probability $1 - q$), and probability $\Pr[\sigma^+ |\ g = 0]$ if his grade is $g = 0$ (which occurs with probability q). On the other hand, a high-type student gets rejected when his signal is σ^-; because $\Pr[\sigma^+ |\ g = 1] = 1$, this happens only when $g = 0$ and the signal is σ^-, i.e., with probability $\Pr[\sigma^- |\ g = 0] \Pr[g = 0 | t = 1]$.

Remark 6.3.6. *While we chose to focus on average population (i.e., school) utility in this chapter, because of space constraints, one can use these derivations of FRP and FNP to calculate the welfare of subpopulations, such as low-type students at a revealing school, which then implies population-level utility comparisons as well. One interesting observation is that, using the above Lemmas and Assumptions 6.2.1 and 6.2.2, one can see that the FPR of a strategic school is* larger *and the FNR* smaller *than that of a revealing school. Thus, while it is intuitively obvious that low-type students prefer a strategic school, these calculations show that high-type students also prefer a strategic school (and the preference is strict unless the assumptions hold with equality).*

Consequences of strategic signaling for access to opportunity

In this section, we quantify the impact of access to strategic signaling and its interaction with accuracy of the information (grades) on which the signals are based. We study both the resulting expected utility of a school as well as the resulting acceptance rates of both types of students. We find that the ability to

strategically signal always has a positive (although bounded) impact, increasing students' acceptance rates and the school's expected utility. The benefit of strategic signaling for both students and the school improves (boundedly so) with the accuracy of the grades, whereas a revealing school and its students receive (potentially dramatically) higher expected utility from noisy grades.

Theorem 6.3.7. *For all $p < 1/2$ and $q > q' \geq 1 - p$, the following hold:*

- *accuracy in grades benefits strategic schools,*

$$\frac{1}{1-p} U_s(p, q') \geq U_s(p, q) \geq U_s(p, q');$$

- *strategic schools have higher expected utility than revealing schools,*

$$2U_r(p, q) \geq U_s(p, q) \geq U_r(p, q);$$

- *and accuracy in grades harms revealing schools,*

$$2(1-p)U_r(p, q) \geq U_r(p, q') \geq U_r(p, q).$$

Further, all above bounds are tight for some q, q'.

Proof. The following theorem is a direct consequence of Lemmas 6.5.1, 6.5.2, 6.5.3, 6.5.4, and 6.5.5, in Section 6.5. □

We see that, perhaps counter-intuitively, adding noise to the grades can help a revealing school get more students admitted, up to a point.[6] This follows from the fact that adding noise to the grade increases the number of students with a high grade overall, by Assumption 6.2.1, as there are more low-type students (whose representation increases as grade accuracy decreases) than high-type students (whose representation decreases as grade accuracy decreases). Adding noise to grades is, however, a blunt instrument, in that it drives up both false negatives and false positives (see Lemma 6.3.4), which limits its utility benefits. The ability to signal strategically is more subtle, driving up false positives (and expected utility), at no cost of false negatives. The power of strategic signaling is maximized when schools have access to highly accurate grades. Accurate information, the ability to control the

[6] A similar observation in a somewhat different setting was made in work of Ostrovsky and Schwarz (Ostrovsky et al., 2010).

noise level of that information, and, most notably, the ability to strategically signal about that information, therefore constitute powerful drivers of unequal access to opportunity in settings where key information is transmitted to a decision-maker on behalf of a population.

We can derive comparisons resulting in similar insights for the false positive and false negative rates of revealing and strategic schools; we do so in the full version of this work (Immorlica et al., 2019).

6.4 Intervention: Standardized Test

The prior sections show that unequal access to strategic signaling can result in unequal access to opportunity. This is driven by high error rates for students accepted from schools with signaling technologies and/or noisy grades. The university has a vested interested in decreasing this error rate as it harms the university's utility. In addition, an outside body or the university itself might be concerned about the resulting unequal access to opportunity. In this section, we explore the impact of a common intervention: the standardized test. While availability of a test score certainly can only improve the expected utility of the university,[7] we find that it has an ambiguous effect on the inequity. In particular, for a large range of parameter settings, the introduction of a test can *increase* the inequality in access to opportunity.

Augmented model

Throughout this section, we augment the model of Section 6.2 to add the requirement that each student must take a test, and the results of that test are visible both to the student's school and to the university. (The school may then incorporate the test results into its subsequent strategic behavior.)

We model the test score $s_i \in \{0, 1\}$ of student i as a noisy estimate of t_i, conditionally independent from the grade g_i, obtained as follows: $\Pr[s_i = 0|t_i = 0] = \Pr[s_i = 1|t_i = 1] = \delta$, for $\delta \in [1/2, 1]$.[8] The score s_i is public, i.e., the school and the university both observe it.

A school has access to a set of signals Σ as before, but now can design a signaling scheme $S : \{0, 1\} \times \{0, 1\} \to \Delta(\Sigma)$ that is a function of both the student's grade

[7] This is because the expected utility of the university from strategic schools without test scores is zero, and so can only increase. For revealing schools, the university gets strictly more information with test scores and hence more utility.

[8] The assumption that $\delta \geq 1/2$ is, as with our analogous assumption about the grades, without loss of generality.

and his test score, i.e., the school designs $\Pr[S|\, g_i, s_i]$ for $\sigma \in \Sigma$. The university again makes accept/reject decisions that maximize its expected utility, but now the university has access to the test score s_i and its distribution δ as well as the signal and the distributions p and q. As before, a *strategic* school chooses a signaling scheme that maximizes the fraction of students accepted whereas a *revealing* school simply transmits the grade to the university as the signal.

As in Section 6.2, we introduce an assumption controlling the noise δ of the test.

Assumption 6.4.1. *The university has non-negative expected utility for accepting a student with a high test score, and negative expected utility for accepting a student with a low test score:*

$$0 \leq p\delta - (1-p)(1-\delta)$$
$$0 > p(1-\delta) - (1-p)\delta.$$

We note that if the expected utility of accepting a student with a high test score were negative, or the expected utility of accepting a student with a low test score were positive, then in the absence of signals, the university would always accept either none or all of the students. Note that regimes when the standardized test is uninformative on its own but becomes informative when coupled with grades may still be interesting. However, even under Assumption 6.4.1, which excludes certain parameter ranges from consideration, we have a rich enough model to illustrate our main findings. In the full version of this work (Immorlica et al., 2019), we show how to relax this assumption, and how doing so affects the optimal signaling scheme.

The following consequence will be useful in our analysis:

Observation 6.4.2. *Under Assumption 6.2.1, Assumption 6.4.1 implies*

$$\delta \geq 1 - p.$$

Fixing p, we denote by $u_{q,\delta}(g, s)$ the expected utility the university derives from admitting a student with score s and grade g:

$$u_{q,\delta}(g, s) := \Pr[t_i = 1|\, g, s] - \Pr[t_i = 0|\, g, s].$$

When $\delta = q = 1$, $u_{q,\delta}(s, g)$ is not defined for $s \neq g$ as in this case s and g are perfectly correlated. For notational convenience, we define $u_{q,\delta}(s, g) = -1$ in these cases.

Lemma 6.4.3. *Assumptions 6.2.2 and 6.4.1 together imply that the university receives non-negative expected utility from accepting a student with both a high grade and a high score, and negative expected utility from a student with both a low grade and a low score:*

$$u_{q,\delta}(1,1) \geq 0 > u_{q,\delta}(0,0).$$

This can be rewritten as

$$pq\delta - (1-p)(1-q)(1-\delta) \geq 0;$$
$$p(1-q)(1-\delta) - (1-p)q\delta < 0.$$

Theorem 6.2.4 (the revelation principle) also holds in this setting, and so we assume for the remainder of this section that $\Sigma = \{\sigma^-, \sigma^+\}$, without loss of generality.

Optimal signaling

We first derive the optimal strategic signaling scheme. Again, a school would like to pack low-quality students together with high quality students, but is now limited in its ability to do so by their test scores. If the expected utility the university receives from a student with a high grade but low test score is negative ($u_{q,\delta}(1,0) < 0$), then this student (and in fact any student with a low test score) will be rejected regardless of the signal from the school. Otherwise ($u_{q,\delta}(1,0) \geq 0$), the school can signal to the university to accept such a student, and can additionally pack in some low-grade-low-score students, subject to maintaining non-negative expected utility for the university.

Theorem 6.4.4. *The optimal signaling scheme for a school with access to grades and a test score, under Assumption 6.4.1, is*

$$\Pr\left[\sigma^+ \middle|\, g=1, s=1\right] = 1$$
$$\Pr\left[\sigma^+ \middle|\, g=0, s=1\right] = 1$$
$$\Pr\left[\sigma^+ \middle|\, g=1, s=0\right] = \begin{cases} 1, & \text{if } u_{q,\delta}(1,0) \geq 0 \\ 0, & \text{if } u_{q,\delta}(1,0) < 0 \end{cases}$$
$$\Pr\left[\sigma^+ \middle|\, g=0, s=0\right] = \begin{cases} \frac{pq(1-\delta)-(1-p)(1-q)\delta}{(1-p)q\delta-p(1-q)(1-\delta)}, & \text{if } u_{q,\delta}(1,0) \geq 0 \\ 0, & \text{if } u_{q,\delta}(1,0) < 0 \end{cases}$$

Proof. We defer the proof to Section 6.6. □

School's utility, false positive and false negative rates

In this section, we calculate the expected utility achieved by both a strategic school and a revealing school as a function of the type distribution, the accuracy of its grades, and the accuracy of the standardized test score. We defer all proofs to Section 6.6.

For a revealing school, the university always accepts high-grade high-score students. If high grades are more informative than low test scores (that is, if $u_{q,\delta}(1,0) \geq 0$, which depends on p as well as q and δ and happens, for instance, if $p = 1/4$, $q = 9/10$, and $\delta = 7/10$), then the university also accepts students with low test scores, benefiting the school. Alternatively, if high test scores are more informative than low grades (i.e., $u_{q,\delta}(0,1) \geq 0$), then the university also accepts students with low grades. These conditions provide additional boosts to the utility of a revealing school.

Lemma 6.4.5 (Revealing school's utility)**.** *The expected utility $U_r(p,q,\delta)$ of a revealing school with access to grades and a test score is*

$$\begin{aligned} U_r(p,q,\delta) = {} & pq\delta + (1-p)(1-q)(1-\delta) \\ & + \mathbb{1}_1\left[u_{q,\delta}(1,0) \geq 0\right](pq(1-\delta) + (1-p)(1-q)\delta) \\ & + \mathbb{1}_1\left[u_{q,\delta}(0,1) \geq 0\right](p(1-q)\delta + (1-p)q(1-\delta)). \end{aligned}$$

For the special case of a revealing school with accurate grades (when $q = 1$), we have

$$U_r(p,1,\delta) = p.$$

As illustrated in Figure 6.1, for fixed p and δ, $U_r(p,q,\delta)$ may not be a decreasing function of q. In fact, when q is small enough, the grades are completely uninformative and the university only admits students with a test score of 1. In that regime, the expected utility for a revealing school is therefore constant in q. For intermediate values of q, the grades are still uninformative on their own but are informative coupled with a high standardized test score; at this point, only students with both a score and a grade of 1 get admitted by the university, and the school's expected utility suddenly drops when compared to smaller q. The school's expected utility in that regime is increasing in q as, under Assumption 6.4.1, increasing the value of q increases the fraction of students with both high scores and high grades. Finally, when q is large enough, the grades are significant enough on their own that only students with high grades are admitted; this leads to a jump in expected utility compared to the intermediate regime. In this regime for high values of q, the school's expected utility

is decreasing as a result of the fact that increasing the value of q now decreases the number of students with a high grade by Assumption 6.2.1, as seen in Section 6.3.

Lemma 6.4.6 (Strategic school's utility)**.** *The expected utility $U_s(p, q)$ when a school signals strategically and $u_{q,\delta}(1, 0) < 0$ is*

$$U_s(p, q, \delta) = p\delta + (1 - p)(1 - \delta);$$

when $u_{q,\delta}(1, 0) \geq 0$, the expected utility is

$$\begin{aligned} U_s(p, q, \delta) = (1 - p(1 - q)(1 - \delta) - (1 - p)q\delta) \\ + (p(1 - q)(1 - \delta) + (1 - p)q\delta)\, \frac{pq(1 - \delta) - (1 - p)(1 - q)\delta}{(1 - p)q\delta - p(1 - q)(1 - \delta)}. \end{aligned}$$

For the special case of a strategic school with accurate grades (when $q = 1$), we have

$$U_s(p, 1, \delta) = 1 - \delta + p.$$

The expected utility of a strategic school is, unsurprisingly, monotone in q (as illustrated in Figure 6.1), as higher-quality information about its students' types allows the school to signal more effectively. For small and intermediate values of q (i.e., insignificant grades), the university bases admission decisions solely on the standardized test score and only admits students with a score of 1 (it has positive expected utility from doing so, by Assumption 6.4.1); in this regime, a strategic school's expected utility is hence constant. When q becomes large enough, i.e., when the grades are significant enough, the university starts having positive expected utility from admitting students with a high grade even if they have a low score, and the school can start bundling these students together with the high score students, leading to a jump in its expected utility. The plotted parameters for the figures are chosen to satisfy Assumptions 6.2.1, 6.2.2, and 6.4.1; the discontinuities occur at q such that $u_{q,\delta}(0, 1) = 0$ and $u_{q,\delta}(1, 0) = 0$.

We also calculate the false positive and false negative rates of strategic and revealing schools; we defer this derivation to (Immorlica et al., 2019).

Impact of Standardized Test

With a perfect standardized test or, in fact, a sufficiently good one, (i.e., high enough δ), it is not hard to see that the university accepts exactly those students with a high test score from strategic as well as revealing schools. Thus, no matter the accuracy of the grades or distribution of types, the standardized test results in equal expected

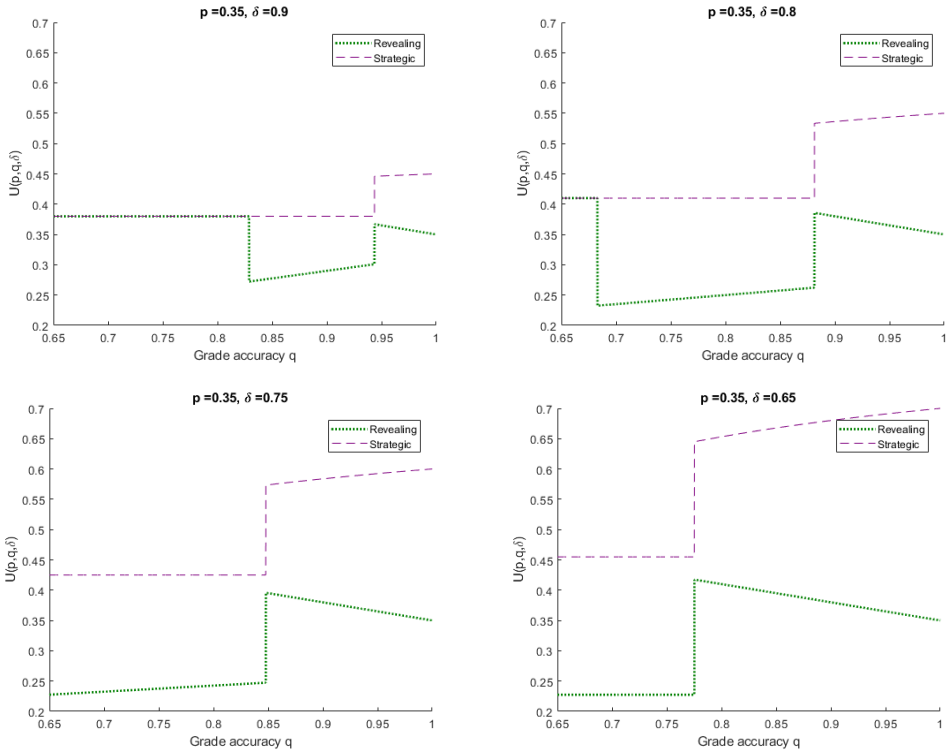

Figure 6.1: Strategic school utility $U_s(p, q, \delta)$ and revealing school utility $U_r(p, q, \delta)$ as a function of the grade accuracy q, for average student type $p = 0.35$. We observe that the expected utility may be non-monotone in q.

utility, and hence equal access to opportunity, for revealing and strategic schools (see Section 6.7 for details). Similarly, if grades are accurate (i.e., $q = 1$), then a revealing school's expected utility is fixed at p whereas a strategic school's expected utility is only diminished (from $2p$ without the test) by the extra constraints introduced by a standardized test. Thus, in this case as well, a standardized test decreases the inequality between the utilities of a strategic and a revealing school, making the ratio of utilities less than 2 (see Section 6.7 for details).

Figure 6.2 plots $U_s(p, q)/U_r(p, q)$, with and without test scores, as a function of q, for $p = 0.35$ and different values of δ. The form of the utility ratio between a strategic and a revealing school in the absence of a test score follows from the fact that both utilities are continuous, and that the expected utility of a strategic school increases while that of a revealing school decreases in q, as we have seen in Section 6.3. The form in the presence of a test score can be explained as follows. First, when in the regime of small values of q, only students with a high standardized test score are admitted by the university, in which case admission decisions do not depend

on how the schools act and both the strategic school and the revealing school have the same expected utility, leading to a ratio of 1. For intermediate values of q, we have previously discussed that the utility for a strategic school remains constant (the university still has positive utility for students with a score of 1 and the strategic school can bundle all such students together, regardless of grade), while the utility for a revealing school suddenly drops (only students with both a high grade and a high score are admitted) and is increasing in q, explaining the sudden drop in ratio of utilities at the change of regime, and the decreasing monotonicity of the ratio in q within the intermediate regime. When q becomes large enough, we have seen that both the revealing and the strategic school experience a jump in utilities, which explains the second discontinuity in the ratio of utilities. Because the revealing school has significantly lower utility than the strategic school for intermediate values of q, the relative jump in the utility of a revealing school is higher than the relative jump in utility of a strategic school. Because in the regime with high values of q, the utility of a strategic school is increasing and that of a revealing school is decreasing, the ratio of utilities is increasing.

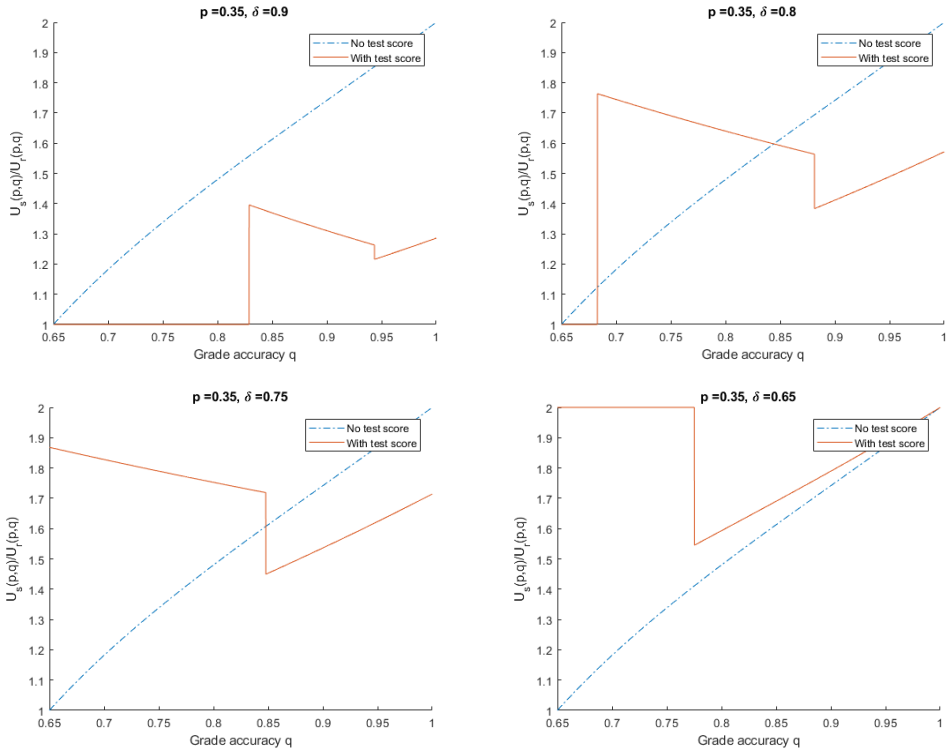

Figure 6.2: The ratio $U_s(p,q)/U_r(p,q)$ of utilities of a strategic school vs. a revealing school, as a function of the grade accuracy q, with and without test score. We observe that the test score intervention may increase inequality.

Interestingly, we observe that the introduction of a standardized test does not always decrease inequity. For noisy grades, when the test score is also sufficiently noisy, the test may have the effect of increasing the ratio of utilities between a strategic school and a revealing school. This is clearly illustrated in Figure 6.2, where the curve with test scores sometimes lies above that without a test. Some intuition for this result is as follows. In the regime for intermediate values of q, as q becomes more and more inaccurate, the ratio of utilities in the presence of a standardized test increases and eventually overtakes the ratio in the absence of a standardized test (which decreases to 1 as the grades become more inaccurate). In the regime for high values of q, the university admits students with a high grade only, independently of what their standardized test scores are; therefore, the utility of a revealing school is the same with or without a standardized test. On the other hand, when the standardized test score becomes more inaccurate, the strategic school can take advantage of the noise in said score to bundle in more students than if there was no standardized test: the university loses in utility from accepting unqualified students with high scores, but at the same time gains in utility from accepting qualified students with low scores, allowing a strategic school to bundle more students when compared to the case with no standardized test. As δ decreases and the standardized test becomes less and less accurate, a strategic school starts losing fewer high-score students to rejection than it gains in admitted low-score students, and its utility increases.

6.5 Proofs: Model Without Standardized Test

Proof of Lemma 6.3.3.

$$\begin{aligned} U_s(p,q) =& \Pr[\sigma^+|g=1]\,(pq+(1-p)(1-q)) + \Pr[\sigma^+|g=0]\,(p(1-q)+q(1-p)) \\ =&(1+2pq-p-q) + (p+q-2pq)\cdot\frac{p+q-1}{q-p} \\ =&1+(p+q-2pq)\cdot\frac{2p-1}{q-p}. \end{aligned}$$

□

Proof of Lemma 6.3.5.

$$\begin{aligned} FPR_s(p,q) =& \Pr[\sigma^+|g=1]\Pr[g=1|t=0] + \Pr[\sigma^+|g=0]\Pr[g=0|t=0] \\ =&1-q+q\cdot\frac{p+q-1}{q-p} \end{aligned}$$

$$\begin{aligned} FNR_s(p,q) &= \big(1 - \Pr[\sigma^+|g=1]\big)\Pr[g=1|t=1] \\ &\quad + \big(1 - \Pr[\sigma^+|g=0]\big)\Pr[g=0|t=1] \\ &= (1-q)\left(1 - \frac{p+q-1}{q-p}\right) \\ &= (1-q)\frac{1-2p}{q-p}. \end{aligned}$$

□

Lemma 6.5.1. *Suppose $p < 1/2$. Then $U_s(p,q)$ and $FPR_s(p,q)$ are increasing functions of $q \in [1-p, 1]$. On the other hand, $U_r(p,q)$, $FPR_r(p,q)$, $FNR_r(p,q)$ and $FNR_s(p,q)$ are decreasing functions of $q \in [1-p, 1]$.*

Proof. We first consider the expected utility of a strategic school.

$$\begin{aligned} \frac{\partial U_s}{\partial q}(p,q) &= \frac{2p-1}{(q-p)^2}\left((1-2p)(q-p) - (p+q-2pq)\right) \\ &= \frac{2p-1}{(q-p)^2}\left(q - p - 2pq + 2p^2 - p - q + 2pq\right) \\ &= 2\frac{2p-1}{(q-p)^2}p(p-1) \\ &> 0, \end{aligned}$$

since $p < 1$ and by Assumption 6.2.2, $2p - 1 < 0$. Therefore, $U_s(p,q)$ is increasing in q.

We next consider the FPR of a strategic school.

$$\frac{\partial FPR_s}{\partial q}(p,q) = \frac{2p(q-p) - (2pq-p)}{(q-p)^2} = \frac{p - 2p^2}{(q-p)^2} > 0$$

as $p < 1/2$ implies $p - 2p^2 = p(1-2p) > 0$.

$U_r(p,q) = pq + (1-p)(1-q) = (2p-1)q + 1 - p$ is decreasing in q as $2p - 1 < 0$. $FPR_r(p,q) = FNR_r(p,q) = 1 - q$ are immediately decreasing in q. $FNR_s(p,q) = (1-q)\frac{1-2p}{q-p}$ is decreasing in q as $\frac{1-q}{q-p}$ is decreasing in q and $1 - 2p > 0$. □

Lemma 6.5.2. *For a revealing school, the impact on expected utility of moving between noisy and accurate grades is quantified by*

$$\frac{1}{2(1-p)} \leq \frac{U_r(p,1)}{U_r(p,q)} \leq 1.$$

A revealing school maximizes its expected utility by setting $q = 1 - p$.

The impact on the FPR and the FNR, when $q \neq 1$, is quantified by

$$FPR_r(p,1) = FNR_r(p,1) = 0,\ FPR_r(p,q) = FNR_r(p,q) = 1 - q.$$

Further, all above bounds are tight for some q.

Proof. For a revealing school, $U_r(p,q) = pq + (1-p)(1-q)$. $U_r(p,q) = pq + (1-p)(1-q) = q(2p-1)+(1-p)$ is a decreasing function of q, so under Assumption 6.2.2 that $pq \geq (1-p)(1-q)$, a revealing school's expected utility is maximized when $pq = (1-p)(1-q)$, i.e., when $q = 1-p$ and minimized when $q = 1$. It is therefore the case that

$$\frac{U_r(p,1)}{U_r(p,q)} \geq \frac{U_r(p,1)}{U_r(p,1-p)} = \frac{p}{2p(1-p)} = \frac{1}{2(1-p)}.$$

The result for false positive and negative rates follow immediately from the fact that they are 0 for accurate and $1-q$ for noisy grades. □

Lemma 6.5.3. *For an strategically signaling school, the impact on expected utility of moving between noisy and accurate grades is quantified by*

$$1 \leq \frac{U_s(p,1)}{U_s(p,q)} \leq \frac{1}{1-p} < 2.$$

An strategically signaling school maximizes its expected utility when $q = 1$.

The impact on the FPR is

$$1 \leq \frac{FPR_s(p,1)}{FPR_s(p,q)} \leq \frac{1}{1-p} < 2.$$

The impact on the FNR, for $q \neq 1$, is

$$FNR_s(p,1) = 0,\ FNR_s(p,q) = 1-q.$$

Further, all above bounds are tight for some q.

Proof. The expected utility of a strategic school with noisy grades is

$$U_s(p,q) = 1 + (p+q-2pq)\cdot\frac{2p-1}{q-p}$$

with $U_s(p,1) = 2p$. Because $U_s(p,q)$ is increasing in q by Lemma 6.5.1, we have that $U_s(p,q) \leq U_s(p,1)$ and $\frac{U_s(p,1)}{U_s(p,q)} \geq 1$. Further, $q \geq 1-p$ implies

$$U_s(p,q) \geq U_s(p,1-p) = 1 + (1-2p(1-p))\cdot\frac{2p-1}{1-2p} = 2p(1-p).$$

Therefore,

$$\frac{U_s(p,1)}{U_s(p,q)} \leq \frac{1}{1-p} < 2,$$

recalling that by Assumption 6.2.1, $1 - p > 1/2$. The ratio of false negative rates is exactly 0, as the false negative rate is 0 when $q = 1$ and non-zero when $q \neq 1$. The false positive rate for accurate grades is $FPR_s(p, 1) = \frac{p}{1-p}$. For noisy grades, $FPR_s(p, q)$ is increasing in q by Lemma 6.5.1, and it must be the case that $FPR_s(p, q) \leq FPR_s(p, 1)$. Further,

$$FPR_s(p, q) \geq FPR_s(p, 1 - p) = \frac{2p(1 - p) - p}{1 - p - p} = \frac{p - 2p^2}{1 - 2p} = p.$$

Hence, as $FPR_s(p, 1) = \frac{p}{1-p}$ we have that

$$\frac{FPR_s(p, 1)}{FPR_s(p, q)} \leq \frac{1}{1 - p} < 2,$$

where the last inequality follows from $p < 1/2$. □

Lemma 6.5.4. *For a school with accurate grades, the impact on expected utility of introducing strategic signaling is*

$$\frac{U_s(p, 1)}{U_r(p, 1)} = 2.$$

The impact on the false positive rate is

$$FPR_s(p, 1) = \frac{p}{1 - p},\ FPR_r(p, 1) = 0.$$

The impact on the false negative rate is

$$FPR_s(p, 1) = FNR_r(p, 1) = 0.$$

The optimal signaling scheme doubles the expected utility of the school by increasing its false positive rate from 0 *to* $\frac{p}{1-p}$*, and keeping its false negative rate constant at* 0*.*

Proof. This follows immediately from the fact that $U_s(p, 1) = 2p$, $FPR_s(p, 1) = \frac{p}{1-p}$, $FNR_s(p, 1) = 0$, $U_r(p, 1) = p$, and $FPR_r(p, 1) = FNR(p, 1) = 0$. □

Lemma 6.5.5. *For a school with noisy grades, the impact on expected utility of introducing strategic signaling is*

$$1 \leq \frac{U_s(p, q)}{U_r(p, q)} \leq 2,$$

and is increasing in q.

The impact on the false positive rate is

$$1 \leq \frac{FPR_s(p,q)}{FPR_r(p,q)} \leq +\infty,$$

and is increasing in $q \in [p-1, 1]$. *The impact on the false negative rate is*

$$\frac{1-2p}{1-p} \leq \frac{FNR_s(p,q)}{FNR_r(p,q)} \leq 1$$

and is decreasing in $q \in [p-1, 1]$. *Further, all above bounds are tight for some* q.

Proof. For a revealing school with noisy grades, the expected utility $U_r(p,q) = pq + (1-p)(1-q) = q(2p-1) + (1-p)$ and the false positive rate $FPR_r(p,q) = 1-q$ are decreasing in q (as $2p - 1 < 0$ by Assumption 6.2.1). By Lemma 6.2.3, we have that $q \geq 1 - p$ and it must be that

$$2p(1-p) = U_r(p, 1-p) \geq U_r(p,q) \geq U_r(p,1) = p.$$

For an strategically signaling school, the expected utility $U_s(p,q)$ is increasing in q by Lemma 6.5.1, hence

$$2p(1-p) = U_s(p, 1-p) \leq U_s(p,q) \leq U_s(p,1) = 2p.$$

The ratio of expected utilities is therefore increasing, and satisfies

$$1 \leq \frac{U_s(p,q)}{U_r(p,q)} \leq 2.$$

The false positive rate $FPR_s(p,q)$ is increasing in q also by Lemma 6.5.1, hence $\frac{FPR_s(p,q)}{FPR_n(p,q)}$ is increasing in q. As $FPR_s(p, 1-p) = p$, $FPR_s(p,q) = p$, $FPR_s(p,1) = \frac{p}{1-p}$ and $FPR_r(p,1) = 0$,

$$1 = \frac{FPR_s(p,1-p)}{FPR_r(p,1-p)} \leq \frac{FPR_s(p,q)}{FPR_n(p,q)} \leq \frac{FPR_s(p,1)}{FPR_r(p,1)} = +\infty.$$

$FNR_s(p,q) = (1-q)\frac{1-2p}{q-p}$ and $FNR_r(p,q) = 1-q$, hence the ratio of false negative rates for $q \neq 1$ is given by

$$H(p,q) = \frac{FNR_s(p,q)}{FNR_r(p,q)} = \frac{1-2p}{q-p},$$

which is a decreasing function of q, and we have

$$1 = H(p, 1-p) \geq \frac{FNR_s(p,q)}{FNR_r(p,q)} \geq \lim_{q \to 1} H(p,q) = \frac{1-2p}{1-p}.$$ □

6.6 Proofs: Model With Standardized Test

Proof of Lemma 6.4.3. By Observation 6.2.3 and Assumption 6.2.1, $q \geq 1-p > 1/2 > 1-q$. Therefore, $pq\delta - (1-p)(1-q)(1-\delta) \geq q\left(p\delta - (1-p)(1-\delta)\right)$, which is non-negative by Assumption 6.4.1, and $p(1-q)(1-\delta) - (1-p)q\delta < (1-q)(p(1-\delta) - (1-p)\delta$, which is negative by Assumption 6.4.1. The rest of the proof follows from the fact that

$$\begin{aligned} u_{q,\delta}(1,1) &= \frac{\Pr\left[t=1, g=1, s=1\right]}{\Pr\left[g=1, s=1\right]} - \frac{\Pr\left[t=0, g=1, s=1\right]}{\Pr\left[g=1, s=1\right]} \\ &= \frac{pq\delta - (1-p)(1-q)(1-\delta)}{\Pr\left[g=1, s=1\right]} \end{aligned}$$

and

$$\begin{aligned} u_{q,\delta}(0,0) &= \frac{\Pr\left[t=1, g=0, s=0\right]}{\Pr\left[g=0, s=0\right]} - \frac{\Pr\left[t=0, g=0, s=0\right]}{\Pr\left[g=0, s=0\right]} \\ &= \frac{p(1-q)(1-\delta) - (1-p)q\delta}{\Pr\left[g=0, s=0\right]} \end{aligned}$$

□

Proof of Theorem 6.4.4. The revelation principle of Lemma 6.2.4 can be extended to the current setting via a nearly identical proof. Therefore, as before, we design σ^+ and σ^- so that every student with signal σ^+ is accepted by the university, and every student with signal σ^- is rejected. The school's goal is then to maximize the probability of a student having signal σ^+, under the constraint that the university gets expected non-negative expected utility from students with signal σ^+, regardless of their score.

We first consider the case in which $s = 1$:

$$\begin{aligned} \Pr\left[t=1|\sigma^+, s=1\right] &= \frac{\Pr\left[\sigma^+, s=1|t=1\right]\Pr\left[t=1\right]}{\Pr\left[\sigma^+, s=1\right]} \\ &= p \cdot \frac{q\delta\Pr[\sigma^+|g=1, s=1] + (1-q)\delta\Pr[\sigma^+|g=0, s=1]}{\Pr\left[\sigma^+, s=1\right]} \end{aligned}$$

We also have that, by similar calculations:

$$\begin{aligned} \Pr\left[t=0|\sigma^+, s=1\right] =& (1-p)\cdot\frac{(1-q)(1-\delta)\Pr[\sigma^+|g=1, s=1]}{\Pr\left[\sigma^+, s=1\right]} \\ &+ (1-p)\cdot\frac{q(1-\delta)\Pr[\sigma^+|g=0, s=1]}{\Pr\left[\sigma^+, s=1\right]}. \end{aligned}$$

Therefore, the university's expected utility for accepting a student with $(\sigma^+, s = 1)$ is non-negative if and only if

$$\begin{aligned}
&pq\delta \Pr[\sigma^+|g = 1, s = 1] + p(1-q)\delta \Pr[\sigma^+|g = 0, s = 1] \\
&\geq (1-p)(1-q)(1-\delta) \Pr[\sigma^+|g = 1, s = 1] \\
&+ (1-p)q(1-\delta) \Pr[\sigma^+|g = 0, s = 1],
\end{aligned}$$

which can be rewritten to give the constraint

$$\begin{aligned}
&\Pr[\sigma^+|g = 1, s = 1] \left(pq\delta - (1-p)(1-q)(1-\delta)\right) \\
&\quad \geq \Pr[\sigma^+|g = 0, s = 1] \left((1-p)q(1-\delta) - p(1-q)\delta\right).
\end{aligned}$$

By Assumption 6.4.1,

$$\begin{aligned}
0 &\leq p\delta - (1-p)(1-\delta) \\
&= \left(pq\delta - (1-p)(1-q)(1-\delta)\right) \\
&\quad - \left((1-p)q(1-\delta) - p(1-q)\delta\right),
\end{aligned}$$

and hence the constraint does not bind, and we are free to set

$$\Pr[\sigma^+|g = 1, s = 1] = \Pr[\sigma^+|g = 0, s = 1] = 1.$$

We now consider the case in which the signal is σ^+ and the score is $s = 0$. Similar calculations to the $s = 1$ case show that the university's expected utility for accepting a student with such a score and signal is non-negative iff

$$\begin{aligned}
&\Pr[\sigma^+|g = 1, s = 0] \left(pq(1-\delta) - (1-p)(1-q)\delta\right) \\
&\quad \geq \Pr[\sigma^+|g = 0, s = 0] \left((1-p)q\delta - p(1-q)(1-\delta)\right). \qquad (6.1)
\end{aligned}$$

Note that by Lemma 6.4.3, $(1-p)q\delta - p(1-q)(1-\delta) \geq 0$.

We split up the case in which $s = 0$ into two sub-cases on $u_{q,\delta}(1, 0)$.

When $u_{q,\delta}(1, 0) \geq 0$, then $pq(1-\delta) - (1-p)(1-q)\delta \geq 0$.

Therefore, to maximize its expected utility, the school should set

$$\begin{aligned}
&\Pr[\sigma^+|g = 1, s = 0] = 1, \\
&\Pr[\sigma^+|g = 0, s = 0] = \min\left(1, \frac{pq(1-\delta) - (1-p)(1-q)\delta}{(1-p)q\delta - p(1-q)(1-\delta)}\right).
\end{aligned}$$

Because by Assumption 6.4.1, $p(1-\delta) - (1-p)\delta < 0$, it must be the case that

$$\frac{pq(1-\delta) - (1-p)(1-q)\delta}{(1-p)q\delta - p(1-q)(1-\delta)} < 1,$$

and the school therefore optimizes its expected utility with

$$\Pr[\sigma^+ | g = 0, s = 0] = \frac{pq(1-\delta) - (1-p)(1-q)\delta}{(1-p)q\delta - p(1-q)(1-\delta).}$$

When $u_{q,\delta}(1,0) < 0$, then the left-hand side of Equation (6.1) is non-positive. The right-hand side non-negative, so

$$\Pr[\sigma^+ | g = 1, s = 0] = \Pr[\sigma^+ | g = 0, s = 0] = 0$$

is required for the inequality to hold. □

Proof of Lemma 6.4.5. By Lemma 6.4.3, $u_{q,\delta}(1,1) \geq 0$, and hence the university accepts students with $g = 1, s = 1$, which occurs with probability $pq\delta + (1-p)(1-q)(1-\delta)$. Similarly, $u_{q,\delta}(0,0) < 0$, so the university rejects students with $g = 0,\ s = 0$. The case $g = 0,\ s = 1$ happens with probability $p(1-q)\delta + (1-p)q(1-\delta)$, and yields the school expected utility 1 iff $u_{q,\delta}(0,1) \geq 0$ (i.e., the university accepts students with $g = 0,\ s = 1$). Similarly, $g = 1,\ s = 0$ happens with probability $pq(1-\delta) + (1-p)(1-q)\delta$ and yields the school expected utility 1 iff $u_{q,\delta}(1,0) \geq 0$.

The second part of the proof has two cases.

When $q = 1$ and $\delta \neq 1$, then $u_{q,\delta}(1,0) = pq(1-\delta) - (1-p)(1-q)\delta = p(1-\delta) \geq 0$. Also, $u_{q,\delta}(0,1) = p(1-q)\delta - (1-p)q(1-\delta) = -(1-p)(1-\delta) < 0$ (recalling that $1-p > 0$ by Assumption 6.2.1 and that $1-\delta > 0$). Therefore, for $\delta \neq 1$,

$$\begin{aligned} U_r(p,1,\delta) =& pq\delta + (1-p)(1-q)(1-\delta) + pq(1-\delta) + (1-p)(1-q)\delta \\ =& p\delta + p(1-\delta) \\ =& p. \end{aligned}$$

When $q = 1$ and $\delta = 1$, then $u_{q,\delta}(1,0) = pq(1-\delta) - (1-p)(1-q)\delta = p(1-\delta) \geq 0$. Also, $u_{q,\delta}(0,1) = p(1-q)\delta - (1-p)q(1-\delta) = 0$. Therefore,

$$\begin{aligned} U_r(p,1,1) =& pq\delta + (1-p)(1-q)(1-\delta) \\ &+ pq(1-\delta) + (1-p)(1-q)\delta \\ &+ p(1-q)\delta + (1-p)q(1-\delta) \\ =& p. \end{aligned}$$

This concludes the proof. □

Proof of Lemma 6.4.6. When $u_{q,\delta}(1,0) < 0$, then

$$\begin{aligned} U_s(p,q,\delta) &= \Pr[s=1] \\ &= p\delta + (1-p)(1-\delta). \end{aligned}$$

When $u_{q,\delta}(1,0) \geq 0$, $\Pr[\sigma^+|g=1,s=1] = \Pr[\sigma^+|g=1,s=0] = \Pr[\sigma^+|g=0,s=1] = 1$, and we have

$$\begin{aligned} &U_s(p,q,\delta) \\ &= 1 - \Pr[g=0,s=0] + \Pr[g=0,s=0]\Pr[\sigma^+|g=0,s=0] \\ &= (1-p(1-q)(1-\delta) - (1-p)q\delta) \\ &\quad + (p(1-q)(1-\delta) + (1-p)q\delta)\frac{pq(1-\delta)-(1-p)(1-q)\delta}{(1-p)q\delta - p(1-q)(1-\delta)}. \end{aligned}$$

When $q = 1$, $u_{q,\delta}(1,0) \geq 0$ (equivalently, $p(1-\delta) \geq 0$), and the expected utility is

$$\begin{aligned} U_s(p,1,\delta) &= (1-(1-p)\delta) + ((1-p)\delta)\frac{p(1-\delta)}{(1-p)\delta} \\ &= 1-(1-p)\delta + p(1-\delta) \\ &= 1-\delta+p. \end{aligned}$$

□

6.7 Supplemetary Material: Impact of Standardized Test

Lemma 6.7.1. *Fix $p > 0$ and $q < 1$. For*

$$\delta > \max\left(\frac{pq}{(pq+(1-p)(1-q)}, \frac{(1-p)q}{p(1-q)+(1-p)q}\right),$$

we have

$$\frac{U_s(p,q,\delta)}{U_r(p,q,\delta)} = 1.$$

Proof. $pq(1-\delta) - (1-p)(1-q)\delta < 0$ and $p(1-q)\delta - (1-p)q(1-\delta) \geq 0$, hence $u_{q,\delta}(1,0) < 0$ and $u_{q,\delta}(0,1) \geq 0$. Therefore, a school's expected utility for revealing is given by

$$pq\delta + (1-p)(1-q)(1-\delta) + p(1-q)\delta + (1-p)q(1-\delta) = p\delta + (1-p)(1-\delta)$$

This is exactly the expected utility of a school that is strategically signaling when $q = 1$, hence the ratio of utilities when strategic over revealing is 1. □

Lemma 6.7.2. *When the grades are accurate, i.e., $q = 1$,*

$$1 \leq \frac{U_s(p,1,\delta)}{U_r(p,1,\delta)} \leq 2,$$

Proof. The expected utility from strategically reporting is $p + (1-p)(1-\delta) + p(1-\delta)$ by Lemma 6.4.6, while the expected utility for revealing is p by Lemma 6.4.5. Therefore,

$$\frac{U_s(p, 1, \delta)}{U_r(p, 1, \delta)} = 2 - \delta + \frac{1-p}{p}(1-\delta) \geq 1.$$

By Assumption 6.2.2, $(1-p)(1-\delta) \leq p\delta$, hence

$$\frac{U_s(p, 1, \delta)}{U_r(p, 1, \delta)} = 2 - \delta + \frac{1-p}{p}(1-\delta) \leq 2 - \delta + \delta = 2. \qquad \square$$

We have by Theorem 6.3.7 that in the absence of standardized test,

$$\frac{U_s(p, 1)}{U_r(p, 1)} = 2.$$

Forcing students to take a standardized test thus improves fairness between two schools with accurate grades.

Part 4

Data and Strategic Behavior

Chapter 7

MECHANISM DESIGN UNDER THIRD-PARTY INFORMATION REVELATION

Yang Cai, Federico Echenique, Hu Fu, Katrina Ligett, Adam Wierman, and Juba Ziani (2018). "Third-Party Data Providers Ruin Simple Mechanisms". In: *arXiv preprint arXiv:1802.07407*, **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** URL: http://arxiv.org/abs/1802.07407.

7.1 Introduction

Information asymmetries are rampant in markets from ad auctions to art auctions, from acquiring a summer home to acquiring a startup. Naturally, whenever significant information asymmetries occur, agents have incentives to acquire information through outside channels. As a result, there is a proliferation of companies that seek to collect information that can be sold to participants in auctions with information asymmetries. Online advertising provides an extreme example. By tracking online behavior, *data providers* are able to sell valuable information about internet users (whose attention is the good for sale) to bidders in online advertising auctions. An FTC report [2014] details the scale and prevalence of such data providers — generating $426 million in annual revenue in 2012 and growing considerably in the years since.

The current chapter focuses on understanding mechanism design in settings with such information asymmetries. We consider a mechanism design setting where a bidder has incomplete information about the good he is bidding on, and obtains information about the good from a third-party data provider.

Specifically, the goal of this chapter is to investigate the impact of third-party data providers on the revenue of simple mechanisms. To do this, we consider a simple market—a single seller, a single bidder, and a single good—and a particular form of information asymmetry—the seller knows the type of the good she is selling, but the bidder has only partial information on the item type; the bidder knows his valuation for each of the m possible item types, but the seller knows only distributional information about the valuations. The key to the model is that, in addition to a prior over the item type, the bidder obtains a signal about the item type from a third party data provider and, while the seller can anticipate the signaling scheme used by the

data provider, the seller does not know the realization of signals. This captures, for example, a simple model of an ad auction where a mechanism designer sells an ad auction slot to an advertiser, who has incomplete information about the users targeted by the slots. The advertiser can get additional, third-party information about the target user(s) via data providers that track, for example, users' cookies.

Our model, though stylized, is already general enough to expose the difficulties created by third-party data providers. When no data provider is present, it is known that simple mechanisms can be used to provide a constant fraction of the revenue of optimal mechanisms; in contrast, our main results show that simple mechanisms cannot provide revenue within a constant factor of the revenue provided by an optimal mechanism, in the presence of a data provider.

In the absence of a data provider, Daskalakis et al. (2016) study optimal auctions in recent work. Daskalakis et al. (2016) look at the design of simple mechanisms in a setting where the only uncertainty about the item type is that it is drawn from a common prior. In this context, the question is whether it is valuable for the seller to share information with the bidder about the item type, or whether mechanisms that do not reveal information can be approximately optimal. Interestingly, Daskalakis et al. (2016) are able to characterize optimal auctions for this setting. Their insights show a direct correspondence between mechanisms for selling a single item of uncertain type and multi-item auctions; in particular, this correspondence implies that the seller does not need to reveal any information about the item type to the bidder in order to maximize his revenue. Further combining this observation with the work of Babaioff et al. (2014) allows them to observe that the better mechanism of two simple approaches—setting a fixed price for the item (the parallel notion of grand bundling, which we term "item-type bundling"), or pricing each item type separately (the parallel notion of item pricing, which we term "item-type pricing")—is guaranteed to yield a constant fraction of the revenue of the optimal mechanism. Thus, in the case where there is no third-party data provider, simple mechanisms are sufficient.

Our results show, in contrast, that the presence of a third-party data provider, who reveals information outside of the control of the seller, complicates the mechanism design task dramatically. We first consider revenue-optimal mechanisms. While Daskalakis et al. (2016)'s characterization of optimal auctions extends naturally to our setting, these optimal mechanisms may be quite complex. Concretely, our setting satisfies a type of revelation principle (Lemma 7.2.10) stating that optimal

mechanisms require only a single round of bidding, followed by a single round of information revelation (full revelation, in fact); however, such a mechanism presents the bidder with a menu of options that includes an option for each possible valuation vector, combined with each possible posterior of the bidder on the item type after receiving the data provider's signal, and requires the seller to condition the price charged on the realization of the item type.

Not only does the presence of a data provider complicate the design of the optimal mechanism, our main results show that it also impacts the revenue achievable via simple mechanisms. Specifically, in the presence of a data provider, the better of item-type bundling and item-type pricing may achieve only an $\Omega(\log m)$ factor approximation of the revenue the seller could have achieved had she offered a richer menu to the bidder, where m is the number of possible item types (see Theorem 7.3.2). In particular, a mechanism that divides the item types into disjoint groups and offers a price on each group can outperform both item-type pricing and item-type bundling by a logarithmic factor. Such mechanisms are known in the multi-item auction literature as partition mechanisms, and are seen as relatively simple mechanisms (see, e.g., Rubinstein, 2016). In our setting, we refer to such mechanisms as *item-type partition mechanisms*.

This separation between the revenue of item-type partitioning and that of item-type pricing and item-type bundling raises a natural question: if we expand our view of what constitutes "simple" mechanisms to include item-type partitioning, which generalizes both item-type pricing and item-type bundling, can we guarantee that simple mechanisms obtain a constant approximation of the optimal revenue in the presence of a data provider?

Our main result uses a more intricate argument to show that this is not the case. We demonstrate that, in the presence of a data provider, optimal mechanisms can outperform the best item-type partition mechanism by an $\Omega(\log\log m)$ factor (Theorem 7.4.2). So, in the presence of a data provider, simple mechanisms truly are no longer optimal. Additionally, our result highlights that the presence of third-party information can simultaneously hurt the optimal revenue achievable by the seller (by a $O(\log\log m/\log m)$ factor, see Theorem 7.4.2). These consequences of our result imply that, in settings where bidders have incomplete information (e.g., ad auctions), it is crucial for a seller whose goal is to maximize revenue via a simple mechanism to have a monopoly on the information available about the good for sale. A seller loses significant revenue if using a simple design without a monopoly on data.

Our discussion so far has focused on the seller, and ignored the data provider's incentives. The results described above do not depend on a specific model of the data provider behavior. However, when interpreting the lower bounds, it is interesting to consider how the data provider may behave. Two particular cases of interest are (i) a *strategic* data provider that seeks to maximize its profit, and (ii) an *adversarial* data provider that seeks to minimize the seller's profit.

We study the case of a *strategic* data provider in Section 7.5, and obtain results about the equilibrium outcome when the seller and the data provider interact strategically. Specifically, we consider a game played between a seller and a data provider. The game has the seller and the data provider each choosing an action simultaneously. The seller proposes a mechanism that the buyer will engage in, a mechanism which depends on the signaling scheme that the seller expects the provider to offer. The data provider chooses a signaling scheme that it offers to the buyer. Both agents, the seller and data provider, seek to maximize profits. Our results highlight that, regardless of the form of the mechanism used by the seller, the strategic data provider chooses to reveal all the information that is available to him (Lemma 7.5.1). Importantly, all of the constructionsused to prove the lower bounds for simple mechanisms apply to the specific case of strategic data providers, and thus the lower bounds discussed above hold in the presence of a strategic data provider.

Finally, we consider the case of an *adversarial* data provider in Section 7.5. In this case, the data provider seeks to minimize the seller's profits, which could be the goal if the data provider were also running a platform that competed with the seller. As in the case of the strategic data provider, our lower bounds can be extended to this setting, and Corollary 7.5.3 highlights that an adversarial data provider can force a revenue of at most $O(1/\log m)$ of the achievable revenue when no data provider is present. Additionally, this setting is of particular interest because it demonstrates behavior that is, perhaps, counter-intuitive. Specifically, in contrast to the case of a strategic data provider, a data provider that is attempting to negatively impact the revenue of the seller *may not* want to fully reveal her information about the item type to the bidder (Lemma 7.5.5). Instead, there may be intermediate signals which, upon revelation, minimize the revenue of the seller. This serves to highlight the complexity of mechanism design in the context of a third-party data provider, motivating the importance of designing mechanisms that have strong lower bounds regardless of the behavior of the data provider.

Summary of contributions

To summarize, in this chapter we make the following contributions. We propose a simple model of an auction in the presence of a third-party data provider, capturing information asymmetry regarding the type of the item for sale. Within this model, we first (Section 7.2) provide a characterization of the optimal auction based on that of Daskalakis et al. (2016), which may require a complex menu of options. Our main results study the potential for simple mechanisms to approximate the revenue of optimal mechanisms. In Section 7.3, we show that the item-type equivalents of item pricing and grand bundling cannot achieve within an $\Omega(\log m)$ factor of the revenue achievable by the optimal mechanism, nor even of the best item-type partition mechanism. Further, in Section 7.4, we show that there may be an $\Omega(\log\log m)$ gap between the revenue of the best item-type partitioning and that of the optimal mechanism. These results highlight that the presence of a data provider significantly reduces the ability of simple mechanisms to approximate the revenue of optimal mechanisms, even in the case of a single seller and a single bidder. Finally, in Section 7.5, we turn to understanding the impact of the behavior of the data provider. We show that our lower bounds also hold for the specific cases of strategic and adversarial data providers. Additionally, we show a contrast between these two cases: strategic, revenue-maximizing data providers always fully reveal the information available to them, while adversarial data providers may only *partially* reveal information. Thus, *partial* revelation may be more damaging to the seller than *full* information revelation.

Related work

There is a rich literature on information and signaling in auctions. One line of research focuses on designing a signaling scheme (on the part of the seller) given a certain auction format such as the second price auction (see, e.g., Lewis et al., 1994; Ganuza, 2004; Eső et al., 2007; Emek et al., 2014; Bro Miltersen et al., 2012; Cheng et al., 2015; Dughmi et al., 2015; Smolin, 2019); another line, closer to our setting, studies the design of *both* the auction and the signaling scheme (again, in such work, there is no data provider; bidders have a prior on their valuation for the item, and any signal on this valuation comes from the seller). Fu et al. (2012) showed that, if the auctioneer commits to a signaling scheme before choosing the form of the auction, full revelation followed by Myerson's auction for the revealed item type is the optimal design. Daskalakis et al. (2016) revealed the subtlety of this order of commitment and showed that, when the design of the auction and that of the signaling scheme are

considered together (without having to commit to one before the other), the optimal strategy is to reveal no information at all, and the overall problem is in fact equivalent to the design of a multi-item auction. In particular, they show that, when the bidders have a publicly known common prior on the type of the item, the optimal revenue for the seller is that of a multi-item auction.

Furthermore, Theorem 2 of Daskalakis et al. (2016) shows a one-to-one correspondence between types when selling a single item of uncertain type and items in a classical multi-item auctions. In particular, item-type pricing, i.e., mechanisms in which the seller first reveals the item type and then charges a take-it-or-leave-it price, is equivalent to selling separately (i.e., item pricing) in the corresponding multi-item auction, and item-type bundling, i.e., mechanisms in which the seller does not reveal any information and offers a single take-it-or-leave-it price, is equivalent to grand bundling in the corresponding multi-item auction. When there is a single bidder, Daskalakis et al. (2016) further combine this correspondence with results of Babaioff et al. (2014) to show that the better of item-type pricing and item-type bundling gives at least $1/6$ of the optimal revenue.

The results described above highlight the connection between our work and the study of simple mechanisms for multi-item auctions. Hart et al. (2017) pioneered this area. They showed that a seller, using item pricing, can extract a $\Omega\left(1/\log^2 m\right)$ fraction of the optimal revenue from an additive bidder whose values for m items are drawn independently, and selling these items as a bundle can achieve a $\Omega\left(1/\log m\right)$-fraction of the optimal revenue if the bidder's values are i.i.d. Li et al. (2013) improved the approximation ratio for item pricing to $O\left(1/\log m\right)$, which is tight. Babaioff et al. (2014) showed that, surprisingly, the better of selling separately and grand bundling can achieve at least $1/6$ of the optimal revenue. Subsequently there has been a surge of results generalizing the results of Babaioff et al. to broader settings (Cai et al., 2013; Yao, 2015; Rubinstein et al., 2015; Cai et al., 2016; Chawla et al., 2016; Cai et al., 2017). At this point, it is known that simple mechanisms such as sequential two-part tariffs can obtain a constant fraction of the optimal revenue for multiple bidders with combinatorial valuations that are, e.g., submodular, XOS (Cai et al., 2017). One might hope to extend these simple deterministic mechanisms to settings where the bidder has correlated values over the items; however, this is impossible. Hart et al. (2013) showed that even for a single additive bidder, when valuations are interdependent, the ratio between the revenue obtainable by a randomized mechanism and that of the best deterministic mechanism can be unbounded.

Finally, there has been a line of work in economics that focuses on mechanism design and revenue guarantees that are robust to uncertainty on the information structure available to the bidders, often in common value auctions (see, e.g., Bergemann et al., 2016; Bergemann et al., 2017; Bergemann et al., 2018; Du, 2018).

7.2 Model and Preliminaries

We consider a single, revenue-maximizing seller selling a single item to a buyer. The item for sale takes one of m possible types, and the buyer's valuation may depend on the item type. The buyer does not know the type i of the item, but has a publicly-known prior π over the item types. We let $\pi(i)$ denote the prior probability that the item is of type i.

The buyer's private value when the good is of each type i is drawn independently from a publicly known distribution $\mathcal{F}(i)$ over the space of non-negative real numbers $\mathbb{R}^+$. We denote by $V(i)$ the buyer's valuation for an item of type i, and denote by $V = (V(1), \ldots, V(m))$ the buyer's valuation vector.

In our setting, there is a third-party data provider who has (potentially imperfect) information on the type of the item, in the form of a random variable X that can be arbitrarily correlated with the type of the item. This is unlike the setting of (Daskalakis et al., 2016) in which only the seller could reveal information about the item type to the bidders. The joint distribution of i and X is publicly known, but the realized value of X is only observed by the data provider.

The data provider designs a *signaling scheme* in the form of a function S that maps X to $\Delta(\Sigma)$, the set of distributions over an alphabet Σ. The data provider is able to commit to such a scheme and, on observing information X, the data provider draws a signal σ from Σ according to the distribution $S(X)$, and sends it to the buyer if the latter purchases from the data provider.

After receiving σ, the buyer updates his prior using Bayes' rule. We denote the resulting posterior by $\pi_\sigma \in \Delta([m])$. The buyer aims to maximize his utility given his posterior on the item type; we assume utilities are quasi-linear. Since the realization of X was not visible to the seller, if the buyer purchases from the data provider, the seller would know only the distribution over the buyer's posteriors, conditioning on the item type i.

Motivated by the sale of online advertisements, where the sale repeats rapidly with the item type redrawn in each round, we assume that the buyer's decision to purchase

from the data provider is made before his value is realized. In this setting, the buyer, seller and data provider act as follows:

1. The seller commits to a information revelation policy and a mechanism. Simultaneously, the data provider commits to a signaling scheme.
2. The buyer decides whether to enter a contract with the data provider and to purchase his information.
3. All the participants receive their private information: the buyer observes his valuation vector, the data provider observes X, and the seller observes the item type.
4. The buyer sees the signal from the data provider. The seller reveals information to the buyer and runs her mechanism.

We remark there is asymmetry between the seller and the data provider, in that the buyer makes the decision of purchasing the data provider's signal before his valuation vector is realized, whereas the purchase decision with the seller is made after the buyer realizes his valuation vector. In the ad auction setting, this asymmetry is motivated by the practice that data sets are often sold in batch, or as "right of access," whereas ads are sold per impression, often via bidding in an auction for each individual ad. This asymmetry is even more marked when the buyer is an agency that bids on behalf of many advertisers in individual auctions but buys data access in batch to inform all such bidding.

In Sections 7.3 and 7.4, we show that there exists a signaling scheme for the data provider such that no simple mechanism can extract a constant fraction of the optimal revenue.[1] In Section 7.5, we take the incentives of the data provider into account and consider two different scenarios: one in which the data provider is strategic and aims to maximize his revenue from selling his signal, and one in which he is adversarial and tries to hurt the seller's revenue.

Simple mechanisms for a single buyer in the absence of a data provider

In the absence of a third-party data provider, in the single seller, single buyer, *multi-item* setting, *where each item has a single type*, Babaioff et al. (2014) show that, although the optimal mechanism may be complex, a simple mechanism achieves a

[1] More specifically, we show that no item-type partitioning mechanism (see Definition 7.2.6) is a constant factor approximation.

constant factor of the optimal revenue. In particular, this mechanism is simply the better of either item pricing or grand bundling. This result was originally stated for multi-item auctions, but the results of Daskalakis et al. (2016) show that the current setting, with a single item that can take on multiple possible types, in fact reduces to the multi-item auction setting. Remark 7.2.1 below explains how this reduction works.

Remark 7.2.1 (Reduction between multi-item, and multi-type auctions)**.** *Consider a single bidder, single item setting with m possible types, prior π and valuation vector $V = (V(1), \ldots, V(m))$, distributed according to joint distribution $\mathcal{F}$. Daskalakis et al. (2016) show that this setting in fact reduces to a multi-item auction with m items, in which the bidder's valuation for the items are distributed as follows: i) draw V according to $\mathcal{F}$, then ii) let the bidder's valuation vector be $V' = (\pi(1)V(1), \ldots, \pi(m)V(m))$, the coordinate-by-coordinate product of V and π.*

The optimal auction in such a single-bidder, multi-item setting can be written without loss of generality as a menu of options $\{(P_o, A_o)\}_o$, such that a bidder either opts out, or selects a single option o in which case he i) must pay price P_o then ii) receives any given item $i \in [m]$ with allocation probability $A_o(i)$. An optimal single-bidder, single-item, multi-type auction is then given by the exact same menu $\{(P_o, A_o)\}_o$, where i) a bidder who picks option o picks price P_o, but now ii) the bidder receives the (here, a single) item with probability $A_o(i^*)$ where i^* is the realized item type. For example, consider the following scenario, studied previously in Hart et al. (2015):

Example 7.2.2. *There are two item types, denoted* 1 *and* 2*. The bidder's prior on the types is uniform, given by $\pi(1) = \pi(2) = 1/2$. The bidder's valuations $V(1)$ and $V(2)$ are i.i.d, and take values* 2*,* 4*, and* 8 *with probabilities* $1/6$*,* $1/2$*, and* $1/3$*. The equivalent multi-item auction is one with two items whose valuations $V'(1), V'(2)$ are i.i.d, and take values distributed as follows: $V'(1) = \pi(1)V(1) = V(1)/2$ and $V'(2) = \pi(2)V(2) = V(2)/2$, i.e.,* 1*,* 2*, and* 4 *with probabilities* $1/6$*,* $1/2$*, and* $1/3$*. As shown by Hart et al. (2015), the optimal mechanism in this case has two menu options:*

1. *Option* 1 *has price* 1*, and gives the first item with probability* $1/2$ *and the second with probability* 0*.*
2. *Option* 2 *has price* 4*, and gives both items with probability* 1*.*

This translates into a menu of options in the multi-type, single-item settings in which a bidder that selects option 1 *gets the item only if it is of type* 1*, with probability* $1/2$*, and a bidder that selects option* 2 *always get the item, independently of the item type.*

As simple mechanisms are important throughout our chapter, we formally define them here, in the context of selling a single item with multiple possible types.

Definition 7.2.3. *An **item-type pricing** mechanism first reveals the type i of the item to the buyer, then offers to sell the item to the buyer at some price $P_{it}(i)$. We also refer to such mechanisms as "selling the types separately," in analogy to the concept of selling separately in the case of multi-item auctions.*

Definition 7.2.4. *An **item-type bundling** mechanism offers the item for sale at some price P_{gr} without revealing any information about the realized type of the item.*

The following result from Babaioff et al. (2014) and Daskalakis et al. (2016) summarizes the power of these simple mechanisms in the single-item, multi-type setting, without a data provider.

Proposition 7.2.5 (Babaioff et al. (2014) and Daskalakis et al. (2016))**.** *In the absence of a data provider, the maximum of item-type pricing and item-type bundling yields at least a $\frac{1}{6}$-approximation to the optimal revenue when there is a single seller, a single buyer, a single item for sale, and the buyer has a publicly known prior over the type of the item.*

A generalization of both item-type bundling and item-type pricing mechanisms is also important for the results in this chapter.

Definition 7.2.6. *An **item-type partition** mechanism first partitions the set of item types into non-empty groups $\mathcal{G}_1$ to $\mathcal{G}_g$, priced (resp.) P_1 to P_g. The mechanism then observes the type i of the item, and offers the item at price P_r, where r is uniquely chosen such that $i \in \mathcal{G}_r$.*

Note that, after observing the offered price P_r, the buyer may infer that the realized item type must belong to group $\mathcal{G}_r$. Item-type pricing is an instantiation of item-type partitioning where the partition contains a separate group for each type; item-type bundling corresponds to item-type partitioning using the trivial partition. Item-type partitioning is, however, significantly more powerful than these other simple mechanisms, as it allows the seller to partition the item types into arbitrarily many groups of arbitrarily many sizes.

The equal revenue distribution

An important tool in the derivation of lower bounds for mechanisms is the so-called equal revenue distribution. This distribution is crucial to a number of our examples in this chapter and is defined as follows.

Definition 7.2.7. *A random variable Y with support $[1, +\infty)$ follows the **equal revenue (ER) distribution** if and only if* $\Pr[Y \leq y] = 1 - \frac{1}{y}$.

The equal revenue distribution gets its name from the fact that it has constant virtual value, and every price in the distribution's support offers the same expected revenue. The equal revenue distribution also has a number of other useful properties, proved by Hart et al. (2017), which we summarize here. Unless otherwise specified, log is taken to be the natural logarithm.

Lemma 7.2.8 (Hart et al. (2017))**.** *Let $m \geq 2$ be an integer, and let $Y_1, \ldots, Y_m$ be m i.i.d random variables that follow the ER distribution. Then:*

$$\Pr\left[\frac{1}{m}\sum_{i=1}^{m} Y_i \geq \frac{\log m}{2}\right] \geq \frac{1}{2},$$

and for any $P \geq 6 \log m$,

$$\Pr\left[\frac{1}{m}\sum_{i=1}^{m} Y_i \geq P\right] \leq \frac{9}{P}.$$

Optimal mechanisms in the presence of a data provider

We provide a characterization of the optimal mechanisms for a single buyer and a single item, with several possible item-types, in the presence of a third-party data provider who knows (possibly imperfect) information about the item type, and who reveals some of this information to the buyer. Note that the data provider is represented via a signaling scheme that, from the model perspective, is subsumed into a probability distribution over posteriors π, representing beliefs of the buyer regarding the item's type. Therefore, a buyer with access to the data provider's signal has private information in the form of a valuation V and a posterior π.

We consider a class of mechanisms that allow the seller to charge the buyer a price that is conditional on the type of the item. We observe that restricting attention to such type-contingent price mechanisms is without loss of generality. The characterization we present is a type of revelation principle, similar to that presented in Daskalakis et al. (2016), where the difference is the presence of a data provider.

First, we need the definition of a conditional price menu.

Definition 7.2.9. *A **menu with conditional prices** is a fixed collection of pairs (A, P), where each $A \in [0,1]^m$ is called an allocation rule, and each $P \in \mathbb{R}_+^m$ is called a pricing rule. The buyer selects at most one pair (A, P). After his choice has been made, the type i of the item is revealed. Given item type i, the buyer pays price $P(i)$, and receives the item with probability $A(i)$.*

We show that there always exists an optimal mechanism that takes the form of a conditional price menu. We postpone the proof of Lemma 7.2.10 and detailed discussion of our characterization to Section 7.6.

Lemma 7.2.10. *For any equilibrium of any mechanism $\mathcal{M}$ in the presence of a data provider, such that the buyer, conditioned on the realization of her valuation vector and posterior beliefs over item types given the signal from the data provider, obtains non-negative payoff in expectation, there is a conditional price menu that is incentive compatible, interim individually rational, and provides the same revenue.*

Lemma 7.2.10 implies the optimal revenue is given by the solution of a linear program whose size is proportional to the number of possible pairs of value vectors and posteriors. We make use of this linear program in Section 7.5. Additionally, note that Lemma 7.2.10 can easily be extended to the multi-buyer setting, in which case one can write the optimal mechanism as an interim individually rational, direct revelation mechanism, with no information revelation by the seller required prior to bidding.

7.3 Revenue of Simple Mechanisms: A Warm-up

Our main results focus on bounding the revenue achievable via simple mechanisms, in the presence of a third party data provider. In this section, we focus on "simple" mechanisms in which the seller runs the better of item-type pricing and item-type bundling. These are particularly interesting mechanisms to consider given Proposition 7.2.5, where Daskalakis et al. (2016), using results of Babaioff et al. (2014), show that this style of mechanism obtains a constant fraction of the optimal revenue when a data provider is not present. To show that this is not the case when a data provider is present, we consider the following construction.

Construction 7.3.1. *Let $m = \eta^2$ be the number of item types, for some integer η. The types are partitioned into η groups $I_1, \ldots, I_\eta$ such that each group contains exactly η*

types. The bidder's prior on the item type is uniform, i.e., the bidder initially believes that each item type is realized with probability $1/m = 1/\eta^2$*, and that the probability that the realized type belongs to group* I_k *is therefore* $1/\eta$*. The bidder's valuation for type* i *in group* I_k *is* $V(i)/k$*, where* $V(i)$ *is a random variable drawn from the equal revenue distribution. The bidder's valuations for different item types are drawn independently of each other.*

In this setting, we allow the data provider to observe to which group the item type belongs. The data provider fully reveals this information to the bidder. We show later (Section 7.5) that this is the signaling scheme that a strategic, revenue-maximizing data provider would sell in this scenario, and that the buyer would always opt to buy the data provider's signaling scheme; therefore, our results extend to the case of a strategic provider.

Given the data provider's signal, the bidder's posterior probability on the item being of type i*, upon observing signal* σ_k *informing him that the group is* I_k*, is given by*

$$\pi_{\sigma_k}(i) = \begin{cases} 0 & i \notin I_k \\ \frac{1}{\eta} & i \in I_k \end{cases}.$$

We use the above construction to prove that the better of item-type pricing and item-type bundling cannot always achieve a constant fraction of the optimal revenue:

Theorem 7.3.2. *There exists a single seller, single bidder, single item (taking one of* m *item types) setting where, in the presence of a data provider who signals information about the item type realization to the bidder, the expected revenue of the better of item-type pricing and item-type bundling is no more than a* $O\left(\frac{1}{\log m}\right)$ *fraction of the expected revenue of the optimal mechanism. More specifically:*

- *In the absence of a data provider, the optimal revenue is* $\Theta\left(\frac{\log^2 m}{\sqrt{m}}\right)$*. The optimal revenue for item-type pricing is* $\Theta\left(\frac{\log m}{\sqrt{m}}\right)$*, and the optimal revenue from item-type bundling is* $\Theta\left(\frac{\log^2 m}{\sqrt{m}}\right)$*.*
- *In the presence of a data provider, the optimal revenue is* $\Theta\left(\frac{\log^2 m}{\sqrt{m}}\right)$*. The optimal revenue from item-type pricing and the optimal revenue from item-type bundling are both* $O\left(\frac{\log m}{\sqrt{m}}\right)$*.*

In particular, this theorem illustrates a setting where the introduction of a data provider does not affect the optimal revenue, but where the data provider's presence is

quite harmful to the optimal revenue of the better of item-type pricing and item-type bundling. We break the proof of this theorem into the following claims, which we prove in Section 7.7.

Claim 7.3.3. *The expected revenue from optimal item-type pricing in Construction 7.3.1 is* $\Theta\left(\frac{\log m}{\sqrt{m}}\right)$*, independently of whether a data provider is present.*

We emphasize that while item-type pricing is unaffected by the introduction of a data provider, the presence of a data provider can harm the optimal revenue of other classes of mechanisms.

Claim 7.3.4. *The expected revenue from optimal item-type bundling in Construction 7.3.1 is* $\Theta\left(\frac{\log^2 m}{\sqrt{m}}\right)$ *in the absence of a data provider.*

Claim 7.3.5. *The expected revenue from item-type bundling in Construction 7.3.1 is* $O\left(\frac{\log m}{\sqrt{m}}\right)$ *in the presence of a data provider.*

Claim 7.3.6. *There exists an item-type partition mechanism that achieves expected revenue* $\Omega\left(\frac{\log^2 m}{\sqrt{m}}\right)$ *in Construction 7.3.1. The optimal revenue in Construction 7.3.1 is* $\Theta\left(\frac{\log^2 m}{\sqrt{m}}\right)$.

7.4 Revenue of Simple Mechanisms: Item-type Partitioning

The previous section shows that neither item-type pricing nor item-type bundling, nor the better of the two, can always achieve a constant fraction of the optimal revenue in the presence of a data provider. However, one may wonder if the result is due to the restrictive nature of the "simple" mechanisms considered. Here, we show that, in the presence of a data provider, even the more general class of item-type partition mechanisms is insufficient to guarantee a constant fraction of the optimal revenue. This is particularly tantalizing due to the fact that Construction 7.3.1 admits an item-type partition mechanism that yields a constant approximation to the optimal revenue. However, in this section, we show a construction where the best item-type partition mechanism only achieves a $O\left(1/\log\log m\right)$ fraction of the optimal revenue. Note that this also implies that our "simpler" simple mechanisms, item-type bundling and item-type pricing, also do not yield a constant fraction of the optimal revenue, since they are special cases of item-type partitioning.

To show this result, we consider the following construction:

Construction 7.4.1. *Given an integer* η*, let* $m = 2^\eta$ *be the number of item types. The bidder's prior on the item type is uniform, i.e., the bidder initially believes the item*

type takes each $i \in [m]$ with probability $1/m$. The bidder's valuation for each type is drawn i.i.d. from an equal revenue distribution.

We consider η possible partitions of the m types. Given a particular $k \in [\eta]$, we partition the set of all types into $m_k = 2^{\eta-k}$ subsets of size $2^k \geq 2$ each. Specifically, for $k \in [\eta]$ we partition the set of types into the subsets $I_{k,1}$ to I_{k,m_k}, where $I_{k,j} = \{(j-1) \cdot 2^k + 1, \ldots, j \cdot 2^k\}$ for all $j \in [m_k]$.

The information we allow the data provider to observe is structured as follows. First, a value of $k \in [\eta]$ is drawn according to the following distribution: for $k \leq \eta - 1$, k is drawn with probability $\frac{1}{k(k+1)}$; $k = \eta$ is drawn with the remaining probability $\frac{1}{\eta}$. The value k is drawn, importantly, independently of the type i of the item. Then, the data provider observes which group of size m_k (i.e., among $I_{k,1}$ to I_{k,m_k}) the item type lies in. Given the observations, the data provider reveals his full information to the bidder, namely, exactly which group of size m_k the item type belongs to. We show later (Section 7.5) that this is what a strategic, revenue-maximizing data provider would reveal in this scenario. We denote by $\sigma_{k,j}$ the realization of the signal that indicates to the bidder that the item belongs to group $I_{k,j}$. We call k the size indicator.

This construction allows us to prove the following main result, that states that the mechanism designer cannot always achieve a constant fraction approximation of the optimal revenue via item-type partition mechanisms:

Theorem 7.4.2. *There exists a single seller, single bidder, single item (taking one of m item types) setting where, in the presence of a data provider who signals information about the item type realization to the bidder, no item-type partition mechanism can achieve revenue higher than $O\left(\frac{1}{\log\log m}\right)$ of the optimal revenue. More specifically:*

- *In the absence of a data provider, the optimal revenue is $\Theta(\log m)$. The optimal revenue from item-type partitioning is $\Theta(\log m)$, and is achieved for the item-type bundling partition.*
- *In the presence of a data provider who signals information about the item type realization to the bidder, the optimal revenue is $\Theta(\log\log m)$, and is achieved by Mechanism 7.4.5 below. The optimal revenue from item-type partition mechanisms is $\Theta(1)$.*

This theorem illustrates a setting where the introduction of a data provider does decrease the revenue of optimal mechanisms, and where restricting to item-type partitioning mechanisms further harms revenue in the presence of a data provider. There therefore could be strong incentives for a seller to be a data monopolist, particularly if the seller has a preference to run simple mechanisms.

The first part of the lemma, when a data provider is not present, is a direct consequence of Hart et al. (2017) and Babaioff et al. (2014). We prove the result in the presence of a data provider via the following sequence of claims, each of which is proven in Section 7.8.

Claim 7.4.3. *The expected revenue from the optimal item-type partition mechanism is $O(1)$ in Construction 7.4.1.*

Claim 7.4.4. *There exists a mechanism that yields revenue $\Omega(\log\log m)$ in Construction 7.4.1. The optimal revenue in Construction 7.4.1 is $\Theta(\log\log m)$.*

To prove Claim 7.4.4, we first construct a mechanism that achieves revenue $\Omega(\log\log m)$ in Construction 7.4.1. In particular, we consider the following design.

Mechanism 7.4.5. *The seller offers a menu of $\sum_{k=1}^{\eta} m_k$ options. For every $\kappa \in [\eta]$, and every $\iota \in [m_\kappa]$, the menu contains the following option $L_{\kappa,\iota}$: the bidder first pays $P_\kappa = \frac{1}{8}\log 2^\kappa = \frac{\log 2}{8}\kappa$, then gets the item if and only if it is in group $I_{\kappa,\iota}$. Note that the price only depends on κ.*

To show that Mechanism 7.4.5 yields revenue $\Omega(\log\log m)$ in Construction 7.4.1, we need the following claim, which characterizes the bidder's behavior in the mechanism. More specifically, we show in the following claim that if the bidder receives signal $\sigma_{k,j}$, he purchases the corresponding option $L_{k,j}$ in Mechanism 7.4.5 with probability almost 1.

Claim 7.4.6. *In the the setting of Construction 7.4.1, suppose the bidder receives signal $\sigma_{k,j}$ (indicating that the item belongs to group $I_{k,j}$ of size 2^k) for $k \geq 2\cdot 10^2 + 1$. Consider the menu of options proposed by the seller in Mechanism 7.4.5. With probability at least $1 - 10^{-3}$, no option $L_{\kappa,\iota}$ with either $\kappa \neq k$ or $\iota \neq j$ yields a higher utility for the bidder than option $L_{k,j}$, and $L_{k,j}$ yields positive utility to the bidder.*

We remark that Mechanism 7.4.5, although it has a concise description, is not "simple" in any of the usual senses, and is in fact carefully tailored to the incentives of the

bidder. We do not know of "simpler" mechanisms that are approximately optimal in this setting.

7.5 Modeling the Behavior of the Data Provider

So far, we have not discussed the behavior of the data provider. The characterization of optimal mechanisms, and our bounds on the achievable revenue of simple mechanisms in the previous sections, do not depend on specific assumptions about the behavior of the data provider. However, it is useful to consider specific models of the data provider when interpreting our lower bounds. In particular, one may wonder if more positive results are possible for some standard game-theoretic models of the interaction between the data provider and the seller.

To this end, it is natural to consider two extreme models for a data provider's incentives: (i) the data provider may be *strategic*, seeking to maximize his revenue from selling his information, or (ii) the data provider may be *adversarial*, seeking to minimize the profits of the seller. In this section, we characterize the behavior of the data provider in each of these models. Our results provide a clear contrast between the two models: strategic data providers always reveal all of their information to the bidder, while adversarial data providers may not reveal all available information to the bidder. Our results have implications for mechanism design, highlighting the importance of mechanisms with good worst-case bounds on revenue, independent of data provider behavior.

Importantly, the proofs of the bounds on the achievable revenue of simple mechanisms in Theorems 7.3.2 and 7.4.2 use constructions that can be interpreted as complete revelation by the data provider. Thus, those bounds apply to the case of a strategic data provider. Concretely, this means that, in the presence of a strategic, revenue-maximizing data provider, simple mechanisms cannot guarantee near-optimal revenue for the seller.

The behavior of a strategic data provider

In this section we show that a third-party data provider, if aiming to maximize his revenue from selling his information, should reveal his information in full. In other words, to maximize his utility, the data provider should send X directly to the buyer as the signal. More formally, the provider should let Σ be the range of X, and let $S(X)$ be the distribution where all probability is point massed on X. In what follows we use S^* to denote this fully-revealing signaling scheme.

We first show that, for any mechanism adopted by the seller, and for any buyer's value vector, a fully revealing signaling scheme maximizes the utility of the buyer. A simple consequence is then that the data provider can retain the added utility as revenue, by pricing his signaling scheme appropriately. So we obtain the conclusion that, no matter what mechanism the seller uses, it is a dominant strategy for the data provider to fully reveal his information.

Let us fix a mechanism $\mathcal{M}$ and valuation vector V. Let S be an arbitrary signaling scheme that maps the data provider's information X to $\Delta(\Sigma)$. Recall that the buyer forms a posterior distribution π_s over the item type i when receiving signal σ drawn from $S(X)$. We denote by $U(V, S)$ the buyer's utility in $\mathcal{M}$ when her value vector is V and he purchases signaling scheme S from the data provider.

To compare $U(V, S^*)$ and $U(V, S)$, in the following we slightly abuse notation and let π_X denote the posterior distribution the buyer forms over i when the signaling scheme is S^* and the buyer receives signal X. A proof of the following lemma appears in Section 7.9.

Lemma 7.5.1. *For any valuation vector V, any information X received by the data provider, and any signaling scheme $S(.)$, $U(V, S^*) \geq U(V, S)$.*

Let $U(V)$ be the buyer's utility if his value vector is V and he does not purchase from the data provider. The ex-ante value of a signaling scheme S for the buyer is then $\mathbb{E}_V[U(V, S) - U(V)]$, and this difference is the highest price the buyer is willing to pay for the scheme S. A rational data provider would charge arbitrarily close to this difference for the signaling scheme, and this would be his revenue. Therefore our lemma immediately implies that S^*, the fully-revealing signaling scheme, maximizes the data provider's revenue.

Corollary 7.5.2. *The fully-revealing signaling scheme is revenue-maximizing for the data provider.*

The corollary above is particularly important because the examples we use to prove Theorems 7.3.2 and 7.4.2 make use of constructions where the data provider uses full revelation. Thus, the implications of those theorems also hold under the assumption of a strategic data provider. In the following corollary we contrast the revenue between the settings where a data provider is or is not present. (Recall that item-type partition mechanisms yield at least $\frac{1}{6}$ of the optimal revenue when no data provider

is present (Babaioff et al., 2014).) This result highlights the damaging impact of a third-party data provider for the seller.

Corollary 7.5.3. *There exists a single seller, single bidder, single item setting where, in the presence of a revenue-maximizing data provider, no item-type partition mechanism can achieve revenue higher than* $O\left(\frac{1}{\log m}\right)$ *of the optimal revenue achievable by an item-type partition mechanism when no data provider is present. More precisely:*

- *The optimal revenue achievable in the absence of a data provider is* $\Theta(\log m)$*, and is attained by item-type bundling.*
- *The optimal revenue achievable in the presence of a data provider is* 1*, and is attained by item-type pricing.*

Proof. Consider a setting with m item types, distributed i.i.d. according to an Equal Revenue distribution. The optimal revenue in the absence of a data provider is $\Theta(\log m)$ by (Hart et al., 2017). As $P \cdot \Pr[V(i) \geq P] = P \cdot \frac{1}{P} = 1$ for all P when $V(i)$ follows an equal revenue distribution, the optimal revenue in the presence of a revenue-maximizing data provider that exactly knows and reveals the item type is 1. On the other hand, for independent valuations, the seller can always guarantee a $\Omega\left(\frac{1}{\log m}\right)$ of the revenue by revealing the item type and pricing optimally, as per (Li et al., 2013). Note that this example is a special case of Construction 7.4.1. □

The behavior of an adversarial data provider

We now move to consider an *adversarial* data provider, who aims to minimize the seller's revenue. The main result in this section shows that revealing *less* information can sometimes be *more* damaging to the seller's revenue. This phenomenon, however, does not occur when the data provider has perfect information about the item type. Our first lemma below shows that he minimizes the expected revenue of the seller by fully revealing the type, if the type is known.

Recall from Section 7.5 that we use S^* to denote the fully-revealing signaling scheme. Throughout this section, we use $REV(S)$ to denote the seller's optimal revenue when the data provider adopts the signaling scheme S.

Lemma 7.5.4. *If the data provider is adversarial and has full information about the type of the item (that is, if X is perfectly correlated with the item type), the optimal*

strategy for the data provider is to reveal X, i.e., to use the fully-revealing signaling scheme S^*.

Proof. Let $SREV$ be the optimal revenue that the seller can achieve when the type of the item is revealed. On the one hand, the seller can always guarantee a revenue of $SREV$ by revealing the type of the item and then selling this type optimally, no matter what signaling scheme is used by the data provider. So $SREV \leq REV(S)$ for any S. On the other hand, when X fully reveals the type, and S^* fully reveals this information, the buyer would know the type, and by the definition of $SREV$ the optimal revenue that can be achieved by the seller is $SREV$. Therefore $SREV \geq REV(S^*)$. Therefore $REV(S^*) \leq REV(S)$ for any scheme S. □

More interestingly, and perhaps counter-intuitively, if the data provider does not have full information, then only *partially* revealing information may minimize the revenue of the seller.

Lemma 7.5.5. *Let the number of item types be $m = 2$. There exists a distribution over the buyer's valuations V, a prior π over the item type and a partially informative distribution over the data provider's information X, such that there is a signaling scheme S, with $REV(S) < REV(S^*)$.*

The proof of this result uses the following example.

Example 7.5.6. *Let the bidder's valuation for each item type be drawn i.i.d., taking value* 1 *with probability* $1/2$ *and value* 2.1 *with probability* $1/2$. *The bidder and the data provider share a common prior $\pi = (3/4, 1/4)$. That is, they both initially believe the item is of type* 1 *with probability* $3/4$ *and of type* 2 *with probability* $1/4$. *The data provider receives information X on some support $\{x_1, x_2\}$. If the item type is* 1*, the provider receives x_1 with probability* $2/3$ *and x_2 with probability* $1/3$*, and if the item type is* 2*, the provider receives x_2 with probability* 1.

Proof. We show that in Example 7.5.6, the data provider has a signaling scheme under which the seller's optimal revenue is less than under the fully-revealing scheme.

If the data provider reveals full information, then with probability $1/2$, the bidder receives x_1 and has posterior $\pi_{x_1} = (1, 0)$ (when receiving x_1, the provider knows the item must be of type 1); with probability $1/2$, he receives x_2 and thus has posterior $\pi_{x_2} = (1/2, 1/2)$. Remember that by the characterization of Section 7.2,

the seller's best response to the data provider can be written as an interim individually rational mechanism that does not require any information revelation; further, the optimal revenue from such mechanisms can be obtained by solving a linear program. Computing the seller's optimal revenue via linear programming, we see that the revenue is $REV(S^*) = 1.1062$.

Consider the following partially revealing signaling scheme: let Σ, the range of the signaling scheme, be $\{\sigma_1, \sigma_2, \sigma_3\}$; let φ be the mapping $\varphi(x_i) = \sigma_i$ for $i = 1, 2$; when the provider receives realization x of X that belongs to $\{x_1, x_2\}$, the provider outputs $\varphi(x)$ w.p. $1 - \varepsilon = 0.86$ and, outputs σ_3 with probability $\varepsilon = 0.14$. Given this signaling scheme, when the bidder receives signal σ_1 (which occurs with probability $\frac{1}{2}(1 - \varepsilon) = 0.43$), he infers $X = x_1$, and so his posterior is $\pi_{\sigma_1} = \pi_{x_1}$. Similarly, when he receives signal σ_2 (which also occurs with probability 0.43), the bidder infers that it must be the case that $X = x_2$, hence he has posterior $\pi_{\sigma_2} = \pi_{x_2}$. Finally, when the bidder receives σ_3 (which occurs with probability $\varepsilon = 0.14$), he infers that $X = x_1$ or $X = x_2$ with equal probability by symmetry, and hence his posterior is $\pi_{\sigma_3} = \frac{1}{2}\left(\pi_{x_1} + \pi_{x_2}\right) = (3/4, 1/4) = \pi$. Computing the optimal revenue of the seller via linear programming, using the the results of Section 7.2, we get that the revenue is only $REV(S) = 1.0991 < 1.1062 = REV(S^*)$. □

It may seem counter-intuitive that the data provider can harm the seller *more* by providing *less* information to the bidder. After all, one consequence of the characterization of Section 7.2 is that the seller can only lower her revenue by revealing more information to the bidder. However, information from the provider and information from the seller are not equivalent from the perspective of the seller, because the seller does not get to see the realization of the signal that the provider sends to the bidder. When the seller reveals information, she knows exactly what the bidder's posterior is, and can act as a function of the realized posterior; she is then faced with exactly the problem solved by Daskalakis et al. (2016) for that realized posterior. When the data provider reveals information, the seller, who only knows the signaling scheme but not the signal, faces a distribution of posteriors and does not know which of them is correct.

In particular, in Example 7.5.6, in the fully-revealing signaling scheme there are two posteriors π_{x_1} and π_{x_2}. Each occurs with probability $1/2$. The seller, intuitively, wishes to design a menu with one option for each of these two posteriors. In the partially revealing signaling scheme there is a third posterior, $\pi_{\sigma_3} = \pi$, which is an average of π_{x_1} and π_{x_2}. In fact, the first signaling scheme is a mean-preserving spread

of the second one. The seller, intuitively, wishes to design a menu with three options, one for each posterior.

This third posterior induces a trade-off in the linear program the seller solves to find the optimal mechanism. The second linear program has more IC constraints for the two posteriors than the linear program given the fully revealing signaling scheme. This makes the revenue the seller gets from bidders with posteriors π_{σ_1} and π_{σ_2} lower than before. The trade-off is that there is now a new posterior π_{σ_3}, from which the seller can make additional revenue. Example 7.5.6 is constructed so that the harm from the additional posterior exceeds the benefit.

Finally, we characterize the revenue the seller loses due to the presence of an adversarial data provider. Note that the revenue the seller can obtain when there is an adversarial data provider is less than what would be achieved under a strategic, revenue-maximizing data provider (who uses a fully-revealing signaling scheme). Thus, it follows from Corollary 7.5.3 that the presence of an adversarial data provider can greatly harm the revenue of the seller.

Corollary 7.5.7. *There exists a single seller, single bidder, single item setting where, in the presence of an adversarial data provider, no item-type partition mechanism can achieve revenue higher than* $O\left(\frac{1}{\log m}\right)$ *of the optimal revenue achievable by an item-type partition mechanism when no data provider is present. More precisely:*

- *The optimal revenue achievable in the absence of a data provider is* $\Theta(\log m)$*, and is attained by item-type bundling.*
- *The optimal revenue achievable in the presence of a data provider is* 1*, and is attained by item-type pricing.*

7.6 Supplementary Material: Revelation Principle, the Full Version

The characterization we present shows that the revenue achievable via any mechanism can be obtained with a conditional price menu.

A few comments about the mechanism are in order. Note that the allocation probability and price may both depend on the realized type of the item. So, one can think of the mechanism as requiring a single round of bidding, followed by a single round of information revelation (in fact, full information revelation), to determine which price $P(i)$ the bidder should pay. Additionally, note that the bidder pays regardless of whether he receives the item. Finally, note that conditional price mechanisms are

strictly more general than item-type partition mechanisms. Item-type partitioning is, in fact, an instantiation of menus with conditional prices in which each $A \in \{0,1\}^m$ (no fractional or probabilistic allocations are allowed), each item type is offered in exactly one option, and the conditional prices within an option are all identical. Each option then corresponds to a single subset of the partition.

Despite allowing prices and allocations to depend on the realization of the item type, the conditional price menus guarantee interim individual rationality, defined as follows.

Definition 7.6.1. *A mechanism is **interim individually-rational** (interim IR) if and only if the bidder's expected utility from participating in the mechanism, conditional on a valuation V and posterior beliefs π over item types, is non-negative.*

Interim IR can be seen as the bidder committing to an option from the menu offered by the mechanism. One justification for this notion is that a bidder might, in theory, be engaged in many auctions simultaneously. Therefore, the bidder might care only about his average payoff across multiple purchases. While for some type realizations such a bidder may lose, with high probability his overall utility is non-negative. Interim IR can always be guaranteed by adding a dummy option with price 0 and allocation probability $A = 0$, such that an agent that gets negative utility from any other option goes for the dummy option.

Proof of Lemma 7.2.10. We treat the pair (V, π_σ), where V is the bidder's valuation vector and π_σ is his posterior given he sees signal σ, as the bidder's type. We follow the same steps as the proof of Theorem 1 and Section A of Daskalakis et al. (2016). Consider a mechanism $\mathcal{M}$ with voluntary participation. $\mathcal{M}$ may use multiple rounds of communication and information revelation to the bidder. For each valuation vector V and posterior π_σ, let $A(V, \pi_\sigma)$ be the (possibly randomized) equilibrium strategy of the bidder when his type is (V, π_σ).

Let $A(i, \zeta)$ be an indicator random variable that indicates whether the bidder gets the item when he chooses strategy ζ and the realized item type is i. Similarly, let $P(i, \zeta)$ denote the price the bidder is asked to pay. The bidder's interim expected utility is then given as follows:

$$\mathbb{E}_{i\sim\pi_\sigma}\left[\mathbb{E}\left[A(i,\zeta)\cdot V(i) - P(i,\zeta)\right]\right],$$

where the first (outer) expectation is with respect to the randomness of the item type, while the second (inner) expectation is with respect to the randomness in the choices of the mechanism, the information revealed and the actions ζ of the bidder.

For all possible types (V, π_σ), and for all possible misreports $(V', \pi_{\sigma'})$ of the bidder, for ζ to be an equilibrium strategy it must be the case that

$$\begin{aligned}&\mathbb{E}_{i\sim\pi_\sigma}\left[\mathbb{E}\left[A\left(i,\zeta\left(V,\pi_\sigma\right)\right)V(i)-P\left(i,\zeta\left(V,\pi_\sigma\right)\right)\right]\right]\\&\geq\mathbb{E}_{i\sim\pi_\sigma}\left[\mathbb{E}\left[A\left(i,\zeta\left(V',\pi_{\sigma'}\right)\right)V(i)-P\left(i,\zeta\left(V',\pi_{\sigma'}\right)\right)\right]\right].\end{aligned}$$

Now, let us abuse notations and define the variables

$$\begin{aligned}A_i\left(V,\pi_\sigma\right)&=\mathbb{E}\left[A\left(i,\zeta\left(V,\pi_\sigma\right)\right)\right],\\P_i\left(V,\pi_\sigma\right)&=\mathbb{E}\left[P\left(i,\zeta\left(V,\pi_\sigma\right)\right)\right].\end{aligned}$$

The equation above can be rewritten as

$$\sum_i \pi_\sigma(i)\left(A_i\left(V,\pi_\sigma\right)V(i)-P_i\left(V,\pi_\sigma\right)\right)\geq\sum_i \pi_\sigma(i)\left(A_i\left(V',\pi_{\sigma'}\right)V(i)-P_i\left(V',\pi_{\sigma'}\right)\right). \tag{IC}$$

Moreover, since the equilibrium ζ respects voluntary participation, the bidder's equilibrium payoff must be non-negative. As a consequence, we have

$$\sum_i \pi_\sigma(i)\left(A_i\left(V,\pi_\sigma\right)V(i)-P_i\left(V,\pi_\sigma\right)\right)\geq 0. \tag{IR}$$

Finally, we note that the revenue of the seller is given by

$$R=\sum_{\pi_\sigma,V}\Pr\left[V,\pi_\sigma\right]\sum_i \pi_\sigma(i)\cdot P_i\left(V,\pi_\sigma\right),$$

where $\Pr\left[V,\pi_\sigma\right]$ is the probability the realized type of the bidder is (V,π_σ).

A mechanism that satisfies constraints (IC) and (IR) and yields revenue R can clearly be implemented as an interim IR menu with conditional prices, in which the options are given by $\left(A\left(V,\pi_\sigma\right),c\left(V,\pi_\sigma\right)\right)$ for each possible type (V,π_σ). Thus, there exists an incentive compatible, individually rational, conditional price menu that provides the same revenue as mechanism $\mathcal{M}$. □

7.7 Proofs: Theorem 7.3.2

Proofs of Claim 7.3.3. In item-type pricing, the seller announces the item type (hence, completely superseding the effect of the data provider's signal) and then offers a price that is a function of the realized item type. The expected revenue of such a mechanism is simply given by

$$\frac{1}{m}\sum_{k=1}^{\eta}\frac{\eta}{k} = \Theta\left(\frac{\log m}{\sqrt{m}}\right),$$

as the expected revenue from selling an item of type i in the kth group is $P \cdot \Pr[V(i) \geq P] = P \cdot \frac{1}{kP} = \frac{1}{k}$, as $k \cdot V(i)$ follows an ER distribution. □

Proof of Claim 7.3.4. This proof follows the same structure as the proof of Proposition 25 of Hart et al. (2017). For all i and all $M \geq 1$, we let $V^M(i) = \min(V(i), M)$. By Hart et al. (2017), $V^M(i)$ has mean $\log M + 1$ and variance upper-bounded by $2M$. In particular, it follows that the expectation and variance of the value of the bundle (renormalized by m), were the bidders valuations truncated at M, satisfy:

$$\begin{aligned}\mathbb{E}\left[\sum_{k=1}\eta\sum_{i\in I_k}\frac{V^M(i)}{k}\right] &= \eta(\log M+1)\cdot\sum_{k=1}^{\eta}\frac{1}{k}\\ &\in \left[\frac{1}{2}(\log M+1)\sqrt{m}\log(m); (\log M+1)\sqrt{m}\left(1+\frac{1}{2}\log m\right)\right]\end{aligned}$$

as well as

$$\mathrm{Var}\left[\sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V^M(i)}{k}\right] \leq 2M\eta\sum_{k}\frac{1}{k^2} \leq \frac{\pi^2}{3}M\sqrt{m}$$

We first give a lower bound on the revenue of the item-type bundling mechanism.

$$\begin{aligned}&\Pr\left[\frac{1}{m}\sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V(i)}{k}\geq P\right]\\ &\geq \Pr\left[\sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V^M(i)}{k}\geq mP\right]\\ &= \Pr\left[\mathbb{E}\left[\sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V^M(i)}{k}\right]-\sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V^M(i)}{k}\leq\eta(\log M+1)\cdot\sum_{k=1}^{\eta}\frac{1}{k}-mP\right]\\ &= 1-\Pr\left[\mathbb{E}\left[\sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V^M(i)}{k}\right]-\sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V^M(i)}{k}>\eta(\log M+1)\cdot\sum_{k=1}^{\eta}\frac{1}{k}-mP\right]\end{aligned}$$

$$\geq 1 - \Pr\left[\left|\mathbb{E}\left[\sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V^M(i)}{k}\right] - \sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V^M(i)}{k}\right| > \eta\left(\log M + 1\right)\cdot\sum_{k=1}^{\eta}\frac{1}{k} - mP\right]$$

$$\geq 1 - \frac{\pi^2 M\sqrt{m}}{3\left(\eta\left(\log M + 1\right)\cdot\sum_{k=1}^{m}\frac{1}{k} - mP\right)^2}$$

$$\geq 1 - \frac{\pi^2 M\sqrt{m}}{3\left(\frac{1}{2}\left(\log M + 1\right)\sqrt{m}\log(m) - mP\right)^2},$$

where the second-to-last step follows from Chebyshev's inequality, in the case when $\eta\left(\log M + 1\right)\cdot\sum_{k=1}^{\eta}\frac{1}{k} - mP \geq 0$. Let $M = \sqrt{m}\log^2 m$ and $P = \frac{\log^2 m}{4\sqrt{m}}$, we obtain that

$$\Pr\left[\frac{1}{m}\sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V(i)}{k} \geq P\right]$$

$$\geq 1 - \frac{\pi^2 m\log^2 m}{3\left(\frac{1}{2}\left(\frac{1}{2}\log m + 2\log\log m + 1\right)\sqrt{m}\log m - \frac{\sqrt{m}\log^2 m}{4}\right)^2}$$

$$\geq 1 - \frac{\pi^2 m\log^2 m}{3\left(\sqrt{m}\log m\cdot\log\log m\right)^2}$$

$$\geq 1 - \frac{\pi^2}{3\left(\log\log m\right)^2}.$$

Therefore, a buyer buys a bundle with price $P = \frac{\log^2 m}{4\sqrt{m}}$ with constant probability (for m large enough), guaranteeing a revenue of $\Omega\left(\frac{\log^2 m}{\sqrt{m}}\right)$.

For the upper bound, we first remark that for $P \leq 2\frac{\log m}{\sqrt{m}}\left(1 + \frac{\log m}{2}\right)$, the revenue is at most $O\left(\frac{\log^2 m}{\sqrt{m}}\right)$. We there assume w.l.o.g that $P > 2\frac{\log m}{\sqrt{m}}\left(1 + \frac{\log m}{2}\right)$. The revenue from grand bundling at price P satisfies, by union bound:

$$P\cdot\Pr\left[\frac{1}{m}\sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V(i)}{k} \geq P\right] \leq P\cdot\Pr\left[\sum_{k=1}^{\eta}\sum_{i\in I_k}\frac{V^{mP}(i)}{k} \geq mP\right]$$
$$+ P\cdot\Pr\left[\exists k\in[\eta], i\in I_k :\ \frac{V(i)}{k} \geq mP\right]$$

By union bound, we have on the one hand that

$$\begin{aligned} P \cdot \Pr\left[\exists k \in [\eta], i \in I_k : \frac{V(i)}{k} \geq mP\right] &\leq P \sum_k \sum_{i \in I_k} \Pr\left[V(i) \geq kmP\right] \\ &= \eta P \sum_k \frac{1}{kmP} \\ &= O\left(\frac{\log m}{\sqrt{m}}\right) \end{aligned}$$

On the other hand, remembering that

$$\begin{aligned} \mathbb{E}\left[\sum_{k=1}^{\eta} \sum_{i \in I_k} \frac{V^{mP}(i)}{k}\right] &\leq (\log(mP) + 1)\sqrt{m}\left(1 + \frac{\log m}{2}\right) \\ \operatorname{Var}\left[\sum_{k=1}^{\eta} \sum_{i \in I_k} \frac{V^{mP}(i)}{k}\right] &\leq \frac{\pi^2}{3} m\sqrt{m}P, \end{aligned}$$

we have by Chebyshev that

$$\begin{aligned} & P \cdot \Pr\left[\sum_{k=1}^{\eta} \sum_{i \in I_k} \frac{V^{mP}(i)}{k} \geq mP\right] \\ &= P \cdot \Pr\left[\sum_{k=1}^{\eta} \sum_{i \in I_k} \frac{V^{mP}(i)}{k} - \mathbb{E}\left[\sum_{k=1}^{\eta} \sum_{i \in I_k} \frac{V^{mP}(i)}{k}\right] \geq mP - \mathbb{E}\left[\sum_{k=1}^{\eta} \sum_{i \in I_k} \frac{V^{mP}(i)}{k}\right]\right] \\ &\leq \frac{\pi^2 m\sqrt{m}P^2}{3\left(mP - (\log m + \log P + 1)\sqrt{m}\left(1 + \frac{\log m}{2}\right)\right)^2} \end{aligned}$$

Using the fact that w.l.o.g, $P \geq 2\frac{\log m}{\sqrt{m}}\left(1 + \frac{\log m}{2}\right)$, i.e.

$$\sqrt{m} \log m \left(1 + \frac{\log m}{2}\right) \leq mP/2,$$

and that

$$(\log P + 1)\sqrt{m}\left(1 + \frac{\log m}{2}\right) = o(mP),$$

we have that

$$\begin{aligned} & mP - (\log m + \log P + 1)\sqrt{m}\left(1 + \frac{\log m}{2}\right) \\ &= mP - (\log P + 1)\sqrt{m}\left(1 + \frac{\log m}{2}\right) - \sqrt{m} \log m \left(1 + \frac{\log m}{2}\right) \\ &\geq mP - o(mP) - \frac{mP}{2} \\ &= \Omega(mP), \end{aligned}$$

which leads to

$$P \cdot \Pr\left[\sum_{k=1}^{\eta} \sum_{i \in I_k} \frac{V^{mP}(i)}{k} \geq mP\right] = O\left(\frac{m\sqrt{m}P^2}{m^2P^2}\right) = O\left(\frac{1}{\sqrt{m}}\right).$$

It follows that

$$P \cdot \Pr\left[\exists k \in [\eta], i \in I_k : \frac{V(i)}{k} \geq mP\right] = O\left(\frac{\log m}{\sqrt{m}}\right),$$

which concludes the proof. □

Proof of Claim 7.3.5. Let P^* be the optimal bundling price, and suppose the data provider announces signal σ_k. There are two cases:

1. For k such that $P^* \geq \frac{6}{k} \log \eta$, by Lemma 7.2.8, the expected revenue is

$$P^* \cdot \Pr\left[\frac{1}{\eta} \sum_{i \in |I_k|} \frac{V(i)}{k} \geq P^*\right] = P^* \cdot \Pr\left[\frac{1}{\eta} \sum_{i \in |I_k|} V(i) \geq kP^*\right] \leq P^* \cdot \frac{9}{kP^*} = \frac{9}{k},$$

   as $|I_k| = \eta$.

2. Otherwise, we have k such that $P^* \leq \frac{6}{k} \log \eta$.

Letting $k^* = \min\{k : P^* > \frac{6}{k} \log \eta\}$, we see that the expected revenue of charging price P^* for the grand bundle is upper-bounded by

$$\begin{aligned}
\frac{1}{\eta}\left(\sum_{k \geq k^*} \frac{9}{k} + \sum_{k < k^*} P^*\right) &\leq \frac{1}{\sqrt{m}}\left(9 \cdot (1 + \log \eta) + \sum_{k < k^*} \frac{6}{k^* - 1} \log \eta\right) \\
&= \frac{1}{\sqrt{m}} (9 \cdot (1 + \log \eta) + 6 \log \eta) \\
&= O\left(\frac{\log m}{\sqrt{m}}\right).
\end{aligned}$$

□

Proof of Claim 7.3.6. Consider the following item-type partition mechanism: the seller first partitions the item types into η groups in the same way as specified in Construction 7.3.1. When the realized item type is in group I_k, she offers to sell the item to the bidder at price $P_k = \frac{\log \eta}{2k}$.

If the bidder receives signal σ_k, then the price offered by the seller must be $\frac{\log \eta}{2k}$, and the bidder knows the item type is from group I_k. By Lemma 7.2.8, as $|I_k| = \eta$, we have:

$$\Pr\left[\frac{1}{\eta} \sum_{i \in I_k} \frac{V(i)}{k} \geq \frac{\log \eta}{2k}\right] \geq \frac{1}{2},$$

and hence with probability at least $1/2$, conditional on $S = \sigma_k$, he accepts the price, yielding expected revenue to the seller of at least $\frac{\log \eta}{4k}$. The total expected revenue for the seller is then given by

$$\frac{1}{\eta} \sum_{k=1}^{\eta} \frac{\log \eta}{4k} = \frac{\log m}{8\sqrt{m}} \sum_{k=1}^{\eta} \frac{1}{k} = \Omega\left(\frac{\log^2 m}{\sqrt{m}}\right).$$

No truthful mechanism can achieve revenue higher than $\log \eta$ times the revenue of item-type pricing conditioned on receiving signal σ_k: Theorem 2 of Li et al. (2013) shows that in traditional multi-item auctions, selling separately achieves at least a $\Omega\left(\frac{1}{\log \eta}\right)$ fraction of the optimal revenue for selling η independent items; this result carries over to single-item, multi-type auctions by the reduction of Daskalakis et al. (2016). Thus, the optimal revenue is at most $O\left(\frac{\log^2 m}{\sqrt{m}}\right)$, and hence the item-type partition mechanism we just described yields a constant approximation to the optimal revenue. □

7.8 Proofs: Theorem 7.4.2

Proof of Claim 7.4.3. Suppose the item-type partition mechanism splits the item types into non-empty groups $\mathcal{G}_1$ to $\mathcal{G}_g$, where $g \leq 2^\eta$ is the number of such groups. Let's assume that the item type i lies in $\mathcal{G}_r$, then the seller offers to sell the item at price at P_r. Suppose the signal is $\sigma_{k,j}$ for some $j \in [m_k]$ with $i \in \mathcal{I}_{k,j}$. In the bidder's posterior, the item type is uniform over $\mathcal{G}_r \cap \mathcal{I}_{k,j}$. Note that $\left|\mathcal{G}_r \cap \mathcal{I}_{k,j}\right| \leq 2^k$. By Lemma 7.2.8, we have

$$P_r \cdot \Pr\left[\frac{1}{\left|\mathcal{G}_r \cap \mathcal{I}_{k,j}\right|} \sum_{t \in \mathcal{G}_r \cap \mathcal{I}_{k,j}} V(t) \geq P_r\right] \leq \begin{cases} 9 & \text{if } P_r \geq 6 \log\left(2^k\right) = 6k \log 2 \\ P_r & \text{if } P_r < 6k \log 2, \end{cases},$$

following from $6 \log\left(2^k\right) \geq 6 \log\left|\mathcal{G}_r \cap \mathcal{I}_{k,j}\right|$.

Let $k^*(r) = \max\{k : P_r \geq 6k \log 2\}$. Further, let us denote by $\Pr[k]$ the probability that the data provider selects a partition of size 2^k. When item type $i \in \mathcal{G}_r$, the

revenue in expectation over the randomness of the signal is upper-bounded by

$$\begin{aligned}
&9 \sum_{k \le k^*(r)} \Pr[k] + P_r \cdot \left(\sum_{k=k^*(r)+1}^{\eta} \Pr[k] \right) \\
&= 9 \sum_{k \le k^*(r)} \frac{1}{k(k+1)} + P_r \cdot \left(\sum_{k=k^*(r)+1}^{\eta-1} \frac{1}{k(k+1)} + \frac{1}{\eta} \right) \\
&\le 9 \left(1 - \frac{1}{k^*(r)+1} \right) + 6 \log 2 \cdot (k^*(r)+1) \left(\frac{1}{k^*(r)+1} - \frac{1}{\eta} + \frac{1}{\eta} \right) \\
&\le 9 + 6 \log 2,
\end{aligned}$$

where the first step follows from the fact that the probability of the data provider selecting a $k \le \eta - 1$ is $\frac{1}{k(k+1)}$, and the probability of him drawing $k = \eta$ is $\frac{1}{\eta}$. Since the upper bound holds for all possible prices, the expected revenue of any item-type partition mechanism is also upper-bounded by $9 + 6 \log 2$. □

Proof of Claim 7.4.6. The bidder's expected utility for $L_{k,j}$ when receiving signal $\sigma_{k,j}$ is given by

$$U_{k,j} = \frac{1}{2^k} \sum_{i \in I_{k,j}} V(i) - \frac{1}{8} \log 2^k,$$

his expected utility for selecting option $L_{\kappa,\iota}$ for $\kappa > k$ is only less (his expected value for the item type is not more, but the price is higher), and his utility for selecting option $L_{\kappa,\iota}$ for $\kappa < k$ is

$$U_{\kappa,\iota} = \frac{1}{2^k} \sum_{i \in I_{\kappa,\iota} \cap I_{k,j}} V(i) - \frac{1}{8} \log 2^\kappa,$$

and his expected utility for selecting any option $L_{\kappa,\iota}$ such that $I_{\kappa,\iota} \cap I_{k,j} = \emptyset$ is negative, since he will pay but never be allocated the item.

Therefore, the bidder prefers $L_{\kappa,\iota}$ to $L_{k,j}$ with $\kappa \le k$ and $I_{\kappa,\iota} \subset I_{k,j}$ only if

$$\frac{1}{2^k} \sum_{i \in I_{k,j} \setminus I_{\kappa,\iota}} V(i) \le \frac{1}{8} \log 2^{k-\kappa}.$$

We want to upper bound the probability of the above event for all $\kappa < k$ and $I_{\kappa,\iota} \subset I_{k,j}$. Let us denote $V^M(i) = \min(V(i), M)$ for any M. We have immediately that $\mathbb{E}\left[V^M(i)\right] = \log M + 1$ and that its variance is upper-bounded by $2M$. Taking $M = 2^{k-1}$ and $W(i) = \log M + 1 - V^M(i)$ yields $|W(i)| \le M$, $\mathbb{E}[W(i)] = 0$ and $\mathbb{E}\left[W(i)^2\right] \le 2 \cdot 2^{k-1} = 2^k$. Recall Bernstein's inequality:

Lemma 7.8.1. *(Bernstein's Inequality): Suppose $Y_1, ..., Y_m$ are independent random variables with zero mean, and $|Y_i| \leq B$ almost surely for all i. Then for any $t > 0$,*

$$\Pr\left[\sum_{i=1} Y_i > t\right] \leq \exp\left(-\frac{\frac{1}{2}t^2}{\sum_{i=1}^{m} E[Y_i^2] + \frac{1}{3}Bt}\right)$$

We can then apply Bernstein's inequality to show that

$$\begin{aligned}
&\Pr\left[\frac{1}{2^k}\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) < \frac{1}{2^k}\cdot\left(\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} (\log M + 1) - t\right)\right] \\
&= \Pr\left[\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) < \sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} (\log M + 1) - t\right] \\
&\leq \Pr\left[\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V^M(i) < \sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} (\log M + 1) - t\right] \\
&= \Pr\left[\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} W(i) > t\right] \\
&\leq \exp\left(-\frac{1}{2}\cdot\frac{t^2}{2^k\cdot|I_{k,j}\setminus I_{\kappa,\iota}| + M\cdot t/3}\right) \\
&= \exp\left(-\frac{1}{2}\cdot\frac{t^2}{2^k\left(2^k - 2^\kappa\right) + M\cdot t/3}\right),
\end{aligned}$$

where Bernstein's inequality is used in the last inequality. Taking

$$t = \left(\frac{3}{4}\right)\left(2^k - 2^\kappa\right)(\log M + 1),$$

we have

$$\begin{aligned}
\frac{1}{2^k}\left(\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} (\log M + 1) - t\right) &= \frac{1}{2^k}\cdot\frac{1}{4}\left(2^k - 2^\kappa\right)(\log M + 1) \\
&\geq \frac{1}{2^k}\cdot\frac{1}{4}\left(2^{k-1}\right)(\log M + 1) \\
&= \frac{1}{8}(\log M + 1),
\end{aligned}$$

and we thus obtain a bound on the probability of the event that a particular menu

option $L_{\kappa,\iota}$ for $\kappa < k$ is better for the bidder than option $L_{k,j}$, given signal $\sigma_{k,j}$:

$$\begin{aligned}
&\Pr\left[\frac{1}{2^k}\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) \le \frac{1}{8}\log 2^{k-\kappa}\right] \\
&< \Pr\left[\frac{1}{2^k}\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) < \frac{1}{8}\left(\log 2^{k-1}+1\right)\right] \\
&\le \Pr\left[\frac{1}{2^k}\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) < \frac{1}{2^k}\left(\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}}\left(\log 2^{k-1}+1\right)-t\right)\right] \\
&\le \exp\left(-\frac{k^2}{2}\cdot\frac{(3/4)^2\left(2^k-2^\kappa\right)^2(\log 2)^2}{2^k\left(2^k-2^\kappa\right)+\frac{1}{4}2^{k-1}\left(2^k-2^\kappa\right)\left(\log 2^{k-1}+1\right)}\right) \\
&\le \exp\left(-\frac{(k-1)^2}{2}\cdot\frac{(3/4)^2\cdot 2^{k-1}\left(2^k-2^\kappa\right)(\log 2)^2}{2^k\left(2^k-2^\kappa\right)+\frac{1}{4}2^{k-1}\left(2^k-2^\kappa\right)\left(\log 2^{k-1}+1\right)}\right) \\
&\le \exp\left(-\frac{k-1}{2}\cdot\frac{(3/4)^2(\log 2)^2}{\frac{2}{k-1}+\frac{1}{4}\left(\log 2+\frac{1}{k-1}\right)}\right),
\end{aligned}$$

For $k \ge 2\cdot 10^2+1$, the above yields

$$\begin{aligned}
&\Pr\left[\frac{1}{2^k}\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) < \frac{1}{8}\left(\log 2^{k-1}+1\right)\right] \\
&\le \exp\left(-(k-1)\cdot\frac{(3/4)^2(\log 2)^2}{\frac{4}{2\cdot 10^2}+\frac{1}{2}\left(\log 2+\frac{1}{2\cdot 10^2}\right)}\right).
\end{aligned}$$

We now let

$$K = \exp\left(\frac{(3/4)^2(\log 2)^2}{\frac{4}{2\cdot 10^2}+\frac{1}{2}\left(\log 2+\frac{1}{2\cdot 10^2}\right)}\right),$$

and note that we then have that for $k \ge 2\cdot 10^2+1$,

$$\Pr\left[\frac{1}{2^k}\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) < \frac{1}{8}\left(\log 2^{k-1}+1\right)\right] \le \left(\frac{1}{K}\right)^{k-1}.$$

Since there are less than 2^k groups $I_{\kappa,\iota}$ such that $I_{\kappa,\iota} \subset I_{k,j}$, a union bound gives us that the probability that the bidder prefers a different option other than $L_{k,j}$ is upper bounded by $2\cdot\left(\frac{2}{K}\right)^{k-1}$. A direct calculation shows that $2\cdot\left(\frac{2}{K}\right)^{k-1} \le 10^{-3}$. □

We are now ready to prove Claim 7.4.4.

Proof of Claim 7.4.4. The proof of Claim 7.4.6 directly implies that the revenue of the considered mechanism is lower-bounded by

$$\frac{(1-10^{-3})\log 2}{8}\left(\sum_{k\geq 2\cdot 10^2+1}^{\eta-1}\frac{k}{k(k+1)}+\frac{\eta}{\eta}\right)=\Omega\left(\log\eta\right)=\Omega\left(\log\log m\right),$$

as a bidder who receives signal $\sigma_{k,j}$ picks option $L_{k,j}$ with price $\frac{\log 2}{8}k$ with probability at least $1-10^{-3}$.

The revenue of the best mechanism is upper-bounded by the optimal revenue the seller could obtain if she knew the realization of the signal. When facing signal $\sigma_{k,j}$, the bidder's posterior is that the item type is taken uniformly at random from group $I_{k,j}$. By Babaioff et al. (2014) and Daskalakis et al. (2016), the better of item-type pricing and item-type bundling (conditioning now on the realization of the signal) yields a constant approximation to the optimal revenue. The revenue from item-type pricing is clearly 1, and the revenue from item-type bundling is $O\left(\log 2^k\right)$ by Lemma 7.2.8 as setting $P>6\log 2^k$ yields constant revenue while setting $P\leq 6\log 2^k$ yields $O\left(\log 2^k\right)$. Therefore, the optimal revenue conditional on the signal being $\sigma_{k,j}$ must be $O\left(\log 2^k\right)=O\left(k\right)$, and the optimal (unconditional) revenue is therefore

$$O\left(\sum_{k=1}^{\eta-1}\frac{k}{k(k+1)}+\frac{\eta}{\eta}\right)=O\left(\log\eta\right)=O\left(\log\log m\right). \qquad \square$$

7.9 Proofs: Lemma 7.5.1

Here, we provide the proof of Lemma 7.5.1:

Proof of Lemma 7.5.1. Consider a lottery of the form given in Section 7.2 and discussed in Section 7.6. Suppose the lottery has $l+1$ options, denoted by L_0, L_1 to L_l, where L_0 is a dummy option with price 0 and allocation $A_0(i)=0$ added to guarantee IR (as in Section 7.6). Further, let $A_k(i)$ denote the probability with which L_k allocates item of type i, and $P_k(i)$ the price at which L_k sells item of type i. The expected utility of the bidder when he has valuation V and signal $S(X)$ is given by

$$U(V,S)=\mathbb{E}_X\left[\mathbb{E}_{\sigma\sim S(X)}\left[\max_k\sum_i\pi_\sigma(i)\left(V(i)A_k(i)-P_k(i)\right)\right]\right].$$

On the other hand, if the data provider fully reveals his information, the bidder possessing this information and with value V would have utility

$$U(V, S^*) = \mathbb{E}_X \left[\max_k \sum_i \pi_X(i)(V(i)A_k(i) - P_k(i)) \right].$$

Since the bidder's posterior when observing the realization of σ is obtained via Bayes update, we have $\pi_\sigma = \mathbb{E}_{\tilde{X} \mid \sigma}[\pi_{\tilde{X}}]$, where on the right hand side the expectation is taken over $\tilde{X}$, the buyer's belief of the data provider's information, drawn from the conditional distribution given the received signal σ. Therefore we can write

$$\begin{aligned}
U(V, S) &= \mathbb{E}_X \left[\mathbb{E}_{\sigma \sim S(X)} \left[\max_k \mathbb{E}_{\tilde{X} \mid \sigma} \left[\sum_i \pi_{\tilde{X}}(i)(V(i)A_k(i) - P_k(i)) \right] \right] \right] \\
&\leq \mathbb{E}_X \left[\mathbb{E}_{\sigma \sim S(X)} \left[\mathbb{E}_{\tilde{X} \mid \sigma} \left[\max_k \sum_i \pi_{\tilde{X}}(i)(V(i)A_k(i) - P_k(i)) \right] \right] \right] \\
&= \mathbb{E}_{\tilde{X}} \left[\max_k \sum_i \pi_{\tilde{X}}(i)(V(i)A_k(i) - P_k(i)) \right] \\
&= \mathbb{E}_X \left[\max_k \sum_i \pi_X(i)(V(i)A_k(i) - P_k(i)) \right] \\
&= U(V, S^*),
\end{aligned}$$

where the inequality follows from Jensen's inequality. Conditional on the signal being s, the distributions of $\tilde{X}|s$ and $X|s$ are identical by definition of Bayes update, which in turn directly implies the distributions of $\tilde{X}$ and X are identical, and the second-to-last equality holds. This concludes the proof. □

Part 5

Conclusion and Discussion

Chapter 8

CONCLUSION AND DISCUSSION

The goal of this thesis is to address some of the challenges related to the exchange and use of data. In particular, in this thesis, we build formal and theoretical foundations for markets for data, for understanding how data is tied with strategic behavior, and for dealing with the societal considerations and issues that arise from the use of data. Yet, many open questions and future directions remain to be explored.

Chapters 3 and 4 of this thesis focus on the study of data markets, and more precisely aim to design mechanisms for data acquisition and aggregation, when this data is held by strategic agents. The work presented in these chapters looks at off-line data acquisition decisions, in which data points come in a batch. In practice, it is often the case that data holders may not be available simultaneously; further, an analyst may want to use information about the data points he has already collected to decide which data will be most useful to him in the future. As such, an important future direction is that of mechanism design for "online" data acquisition, when agents participate sequentially in the data analyst's mechanism, and the analyst uses the history of acquired data points to inform future data acquisition decisions.

Further, most of the current line of work on data acquisition (including the work presented in this thesis) is tailored to simple statistical tasks, such as moment and parameter estimation. With the growing importance of machine learning and automated decision-making, future work should aim to develop mechanisms that efficiently purchase data from data holders for classification tasks.

Another practical consideration is that many settings of interest are complex and involve many entities that may bid or compete over agents' data, as well as intermediaries that must decide how to acquire data from many sources, how to allocate and sell this data to a variety of interested buyers, and what services to provide on said data; therefore, optimal mechanism design for data acquisition and aggregation in multi-sided settings, with many (possibly competing) agents that aim to buy and/or sell data, is a natural direction to explore.

When it comes to data acquisition, it is also important to guarantee the privacy of individuals whose sensitive data is used. We show how to acquire data in a differentially private manner, and how to compensate agents for any remaining

privacy loss, in Chapter 5. We focus on the case in which agent data and costs are independent. However, there is a need to understand data acquisition when privacy costs are correlated with how sensitive the data is, as is often the case in real life; a concrete question in this direction is how to use differential privacy to prevent information about the agents' costs from being leaked to an adversary that observes data purchasing decisions and the outcome of the mechanism. While Ghosh et al. (2015) suggest this might be an impossible task in the worst case, Nissim et al. (2014) show that restricting attention to special cases, many of which could potentially capture practical applications, helps.

Further, this thesis and most of the line of work on data acquisition with privacy constraints consider settings where an individual's perception of how much he values privacy is fixed. Realistically, this may not be the case; an individual may value the same differential privacy level differently, depending on how many computations are performed on his data, and how widely his data is shared. As such, data acquisition with privacy becomes a much more relevant (and challenging) task when studied in the context of multi-sided markets for data, where one individual's data may be acquired, used, and sold by many different entities.

Chapter 6 aims to understand how various individuals or populations may be treated unfairly based on sensitive attributes such as gender or race, and to prevent such disparate treatment. There, we focus on fairness concerns in the context of university admissions, and identify unequal access to signaling as a possible source of unfairness, that may be exacerbated by classical interventions (e.g., forcing students to pass a standardized test). There may be many other sources of unfairness between individuals or populations that are not yet well understood; as such, it is crucial that future work aims to identify and study such sources, so as to understand what interventions are efficient, in what contexts.

Additionally, long-term considerations are central when it comes to fairness in decision-making. Decisions are rarely made in isolation, and it is often the case that decisions made today inform, compose with, and affect decisions made in the future. In such cases, fairness of individual decisions does not guarantee fairness of a combination of decisions, as observed by Dwork et al. (2018). The work of Kannan et al. (2019) mentioned in this thesis studies such long-term effects, when decisions made at the university level affect decision-making on the job market. However, there is still a need to develop broader scope decision-making algorithms that do not introduce bias and prevent bias from propagating, even when they are composed

with other and possibly future decisions, and even when these other decisions are outside of the control of the decision maker.

A recurrent theme in this thesis is that of agents acting strategically, and a question of importance is fairness in the context of strategic behavior. In particular, can one design mechanisms that treat individuals or populations fairly, even when these agents may have incentives to misreport their data to obtain better outcomes, and different agents may have different abilities to conduct such manipulations? How does one design mechanisms that incentivize decision-makers to explore and give a chance to populations that have been historically discriminated against, and may in turn appear unappealing (either due to lack of accurate data on, or to persisting bias against such populations)?

Finally, Chapter 7 studies the role of data as a source of information that explains but also affects strategic behavior, and in turn mechanism design. More precisely, the chapter examines settings of incomplete information with third-party information revelation, and shows that there, standard, simple mechanisms do not capture a significant fraction of the achievable revenue. A natural open question, from the mechanism designer's point of view, is to understand what *simple* and *near-optimal* mechanisms look like, if they even exist in the first place. On the other hand, a data provider with access to information about a strategic setting may wonder how to release some of this information in order to affect strategic behavior and mechanism design, and to incentivize specific outcomes—in particular, socially better, and fairer ones.

BIBLIOGRAPHY

Jacob Abernethy, Yiling Chen, Chien-Ju Ho, and Bo Waggoner (2015). "Low-Cost Learning via Active Data Procurement". In: *Proceedings of the 16th ACM Conference on Economics and Computation*, pp. 619–636.

John M Abowd (2018). "The U.S. Census Bureau Adopts Differential Privacy". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2867–2867.

Ricardo Alonso and Odilon Camara (2016). "Persuading Voters". In: *American Economic Review* 106.11, pp. 3590–3605.

Simon P. Anderson and Regis Renault (2006). "Advertising Content". In: *American Economic Review* 96.1, pp. 93–113.

Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner (2016). "Machine Bias: There's Software Used Across The Country to Predict Future Criminals, And It's Biased Against Blacks." In: *ProPublica* 23.

Itai Arieli and Yakov Babichenko (2016). "Private Bayesian Persuasion". In: *Manuscript*.

Moshe Babaioff, Nicole Immorlica, Brendan Lucier, and S Matthew Weinberg (2014). "A Simple and Approximately Optimal Mechanism for an Additive Buyer". In: *Proceedings of the 2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pp. 21–30.

Dirk Bergemann, Benjamin A Brooks, and Stephen Morris (2016). "Informationally Robust Optimal Auction Design". In: *Manuscript*.

Dirk Bergemann, Benjamin Brooks, and Stephen Morris (2015). "The Limits of Price Discrimination". In: *American Economic Review* 105.3, pp. 921–57.

– (2017). "First-Price Auctions with General Information Structures: Implications for Bidding and Revenue". In: *Econometrica* 85.1, pp. 107–143.

– (2018). "Revenue Guarantee Equivalence". In: *Manuscript*.

Dirk Bergemann and Stephen Morris (2019). "Information Design: A Unified Perspective". In: *Journal of Economic Literature* 57.1, pp. 44–95.

Dirk Bergemann and Martin Pesendorfer (2007). "Information Structures in Optimal Auctions". In: *Journal of Economic Theory* 137.1, pp. 580–609.

Stephen Boyd and Lieven Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.

Peter Bro Miltersen and Or Sheffet (2012). "Send Mixed Signals: Earn More, Work Less". In: *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 234–247.

Isabelle Brocas and Juan D. Carrillo (2007). "Influence Through Ignorance". In: *RAND Journal of Economics* 38.4, pp. 931–947.

Yang Cai, Constantinos Daskalakis, and Christos Papadimitriou (2015). "Optimum Statistical Estimation with Strategic Data Sources". In: *Conference on Learning Theory*, pp. 280–296.

Yang Cai, Nikhil R. Devanur, and S. Matthew Weinberg (2016). "A Duality Based Unified Approach to Bayesian Mechanism Design". In: *The 48th Annual ACM Symposium on Theory of Computing*.

Yang Cai, Federico Echenique, Hu Fu, Katrina Ligett, Adam Wierman, and Juba Ziani (2018). "Third-Party Data Providers Ruin Simple Mechanisms". In: *arXiv preprint arXiv:1802.07407*, **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** URL: http://arxiv.org/abs/1802.07407.

Yang Cai and Zhiyi Huang (2013). "Simple and Nearly Optimal Multi-Item Auctions". In: *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 564–577.

Yang Cai and Mingfei Zhao (2017). "Simple Mechanisms for Subadditive Buyers via Duality". In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 170–183.

Ioannis Caragiannis, Ariel Procaccia, and Nisarg Shah (2016). "Truthful Univariate Estimators". In: *International Conference on Machine Learning*, pp. 127–135.

Venkat Chandrasekaran, Katrina Ligett, and Juba Ziani (2016). "Efficiently Characterizing Games Consistent with Perturbed Equilibrium Observations". In: *arXiv preprint arXiv:1603.01318.* **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** URL: https://arxiv.org/abs/1603.01318.

Shuchi Chawla and J Benjamin Miller (2016). "Mechanism Design for Subadditive Agents via an Ex-Ante Relaxation". In: *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 579–596.

Yiling Chen, Nicole Immorlica, Brendan Lucier, Vasilis Syrgkanis, and Juba Ziani (2018a). "Optimal Data Acquisition for Statistical Estimation". In: *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 27–44. **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** DOI: 10.1145/3219166.3219195. URL: http://doi.acm.org/10.1145/3219166.3219195.

Yiling Chen, Chara Podimata, Ariel D Procaccia, and Nisarg Shah (2018b). "Strategyproof Linear Regression in High Dimensions". In: *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 9–26.

Yiling Chen and Shuran Zheng (2018c). "Prior-free Data Acquisition for Accurate Statistical Estimation". In: *arXiv preprint arXiv:1811.12655*.

Yu Cheng, Ho Yee Cheung, Shaddin Dughmi, Ehsan Emamjomeh-Zadeh, Li Han, and Shang-Hua Teng (2015). "Mixture Selection, Mechanism Design, and Signaling". In: *Proceedings of the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pp. 1426–1445.

Alexandra Chouldechova (2017). "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments". In: *Big data* 5.2, pp. 153–163.

Stephen Coate and Glenn C Loury (1993). "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" In: *American Economic Review*, pp. 1220–1240.

Ethan Cohen-Cole (2011). "Credit Card Redlining". In: *Review of Economics and Statistics* 93.2, pp. 700–713.

Rachel Cummings, Stratis Ioannidis, and Katrina Ligett (2015a). "Truthful Linear Regression". In: *Conference on Learning Theory*, pp. 448–483.

Rachel Cummings, Katrina Ligett, Aaron Roth, Zhiwei Steven Wu, and Juba Ziani (2015b). "Accuracy for Sale: Aggregating Data with a Variance Constraint". In: *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pp. 317–324. **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** DOI: 10.1145/2688073.2688106. URL: http://doi.acm.org/10.1145/2688073.2688106.

Constantinos Daskalakis, Christos Papadimitriou, and Christos Tzamos (2016). "Does Information Revelation Improve Revenue?" In: *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 233–250.

Ofer Dekel, Felix Fischer, and Ariel D Procaccia (2010). "Incentive Compatible Regression Learning". In: *Journal of Computer and System Sciences* 76.8, pp. 759–777.

Shahar Dobzinski and Shaddin Dughmi (2009). "On the Power of Randomization in Algorithmic Mechanism Design". In: *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pp. 505–514.

Jinshuo Dong, Aaron Roth, Zachary Schutzman, Bo Waggoner, and Zhiwei Steven Wu (2018). "Strategic Classification from Revealed Preferences". In: *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 55–70.

Songzi Du (2018). "Robust Mechanisms Under Common Valuation". In: *Econometrica* 86.5, pp. 1569–1588.

Shaddin Dughmi (2014). "On the Hardness of Signaling". In: *Proceedings of the 2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pp. 354–363.

– (2017). "Algorithmic Information Structure Design: a Survey". In: *ACM SIGecom Exchanges* 15.2, pp. 2–24.

Shaddin Dughmi, Nicole Immorlica, Ryan O'Donnell, and Li-Yang Tan (2015). "Algorithmic Signaling of Features in Auction Design". In: *International Symposium on Algorithmic Game Theory*, pp. 150–162.

Shaddin Dughmi, Nicole Immorlica, and Aaron Roth (2014). "Constrained Signaling in Auction Design". In: *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1341–1357.

Shaddin Dughmi and Haifeng Xu (2016). "Algorithmic Bayesian Persuasion". In: *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, pp. 412–425.

– (2017). "Algorithmic Persuasion with No Externalities". In: *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 351–368.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel (2012). "Fairness Through Awareness". In: *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pp. 214–226.

Cynthia Dwork and Christina Ilvento (2018). "Fairness Under Composition". In: *arXiv preprint arXiv:1806.06122*.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith (2006). "Calibrating Noise to Sensitivity in Private Data Analysis". In: *Proceedings of the 3rd Conference on Theory of Cryptography*, pp. 265–284.

Cynthia Dwork and Aaron Roth (2014). "The Algorithmic Foundations of Differential Privacy". In: *Foundations and Trends® in Theoretical Computer Science* 9.3–4, pp. 211–407.

Yuval Emek, Michal Feldman, Iftah Gamzu, Renato PaesLeme, and Moshe Tennenholtz (2014). "Signaling Schemes for Revenue Maximization". In: *ACM Transactions on Economics and Computation* 2.2, 5:1–5:19.

Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova (2014). "Rappor: Randomized Aggregatable Privacy-Preserving Ordinal Response". In: *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1054–1067.

Péter Eső and Balazs Szentes (2007). "Optimal Information Disclosure in Auctions and the Handicap Auction". In: *The Review of Economic Studies* 74.3, pp. 705–731.

Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian (2015). "Certifying and Removing Disparate Impact". In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 259–268.

Lisa K. Fleischer and Yu-Han Lyu (2012). "Approximately Optimal Auctions for Selling Privacy when Costs Are Correlated with Data". In: *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 568–585.

Dean P Foster and Rakesh V Vohra (1992). "An Economic Argument for Affirmative Action". In: *Rationality and Society* 4.2, pp. 176–188.

Yoav Freund and Robert E. Schapire (1999). "Adaptive Game Playing Using Multiplicative Weights". In: *Games and Economic Behavior* 29.1, pp. 79–103.

Sorelle A Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian (2016). "On the (im) possibility of fairness". In: *arXiv preprint arXiv:1609.07236*.

FTC (2014). *FTC Recommends Congress Require the Data Broker Industry to be More Transparent and Give Consumers Greater Control Over Their Personal Information*. URL: https://www.ftc.gov/news-events/press-releases/2014/05/ftc-recommends-congress-require-data-broker-industry-be-more.

Hu Fu, Patrick Jordan, Mohammad Mahdian, Uri Nadav, Inbal Talgam-Cohen, and Sergei Vassilvitskii (2012). "Ad Auctions with Data". In: *International Symposium on Algorithmic Game Theory*, pp. 168–179.

Juan-Jo sé Ganuza (2004). "Ignorance Promotes Competition: An Auction Model with Endogenous Private Valuations". In: *Rand Journal of Economics*, pp. 583–598.

Matthew Gentzkow and Emir Kamenica (2014). "Costly Persuasion". In: *American Economic Review* 104.5, pp. 457–62.

– (2017). "Competition in Persuasion". In: *The Review of Economic Studies* 84.1, pp. 300–322.

Arpita Ghosh and Katrina Ligett (2013). "Privacy and Coordination: Computing on Databases with Endogenous Participation". In: *Proceedings of the 14th ACM conference on Electronic Commerce*, pp. 543–560.

Arpita Ghosh, Katrina Ligett, Aaron Roth, and Grant Schoenebeck (2014). "Buying Private Data without Verification". In: *Proceedings of the 15th ACM conference on Economics and computation*, pp. 931–948.

Arpita Ghosh and Aaron Roth (2015). "Selling Privacy at Auction". In: *Games and Economic Behavior* 91, pp. 334–346.

Jerry R. Green and Jean-Jacques Laffont (1977). "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods". In: *Econometrica* 45.2, pp. 427–438.

Andy Greenberg (2016). "Apple's 'Differential Privacy' Is About Collecting Your Data–But Not Your Data". In: *Wired Magazine*.

Mingyu Guo and Argyrios Deligkas (2013). "Revenue Maximization via Hiding Item Attributes". In: *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pp. 157–163.

Sara Hajian and Josep Domingo-Ferrer (2013). "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining". In: *IEEE Transactions on Knowledge and Data Engineering* 25.7, pp. 1445–1459.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters (2016a). "Strategic Classification". In: *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pp. 111–122.

Moritz Hardt, Eric Price, and Nathan Srebro (2016b). "Equality of Opportunity in Supervised Learning". In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 3323–3331.

Sergiu Hart and Noam Nisan (2013). "The Menu-Size Complexity of Auctions". In: *Proceedings of the 14th ACM Conference on Electronic commerce*, pp. 565–566.

– (2017). "Approximate Revenue Maximization with Multiple Items". In: *Journal of Economic Theory* 172, pp. 313–347.

Sergiu Hart and Philip J Reny (2015). "Maximal Revenue with Multiple Goods: Nonmonotonicity and Other Observations". In: *Theoretical Economics* 10.3, pp. 893–922.

Bengt Holmström (1979). "Groves' Scheme on Restricted Domains". In: *Econometrica* 47.5, pp. 1137–1144.

Daniel G Horvitz and Donovan J Thompson (1952). "A Generalization of Sampling without Replacement from a Finite Universe". In: *Journal of the American statistical Association* 47.260, pp. 663–685.

Lily Hu and Yiling Chen (2017). "Fairness at Equilibrium in the Labor Market". In: *arXiv preprint arXiv:1707.01590*.

Nicole Immorlica, Katrina Ligett, and Juba Ziani (2019). "Access to Population-Level Signaling As a Source of Inequality". In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 249–258. **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** DOI: 10.1145/3287560.3287579. URL: http://doi.acm.org/10.1145/3287560.3287579.

"Implementing the "Wisdom of the Crowd"" (2014). In: *Journal of Political Economy* 122.5, pp. 988–1012.

Justin P. Johnson and David Myatt (2006). "On the Simple Economics of Advertising, Marketing, and Product Design". In: *American Economic Review* 96.3, pp. 756–784.

Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth (2016). "Fairness in Learning: Classic and Contextual Bandits". In: *Advances in Neural Information Processing Systems*, pp. 325–333.

Emir Kamenica and Matthew Gentzkow (2011). "Bayesian Persuasion". In: *American Economic Review* 101.6, pp. 2590–2615.

Faisal Kamiran and Toon Calders (2012). "Data Preprocessing Techniques for Classification without Discrimination". In: *Knowledge and Information Systems* 33.1, pp. 1–33.

Sampath Kannan, Michael Kearns, Jamie Morgenstern, Mallesh Pai, Aaron Roth, Rakesh Vohra, and Zhiwei Steven Wu (2017). "Fairness Incentives for Myopic Agents". In: *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 369–386.

Sampath Kannan, Aaron Roth, and Juba Ziani (2019). "Downstream Effects of Affirmative Action". In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 240–248. **J. Ziani is the primary author, came up with and proved most of the results, and contributed to writing the manuscript.** DOI: 10.1145/3287560.3287578. URL: http://doi.acm.org/10.1145/3287560.3287578.

Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu (2018). "Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80, pp. 2564–2572.

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan (2016). "Inherent Trade-offs in the Fair Determination of Risk Scores". In: *arXiv preprint arXiv:1609.05807*.

Anton Kolotilin, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li (2017). "Persuasion of a Privately Informed Receiver". In: *Econometrica* 85.6, pp. 1949–1964.

Yuqing Kong, Katrina Ligett, and Grant Schoenebeck (2016a). "Putting Peer Prediction under the Micro (Economic) Scope and Making Truth-Telling Focal". In: *International Conference on Web and Internet Economics*, pp. 251–264.

Yuqing Kong and Grant Schoenebeck (2016b). "Equilibrium Selection in Information Elicitation without Verification via Information Monotonicity". In: *arXiv preprint arXiv:1603.07751*.

– (2018). "Eliciting Expertise without Verification". In: *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 195–212.

Tracy R Lewis and David EM Sappington (1994). "Supplying Information to Facilitate Price Discrimination". In: *International Economic Review*, pp. 309–327.

Xinye Li and Andrew Chi-Chih Yao (2013). "On Revenue Maximization for Selling Multiple Independently Distributed Items". In: *Proceedings of the National Academy of Sciences* 110.28, pp. 11232–11237.

Katrina Ligett and Aaron Roth (2012). "Take It or Leave It: Running a Survey When Privacy Comes At a Cost". In: *International Workshop on Internet and Network Economics*, pp. 378–391.

Yang Liu and Yiling Chen (2016). "Learning to Incentivize: Eliciting Effort via Output Agreement". In: *arXiv preprint arXiv:1604.04928*.

Yang Liu and Yiling Chen (2017). "Sequential Peer Prediction: Learning to Elicit Effort Using Posted Prices". In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence*.

– (2018). "Surrogate Scoring Rules and a Dominant Truth Serum". In: *arXiv preprint arXiv:1802.09158*.

Andreu Mas-Colell, Michael Dennis Whinston, Jerry R Green, et al. (1995). *Microeconomic Theory*. Vol. 1. Oxford University Press.

Reshef Meir, Shaull Almagor, Assaf Michaely, and Jeffrey S Rosenschein (2011). "Tight Bounds for Strategyproof Classification". In: *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pp. 319–326.

Reshef Meir, Ariel D Procaccia, and Jeffrey S Rosenschein (2012). "Algorithms for Strategyproof Classification". In: *Artificial Intelligence* 186, pp. 123–156.

Roger B Myerson (1981). "Optimal Auction Design". In: *Mathematics of Operations Research* 6.1, pp. 58–73.

John F Nash et al. (1950). "Equilibrium Points in n-person Games". In: *Proceedings of the National Academy of Sciences* 36.1, pp. 48–49.

George L. Nemhauser and Laurence A. Wolsey (1988). *Integer and Combinatorial Optimization*. Wiley, pp. I–XIV, 1–763.

Whitney K Newey and Daniel McFadden (1994). "Large Sample Estimation and Hypothesis Testing". In: *Handbook of econometrics* 4, pp. 2111–2245.

Noam Nisan and Amir Ronen (2001). "Algorithmic Mechanism Design". In: *Games and Economic behavior* 35.1-2, pp. 166–196.

Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani (2007). *Algorithmic Game Theory*. Cambridge University Press.

Kobbi Nissim, Salil Vadhan, and David Xiao (2014). "Redrawing the Boundaries on Purchasing Data from Privacy-sensitive Individuals". In: *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, pp. 411–422.

Michael Ostrovsky and Michael Schwarz (2010). "Information Disclosure and Unraveling in Matching Markets". In: *American Economic Journal: Microeconomics* 2.2, pp. 34–63.

Juan Perote-Peña and Javier Perote (2003). "The Impossibility of Strategy-Proof Clustering". In: *Economics Bulletin* 4.23, pp. 1–9.

Zinovi Rabinovich, Albert Xin Jiang, Manish Jain, and Haifeng Xu (2015). "Information Disclosure as a Means to Security". In: *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pp. 645–653.

Luis Rayo and Ilya Segal (2010). "Optimal Information Disclosure". In: *Journal of Political Economy* 118.5, pp. 949–987.

Aaron Roth and Grant Schoenebeck (2012). "Conducting Truthful Surveys, Cheaply". In: *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 826–843.

Aviad Rubinstein (2016). "On the Computational Complexity of Optimal Simple Mechanisms". In: *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pp. 21–28.

Aviad Rubinstein and S. Matthew Weinberg (2015). "Simple Mechanisms for a Subadditive Buyer and Applications to Revenue Monotonicity". In: *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pp. 377–394.

Alex Smolin (2019). "Disclosure and Pricing of Attributes". In: *Manuscript*.

Michael Spence (1973). "Job Market Signaling". In: *The Quarterly Journal of Economics* 87.3, pp. 355–374.

Vasilis Syrgkanis, Elie Tamer, and Juba Ziani (2017). "Inference on Auctions with Weak Assumptions on Information". In: *arXiv preprint arXiv:1710.03830.* **V. Syrgkanis is the primary author. J. Ziani contributed to part of the theoretical results and the simulations.** URL: http://arxiv.org/abs/1710.03830.

Haifeng Xu, Zinovi Rabinovich, Shaddin Dughmi, and Milind Tambe (2015). "Exploring Information Asymmetry in Two-stage Security Games". In: *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pp. 1057–1063.

Andrew Chi-Chih Yao (2015). "An n-to-1 Bidder Reduction for Multi-Item Auctions and Its Applications". In: *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 92–109.

Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi (2017). "Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment". In: *Proceedings of the 26th International Conference on World Wide Web*, pp. 1171–1180.