The Value of Data Records

Simone Galperti — UC, San Diego
Aleksandr Levkun — UC, San Diego
Jacopo Perego — Columbia University

September 9, 2021

Abstract

Many online platforms intermediate trade between sellers and buyers relying on individual data records of their personal characteristics. A key question is how much value a platform derives from each record. Is this value higher for one buyer than for another? What are its properties? We answer these questions by combining a modern information-design perspective with classic duality methods. We show that the value of a buyer's record cannot be correctly assessed by focusing only on the payoff that a platform directly earns from the trade between that buyer and a seller. This is because of a novel externality between records, which arises when a platform pools records to withhold information from the sellers. We then characterize how much a platform is willing to pay for more records—e.g., for getting new buyers to join it—and for better records—e.g., for more information about existing buyers. Our analysis establishes basic properties of the demand side of data markets. Our methods apply generally to a large class of principal-agent problems.

JEL Classification Numbers: C72, D82, D83

Keywords: Data, Information, Records, Duality, Externalities

We are thankful to S. Nageeb Ali, Alessandro Bonatti, Wouter Dessein, Laura Doval, Elliot Lipnowski, Xiaosheng Mu, Andrea Prat, Joel Sobel, Denis Shishkin, Rakesh Vohra, Glen Weyl and seminar participants at Bonn University, Cambridge University, Columbia University, Penn State University, Pittsburgh University & CMU, UC Davis, UC San Diego, University of Chicago, and National University of Singapore for useful comments.

> The first step toward valuing individual contributions to the data economy is measuring these (marginal) contributions. (Posner and Weyl, 2018, p. 244)

1 Introduction

Personal data is the "new oil" in modern economies. It can enhance firms' productivity and be a source of market power. How it is collected, exchanged, and used has been at the center of contentious debates on policy and regulation (Stigler Report, 2019). While markets for data have been rapidly developing, they often lack transparency and are far from being well understood (Federal Trade Commission, 2014). This paper asks a basic question: What is the value of an individual data record? Is one consumer's record more valuable than another's for an e-commerce platform? How much should each be paid? Addressing such questions is important to better understand the *demand side* of data markets, how they work, and how to fairly compensate data sources for their specific contribution to the digital economy (Lanier, 2013; Acquisti et al., 2016; Arrieta-Ibarra et al., 2018).

This paper breaks new ground in two ways. First, we focus on the value of an *individual* data record, taking into account its specific realized content. Second, we study this value for *intermediation* problems: An intermediary uses its data to strategically direct interactions between agents with conflicting interests by providing them with information or by affecting their incentives (e.g., with monetary transfers). Such problems are ubiquitous due to the rise of digital platforms and data brokers ("info-mediaries" as in Acquisti et al., 2016): Besides e-commerce, examples include matching markets (like ride-sharing and navigation services) and auction-based markets (like ad auctions or eBay). However, the value of data when it is used by an intermediary needs to be established in fundamentally different ways than when it is used by a standard decision-maker to identify his best choices. This is because the intermediary may withhold information by pooling data records together, thus creating externalities between them. Our solution involves modeling intermediation as an information-design problem and leveraging its structure as a linear program.

Consider an example. An online platform mediates the interactions between a population of buyers and a monopolist, which produces a good at zero marginal cost. For each buyer, the platform owns a data *record* which consists of a list of her personal characteristics. There are two types of records, denoted by ω_1 and ω_2, depending on what the platform knows about the buyer. Concretely, type ω_k reveals that the buyer's valuation for the seller's good is k. We refer to the collection of buyers' records as the platform's database. It contains 3 million records of type ω_1 and 6 million of type ω_2. The seller knows only the database composition, denoted by $q = (3M, 6M)$. For each interaction, the platform sends a signal about ω to the seller so as to influence the price he charges (as in Bergemann et al., 2015). Concretely, it divides the buyers into market segments using their record—like their gender or age—and then tells the seller to which segment each buyer belongs.

In this context, our main question can be formulated as follows: How much value does the platform derive from each buyer's record and why? The individual contribution of each record, denoted by $v^*(\omega)$, can also offer a benchmark for compensating each buyer for her specific data. Moreover, it is a necessary stepping stone to studying how much the platform is willing to pay for having more data. This can have two meanings in our context: (*i*) having *more* records in the database and hence the ability to mediate more interactions between the corresponding buyers and the seller (e.g., because new buyers join the platform); (*ii*) having *better* records by observing more informative characteristics about existing buyers (e.g., because they are more active on the platform).[1] Answering these questions is standard if the platform itself is the seller and thus directly sets a price for each buyer knowing ω. In this case, it effectively solves a decision problem and $v^*(\omega)$ is the profit for each interaction. Things change significantly for intermediation problems due to the conflict of interests.

To illustrate, suppose the platform maximizes the buyers' consumer surplus to the detriment of the seller's profits. To do so, the platform assigns each buyer whose record is of type ω_1 to a subprime segment, s'; instead, it assigns each buyer whose record is of type ω_2 to s' or to a prime segment, s'', with equal probability. As a result, the seller optimally sets a price of 1 for segment s' and of 2 for s''. Thus, the expected surplus—hence, the platform's payoff—is ½ for each type-ω_2 record and 0 for each type-ω_1 record. Are these payoffs the actual value the platform derives from each record? The answer is no. Perhaps counterintuitively, the most valuable records are those of buyers whose interaction with the seller yields the lowest payoff. To see why, imagine two buyers, Ann and Bonnie, whose records are of type ω_1 and ω_2 respectively. Bonnie's interaction with the seller yields a positive surplus only when pooled with Ann's interaction through s'. If so, Ann's record "helps" to persuade the seller to set a low price for Bonnie. Hence, it should not be worthless, even though Ann's interaction by itself yields zero surplus. Indeed, we show that $v^*(\omega_1) = 1$, which reflects that Ann's record exerts a positive information externality on Bonnie's interaction. By contrast, $v^*(\omega_2) = 0$

[1]The distinction between more and better records reflects the difference between *marketing lists* and *data appends*, the two main products traded in the data-brokerage industry (see, Federal Trade Commission, 2014). The former allows companies to identify new customers who share a given set of characteristics. The latter allows companies to learn new characteristics about existing customers.

because Bonnie's record has to "repay" this externality to Ann's record. Each $v^*(\omega)$ equals the marginal effect on the platform's total payoff of adding a record of type ω to the database. At the same time, $v^*(\omega)$ is also the unit value of every type-ω record already in the database due to the linear structure of the platform's problem.

Our main contribution is to characterize what determines these unit values of data records. As the example showed, the payoff the platform *directly* obtains from a record is a biased measure of its value—hence, of the willingness to pay for more data. This is in sharp contrast with decision problems, such as when the platform maximizes the seller's profits. In intermediation problems, conflicts of interest lead the platform to pool records so as to produce partially informative signals, thereby using data of one interaction to influence the outcomes of other interactions. We show that, as a result, $v^*(\omega)$ is the sum of the platform's direct payoff from each type-ω record and the externalities caused by that record (like Ann's) on other records and their interactions (like Bonnie's). We characterize which records generate positive and which negative externalities. We relate this to how the platform exploits the seller's incentives across interactions with its signals.[2] Using a general version of the above example, we show that as long as the platform cares more about buyers' surplus than the seller's profit, our externalities satisfy a single-crossing property: They are positive for low-valuation buyers and negative for high-valuation buyers. If ignored, this could lead to overcompensating the latter for their data at the expense of the former.

We then use our framework to study the problem of acquiring more records. We can evaluate its effects using the platform's overall preference over databases, which is pinned down by v^* as a function of their composition q. The platform essentially uses each buyer's record as an input to produce the information aimed at influencing the seller. As in textbook consumer theory, we can treat $v^*(\omega)$ as akin to the marginal utility from type-ω records and study the substitutability between types of records. We show that these marginal utilities are step-wise diminishing and record types are imperfect substitutes (or even complements). Together, these properties establish a "scarcity principle" for data: Scarcer types of records are more valuable both in absolute and in relative terms. They also enable standard analysis of the platform's demand functions.

Alternatively, the platform could obtain better records by refining the information it has about the corresponding buyers. For example, imagine some buyer Cindy who has just joined the

[2]Importantly, these externalities arise even when data records are statistically independent. As such, they fundamentally differ from "learning" externalities highlighted by the literature, which depend on the correlation between records (see the discussion of the related literature).

platform, which knows her IDs but nothing else. At some point the platform may observe some of Cindy's characteristics, which refines her record. We show that by changing the database composition, such refinements have direct effects—Cindy's record becomes more valuable in expectation—and indirect effects—unrefined records can become more or less valuable. These effects are due to the aforementioned externalities and exist even if each disclosing buyer provides no information about any other buyer (i.e., refinements are independent). Thus, by disclosing her characteristics, Cindy may cause the compensation for other buyers' records to change—even if the platform does not learn anything about them.

Given these complex effects of refinements on the unit value of all records, it is unclear whether they benefit the platform overall and hence command a positive willingness to pay. We show that this benefit is positive, but decreasing at the margin in the scale of records of a certain type that are refined. We provide a sharp condition for the benefit to be zero, which depends only on how the platform uses its current database. This can happen even if the platform would use both the refined and the unrefined records differently, acting on the new information it receives. This is in sharp contrast with decision problems, where the benefit from obtaining information is positive if and only if it changes optimal behavior. In fact, we show that refinements that are correlated between records can *lower* their unit values and result in a *negative* willingness to pay.

Our analysis applies more generally to any setting where a principal mediates the interactions between multiple agents by providing them with information or by affecting incentives with actions under its control. We can also let the agents have some payoff-relevant data, under the assumption that the principal has direct access to all parties' data.[3] We can continue to view it as using each interaction's data as a physical input to produce information or choose its actions. This perspective allows us to use linear-programming duality to characterize the unit value v^* of these inputs, adapting classic work of Dorfman et al. (1987) and Gale (1989) to our class of problems. This general and flexible approach can handle the complexities of assessing the value of data for any intermediation problem.

Overall, this paper aims to improve our understanding of the *demand side* of data markets. We do so by studying the platform's willingness to pay for data records and its properties. These can be useful for demand estimation when studying, for instance, the welfare effects of privacy regulations and antitrust interventions.[4] A better understanding of the value of people's data may help improve the status quo where they are not compensated for sharing it. This has

[3] In a related project, we analyze the case where the principal has to first elicit the data from its sources.

[4] See, e.g., Scott-Morton et al. (2019); Crémer et al. (2019); Goldberg et al. (2021).

many pitfalls (Lanier, 2013; Arrieta-Ibarra et al., 2018) and improvements may have significant benefits (Jones and Tonetti, 2020). They depend, however, on figuring out how much each person should be paid for the data he or she actually provides.

Related Literature. This paper contributes to the burgeoning literature on data markets, comprehensively reviewed by Bergemann and Bonatti (2019) and Bergemann and Ottaviani (2021).

One strand of this literature studies the optimal "use" of a database. This often involves a single party—such as a platform or data broker—who owns a database and designs information products for some agents—such as sellers, advertisers, or decision makers—to either charge a price or influence their behavior (or both). In Admati and Pfleiderer (1986, 1990), a platform sells signals (i.e., Blackwell experiments) about an asset value to market traders. In Bergemann and Bonatti (2015), a platform sells segments of buyers to advertisers and charges a linear price based on each segment's size. In Bergemann et al. (2018), a platform designs menus of experiments to screen information buyers with heterogeneous priors. Yang (2020) studies a related selling problem in considerably richer settings. Our platform also owns a database and uses it to designs information. However, our focus is not on the properties of the optimal information products and their prices. Indeed, we primarily analyze a simpler class of problems in which the platform provides information for free (we discuss extensions in Section 5). This is because our main interest is in the upstream market, where the platform obtains the data records that will form its database. These records have two key features: They give the platform access to the buyers, on top of providing information about them; they can be valued ex post, taking into account their specific content (i.e., their type).

A separate strand of the literature on data markets studies how platforms can incentivize consumers to disclose their personal data. For example, Choi et al. (2019), Acemoglu et al. (2021), and Ichihashi (2021) study the "learning" externalities that one consumer's disclosure can exert on others, when their data is correlated. Bergemann et al. (2020) study the flow and allocation of personal data between consumers, firms, and data intermediaries. They examine how the correlation in consumers' data affects their incentives to participate in the data market and other market observables. The main differences from our paper are two. First, our platform is assumed to already have the database.[5] This offers a useful benchmark to study the effects of privacy regulations.[6] Second, we isolate a new data externality, which stems from the platform's

[5]The vast majority of data-brokers' transactions of information products happens without the consumer's knowledge. See, e.g., Federal Trade Commission (2014).

[6]For example, in a related project we use this benchmark to study how privacy affects the value of individual records.

pooling records to withhold information from the sellers. It arises even if consumers' records are statistically independent, so is orthogonal to the aforementioned learning externalities.

Our work is also related to the literature on data privacy, which is reviewed by Acquisti et al. (2016). Calzolari and Pavan (2006) analyze information externalities between sequential interactions, which we do not consider in this paper. Ali et al. (2020) examine when giving consumers control over their data can help them benefit from personalized pricing. Ichihashi (2020) finds that a multi-product platform can prefer to commit to not using consumers' data for personalized pricing and maximizes profits via product recommendations.

Our methods build on the information-design literature (see Bergemann and Morris, 2019, for a review). After formulating our "data-use" problem as a linear program using standard arguments (Bergemann and Morris (2016)), we consider its dual to obtain our "data-value" problem. Others have used duality to study information design (Kolotilin (2018), Galperti and Perego (2018), Dworczak and Martini (2019), Dworczak and Kolotilin (2019), and Dizdar and Kováč (2020)). The main differences of our paper are two. While they exploit the dual to solve the primal design problem, we use it to address an distinct economic question of independent interest—what is the value of data? And while those papers focus on single-receiver problems or employ a "belief approach" (as in Kamenica and Gentzkow, 2011), we study problems with multiple agents interacting strategically through the notion of Bayes-correlated equilibrium. This links our work to an earlier literature on dual analysis of correlated equilibria (Nau and McCardle (1990), Nau (1992), Myerson (1997)). Finally, the mechanism-design literature has also used duality methods at least since Myerson (1983) and Myerson (1984), and recently to study informationally robust mechanisms (e.g., Du, 2018; Brooks and Du, 2020, 2021).

2 Model

For ease of exposition, we present the model and analysis in the context where an e-commerce platform mediates the interactions of a group of sellers with a population of buyers, similarly to the introduction example.[7] Nonetheless, we keep our description fairly general, as the details are irrelevant and may distract from our main points. In fact, our approach and results apply much more broadly. We discuss this and other aspects of the model in Section 5.

Let $i = 0$ denote the platform, which plays the role of the principal (it). Let $I = \{1, \ldots, n\}$ be a set of sellers, who play the role of the agents (he). Let A_i be the finite set of seller i's

[7] Besides Bergemann et al. (2015), see also Elliott et al. (2020) who study a platform that tries to influence which prices competing sellers offer to their customers.

actions. We can interpret a_i as seller i's choosing his product's price, features, or quality. The platform is used by a continuum of buyers (she), each interested in buying a product from the sellers. Each buyer's preference over the sellers' products is pinned down by a random variable θ, which is independently and identically distributed across buyers over a finite set Θ.

The platform has exclusive access to some data about each buyer. We think of this data as a concrete *record* of personal characteristics that is informative about θ—perhaps only partially. We assume throughout that a buyer's record provides information only about her θ, but not about any other buyer's θ.[8] There are different *types* of records—denoted by ω in some finite set Ω—depending on what the platform knows about the buyer. Let ω° denote the records whose data is fully uninformative about θ. Only the platform observes ω, which gives it an informational advantage over the sellers. Let $q \in \mathbb{R}_+^\Omega$ denote the platform's collection of buyers' records, where $q(\omega)$ are of type ω. We refer to q as the platform's *database*.

For each interaction between a buyer and the sellers, we leave her purchase decision given their actions implicit and embed it in the payoff functions of the sellers and the platform. For every ω and action profile $a = (a_1 \ldots, a_n)$, let $u_i(a, \omega)$ be i's expected payoff conditional on the buyer's record. Let $\Gamma_\omega = \{I, (A_i, u_i(\cdot, \omega))_{i=0}^n\}$, which defines a complete-information game between the sellers. Thus, we may also refer to Γ_ω as a buyer-sellers interaction of type ω. The primitives $\Gamma = \{\Gamma_\omega\}_{\omega \in \Omega}$ and q are common knowledge.

Using its data, the platform mediates each interaction by privately conveying some information about its type to each seller so as to influence their actions. The sellers combine this information with Γ and q to form beliefs and choose actions. Our platform has full commitment power, similarly to the omniscient information designer in Bergemann and Morris (2019). Formally, the platform publicly commits to an information structure that, for each interaction, produces a private signal about ω for each seller i. By standard arguments (Myerson, 1983, 1984; Bergemann and Morris, 2016), we can focus on information structures in the form of recommendation mechanisms, where the platform privately recommends an action to each seller which he must find optimal to follow (obedience). A mechanism is then a function $x : \Omega \to \Delta(A)$, where $x(a|\omega)$ can be interpreted as the share of interactions of type ω that lead to recommendation profile a.[9] The formal problem is

$$\mathcal{U}_q : \qquad \max_x \sum_{\omega \in \Omega, a \in A} u_0(a, \omega) x(a|\omega) q(\omega)$$

[8] We make this assumption mostly for ease of exposition. Our model can accommodate correlation among records, as discussed in Section 4.2.2.

[9] Note that we do not allow $x(\cdot|\omega)$ to differ between records of the same type ω, but this is without loss of generality.

s.t. for all $i \in I$ and $a_i, a_i' \in A_i$,

$$\sum_{\omega \in \Omega, a_{-i} \in A_{-i}} \left(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega)\right) x(a_i, a_{-i}|\omega) q(\omega) \geq 0. \tag{1}$$

Each constraint (1) is equivalent to requiring that a_i maximizes seller i's expected utility conditional on the information conveyed by a_i given x and the database q. We denote any optimal mechanism by x_q^*. We define the *direct payoff* generated by each record of type ω as

$$u_q^*(\omega) \triangleq \sum_{a \in A} u_0(a, \omega) x_q^*(a|\omega),$$

and the *total payoff* generated by the database as

$$U^*(q) \triangleq \sum_{\omega \in \Omega} u_q^*(\omega) q(\omega). \tag{2}$$

We assume that $\mathcal{U}_q$ satisfies the following minor regularity property, which holds generically in the space of sellers' payoff functions.[10]

Assumption 1 (Non-degeneracy)**.** *Of the constraints (1) that define the feasible set of mechanisms x for $\mathcal{U}_q$, no more than $|A \times \Omega|$ are ever active at the same time.*

3 The Unit Value of Data

This section addresses our main question: How much value does the platform derive from each buyer's record and why? This individual contribution of each record can offer a benchmark for compensating each buyer for her specific data. This is also a necessary stepping stone to addressing the question of how much the platform is willing to pay for "having more data." This colloquial expression can have two meanings in our context: (1) having *more* records in the database and hence the ability to mediate more interactions between the corresponding buyers and the sellers; (2) having *better* records by observing more informative characteristics about existing buyers. We analyze both in Section 4.

To develop and explain our answers, we find it useful to think of the platform's database as defining a collection of decisions. For each interaction between a buyer and the sellers, the platform has to decide what to disclose about the buyer so as to influence the sellers' actions. As such, these decisions are fundamentally different from standard decision problems where the platform would directly control the actions that determine payoffs. This difference will affect how we approach our questions and their answers.

[10] For more details, see Remark 1 in Appendix B.

To see the point, suppose for a moment that the platform itself were a monopolistic seller (namely, there is no other seller). For each interaction with a buyer, it would directly choose a to maximize $u_0(a,\omega)$ based on her record, effectively solving a standard decision problem. What is then the value of a record of type ω? The answer would simply be the payoff from optimally using it as an input in its associated decision, namely, $u^*(\omega) = \max_{a\in A} u_0(a,\omega)$. How much is the platform willing to pay for one more record of type ω? The answer would again be $u^*(\omega)$. Finally, what is its willingness to pay for better records? We would evaluate the information contained in a record by netting out the payoff of the optimal a if the platform had to use a record with fully uninformative data, namely, $u^*(\omega^\circ)$.[11] We could then aggregate $u^*(\omega) - u^*(\omega^\circ)$ across ω, as we normally do to evaluate information structures. All these answers should look familiar, but they crucially rely on the premise that $u^*(\omega)$ is the right measure of the value of a record of type ω.

This premise is no longer valid when the platform and the sellers are distinct entities with conflicting interests. In contrast to the previous case where the platform directly chose a, the decision of which signals to send to the sellers given one buyer's record is no longer independent of such decisions for other records in the database. What information a signal conveys or withholds about one record depends on which other records lead to the same signal.[12] This implies that while the value of each record continues to be determined by how it is used to guide decisions, this use is not confined to the interaction physically attached to that record. This complicates the analysis. Although by (2) the direct payoffs u^* provide an immediate way to divide the total payoff $U^*(q)$ across the types of records, this division does not reflect the actual value of each record as shown in our introduction example. We instead have to systematically keep track of all the ways the platform uses each record to mediate all interactions and the resulting interdependencies. This will allow us to simultaneously assess how much value it derives from each record and how it does so.

For explanations, it will be convenient to refer to usual intuitions and results for standard decision problems. To help this comparison, note that the previous case where the platform itself is a monopolistic seller is formally equivalent to our model where the platform and the monopolistic seller are distinct parties and the platform maximizes his payoff ($n = 1$ and $u_0 = u_1$). More generally, if all parties have aligned interests (i.e., u_i is an affine transformation of u_0

[11] The same point applies for other netting-out procedures, such as in Frankel and Kamenica (2019).

[12] This dependence between signal decisions is orthogonal to our commitment assumption. On the one hand, it would arise even if our platform could not commit and we had to rely on some equilibrium notion. On the other hand, these decisions are fundamentally different from those in the previous case of a monopolist platform even if we allow it to commit (which is of course irrelevant in that scenario).

for all $i = 1, \ldots, n$), constraints (1) can be omitted, so it is as if the platform directly chooses all actions. Thus, even if in reality each interaction involves multiple parties, for each record in the database the platform effectively faces an independent, standard, decision problem to maximize $u_0(a, \omega)$. This implies that $\mathcal{U}_q$ is separable across record types. For convenience, we will slightly abuse terminology and use the terms *standard-decision problem* and *intermediation problem* to refer to our model with aligned and conflicting interests, respectively.

3.1 The Data-Value Problem

Our approach builds on the observation that any information-design problem is a linear program. A standard economic interpretation is that linear programs describe the problem of optimally using some scarce inputs to produce some output (Dorfman et al., 1987, p. 39). We think of information design as a "data-use" problem, where the inputs are the records in the database and the output is the information conveyed by each mechanism in the form of recommendations. Following Dorfman et al. (1987, p. 39), we then exploit the dual of this data-use problem to evaluate each record.

We call this assessment task the *data-value* problem. Let $\lambda = (\lambda_1, \ldots, \lambda_n)$ where $\lambda_i : A_i \times A_i \to \mathbb{R}_+$ for all $i \in I$. For each i and (a, ω) define

$$t_i(a, \omega) \triangleq \sum_{a_i' \in A_i} \left(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega) \right) \lambda_i(a_i' | a_i) \tag{3}$$

and $t(a, \omega) \triangleq \sum_{i \in I} t_i(a, \omega)$. The data-value problem is

$$\begin{aligned} \mathcal{V}_q : \quad & \min_{v, \lambda} \sum_{\omega \in \Omega} v(\omega) q(\omega) \\ & \text{s.t. for all } \omega \in \Omega, \\ & v(\omega) = \max_{a \in A} \left\{ u_0(a, \omega) + t(a, \omega) \right\}, \end{aligned} \tag{4}$$

We denote any optimal solution by (v_q^*, λ_q^*) and the induced functions t by t_q^*. By standard linear-programming arguments v_q^* is unique generically with respect to q (i.e., except on a set of qs with measure zero). We refer to equation (4) as the *value formula*, which defines our main object of interest. The reason hinges on the next relation between the data-use and data-value problems and on the following interpretation.

Lemma 1. *For any database q, $\mathcal{V}_q$ is equivalent to the dual of $\mathcal{U}_q$. Thus, for every x_q^* and (v_q^*, λ_q^*)*

$$\sum_{\omega \in \Omega} v_q^*(\omega) q(\omega) = U^*(q) \triangleq \sum_{\omega \in \Omega} u_q^*(\omega) q(\omega). \tag{5}$$

All proofs are in Appendix B.

This duality relation follows from basic linear-programming results. Yet, applied to our specific problem, it becomes the key to answering our economic questions. In $\mathcal{U}_q$, every x determines a joint measure $\chi \in \mathbb{R}_+^{\Omega \times A}$, which must satisfy $\sum_{a \in A} \chi(a, \omega) = q(\omega)$; that is, the use of type-ω records to produce recommendations must exhaust their stock $q(\omega)$ in the database. Formally, $v(\omega)$ is the multiplier of this constraint, which is usually interpreted as the shadow price of the corresponding input through the thought experiment of adding a marginal unit of it. In fact, $v_q^*(\omega)$ is the derivative of $U^*(q)$ with respect to $q(\omega)$, as for any constrained optimization problem. However, it would be misleading to think that $v_q^*(\omega)$ captures only the value of a marginal record of type ω. In fact, Lemma 1 demonstrates that $v_q^*(\omega)$ captures the value of *each* record of type ω in the database, not just the marginal one. We will then refer to $v_q^*(\omega)$ as the *unit value* of a record of type ω (see also Gale (1989), p. 12). Note that $\mathcal{V}_q$ assigns such values to all records simultaneously and does not require to find x_q^*. Also, v_q^* can depend on q for intermediation problems but not for standard-decision problems, as in this case $v_q^*(\omega) = \max_{a \in A} u_0(a, \omega)$ for all ω.

The rest of the paper characterizes the properties of v_q^*. Here, we begin by establishing a lower bound for the value of a record. For $\omega \in \Omega$, let $CE(\Gamma_\omega)$ be the set of correlated equilibria of the game Γ_ω.

Lemma 2 (Lower Bound)**.** *For every q,*

$$v_q^*(\omega) \geq \bar{u}(\omega) \triangleq \max_{y \in CE(\Gamma_\omega)} \sum_{a \in A} u_0(a, \omega) y(a), \qquad \omega \in \Omega.$$

Note that by Lemma 1 and 2 there is an optimal x_q^* that satisfies $x_q^*(\cdot|\omega) \in CE(\Gamma_\omega)$ for all ω if and only if $v_q^*(\omega) = \bar{u}(\omega)$ for all ω. In words, for such an x_q^* the platform fully discloses the buyer's record to the sellers for all interactions.

Interpretations and Uses of v_q^*

We pause briefly to clarify how v_q^* can be interpreted and used in relation to our main questions.

Individual Compensation. By quantifying how much a record contributes to the total payoff $U^*(q)$, v_q^* offers a benchmark for individually compensating each buyer as the "owner" of her record. In $\mathcal{V}_q$, we can view the platform as choosing v to minimize the total expenditure to compensate the buyers. However, the platform is constrained by equation (4), which determines a lower bound for each buyer's compensation that takes into account how her record is used.[13] Paraphrasing Dorfman et al. (1987, p. 43), this interpretation is reminiscent of the operation of a competitive market where competition forces the platform to offer the "owner" of a record the full value to which her input gives rise, while competition among these "owners" drives down this value to the minimum consistent with this limitation. Gale (1989, Chapter 3.5) also shows how dual problems can deliver competitive prices of scarce inputs.

To see the importance of compensating data owners based on v_q^* and not u_q^*, consider again our introduction example of a surplus-maximizing platform. Suppose it decides—perhaps forced by some regulation—to compensate the buyers for their contribution to $U^*(q)$ by giving back some share δ to them. How δ is chosen and the compensations implemented is irrelevant here. The more fundamental question is how much each buyer should get. It seems that the answer should take into account the buyer's specific record. One could use u_q^*, which would result in incorrectly allocating $\delta U^*(q)$ only to the buyers with $\omega = \omega_2$ (each receiving $\delta u_q^*(\omega_2) = \delta\, 0.5$) because $u_q^*(\omega_1) = 0$. In fact, only the buyers with $\omega = \omega_1$ contribute to $U^*(q)$ because $v_q^*(\omega_2) = 0$. Thus, $\delta U^*(q)$ should be allocated only to these buyers (each receiving $\delta v_q^*(\omega_1) = \delta$).

More Records. Suppose the platform is offered a new buyer's record. Let $\rho \in \Delta(\Omega)$ be its belief about the record's type, which it will observe after acquiring the record. We can then define its willingness to pay for the new record as

$$v_q^*(\rho) \triangleq \sum_{\omega} \rho(\omega) v_q^*(\omega).$$

In fact, this is the expected marginal increase of the platform's total payoff from the new record.

Better Records. Imagine that type-ω° records correspond to buyers who blocked tracking. How much would be platform be willing to pay to learn more about them? In other words, how can we quantify the value of the information contained in a record when the platform may use it to mediate multiple interactions? Similarly to standard-decision problems, one way is to use the value of type-ω records net of the value of type-ω° records:

$$v_q^*(\omega) - v_q^*(\omega^\circ), \qquad \omega \in \Omega. \tag{6}$$

Despite the superficial similarity, this value of information for intermediation problems differs in important ways from standard-decision problems, as we explain in Section 4.2.

[13] In fact, by complementary slackness $v_q^*(\omega) = u_0(a, \omega) + t_q^*(a, \omega)$ if $x_q^*(a|\omega) > 0$. We provide another independent economic interpretation of the data-value problem in Section A.

3.2 Value Decomposition and Data Externalities

This section delves deeper into what determines the unit value of data records. Key to this will be comparing v_q^* with the direct payoffs u_q^*. We will show that the gap between them quantifies an externality between records, which is a defining feature of intermediation problems.

The value of each record can be decomposed into two parts: The direct payoff $u_q^*(\omega)$ and another component, $t_q^*(\omega)$, which captures the indirect effects a record of type ω generates. This decomposition formalizes why and how u_q^* can bias our assessment of the actual value of each record for the platform.

Proposition 1. *For all $\omega \in \Omega$, $v_q^*(\omega) = u_q^*(\omega) + t_q^*(\omega)$ where*

$$t_q^*(\omega) \triangleq \sum_{a\in A} t_q^*(a,\omega)x_q^*(a|\omega) \overset{a.e.}{=} \sum_{\omega'\in\Omega} \frac{\partial u_q^*(\omega')}{\partial q(\omega)} q(\omega') \tag{7}$$

This result highlights two aspects of the value of records. The first is that the indirect effects t_q^* are akin to an externality. By the last part of (7), $t_q^*(\omega)$ captures the marginal effect of a type-ω record on the direct payoff of *all* records. This externality is purely informational: By being in the database, each record affects the platform's informational advantage and hence its decisions with all other records through the optimal x_q^*. In fact, $\frac{\partial}{\partial q(\omega)} u_q^*(\omega') = \sum_a u_0(a,\omega') \frac{\partial}{\partial q(\omega)} x_q^*(a|\omega')$. Adjustments in x_q^* can arise because changing $q(\omega)$ can render x_q^* no longer feasible (i.e., obedient) or optimal.

Which records generate positive and which negative externalities?

Corollary 1. *$t_q^*(\omega) < 0$ for some ω if and only if $t_q^*(\omega') > 0$ for some ω'. Moreover, $t_q^*(\omega) < 0$ implies $u_q^*(\omega) > \bar{u}(\omega)$, while $u_q^*(\omega) < \bar{u}(\omega)$ implies $t_q^*(\omega) > 0$.*[14]

The first part shows that the externalities lead to cross-subsidization of value from records with $t_q^*(\omega) < 0$ to records with $t_q^*(\omega') > 0$. The latter's $v_q^*(\omega')$ exceeds $u_q^*(\omega')$, so they must extract this extra value from records with $t_q^*(\omega) < 0$. The second part of the corollary explains this cross-subsidization. Records with $t_q^*(\omega) < 0$ generate a direct payoff that exceeds the full-information payoff $\bar{u}(\omega)$, which requires that $u_0(a,\omega) > \bar{u}(\omega)$ and $x_q^*(a|\omega) > 0$ for some a. That is, the platform earns a payoff with type-ω records that would never be possible by fully disclosing such records, so it relies on pooling them with records of different types. This help from type-ω' records justifies why $t_q^*(\omega') > 0$ and their value exceeds $u_q^*(\omega')$. Conversely, if

[14] The corollary follows because Lemma 1 implies $\sum_{\omega\in\Omega} t_q^*(\omega)q(\omega) = 0$, and Lemma 2 and Proposition 1 imply $t_q^*(\omega) \geq \bar{u}(\omega) - u_q^*(\omega)$ for all ω.

$t_q^*(\omega) < 0$, a record of type ω benefits from externalities caused by other records and hence has to "repay" them, which lowers its value. For the last part, we can interpret $u_q^*(\omega) < \bar{u}(\omega)$ as "sacrificing" type-ω records, as the platform could fully disclose them and ensure a payoff $\bar{u}(\omega)$. For this sacrifice to be worthwhile, such records must receive a compensation, explaining $t_q^*(\omega) > 0$. This last part offers a sufficient condition for $t_q^* \neq 0$ that is simple to check, but is not necessary. We provide another sufficient condition in Appendix C.

It is worth emphasizing that this externality arises even though records are statistically independent. As such, it is fundamentally different from the "learning externalities" studied in the literature (e.g., see Choi et al., 2019; Acemoglu et al., 2021; Bergemann et al., 2020). The latter arise as the result of an inherent statistical fact: A buyer's record may be informative about the θ of another buyer. This channel is purposefully absent in our paper. We instead emphasize externalities that arise endogenously when the platform pools some records to influence the sellers' actions. Indeed, they are a distinctive feature of intermediation problems and are absent in standard-decision problems.[15]

A second aspect highlighted by Proposition 1 is that the externalities through u_q^* are tightly related to how the platform exploits the sellers' primitive incentives. By the first part of (7), we can view $t_q^*(\omega)$ as aggregating externalities that type-ω records generate by inducing specific actions a. These are inversely related to the payoff the platform gets, in the following sense.

Corollary 2. *Suppose $x_q^*(a|\omega) > 0$ and $x_q^*(a'|\omega) > 0$. Then, $u_0(a,\omega) > u_0(a',\omega)$ if and only if $t_q^*(a,\omega) < t_q^*(a',\omega)$.*[16]

Thus, inducing actions whose payoff exceeds $\bar{u}$ by more, for instance, requires paying larger externalities to other records. Since $t_q^*(a,\omega) \triangleq \sum_{i\in I} t_{q,i}^*(a,\omega)$, we can view $t_{q,i}^*(a,\omega)$ as how much seller i contributes to the externality. Recall that $t_{q,i}^*(a,\omega)$ differs from zero only if $\lambda_{q,i}^*(a_i'|a_i) > 0$ for some a_i' (see (3)). By standard arguments (complementary slackness), $\lambda_{q,i}^*(a_i'|a_i) > 0$ only if

$$\sum_{\omega,a_{-i}} \Big(u_i(a_i,a_{-i},\omega) - u_i(a_i',a_{-i},\omega)\Big) x_q^*(a_i,a_{-i}|\omega)q(\omega) = 0; \tag{8}$$

the converse also holds generically in q. In words, $\lambda_{q,i}^*(a_i'|a_i) > 0$ if and only if seller i is indifferent between a_i and a_i' conditional on receiving recommendation a_i from x_q^*.

[15]In standard-decision problems full disclosure is always optimal. In this case, $u_q^*(\omega) = \bar{u}(\omega)$ and, as discussed after Lemma 2, $v_q^*(\omega) = \bar{u}(\omega)$. Therefore, $t_q^*(\omega) = 0$. Note that the converse is not true: It is possible to construct examples where $t_q^*(\omega) = 0$ and $v_q^*(\omega) > \bar{u}(\omega)$ for all ω.

[16]This follows from complementary slackness, namely, $v_q^*(\omega) = u_0(a,\omega) + t_q^*(a,\omega)$ if $x_q^*(a|\omega) > 0$.

Corollary 3. *The sellers who contribute to the externality $t_q^*(\omega)$ are only those whom x_q^* renders indifferent with the actions it recommends using records of type ω (i.e., (8) holds).*

Note that this result differs from the immediate fact that optimal solutions of linear programs occur on the boundary of the feasible set, which here means that some obedience constraint must bind. As q varies, x_q^* and hence t_q^* may change. However, as long as λ_q^* does not change (see Proposition 3 below), how each seller contributes to $t_q^*(\omega)$ does not change. Section A explains further how the platform exploits the sellers to determine their contributions to these externalities.

3.2.1 Application – Data and Price Discrimination (Part I)

To illustrate the importance of these data externalities, we consider a more general version of our introduction example, following Bergemann et al. (2015). There is only one seller ($n = 1$) who chooses the price a_1 for his product. For each buyer, θ is her valuation for the seller's product. Let $\Omega = \{\omega_1, \ldots, \omega_K\} \subset \mathbb{R}_+$, $K \geq 2$, and ω_k is strictly increasing in the index k. Each record of type ω_k fully reveals that $\theta = \omega_k$. Normalizing the seller's constant marginal cost to zero, his profit is a_1 if $\omega \geq a_1$ and zero otherwise: $u_1(a_1, \omega) = a_1 \mathbb{I}\{\omega \geq a_1\}$. Suppose the platform maximizes a weighted sum of profits and consumer surplus: $u_0(a_1, \omega) = \pi a_1 \mathbb{I}\{\omega \geq a_1\} + (1-\pi)\max\{\omega - a_1, 0\}$, where $\pi \in [0,1]$. Finally, let a_q be the price the seller would set conditional on knowing only the database composition q.

Proposition 2. *For $\pi \leq \frac{1}{2}$,*

$$v_q^*(\omega) = \begin{cases} (1-\pi)\omega & \text{if } \omega < a_q \\ \pi a_q + (1-\pi)(\omega - a_q) & \text{if } \omega \geq a_q; \end{cases}$$

moreover, $t_q^*(\omega) > 0$ for $\omega < a_q$ and $t_q^*(\omega) \leq 0$ for $\omega \geq a_q$. For $\pi \geq \frac{1}{2}$ we have $v_q^*(\omega) = u_q^*(\omega) = \pi\omega$ for all ω.

To understand this result, we note that the optimal x_q^* takes only two forms depending on π (see Appendix B). If $\pi \leq \frac{1}{2}$, the platform always maximizes the buyers' surplus subject to holding the seller's expected profits at a_q, as if $\pi = 0$. Thus, it is as if trade happens for every interaction, generating total surplus equal to ω, and only the buyers with valuation at least a_q contribute to guaranteeing this reservation profits for the seller. If $\pi \geq \frac{1}{2}$, the platform fully discloses all records. This allows the seller to perfectly price discriminate between buyers, so profits always equal the buyer's valuation and her surplus is zero.

Whenever the platform cares more about the buyers' surplus than the seller's profits, the direct payoff u_q^* provides a biased account of the value of each record. The result shows that this bias has a specific structure: t_q^* satisfies a single-crossing property in ω and this holds generally across q. That is, u_q^* is biased downward for low-valuation buyers (i.e., $\omega < a_q$) and upward for high-valuation buyers (i.e., $\omega \geq a_q$). This illustrates that ignoring the externalities we highlight may lead the platform to overcompensate high-valuation buyers for their data at the expense of low-valuation ones.

How does caring more about the buyers' surplus affect the value of their records? By simple algebra, lowering $\pi \leq \frac{1}{2}$ decreases $v_q^*(\omega)$ if and only if the buyer has an intermediate valuation ($a_q \leq \vartheta(\omega) < 2a_q$). Intuitively, for such records a larger share of the buyers' valuation goes to fund the seller's guaranteed profits of a_q, which becomes more costly as their surplus becomes more important. By contrast, the records of buyers with low valuation help to achieve positive surplus with other buyers, and the records of buyers with high valuation just yield a large surplus. For $\pi \geq \frac{1}{2}$, $v_q^*(\omega)$ increases in π independently of ω. This is because the platform helps the seller extract the full surplus from each interaction, and it cares more about doing so.

4 Acquiring More Data

This section studies how the values of buyers' records v_q^* depend on the database composition q. This dependence allows us to assess the willingness to pay of a given platform for acquiring more data in the sense of more records (Section 4.1) and better records (Section 4.2). Alternatively, we can interpret this analysis as comparative static exercises that show how the unit values of records vary between platforms that differ only in their databases.

4.1 More Records: Preferences over Databases

How do changes in the quantity of type-ω records affect their value? For example, do records become less valuable as they become more abundant in the database? To address these questions, we can think of the platform as a consumer of different types of goods called data records, where U^* is its utility function over bundles q of such goods. Then, $v_q^*(\omega)$ is akin to the marginal utility of type-ω records at q. We can also measure the marginal rate of substitution between records of type ω and ω' at q in the usual way, by letting $MRS_q(\omega, \omega') \overset{\text{a.e.}}{=} -\frac{v_q^*(\omega)}{v_q^*(\omega')}$. Thus, v_q^* fully characterizes the platforms' preferences over databases.

A classic property in standard consumer theory is that marginal utilities are diminishing.

Does the same hold for our platform? More generally, how does v_q^* vary with q? We first show that v_q^* is constant with respect to local, yet discrete, changes in q.

Proposition 3 (Stability)**.** *There exists a finite collection $\{Q_1, \ldots, Q_M\}$ of open, convex, and disjoint subsets of $\mathbb{R}_+^\Omega$ such that $\cup_m Q_m$ has full measure and, for every m, v_q^* is unique and constant for $q \in Q_m$.*

In fact, each Q_m is the interior of a cone in the space of databases $\mathbb{R}_+^\Omega$.[17] Importantly, v_q^* is constant even though the platform may adjust how she uses her data when q changes. Indeed, we can show that within each cone, while $v_q^*(\omega)$ is constant, the optimal $x_q^*(\omega)$ changes as a function of q (see Remark 1 in Appendix B). Intuitively, this is because x_q^* has to be fine-tuned to maximally exploit the sellers' incentives. By contrast, v_q^* depends only on which sellers' incentives are exploited, but not on how much (recall equation (3) and Corollary 3).

For global changes in q, we find that records of a given type become more valuable as they become scarcer. This establishes a "scarcity principle" for data and implies diminishing marginal utilities. For every q, define the share of records of each type by

$$\mu_q(\omega) \triangleq \frac{q(\omega)}{\sum_{\omega'} q(\omega')}, \qquad \omega \in \Omega.$$

Proposition 4 (Scarcity Principle)**.** *Consider databases q and q'. Fix $\omega \in \Omega$. If $\mu_q(\omega) < \mu_{q'}(\omega)$, then $v_q^*(\omega) \geq v_{q'}^*(\omega)$. Moreover, there exists $\bar{\mu}(\omega) < 1$ such that, if $\mu_q(\omega) > \bar{\mu}(\omega)$, then $v_q^*(\omega) = \bar{u}(\omega)$.*

Although each type of records can contribute to the platform's informational advantage in different ways, it always becomes more valuable as it becomes scarcer. This implies that $v_q^*(\omega)$ is weakly decreasing in $q(\omega)$—for any selection from the optimal solution correspondence of $\mathcal{V}_q$. Holding fixed the quantity of all other types of records, the platform's demand for type-ω records is downward sloping and converges to $\bar{u}(\omega)$ when $q(\omega)$ is sufficiently large. Equivalently, the individual contribution of type-ω records to the platform's payoff—hence, their owners' benchmark compensation—decreases as their quantity increases.

Another classic property in standard consumer theory is that marginal rates of substitution are diminishing, capturing imperfect substitutability between goods. Again, does the same hold for our platform? The answer is yes, unless it faces a trivial intermediation problem—that is, it is optimal to always fully disclose all records to the sellers, independently of q. The platform's

[17] It is easy to see that unit values are constant along the rays in the space of databases: If $q' = \alpha q$ for $\alpha > 0$, then $v_q^* = v_{q'}^*$. This is because only the frequency of record types matters for the sellers' incentives.

preferences are always weakly convex, because $U^*(q)$ is always a weakly concave function of q.[18] However, in standard-decision problems fully disclosing every buyer's record is optimal regardless of q, which implies that $MRS_q(\omega, \omega') = -\frac{\bar{u}(\omega)}{\bar{u}(\omega')}$ for all ω and ω' and hence all types of records are perfect substitutes. Even in an intermediation problem full disclosure can be optimal for some particular q. The next result shows that, when this is the case, full disclosure is optimal for all q, so all types of records are again perfect substitutes.

Proposition 5. *Suppose it is optimal for the platform to fully disclose every record for some database $q \in \mathbb{R}^{\Omega}_{++}$. Then, $v^*_{q'}(\omega) = \bar{u}(\omega)$ for all ω and $q' \in \mathbb{R}^{\Omega}_{+}$, which implies that full disclosure is optimal for all q'.*

This has several implications. First, some types of records are imperfect substitutes if and only if it is never optimal to fully disclose all records. In this case, MRSs vary and $v^*_q \neq \bar{u}$ for all $q \in \mathbb{R}^{\Omega}_{++}$. Second, showing that it is not optimal to fully disclose some ω for some $q \in \mathbb{R}^{\Omega}_{++}$ suffices to show that full disclosure is never optimal for all $q \in \mathbb{R}^{\Omega}_{++}$. More generally, one needs to know only the possible types of interactions (i.e., Γ) and not their quantities to establish the (sub)optimality of full disclosure.[19] Last but not least, we can identify whether the platform faces a non-trivial intermediation problem by detecting that it treats some types of records as imperfect substitutes.

Convexity of preferences implies that all intermediation problems lead to standard demand analysis. Choosing an optimal database subject to a budget constraint is a well-behaved problem. Given market price $p(\omega) > 0$ for every record type ω, the optimal q is characterized by

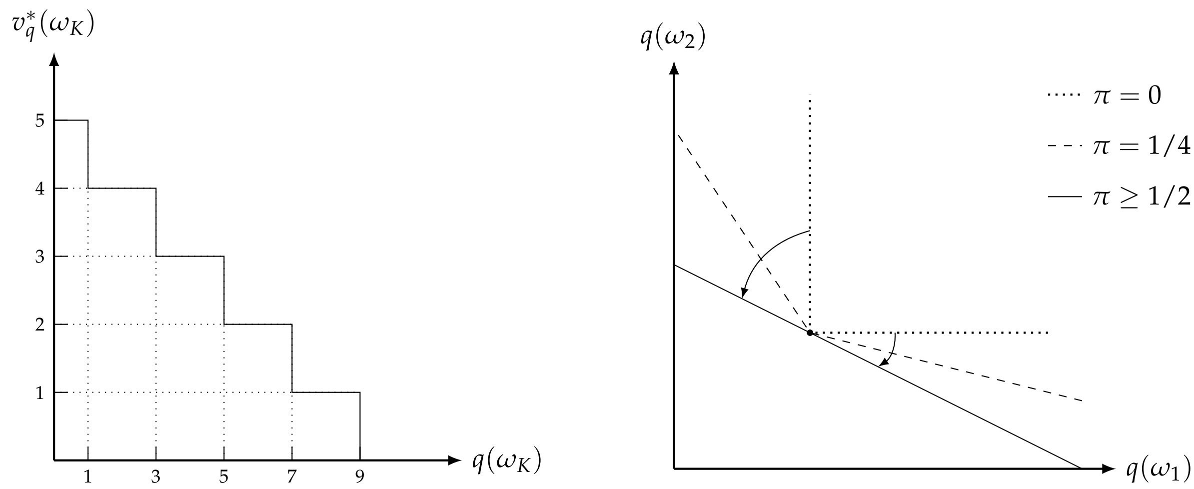
$$\max_{v \in v^*_q} \frac{v(\omega)}{v(\omega')} \geq \frac{p(\omega)}{p(\omega')} \geq \min_{v \in v^*_q} \frac{v(\omega)}{v(\omega')}, \qquad \omega, \omega' \in \Omega.^{20}$$

In this way, we can use v^*_q to characterize the platform's *demand functions* for data records, thus enabling a general study of the demand side of the "market for data." Which prices will prevail in this market is of course determined by the interplay of demand and supply. Under perfect competition, Dorfman et al. (1987) and Gale (1989) provide arguments for equality between v^*_q and equilibrium prices.

[18]Concavity follows because, by (5), we can view U^* as the result of minimizing a family of functions that are linear in q (see, e.g., Theorem 5.5 in Rockafellar, 1970). It is related directly to the concavification results in Mathevet et al. (2020) and indirectly to the individual-sufficiency results in Bergemann and Morris (2016).

[19]We provide sufficient conditions for this in terms of Γ in Appendix C.

[20]We slightly abuse notation by letting v^*_q stand for the *set* of optimal solutions at q. This condition is equivalent to $p \in \partial U^*(q)$, where $\partial U^*(q)$ is the superdifferential of U^* at q. Note that in the special case with a unitary budget and $p(\omega) = 1$ for all ω, choosing q is isomorphic to choosing an optimal prior in $\Delta(\Omega)$.



(a) Example of a demand curve: $\pi = 0$, $K = 10$, $\theta_k = k$ ($\forall k$), $q(\omega_k) = 1$ ($\forall k < K$).

(b) Example of indifference curves becoming less convex: $K = 2$, $\theta_k = k$ ($\forall k$)

Figure 1: Platform's demand and indifference curves

4.1.1 Application – Data and Price Discrimination (Part II)

We illustrate some of these concepts by specializing our analysis to the setting of Section 3.2.1. Recall that there is a single price-setting seller and the platform maximizes a weighted sum of his profits and the buyers' surplus, where the former receives weight $\pi \in [0, 1]$.

We first show an example of a downward-sloping demand curve. Figure 1(a) shows the value of records of type ω_K calculated using Proposition 2. This value is locally constant—as discussed in Proposition 3—and decreases as these records become more abundant—as discussed in Proposition 4. The figure also shows that, as $q(\omega_K)$ becomes sufficiently large, $v_q^*(\omega_K)$ reaches a lower bound, which in this case is 0.

Next, we explore how the substitutability between records changes as a function of π. When $\pi < \frac{1}{2}$, records become less substitutable as the platform cares more about the buyers' surplus (π decreases). When $\pi \geq \frac{1}{2}$, all types of records are perfect substitutes and $MRS_q(\omega, \omega') = -\frac{\omega}{\omega'}$ for all q, which is thus constant in π. Recall that a_q is the seller's optimal price if he knows only the database composition q.

Corollary 4. *Fix q and increase $\pi < \frac{1}{2}$. If $\omega, \omega' < a_q$, $MRS_q(\omega, \omega')$ is constant at $-\frac{\omega}{\omega'}$. If $\omega < a_q \leq \omega'$, $MRS_q(\omega, \omega')$ increases monotonically towards $-\frac{\omega}{\omega'}$ from below. If $\omega' > \omega \geq a_q$, $MRS_q(\omega, \omega')$ decreases monotonically towards $-\frac{\omega}{\omega'}$ from above.*

In words, as π increases towards $\frac{1}{2}$, for record types on the opposite side of a_q the platform's indifference curves rotate counter-clockwise in the direction of perfect substitutability. For records on the same side of a_q, its indifference curves rotate clockwise in the direction of perfect substitutes. Thus, the indifference curves become “less convex” around the dimension ω such that $\omega = a_q$. In particular, at $\pi = 0$ records of type ω such that $\omega = a_q$ are perfect complements with every other type. These patterns are illustrated in the right panel of Figure 1, which shows the platform's indifference curves in the case with two types of records.

4.2 Better Records and Willingness to Pay for Information

A platform can also change its database by refining some its existing records with better information. For example, this could involve observing new personal characteristics about a subset of the buyers. Intuitively, refining a buyer's record changes its type according to what the platform learns. How do such refinements change the value that the platform derives from each buyer's record? Do they always improve the platform's total payoff—hence lead to a positive willingness to pay for them?

We begin by formalizing what a refinement is. Recall that every buyer's record of type $\omega \in \Omega$ is informative about her θ, so it induces a belief $p_\omega \in \Delta(\Theta)$. A refinement of a record of type ω is a distribution $\sigma_\omega \in \Delta(\Omega)$ that satisfies the usual Bayes' consistency condition $p_\omega = \sum_{\omega' \in \Omega} \sigma_\omega(\omega') p_{\omega'}$. That is, any such refinement is equivalent to observing an exogenous signal that transforms a record of type ω into a record of type ω' with probability $\sigma_\omega(\omega')$. For instance, the original record may contain only the buyer's age, while the refined record may also contain her gender. When refining multiple records of type ω—in particular, a *share* $\alpha \in [0,1]$ of $q(\omega)$—each is refined independently according to σ_ω.[21] Importantly, implicit in the definition there is the assumption that Ω is “rich” in the sense that it already contains all record types which can result from refinement σ_ω. This allows us to use the platform's preferences over all databases in $\mathbb{R}^\Omega_+$ characterized by v^*_q to assess the consequences of refinement σ_ω.

Consider refining a share α of type-ω records according to σ_ω. How does this change the unit value that the platform derives from its records? Such a refinement has both direct and indirect effects, as it affects both the records that are being refined and those that are not. The root of these interdependencies is the externality discussed in Section 3.2. Refining α of type-ω records changes the original database q into a new one, denoted by q_α, which contains fewer records of type ω and more records of the types ω' that result from the refinement (i.e., $\omega' \in \operatorname{supp} \sigma_\omega$). Thus, the unit value of the former records may increase and that of the latter decrease by the scarcity principle (Proposition 4). Formally, given $\alpha \in [0,1]$ and refinement σ_ω, by the

[21] We discuss refinements that are correlated among records in Section 4.2.2.

Law of Large Numbers $q_\alpha(\omega) = (1-\alpha)q(\omega)$ and $q_\alpha(\omega') = q(\omega') + \alpha\sigma_\omega(\omega')q(\omega)$ (where we can interpret $\alpha = 0$ as refining only one record since it is infinitesimal and $q_0 = q$). Note that the composition q_α of the new database is certain, even though it is uncertain which records of type ω become of type ω'. Thus, it suffices that the sellers know that a database q has been refined according to σ_ω and α for them to know the resulting composition q_α.[22]

Corollary 5. *Fix q and suppose that a share α of type-ω records is refined according to σ_ω.*

(Direct Effects:) *The value of the* refined *records increases in expectation. That is, $\sum_{\omega'\in\Omega} v^*_{q_\alpha}(\omega')\sigma_\omega(\omega') - v^*_q(\omega) \geq 0$. This increase shrinks as α gets larger.*

(Indirect Effects:) *The value of the* unrefined *records of type ω increases: $v^*_{q_\alpha}(\omega) \geq v^*_q(\omega)$. The value of* unrefined *records of type $\omega' \in \operatorname{supp}\sigma_\omega$ decreases: $v^*_{q_\alpha}(\omega') \leq v^*_q(\omega')$. Both these effects are larger as α gets larger.*

With regard to the refined records, the expected gain in their value can shrink but never turn into a loss, even if refining more records lowers the value of the record types that result from it. Intuitively, for each refined record the platform knows more about the corresponding interaction, so it can better tailor its signals for the sellers and achieve more with that record. However, the externalities that contribute to its value may now be smaller. The former positive aspect dominates the latter because we are considering independent refinements. This is no longer true if refinements are correlated between records (see Section 4.2.2).

Corollary 5 highlights a novel implication of people's decisions to disclose their personal data. Recall that v^*_q is a benchmark for individually compensating buyers for their specific record. We can interpret a refinement as a buyer's decision of whether to disclose more of her personal characteristics. Imagine a group of similar buyers—i.e., whose records are of the same type—which includes Ann and Bonnie. Ann is among those who decide to disclose, expecting that her record will become more valuable and hence may result in a higher compensation (direct effect). Bonnie instead decides *not* to disclose, yet her record may also become more valuable but for different reasons (indirect effect). Moreover, a larger group of disclosing buyers decreases Ann's expected gain in value, but increases Bonnie's. The disclosing buyers can also cause the value of, say, Charlie's record to fall because it is of one of the types that can result from the refinement, which can lead to a lower compensation for her (indirect effect). Importantly, these effects happen even if the platform does not learn anything new about

[22] Of course, in reality the platform may refine its database privately without the sellers' knowing exactly how. Allowing for this requires enriching the model accordingly and introduces further complications, which we leave for future research.

Bonnie and Charlie from what Ann and the other buyers disclosed—in contrast to the learning externalities discussed in the literature (see Section 3.2).

Given these complex positive and negative effects of refinements on the unit value of all records, it is unclear whether they benefit the platform overall. In fact, those effects reflect a fundamental trade-off that refining records can generate in intermediation problems (but not in standard-decision problems). On the one hand, knowing more about each refined record allows the platform to better tailor its signals for the sellers and possibly achieve more in those interactions. On the other hand, changing the database q can change the sellers' beliefs about the buyers they interact with, which in turn can weaken the platform's informational advantage and hence ability to influence their actions. A mechanism x may be obedient before the refinement but not after it, which can hurt the platform. Nonetheless, we obtain the following.

Proposition 6. *Fix q and suppose that a share α of type-ω records is refined according to σ_ω. The platform weakly benefits from this refinement: $U^*(q_\alpha) - U^*(q) \geq 0$. Moreover, the benefit is zero for all $\alpha \in [0,1]$ if (and only if generically in q) there exists $a \in supp\, x^*_q(\cdot|\omega'')$ for $\omega'' = \omega$ and all $\omega'' \in supp\, \sigma_\omega$. Finally, the refinement's marginal benefit decreases in α.*

This implies that the platform's willingness to pay for a refinement is always weakly positive, so the positive effects on refined records always dominate the negative effects on other records. The platform's willingness to pay can be strictly negative for refinements that are correlated between records (see Section 4.2.2).

Proposition 6 provides a sharp condition for the willingness to pay for refinements to be zero, which depends only on the initial q. Given this q, there must be a *common* action profile that the platform induces with positive probability both for the original record to be refined and for every type that it can turn into when refined. Intuitively, this means that the platform is exploiting its informational advantage to *sometimes* use the original record as if it was already refined, so refining it does not make it more valuable. In fact, under this condition all direct and indirect effects in Corollary 5 are zero. Importantly, note that the direct effect of refining a record can be zero even if the platform uses it differently after the refinement (i.e., even if $x^*_q(\cdot|\omega) \neq x^*_q(\cdot|\omega')$ and $u^*_q(\omega) \neq u^*_q(\omega')$ for some $\omega' \in \operatorname{supp} \sigma_\omega$). And overall the platform may be unwilling to pay a strictly positive price for refining its records, despite acting on the information it receives (i.e., changing x^*_q). This is different from standard-decision problems, for which a key insight is that more information is strictly beneficial if it changes the optimal choices.

Finally, Proposition 6 shows that the marginal benefit of a refinement is diminishing in the share of refined records. This may be reminiscent of classic results in standard decision prob-

$x_q^*(a\|\omega)$	$\omega = 1$	$\omega = 2$	$\omega = \omega^\circ$
$a = 1$	1	$\frac{q(1)-(2h-1)q(\omega^\circ)}{q(2)}$	1
$a = 2$	0	$1 - \frac{q(1)-(2h-1)q(\omega^\circ)}{q(2)}$	0

Table 1: Optimal x_q^*.

lems where information has decreasing marginal returns. However, there is an important difference. Our exercise is not to gradually give the platform more information about one fixed interaction so that it can better learn the buyer's preferences. Focusing on this intensive margin is perhaps the most typical way of studying returns from information, especially in standard decision problems (see Bergemann and Ottaviani, 2021, Section 2.5). We instead fix the amount of information we give the platform for each interaction (i.e., σ_ω) and vary how many interactions we *independently* refine in this way (i.e., α). As such, this extensive-margin exercise has constant returns for standard-decision problems, but not for intermediation problems—again, due to the externalities documented in Section 3.2.[23]

4.2.1 Application – Data and Price Discrimination (Part III)

We illustrate some of these points using the setting of Section 3.2.1 with a single price-setting seller. Suppose the platform maximizes the buyers' surplus ($\pi = 0$). As before, ω_1 and ω_2 are the record types that correspond to buyers whose valuation θ is 1 and 2; instead, ω° corresponds to buyers' whose valuation is believed to be $\theta = 2$ with probability $h > \frac{1}{2}$ and $\theta = 1$ otherwise. Fix any q that satisfies $q(\omega^\circ) < q(\omega_1) < q(\omega_2)$ in which case we have that $v_q^*(\omega_1) = 1$, $v_q^*(\omega_2) = 0$, and $v_q^*(\omega^\circ) = 1 - h$. Now, suppose we refine a share α of type-ω° records with a refinement σ_{ω° such that $\sigma_{\omega^\circ}(\omega_2) = h$ and $\sigma_{\omega^\circ}(\omega_1) = 1 - h$. As shown in Table 1, the platform changes how it uses the refined records—compare $x_q^*(\cdot|\omega^\circ)$ and $x_q^*(\cdot|\omega_2)$—as well as the unrefined records of type ω_2—note that $x_q^*(\cdot|\omega_2)$ depends on q. Nonetheless, the "if" condition in Proposition 6 holds. Therefore, for any $\alpha \in [0, 1]$ the platform's willingness to pay for the refinement as well as the expected increase in unit value of each refined record are zero: $U^*(q) = U^*(q_\alpha)$ and $v^*(\omega_1)\sigma_{\omega^\circ}(\omega_1) + v^*(\omega_2)\sigma_{\omega^\circ}(\omega_2) = v^*(\omega^\circ)$. Both are instead strictly positive if $q(\omega_1) < q(\omega^\circ)$ and $\alpha > 0$ is sufficiently small. See Appendix D for more details.

[23] In fact, given σ_ω the marginal effect of changing α on $U^*(q_\alpha)$ equals $\sum_{\omega' \in \Omega} v_{q_\alpha}^*(\omega')\sigma_\omega(\omega') - v_{q_\alpha}^*(\omega)$ (see the proof of Proposition 6 for details).

4.2.2 Correlation and General Refinements

One may wonder whether the results in this section depend on how records are refined. For instance, records may not be refined independently. Observing new characteristics about some buyer may provide information about similar unobserved characteristics of other buyers. Note that this is consistent with our initial assumption that, for a fixed database, each buyer's record provides information only about her θ. Each record should contain all the data available to the platform that is informative about that buyer, which may involve observations of variables that belong to other individuals. Once this assignment is done for each buyer, conditional on her record any other buyer's record provides no additional information about her θ. Given this, our analysis is unchanged for a fixed database. However, observing new characteristics for one individual, may require to update the records of multiple buyers in a correlated way and change the database accordingly.

Nonetheless, we can continue to view general refinements as changing the platform's database and analyze their consequences using the tools developed above. For every q and $\omega \in \Omega$, let $Q(\omega, q)$ be the set of databases we can reach from q by refining type-ω records: $Q(\omega, q)$ contains all $q' \in \mathbb{R}_+^\Omega$ that have fewer records of type ω, more records of the types ω' that can result from the refinement, the same quantity of records of all other types, and the same total quantity of records. Acquiring better data about type-ω records involves transitioning from q to some $q' \in Q(\omega, q)$, possibly with randomness. That is, it defines a distribution $\rho(\omega, q) \in \Delta(Q(\omega, q))$, which can exhibit any correlation in how records are refined.

Proposition 5 and Corollary 6 do not extend to general refinements. It is easy to see that, if given q it is optimal to fully disclose every buyer's records, then *any* refinement $\rho(\omega, q)$ increases both the unit value of the refined records and the platform's total payoff in expectation.[24] This is analogous to the result that the value of information in standard-decision problems is always non-negative. However, this can fail for intermediation problems, as the next example illustrates. The reason is again that refining records changes not only how much the platform knows about each buyer-sellers interaction, but also the degree and nature of the asymmetric information between the platform and the sellers implied by the commonly-known database composition. This highlights another important distinctive feature of the value of data in intermediation problems. In practice, for large databases it may be more reasonable that acquiring better data takes the form of random draws from a large population. The conceptual point remains that, unlike for decision-makers, for intermediaries information can have strictly

[24] This follows from the fact that $v_q^*(\omega)$ is independent of q for all $\omega \in \Omega$ and therefore only the marginal of ρ for each refined record matters.

negative value.

To illustrate, we use again the setting of Section 4.2.1 with a surplus-maximizing platform ($\pi = 0$). In addition, assume that $q(\omega_1) + q(\omega^\circ)(1-h) > q(\omega_2)$. Consider the following $\rho(\omega^\circ, q)$, which is arguably extreme but serves to make our point as clearly as possible. The platform is told that all its type-ω° records involve buyers with the same valuation. Thus, if refined, with probability $1-h$ they *all* become records of type ω_1 and with probability h they *all* become records of type ω_2. That is, $\rho(\omega^\circ, q)$ puts probability only on q' and q'', where $q'(\omega_1) = q(\omega_1) + q(\omega^\circ)$, $q'(\omega_2) = q(\omega_2)$, $q''(\omega_1) = q(\omega_1)$, $q''(\omega_2) = q(\omega_2) + q(\omega^\circ)$, and $q'(\omega^\circ) = q''(\omega^\circ) = 0$. This refinement has a strictly negative effect not only on the unit value of the refined records, but also on the platform's total payoff. Since $v^*_{q'}(\omega_1) = 0$ and $v^*_{q''}(\omega_2) = 0$ (see Appendix D for the calculations), we have

$$(1-h)v^*_{q'}(\omega_1) + hv^*_{q''}(\omega_2) - v^*_q(\omega^\circ) = -(1-h) < 0$$

and

$$U^*(q) = q(\omega_1) + q(\omega^\circ)(1-h) > \max\{q(\omega_2), q(\omega_1)\} = \max\{U^*(q'), U^*(q'')\}.$$

By contrast, if the platform maximizes the seller's profits ($\pi = 1$), we have $v^*_{\hat{q}}(\omega_1) = 1$, $v^*_{\hat{q}}(\omega_2) = 2$, and $v^*_{\hat{q}}(\omega^\circ) = 2h$ for all $\hat{q}$. The *same* refinement has a strictly positive effect on both the unit value of the refined records and the platform's total payoff.

The key is that a profit-maximizing platform treats each buyer-seller interaction as an independent decision problem, so it does not care about correlation in how it learns about records. By contrast, a surplus-maximizing platform cares about such correlation, because it can have profound consequences on its informational advantage through the composition of its database.

5 Discussion

Our framework and results apply more broadly to any setting where a principal mediates interactions between multiple agents using data. We briefly explain this applicability here.

Principal's Actions and Agents' Data. For ease of exposition, we simplified the model in several ways. Neither changes the analysis or its interpretations. First, we can allow the principal to also choose an action $a_0 \in A_0$ for each mediated interaction. In this case, a mechanism x also has to specify a_0 for each ω. Second, we can allow each agent i to also observe privately some own data about the interaction he is in. For example, in our leading e-commerce context each seller can observe the quality or the history of customer reviews of his product.

We can again model the realizations of such data with some finite set Ω_i, where each ω_i is ultimately an exogenous signal about some underlying payoff-relevant θ. We chose not to use "signal" in reference to ω_i to avoid confusion with the principal's endogenous signals. Let $\Omega = \Omega_0 \times \ldots \times \Omega_n$ with typical element $\omega = (\omega_0, \ldots, \omega_n)$. The key assumption is that the principal also observes the private data of each agent—i.e., the entire $\omega = (\omega_0, \ldots, \omega_n)$—as does the omniscient designer in Bergemann and Morris (2016). Thus, now the whole vector ω defines a type of data record in the principal's database and characterizes each interaction that she mediates. Our proofs in Appendix B already take this more general setting into account.

A Comment on Terminology. Each Ω_i is akin to what is often called the set of party i's types. We can then view an interaction as being characterized by the profile ω of all its participants' types. We instead use "type" to refer to ω itself because our analysis focuses on how the principal mediates interactions based on what she knows about their characterizing data *as a whole*. This use is also consistent with viewing ω_0 as the principal's type when she alone observes data, as this type is defined by what data characterizes the interaction she is mediating.

Simultaneous vs Sequential Mediation. We interpret the principal as mediating all interactions in her database simultaneously. For instance, in the introduction example the platform may mediate all buyer-seller interactions simultaneously, where the seller sets one price for each market segment created by the platform. An equally valid interpretation is that the principal commits to a mechanism for the whole database and then interactions are drawn independently and mediated one at a time. In our example, the platform may create the market segments once and for all, then draw a buyer one at a time and tell the seller to which segment he or she belongs. Depending on the application, one interpretation may fit better.

Another Example: Routing Games. Our leading example has been that of an e-commerce platform mediating the interactions between buyers and sellers. We conclude this discussion by sketching another applications of the model. A navigation app uses data about routes' conditions to direct traffic by providing drivers with information—such as recommended routes and travel times. Das et al. (2017) propose a simple way to model this complex problem. Suppose the app (principal) seeks to minimize congestion. We can think of an interaction as consisting of a group of drivers (agents) in some city who simultaneously choose, say, one of two routes between its residential and business district. For each route, the travel time increases in how many drivers choose it but at different rates (e.g., because one is a highway and one is surface streets); travel times also depend on some uncertain event (e.g., construction work), which is observed only by the app. A database is then the collection of the realized events for all interactions across the cities served by the app. For simplicity, suppose each interaction happens in a different city so that they are independent in all respects, including their uncertain event. If the database is large, its composition q should reflect the primitive distribution of this event (i.e., the probability of construction on a given route).

References

ACEMOGLU, D., A. MAKHDOUMI, A. MALEKIAN, AND A. OZDAGLAR (2021): "Too Much Data: Prices and Inefficiencies in Data Markets," *Americal Economic Journal: Microeconomics*, Forthcoming.

ACQUISTI, A., C. TAYLOR, AND L. WAGMAN (2016): "The Economics of Privacy," *Journal of Economic Literature*, 54, 442–92.

ADMATI, A. AND P. PFLEIDERER (1990): "Direct and Indirect Sale of Information," *Econometrica*, 58, 901–28.

ADMATI, A. R. AND P. PFLEIDERER (1986): "A Monopolistic Market for Information," *Journal of Economic Theory*, 39, 400–438.

ALI, S. N., G. LEWIS, AND S. VASSERMAN (2020): "Voluntary Disclosure and Personalized Pricing," *arXiv preprint arXiv:1912.04774v2*.

ARRIETA-IBARRA, I., L. GOFF, D. JIMÉNEZ-HERNÁNDEZ, J. LANIER, AND E. G. WEYL (2018): "Should We Treat Data as Labor? Moving beyond "Free"," *AEA Papers and Proceedings*, 108, 38–42.

BERGEMANN, D. AND A. BONATTI (2015): "Selling Cookies," *American Economic Journal: Microeconomics*, 7, 259–94.

——— (2019): "Markets for Information: An Introduction," *Annual Review of Economics*, 11, 85–107.

BERGEMANN, D., A. BONATTI, AND T. GAN (2020): "The Economics of Social Data," *Cowles Foundation Discussion Papers*.

BERGEMANN, D., A. BONATTI, AND A. SMOLIN (2018): "The Design and Price of Information," *American Economic Review*, 108, 1–48.

BERGEMANN, D., B. BROOKS, AND S. MORRIS (2015): "The Limits of Price Discrimination," *American Economic Review*, 105 (3).

BERGEMANN, D. AND S. MORRIS (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

——— (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57(1), pp. 44-95).

Bergemann, D. and M. Ottaviani (2021): "Information Markets and Nonmarkets," *Cowles Foundation Discussion Papers*.

Bertsimas, D. and J. N. Tsitsiklis (1997): *Introduction to Linear Optimization*, Athena Scientific.

Brooks, B. and S. Du (2020): "A Strong Minimax Theorem for Informationally-Robust Auction Design," *Working Paper*.

——— (2021): "Optimal Auction Design with Common Values: An Informationally-Robust Approach," *Econometrica*, 89(3), 1313–1360.

Calzolari, G. and A. Pavan (2006): "On the Optimality of Privacy in Sequential Contracting," *Journal of Economic Theory*, 130, 168–204.

Choi, J. P., D.-S. Jeon, and B.-C. Kim (2019): "Privacy and personal data collection with information externalities," *Journal of Public Economics*, 173, 113–124.

Crémer, J., Y.-A. de Montjoye, and H. Schweitzer (2019): "Competition policy for the digital era," *European Commission*.

Das, S., E. Kamenica, and R. Mirka (2017): "Reducing Congestion through Information Design," in *2017 55th annual allerton conference on communication, control, and computing (allerton)*, IEEE, 1279–1284.

Dizdar, D. and E. Kováč (2020): "A Simple Proof of Strong Duality in the Linear Persuasion Problem," *Games and Economic Behavior*, 122, 407–412.

Dorfman, R., P. A. Samuelson, and R. M. Solow (1987): *Linear Programming and Economic Analysis*, Courier Corporation.

Du, S. (2018): "Robust Mechanisms Under Common Valuation," *Econometrica*, 86(5), 1569–1588.

Dworczak, P. and A. Kolotilin (2019): "The Persuasion Duality," *Available at SSRN 3474376*.

Dworczak, P. and G. Martini (2019): "The Simple Economics of Optimal Persuasion," *Journal of Political Economy*, 127, 1993–2048.

Elliott, M., A. Galeotti, and A. Koh (2020): “Market Segmentation through Information,” *Working Paper*.

Federal Trade Commission (2014): *Data Brokers: A Call for Transparency and Accountability*, A Report by the Federal Trade Commission, May.

Frankel, A. and E. Kamenica (2019): “Quantifying information and uncertainty,” *American Economic Review*, 109, 3650–80.

Gale, D. (1989): *The Theory of Linear Economic Models*, University of Chicago press.

Galperti, S. and J. Perego (2018): “A Dual Perspective on Information Design,” *Available at SSRN 3297406*.

Goldberg, S., G. Johnson, and S. Shriver (2021): “Regulating privacy online: An economic evaluation of the GDPR,” *Available at SSRN*.

Ichihashi, S. (2020): “Online Privacy and Information Disclosure by Consumers,” *American Economic Review*, 110, 569–95.

——— (2021): “The Economics of Data Externalities,” *Journal of Economic Theory*.

Jones, C. I. and C. Tonetti (2020): “Nonrivalry and the Economics of Data,” *American Economic Review*, 110, 2819–58.

Kamenica, E. and M. Gentzkow (2011): “Bayesian Persuasion,” *American Economic Review*, 101, 2590 – 2615.

Kolotilin, A. (2018): “Optimal Information Disclosure: A Linear Programming Approach,” *Theoretical Economics*, 13, 607 – 635.

Lanier, J. (2013): *Who Owns the Future?*, Simon & Schuster.

Mathevet, L., J. Perego, and I. Taneva (2020): “On Information Design in Games,” *Journal of Political Economy*, 128, 1370–1404.

Milgrom, P. and C. Shannon (1994): “Monotone Comparative Statics,” *Econometrica*, 157–180.

Myerson, R. B. (1983): “Mechanism Design by an Informed Principal,” *Econometrica*, 51, 1767–1797.

——— (1984): "Two-Person Bargaining Problems with Incomplete Information," *Econometrica*, 52, 461–488.

——— (1997): "Dual Reduction and Elementary Games," *Games and Economic Behavior*, 21, 183–202.

NAU, R. F. (1992): "Joint Coherence in Games of Incomplete Information," *Management Science*, 38, 374–387.

NAU, R. F. AND K. F. MCCARDLE (1990): "Coherent Behavior in Noncooperative Games," *Journal of Economic Theory*, 50, 424–444.

POSNER, E. AND E. G. WEYL (2018): *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*, Princeton University Press.

ROCKAFELLAR, R. T. (1970): *Convex analysis*, vol. 36, Princeton university press.

SCOTT-MORTON, F., P. BOUVIER, A. EZRACHI, B. JULIEN, R. KATZ, G. KIMMELMAN, A. MELAMED, AND J. MORGENSTERN (2019): "Report of the Committee for the study of digital platforms, market structure and antitrust subcommittee," *George Stigler Center for the study of the economy and the state, University of Chicago Booth School of Business*.

STIGLER REPORT (2019): "Stigler Committee on Digital Platforms, Final Report," September.

YANG, K. H. (2020): "Selling Consumer Data for Profit: Optimal Market-Segmentation Design and its Consequences," *Working Paper*.

Appendix

A A Gambling Perspective on Data Value and Externalities

To better understand the value of data records and the externalities between them, it will help to provide a stand-alone interpretation of the data-value problem $\mathcal{V}_q$. With minor adjustments, this interpretation applies to problems where the sellers also observe some data and the platform takes some action. In this part, we will fix $q \in \mathbb{R}^{\Omega}_{++}$ and so drop it from notation.

For our interpretation, it is convenient to rewrite the data-value problem in the following equivalent way. Let $b = (b_1, \ldots, b_n)$ be a profile such that $b_i : A_i \to \mathbb{R}_{++}$ for all i and $\ell = (\ell_1, \ldots, \ell_n)$ be a profile such that $\ell_i : A_i \to \Delta(A_i)$ for all i. Given (b, ℓ), for each $i \in I$ and (a, ω) define

$$t_i(a,\omega) = b_i(a_i) \sum_{a_i' \in A_i} \left(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega) \right) \ell_i(a_i'|a_i)$$

and $t(a,\omega) = \sum_{i \in I} t_i(a,\omega)$. The data-value problem can be rewritten as

$$\begin{aligned} \mathcal{V}_q : \quad & \min_{v,b,\ell} \sum_{\omega \in \Omega} v(\omega) q(\omega) \\ & \text{s.t. for all } \omega \in \Omega, \\ & v(\omega) = \max_{a \in A} \left\{ u_0(a,\omega) + t(a,\omega) \right\}, \end{aligned} \tag{9}$$

As in the main text, we will denote any optimal solution of $\mathcal{V}_q$ by (v_q^*, b_q^*, ℓ_q^*).

A.1 Gambles Against the Agents

Our interpretation hinges on unpacking how the platform determines the sellers' contributions to the externalities between buyers' records. The value formula (9) reveals that she does so through her choice of b and ℓ, which fully pin down each $t(a,\omega)$ and hence ultimately $v(\omega)$. Recall that the platform wants to *minimize* the values of her records, so she would like to lower $t(a,\omega) = \sum_{i \in I} t_i(a,\omega)$ as much as possible for all (a,ω). Each term of $t_i(a,\omega)$ takes the form

$$b_i(a_i)\ell_i(a_i'|a_i) \left(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega) \right),$$

which contributes to lowering $t_i(a,\omega)$ if and only if $\ell_i(a_i'|a_i) > 0$ and $u_i(a_i, a_{-i}, \omega) < u_i(a_i', a_{-i}, \omega)$. In words, if seller i knew his interaction's type ω and his opponents' offers a_{-i}, he would strictly prefer a_i' to a_i. In this case, offering a_i amounts to making a mistake from an ex-post viewpoint. We will then say that seller i regrets offering a_i.

Thus, inducing sellers to play actions they will regret emerges as an intrinsic goal of the platform's problem—together with maximizing her payoff u_0 of course. In this view, (b_i, ℓ_i) and the corresponding t_i become an exploitation strategy on the part of the platform against seller i. In order to induce regrettable actions, she must withhold some information from seller i about ω or a_{-i}. This explains why the platform may prefer partial disclosure, from the perspective of her data-value problem. In the end, the value $v(\omega)$ results from a trade-off between the payoff $u_0(a, \omega)$ and the return from inducing sellers to choose regrettable actions.

This return depends on the structure of b and ℓ, which define a family of gambles against the sellers. To see this, fix any (a, ω) and seller i. Then, $\ell_i(\cdot|a_i) \in \Delta(A_i)$ defines a lottery over prizes, where for each a_i' the prize is $u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega)$ for the platform; the scaling term $b_i(a_i)$ defines the stakes that she bets on this lottery. Given her objective, the platform "wins" when $u_i(a_i, a_{-i}, \omega) < u_i(a_i', a_{-i}, \omega)$ and "loses" otherwise. Thus, we can interpret $t(a, \omega)$ as the overall expected prize from (b, ℓ). We can then think of $\mathcal{V}$ as a fictitious environment where money is a medium of exchange and the platform can write monetary gambling contracts with each seller. Such contracts are enforced through contingent-claim markets that determine prizes based on the interaction's type ω and outcome a.[25]

We can then link the externalities between buyers' records with how the platform chooses these gambles in $\mathcal{V}$. Negative externalities $t^* < 0$ correspond to gambles favorable to the platform, in the sense that wins exceed losses in expectation. This requires the help of other records to withhold information and thus induce the sellers to play actions they will regret. Conversely, positive externalities $t^* > 0$ correspond to gambles unfavorable to the platform, in the sense that losses exceed wins in expectation. Corollary 1 implies that, at the optimum, the platform chooses gambles that favor her for some records, but not for others. In fact, this stems from deeper constraints and trade-offs in the use of such gambles against the sellers.

A.2 Feasible Gambles and Trade-offs

The gambles the platform can use to exploit the sellers in $\mathcal{V}$ have specific features that help us understand the nature of the data-value problem.

Some of these features reflect structural properties of $\mathcal{V}$. While the prizes of each gamble are contingent on both ω and the entire a, for each seller i both b_i and ℓ_i can depend only of his a_i.

[25] See Nau (1992) for a related interpretation.

This constrains the platform's ability to tailor her gambles with each seller across records. These properties reflect in $\mathcal{V}$ key interdependences in $\mathcal{U}$: The independence of (b_i, ℓ_i) from a_{-i} reflects the interdependence in $\mathcal{U}$ between sellers' incentives; the independence of (b_i, ℓ_i) from ω reflects the non-separability of $\mathcal{U}$ across data records. To see this, suppose $\ell_i(\hat{a}_i|a_i) > 0$. Then, (b_i, ℓ_i) links the right-hand side of the value formula (9) for (a_i, a_{-i}, ω) and (a_i, a'_{-i}, ω'). In particular, if $u_i(a_i, a_{-i}, \omega) < u_i(\hat{a}_i, a_{-i}, \omega)$ but $u_i(a_i, a'_{-i}, \omega') > u_i(\hat{a}_i, a'_{-i}, \omega')$, the platform faces a trade-off in determining v, as she may not be able to use (b_i, ℓ_i) to lower $v(\omega)$ without also raising $v(\omega')$. This is another way to see why and how externalities arise between records. When committing to (b, ℓ) the platform has to take into account these effects of each (b_i, ℓ_i) across records. How she solves these trade-offs depends on the relative frequency of records in the database (hence q). Importantly, this transformation of non-separabilities in $\mathcal{U}$ into independence properties of (b, ℓ) is what enables $\mathcal{V}$ to assign values individually to each record.

In fact, the platform faces other restrictions in her ability to *jointly* exploit the sellers. It is intuitive that she would want to design (b, ℓ) so that $t(a, \omega) \leq 0$ for all (a, ω) with some strict inequality. This would guarantee a sure arbitrage against the sellers. Such gambles, however, are infeasible in the following sense. Recall that by complementary slackness $x^*(a|\omega) > 0$ implies $v^*(\omega) = u_0(a, \omega) + t^*(a, \omega)$. Thus, since every ω must induce some action profile for every x, action profiles that cannot be in the support of any obedient $x(\cdot|\omega)$ are irrelevant for determining $v^*(\omega)$. Given this, define

$$\mathbf{X} = \{(a, \omega) \in A \times \Omega : x(a|\omega) > 0 \text{ for some obedient } x\}.$$

Let $G(\mathbf{X})$ be the set of gambles that can be contingent only on pairs $(a, \omega) \in \mathbf{X}$ (formally, we restrict the functions b and ℓ to the subdomain $\mathbf{X}$). Note that restricting the platform to choosing from $G(\mathbf{X})$ in $\mathcal{V}$ is immaterial for its optimal solution, in the same way that restricting x to the domain $\mathbf{X}$ is immaterial in $\mathcal{U}$.

Proposition 7. *For every gamble $(b, \ell) \in G(\mathbf{X})$, if $t(a, \omega) < 0$ for some (a, ω), there must exist (a', ω') such that $t(a', \omega') > 0$.*

This property is closely related to a similar result in Nau (1992). For completeness Appendix B provides a proof, which relies on a dual characterization of $\mathbf{X}$ using Farkas' lemma.

The economic takeaway is that in her attempt to minimize values v by exploiting the sellers with (b, ℓ), the platform faces a fundamental trade-off, which is a hallmark of problem $\mathcal{V}$. Successfully exploiting the sellers for records of type ω with some outcome a requires paying the cost of losing against them for records of some other type ω' or outcome a'. Note that this result is stronger than Corollary 1, as it refers to the deep structure of data-value problems for intermediaries. It also sheds light on how and how much they can actually manipulate sellers by conveying information.

B Proofs

All proofs in this appendix are for the general case where the agents observe private data in the form of $\omega_i \in \Omega_i$ and hence $\omega = (\omega_0, \omega_1, \ldots, \omega_n)$ (see Section 5). The special case where only the principal observes data obtains by having $|\Omega_i| = 1$ for all $i \in I$.

Proof of Lemma 1. We will formulate the problem directly in terms of choosing a measure $\chi \in \mathbb{R}_+^{A\times\Omega}$. Formally, the problem is

$$\begin{aligned}
\mathcal{U}_q: \qquad & \max_{\chi} \sum_{\omega\in\Omega, a\in A} u_0(a,\omega)\chi(a,\omega) \\
& \text{s.t. for all } i \in I, \omega_i \in \Omega_i, \text{ and } a_i, a_i' \in A_i, \\
& \sum_{\omega_{-i}\in\Omega_{-i}, a_{-i}\in A_{-i}} \left(u_i(a_i,a_{-i},\omega) - u_i(a_i',a_{-i},\omega)\right)\chi(a_i,a_{-i},\omega) \geq 0, \qquad (10) \\
& \text{and for all } \omega \in \Omega, \qquad (11) \\
& \qquad \sum_{a\in A} \chi(a,\omega) = q(\omega).
\end{aligned}$$

It is convenient to express this problem in matrix form. Fix an arbitrary total ordering of the set $A \times \Omega$. We denote by $\mathbf{u}_0 \in \mathbb{R}^{A\times\Omega}$ the vector whose entry corresponding to (a,ω) is $u_0(a,\omega)$. For every player i, let $\mathbf{U}_i \in \mathbb{R}^{(A_i\times A_i\times\Omega_i)\times(A\times\Omega)}$ be a matrix thus defined: For each row $(a_i', a_i'', \omega_i') \in A_i \times A_i \times \Omega_i$ and column $(a,\omega) \in A\times\Omega$, let the corresponding entry be

$$\mathbf{U}_i((a_i',a_i'',\omega_i'),(a,\omega)) = \begin{cases} u_i(a_i',a_{-i},\omega) - u_i(a_i'',a_{-i},\omega) & \text{if } a_i' = a_i, \omega_i' = \omega_i \\ 0 & \text{else.} \end{cases}$$

Thus, $\mathbf{U}_i(a_i',a_i'',\omega_i')$ denotes the row labeled by (a_i',a_i'',ω_i') (which defines the corresponding obedience constraint) and $\mathbf{U}_i(a,\omega)$ denotes the column labeled by (a,ω). Define the matrix $\mathbf{U}$ by stacking all the matrices $\{\mathbf{U}_i\}_{i\in I}$ on top each other. Finally, define the indicator matrix $I \in \{0,1\}^{\Omega\times(A\times\Omega)}$ such that, for each row ω' and column (a,ω'),

$$I(\omega',(a,\omega)) := \begin{cases} 1 & \text{if } \omega' = \omega \\ 0 & \text{else.} \end{cases}$$

With this notation and treating q as a vector, $\mathcal{U}_q$ can be written as follows:

$$\begin{array}{lrcl} \max_{\chi} & \mathbf{u}_0^T \chi & & \\ \text{s.t.} & \mathbf{U}\chi & \geq & \mathbf{0}, \\ & I\chi & = & q, \\ & \chi & \geq & \mathbf{0}. \end{array} \tag{12}$$

Given this, by standard linear-programming arguments (Bertsimas and Tsitsiklis (1997)) the dual of $\mathcal{U}_q$ can be written as

$$\min_{\lambda, v} \mathbf{0}^T \lambda + q^T v$$

subject to for all $i = 1, \ldots, n$, $a_i, a_i' \in A_i$, and $\omega_i \in \Omega_i$,

$$\lambda_i(a_i'|a_i, \omega_i) \geq 0,$$

$v(\omega) \in \mathbb{R}$ for all $\omega \in \Omega$ (i.e., it is unconstrained), and for all $(a, \omega) \in A \times \Omega$

$$u_0(a, \omega) \leq v(\omega) - \sum_{i \in I} \left\{ \sum_{a_i' \in A_i} \left(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega)\right) \lambda_i(a_i'|a_i, \omega_i) \right\}.$$

The objective simplifies to

$$\min_{\lambda, v} \sum_{\omega \in \Omega} v(\omega) q(\omega).$$

The second set of constraints can be written as

$$v(\omega) \geq u_0(a, \omega) + \sum_{i \in I} \left\{ \sum_{a_i' \in A_i} \left(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega)\right) \lambda_i(a_i'|a_i, \omega_i) \right\}.$$

Finally, we express this dual in a form that is equivalent to $\mathcal{V}_q$ by exploiting the structure of the specific problem at hand. To this end, for every i and ω_i we can set $\lambda_i(a_i|a_i, \omega_i) = 1$ (or any strictly positive value) for all $a_i \in A_i$. Given this, for every i and $(a_i, \omega_i) \in A_i \times \Omega_i$, define

$$b_i(a_i, \omega_i) = \sum_{a_i' \in A_i} \lambda_i(a_i'|a_i, \omega_i),$$

which is strictly positive by construction. Also, for every i and $(a_i', a_i, \omega_i) \in A_i \times A_i \times \Omega_i$ define

$$\ell_i(a_i'|a_i, \omega_i) = \frac{\lambda_i(a_i'|a_i, \omega_i)}{b_i(a_i, \omega_i)},$$

which implies that $\ell_i(\cdot|a_i, \omega_i) \in \Delta(A_i)$. Letting $b = (b_1, \ldots, b_n)$ and $\ell = (\ell_1, \ldots, \ell_n)$ so defined, we have the dual of $\mathcal{U}_q$ is equivalent to

$$\min_{v, b, \ell} \sum_{\omega \in \Omega} v(\omega) q(\omega)$$

subject to for all (a, ω)

$$v(\omega) \geq u_0(a, \omega) + \sum_{i \in I} t_i(a, \omega),$$

where

$$t_i(a, \omega) = b_i(a_i, \omega_i) \sum_{a'_i \in A_i} \left(u_i(a_i, a_{-i}, \omega) - u_i(a'_i, a_{-i}, \omega) \right) \ell_i(a'_i | a_i, \omega_i).$$

Since for every $\omega \in \Omega$ this constraint has to hold for all $a \in A$ and the data-value problem is the minimization problem, we conclude that each $v(\omega)$ has to satisfy

$$v(\omega) = \max_{a \in A} \left\{ u_0(a, \omega) + t(a, \omega) \right\},$$

where $t(a, \omega) = \sum_{i \in I} t_i(a, \omega)$. Thus, we obtain problem $\mathcal{V}_q$. □

Remark 1. *We can transform $\mathcal{U}_q$ to the standard form $\mathcal{U}_q^S$ which can be written as follows:*

$$\begin{array}{rrcl} \max\limits_{\chi, s} & \mathbf{u}_0 \chi & & \\ s.t. & \mathbf{U}\chi - s & = & \mathbf{0}, \\ & I\chi & = & q, \\ & \chi, s & \geq & \mathbf{0}, \end{array} \tag{13}$$

where each $s_i(a'_i | a_i, \omega_i)$ is a nonnegative slack variable. The dual of $\mathcal{U}_q^S$ coincides with the data-value problem $\mathcal{V}_q$. Note that $\mathcal{U}_q$ always has an optimal solution χ_q^*, which is generically unique and hence corresponds to an extreme point of the polyhedron of feasible χ. Moreover, this χ_q^* is an optimal solution of $\mathcal{U}_q^S$ as well. The extreme point χ_q^* is nondegenerate by Assumption 1 and characterized by a square nonsingular active-constraint submatrix $\mathbf{B}$ consisting of linearly independent rows of the stacked matrix $\left[\begin{array}{c|c} \mathbf{U} & -\mathbf{1} \\ \hline I & \mathbf{0} \end{array} \right]$, where $\mathbf{1}$ is the identity matrix. As illustrated in Chapter 4 of Bertsimas and Tsitsiklis (1997), given $\mathbf{B}$, we have

$$\left[\begin{array}{c} \chi_q^* \\ \hline s_q^* \end{array} \right] = \mathbf{B}^{-1} \left[\begin{array}{c} \mathbf{0} \\ \hline q \end{array} \right], \tag{14}$$

where s_q^* is the vector of optimal slack variables in $\mathcal{U}_q^S$. A corresponding solution of $\mathcal{V}_q$ is given by

$$\left[\begin{array}{c} v_q^* \\ \hline \lambda_q^* \end{array} \right] = \mathbf{u}_0 \mathbf{B}^{-1}. \tag{15}$$

It follows that as long as the optimal solutions of $\mathcal{U}_q$ and $\mathcal{V}_q$ are defined by the same extreme point defined by $\mathbf{B}$, χ_q^* varies with q, but (v_q^*, λ_q^*) does not.

Proof of Lemma 2. Fix an optimal solution (v_q^*, b_q^*, ℓ_q^*) of $\mathcal{V}_q$. For every q, $\omega \in \Omega$, and $x(\cdot|\omega) \in CE(\Gamma_\omega)$, by (4) we have

$$\begin{aligned}
v_q^*(\omega) &\geq \sum_{a \in A} u_0(a,\omega)x(a|\omega) + \sum_{a \in A} t(a,\omega)x(a|\omega) \\
&= \sum_{a \in A} u_0(a,\omega)x(a|\omega) \\
&\quad + \sum_{a \in A} \left\{ \sum_{i \in I} b_i^*(a_i,\omega_i) \sum_{\hat{a}_i \in A_i} \left(u_i(a_i,a_{-i},\omega) - u_i(\hat{a}_i,a_{-i},\omega)\right) \ell_i^*(\hat{a}_i|a_i,\omega_i) \right\} x(a|\omega) \\
&= \sum_{a \in A} u_0(a,\omega)x(a|\omega) \\
&\quad + \sum_{i \in I} \sum_{a_i,\hat{a}_i \in A_i} b_i^*(a_i,\omega_i)\ell_i^*(\hat{a}_i|a_i,\omega_i) \left\{ \sum_{a_{-i} \in A_{-i}} \left(u_i(a_i,a_{-i},\omega) - u_i(\hat{a}_i,a_{-i},\omega)\right) x(a|\omega) \right\} \\
&\geq \sum_{a \in A} u_0(a,\omega)x(a|\omega),
\end{aligned}$$

where the last inequality follows because any $x(\cdot|\omega) \in CE(\Gamma_\omega)$ is defined by the property that, for all $i \in I$ and $a_i, a_i' \in A_i$,

$$\sum_{a_{-i} \in A_{-i}} \left(u_i(a_i,a_{-i},\omega) - u_i(a_i',a_{-i},\omega)\right) x(a_i,a_{-i}|\omega) \geq 0.$$

Since $x(\cdot|\omega)$ is an arbitrary element of $CE(\Gamma_\omega)$, we conclude that $v_q^*(\omega) \geq \bar{u}(\omega)$. $\square$

Proof of Proposition 1. By complementary slackness, $x_q^*(a,\omega) > 0$ implies $v_q^*(\omega) = u_0(a,\omega) + t_q^*(a,\omega)$. Hence,

$$v_q^*(\omega) = \sum_{a \in A} u_0(a,\omega)x_q^*(a|\omega) + \sum_{a \in A} t_q^*(a,\omega)x_q^*(a|\omega) = u_q^*(\omega) + t_q^*(\omega).$$

Suppose we start from database q, with $q(\omega) > 0$, and we increase the quantity of ω-datapoints from $q(\omega)$ to $\hat{q}(\omega)$, thus obtaining the database $\hat{q}$. We can write

$$U^*(\hat{q}) - U^*(q) = u_{\hat{q}}^*(\omega)[\hat{q}(\omega) - q(\omega)] + \sum_{\omega' \in \Omega} [u_{\hat{q}}^*(\omega') - u_q^*(\omega')]\hat{q}(\omega')$$

Dividing both sides by $\hat{q}(\omega) - q(\omega)$, taking limits as $\hat{q}(\omega) \to q(\omega)$, and using Lemma 1, we obtain that

$$\begin{aligned}
t_q^*(\omega) &= v_q^*(\omega) - u_q^*(\omega) = \frac{\partial U^*(q)}{\partial q(\omega)} - u_q^*(\omega) \\
&= \lim_{\hat{q}(\omega) \to q(\omega)} \frac{\sum_{\omega' \in \Omega} [u_{\hat{q}}^*(\omega') - u_q^*(\omega')]\hat{q}(\omega')}{\hat{q}(\omega) - q(\omega)} = \sum_{\omega' \in \Omega} \frac{\partial u_q^*(\omega')}{\partial q(\omega)} q(\omega')
\end{aligned}$$

$$= \sum_{\omega' \in \Omega, a \in A} u_0(a, \omega') \left(\lim_{\hat{q}(\omega) \to q(\omega)} \frac{[x^*_{\hat{q}}(a|\omega') - x^*_q(a|\omega')]}{\hat{q}(\omega) - q(\omega)} \right) \hat{q}(\omega') =$$
$$= \sum_{\omega' \in \Omega, a \in A} u_0(a, \omega') \frac{\partial x^*_q(a|\omega')}{\partial q(\omega)} q(\omega'),$$

where the existence of the derivative $\frac{\partial x^*_q(a|\omega')}{\partial q(\omega)}$ almost everywhere follows from (14). $\square$

Proof Proposition 2 First, note that we can write $u_0(a_1, \omega) = \pi a_1 \mathbb{I}\{\omega \geq a_1\} + (1 - \pi) \max\{\omega - a_1, 0\}$ as

$$[a(2\pi - 1) + (1 - \pi)\omega] \mathbb{I}\{\omega \geq a\},$$

which is strictly increasing in a if and only if $\pi > \frac{1}{2}$. Let $\bar{x}^*$ be the profit-maximizing solution (i.e., for $\pi = 1$) and $\underline{x}^*$ be the surplus-maximizing solution (i.e., for $\pi = 0$).

Lemma 3. *$\bar{x}^*$ is optimal for all $\pi \geq \frac{1}{2}$ and $\underline{x}^*$ is optimal for all $\pi \leq \frac{1}{2}$.*

Proof. Fix any (non-trivial) q and $\pi \in (0, 1)$. Problem $\mathcal{U}_q$ involves maximizing

$$\sum_{\omega, a} u_\pi(a, \omega) x(a|\omega) q(\omega) = \sum_{\omega \geq a} [a(2\pi - 1) + (1 - \pi)\omega] x(a|\omega) q(\omega)$$

subject to constraints (1).

Suppose that $\pi > \frac{1}{2}$. Note that $\bar{x}^*$ is feasible and maximizes the objective function pointwise for every ω. Indeed, since $\bar{x}^*(\omega|\omega) = 1$, for every ω we have that $\bar{x}^*$ selects the highest $a \leq \omega$ for every ω, thereby maximizing $a(2\pi - 1)\mathbb{I}\{\omega \geq a\}$; it also maximizes $\sum_{a \leq \omega} \omega x(a|\omega)$ for every ω. We can invoke the Theorem of the Maximum to extend the optimality of $\bar{x}^*$ at $\pi = \frac{1}{2}$.

Suppose now that $\pi < \frac{1}{2}$. Now for each ω the objective is to pair ω with the smallest possible a and do so with the highest probability allowed by (1). This is what $\underline{x}^*$ essentially does. We can again invoke the Theorem of the Maximum to extend the optimality of $\underline{x}^*$ at $\pi = \frac{1}{2}$. $\square$

We now derive the expression of $v^*_q(\omega)$ in the statement of the proposition. The case of $\pi \geq \frac{1}{2}$ follows immediately from the fact that $\bar{x}^*$ is full disclosure. Now suppose $\pi < \frac{1}{2}$. We will construct a candidate v^*_q and prove it solves $\mathcal{V}_q$ using strong duality. First, under $\underline{x}^*$ we have

$$\begin{aligned} U^*(q) &= \sum_{\omega, a} [\pi u_1(a, \omega) + (1 - \pi) u_0(a, \omega)] \underline{x}^*(a|\omega) q(\omega) \\ &= \pi \sum_{\omega, a} a \mathbb{I}\{\omega \geq a\} \underline{x}^*(a|\omega) q(\omega) \\ &\quad + (1 - \pi) \left[\sum_{\omega < a_q} \omega q(\omega) + \sum_{\omega \geq a_q} (\omega - a_q) q(\omega) \right]. \end{aligned}$$

Note that

$$\sum_{\omega,a} a\mathbb{I}\{\omega \geq a\}\underline{x}^*(a|\omega)q(\omega) = a_q \sum_{\omega \geq a_q} q(\omega),$$

because the left-hand side is the seller's expected profits under $\underline{x}^*$, which by construction equal to the expected profit from the fixed uninformed price a_q. Therefore, we can write

$$\begin{aligned} U^*(q) &= \pi a_q \sum_{\omega \geq a_q} q(\omega) + (1-\pi)\left[\sum_{\omega < a_q} \omega q(\omega) + \sum_{\omega \geq a_q} (\omega - a_q)q(\omega)\right] \\ &= (2\pi - 1)a_q \sum_{\omega \geq a_q} q(\omega) + (1-\pi)\sum_{\omega} \omega q(\omega). \end{aligned}$$

Now we construct (v_q^*, λ_q^*), we show that it satisfies all dual constraints and that it yields $\sum_\omega v_q^*(\omega)q(\omega) = U^*(q)$, which proves that (v_q^*, λ_q^*) is optimal by strong duality. Recall that, in general, for all (a, ω) the dual constraint reads as

$$v(\omega) \geq u_\pi(a,\omega) + \sum_{a'}[u_1(a,\omega) - u_1(a',\omega)]\lambda(a'|a).$$

Let $\lambda_q^*(a'|a) = 0$ for all $a' \neq a_q$. Let $\lambda_q^*(a_q|a) = 1 - 2\pi$ for all $a \in \operatorname{supp} \underline{x}(\cdot|\omega)$ for some ω and $\lambda_q^*(a_q|a) = 0$ otherwise. Given this, for $\omega < a_q$, the right-hand side of the dual constraint equals

$$\begin{cases} \pi a + (1-\pi)(\omega - a) + a\lambda_q^*(a_q|a) & \text{if } a \leq \omega \\ 0 & \text{if } a > \omega. \end{cases}$$

Given $\lambda_q^*(a_q|a)$, the first line always equals $(1-\pi)\omega > 0$. Therefore, for $\omega < a_q$ define

$$v_q^*(\omega) = (1-\pi)\omega.$$

For $\omega \geq a_q$, the right-hand side of the dual constraint equals

$$\begin{cases} \pi a + (1-\pi)(\omega - a) + (a - a_q)\lambda_q^*(a_q|a) & \text{if } a \leq \omega \\ -a_q\lambda_q^*(a_q|a) & \text{if } a > \omega. \end{cases}$$

Given $\lambda_q^*(a_q|a)$, the first line always equals

$$(2\pi - 1)a_q + (1-\pi)\omega = \pi a_q + (1-\pi)(\omega - a_q) > 0.$$

Therefore, for $\omega \geq a_q$ define

$$v_q^*(\omega) = (2\pi - 1)a_q + (1-\pi)\omega.$$

Note that by construction v_q^* satisfies all dual constraint and $\sum_\omega v_q^*(\omega)q(\omega) = U^*(q)$, as desired.

It follows immediately that for $\pi < \frac{1}{2}$ we have $t_q^*(\omega) > 0$ for $\omega < a_q$ and $t_q^*(\omega) \leq 0$ for $\omega \geq a_q$. $\square$

Proof Proposition 3. By the formulation of $\mathcal{V}_q$ and Lemma 2, the polyhedron of feasible solutions of $\mathcal{V}_q$, denoted by $F(\mathcal{V}_q)$ does not contain a line because all dual variables are bounded from below. By Theorem 2.6 in Bertsimas and Tsitsiklis (1997), $F(\mathcal{V}_q)$ has at least one extreme point and at most finitely many of them by Corollary 2.1 in Bertsimas and Tsitsiklis (1997). By Theorem 4.4 in Bertsimas and Tsitsiklis (1997), $\mathcal{V}_q$ has at least one optimal solution. By Theorem 2.7 in Bertsimas and Tsitsiklis (1997), we can focus on solutions that are extreme points of $F(\mathcal{V}_q)$.

Fix q and suppose that the optimal solution (v_q^*, λ_q^*) of the dual of $\mathcal{U}_q$ is unique. As explained in Remark 1, there exists a submatrix $\mathbf{B}$ such that (v_q^*, λ_q^*) satisfies (15). Given Assumption 1, Theorem 3.1 and Exercise 3.6 in Bertsimas and Tsitsiklis (1997) imply that

$$\left[\begin{array}{c|c} \mathbf{U} & -\mathbf{1} \\ \hline I & \mathbf{0} \end{array}\right] \mathbf{B}^{-1} \left[\frac{\mathbf{0}}{q}\right] \geq \left[\frac{\mathbf{0}}{q}\right].$$

The inequality is strict for each row of $\mathbf{U}$ that corresponds to $\lambda_{q,i}^*(a_i'|a_i,\omega_i) = 0$ (or, equivalently, $\ell_{q,i}^*(a_i'|a_i,\omega_i) = 0$):

$$\left[\mathbf{U}_i(a_i,a_i',\omega_i) \mid -\mathbf{1}_i(a_i,a_i',\omega_i)\right] \mathbf{B}^{-1} \left[\frac{\mathbf{0}}{q}\right] > 0, \tag{16}$$

where $\mathbf{1}_i(a_i,a_i',\omega_i)$ is the row of the identity matrix $\mathbf{1}$ that corresponds to (i,a_i,a_i',ω_i). Note that for each row ω of the indicator matrix I (i.e., $I(\omega)$), which corresponds to variable $v_q^*(\omega)$, it automatically holds that $\left[I(\omega) \mid \mathbf{0}\right] \mathbf{B}^{-1} \left[\frac{\mathbf{0}}{q}\right] = q(\omega)$. Similarly, for each row of $\mathbf{U}$ that corresponds to $\lambda_{q,i}^*(a_i'|a_i,\omega_i) > 0$ (or, equivalently, $\ell_{q,i}^*(a_i'|a_i,\omega_i) > 0$), it holds that $\left[\mathbf{U}_i(a_i,a_i',\omega_i) \mid -\mathbf{1}_i(a_i,a_i',\omega_i)\right] \mathbf{B}^{-1} \left[\frac{\mathbf{0}}{q}\right] = 0$ as long as $\mathbf{B}$ identifies the optimal extreme point.

Now consider changes in q and note that it only enters the objective of $\mathcal{V}_q$. Each condition (16) defines an open set of q's in $\mathbb{R}_+^\Omega$ that satisfy it. Define $(v_\mathbf{B}^*, \lambda_\mathbf{B}^*)$ identified by $\mathbf{B}$ as in (15) and

$$Q(\mathbf{B}) = \{q : (16) \text{ holds for all } i \in I \text{ and } (a_i,a_i',\omega_i) \text{ s.t. } \lambda_{\mathbf{B},i}^*(a_i|a_i',\omega_i) = 0\}.$$

Note that $Q(\mathbf{B})$ is an open set because it is the intersection of finitely many open sets.

Now recall that there are only finitely many extreme points of the dual polyhedron of feasible solutions. Therefore, there are finitely many submatrices $\{\mathbf{B}_1, \ldots, \mathbf{B}_K\}$ such that each identifies an optimal $(v^*_{\mathbf{B}_k}, \lambda^*_{\mathbf{B}_k})$ that is unique for all $q \in Q(\mathbf{B}_k)$. For all $k = 1, \ldots, K$, define $Q_k = Q(\mathbf{B}_k)$. By construction, each Q_k is open and $q, q' \in Q_k$ implies that $(v^*_q, \lambda^*_q) = (v^*_{q'}, \lambda^*_{q'})$. Since (v^*_q, λ^*_q) is generically unique with respect to q, it follows that $\mathbb{R}^\Omega_+ \setminus \cup_k Q_k$ has Lebesgue measure zero. $\square$

Proof of Proposition 4. Fix $\mu_1, \mu_2 \in \Delta(\Omega)$. Let $\Omega^i = \{\omega \in \Omega : \mu_i(\omega) > \mu_j(\omega), j \neq i\}$, $i \in \{1, 2\}$, and $\Omega^3 = \Omega \setminus \Omega_1 \setminus \Omega_2$.

Let $X = \mathbb{R}^\Omega \times \mathbb{R}^{A_1 \times A_1}_+ \times \ldots \times \mathbb{R}^{A_n \times A_n}_+$. Associate the canonical component-wise order with X, with an exception that the order is reversed for $\omega \in \Omega^1$. X is a lattice, with a typical element (v, λ), where $v \in \mathbb{R}^\Omega$ and $\lambda \in \mathbb{R}^{A_1 \times A_1}_+ \times \ldots \times \mathbb{R}^{A_n \times A_n}_+$.

The data-value problem is equivalent to the problem $\max_{(v,\lambda) \in S} f(v, \lambda; \mu)$, where $f(v, \lambda; \mu) = -\sum_{\omega \in \Omega} v(\omega)\mu(\omega)$ and the feasible set $S \subset X$ is given by the inequalities

$$v(\omega) \geq u_0(a, \omega) + \sum_{i \in I} \sum_{a'_i \in A_i} (u_i(a_i, a_{-i}, \omega) - u_i(a'_i, a_{-i}, \omega))\lambda_i(a'_i | a_i).$$

We treat μ as a parameter. Note that S does not depend on μ. Furthermore, μ is an element of $(|\Omega| - 1)$-dimensional simplex, with which we associate the following partial order: $\mu' \geq \mu$ if $\mu'(\omega) \geq \mu(\omega)$ for $\omega \in \Omega^1$, $\mu'(\omega) \leq \mu(\omega)$ for $\omega \in \Omega^2$, and $\mu'(\omega) = \mu(\omega)$ for $\omega \in \Omega^3$. Note that $\mu_1 \geq \mu_2$ in accordance with this partial order.

We want to show that f is supermodular in (v, λ) and has increasing differences in $(v, \lambda; \mu)$. Observe that

$$\begin{aligned}
f(v', \lambda'; \mu) + f(v'', \lambda''; \mu) &= -\sum_{\omega \in \Omega} v'(\omega)\mu(\omega) - \sum_{\omega \in \Omega} v''(\omega)\mu(\omega) \\
&= -\sum_{\omega \in \Omega} (v'(\omega) + v''(\omega))\mu(\omega) \\
&= -\sum_{\omega \in \Omega} (\max\{v'(\omega), v''(\omega)\} + \min\{v'(\omega), v''(\omega)\})\mu(\omega) \\
&= f((v', \lambda') \wedge (v'', \lambda''); \mu) + f((v', \lambda') \vee (v'', \lambda''); \mu).
\end{aligned}$$

Then f is supermodular in (v, λ).

Fix $(v', \lambda') \geq (v, \lambda)$ and $\mu' \geq \mu$. Observe that

$$\begin{aligned}
& (f(v', \lambda', \mu') - f(v, \lambda, \mu')) - (f(v', \lambda', \mu) - f(v, \lambda, \mu)) \\
= & \sum_{\omega \in \Omega} (v(\omega) - v'(\omega))(\mu'(\omega) - \mu(\omega)) \\
= & \sum_{\omega \in \Omega^1} (v(\omega) - v'(\omega))(\mu'(\omega) - \mu(\omega)) + \sum_{\omega \in \Omega^2} (v(\omega) - v'(\omega))(\mu'(\omega) - \mu(\omega)) \geq 0,
\end{aligned}$$

where the inequality follows from the adapted partial orders. Then f has increasing differences in $(v, \lambda; \mu)$.

Finally, by Theorem 5 in Milgrom and Shannon (1994), $\arg\max_{(v,\lambda)\in S} f(v, \lambda; \mu)$ is monotone nondecreasing in μ. This monotone comparative statics coupled with generic uniqueness of (v_q^*, b_q^*, ℓ_q^*) with respect to q imply that if $\mu_q(\omega) > \mu_{q'}(\omega)$ for two databases q and q' then $v_q^*(\omega) \leq v_{q'}^*(\omega)$.

When only interactions of type ω are present in the database, that is, $\mu_q(\omega) = 1$, we have $v_q^*(\omega) = \bar{u}(\omega)$. Indeed, the definition of $\bar{u}(\omega)$ implies that it can be written as

$$\bar{u}(\omega) = \min_{b_\omega, \ell_\omega} \max_{a \in A} \left\{ u_0(a, \omega) + t_{b_\omega, \ell_\omega}(a, \omega) \right\},$$

where $t_{b_\omega,\ell_\omega}(a, \omega) = \sum_{i\in I} b_{i,\omega}(a_i) \sum_{a_i' \in A_i} (u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega))\ell_{i,\omega}(a_i'|a_i)$, $b_\omega = (b_{1,\omega}, \ldots, b_{n,\omega})$, with $b_{i,\omega} : A_i \to \mathbb{R}_{++}$, and $\ell_\omega = (\ell_{1,\omega}, \ldots, \ell_{n,\omega})$, with $\ell_{i,\omega} : A_i \to \Delta(A_i)$.

For $\varepsilon > 0$, consider a set $M_\varepsilon(\omega)$ defined as $M_\varepsilon(\omega) = \{\mu \in \Delta(\Omega) : \mu(\omega') \in (0, \varepsilon) \text{ for } \omega \neq \omega', \mu(\omega) < 1\}$. By Proposition 3, there exists a finite collection $\{\mathcal{P}_1, \ldots, \mathcal{P}_K\}$ of open, convex, and disjoint subsets of $\Delta(\Omega)$ such that $\cup_k \mathcal{P}_k$ has measure one and, for every k, (v_q^*, b_q^*, ℓ_q^*) is unique and constant for q, with $\mu_q \in \mathcal{P}_k$. Therefore, we can always find $\mathcal{P}_m \in \{\mathcal{P}_1, \ldots, \mathcal{P}_K\}$, such that $\mathcal{P}_m \cap M_\varepsilon(\omega)$ is nonempty, open, and convex for all $0 < \varepsilon \leq \delta$, where $\delta > 0$. Then $v_q^*(\omega)$ is unique and constant for all $q \in \mathbb{R}_{++}^\Omega$, with $\mu_q \in \mathcal{P}_m \cap M_\delta(\omega)$. Let us refer to this constant as $\hat{u}(\omega)$. If $\hat{u}(\omega) = \bar{u}(\omega)$, then the result follows. Suppose, on the contrary, that $\hat{u}(\omega) \neq \bar{u}(\omega)$. We can always pick a sequence μ^n, $n \in \mathbb{N}$, from $\mathcal{P}_m \cap M_\delta(\omega)$ that converges to $\tilde{\mu}$, with $\tilde{\mu}(\omega) = 1$. Then for every $n \in \mathbb{N}$, $v_q^*(\omega) = \hat{u}(\omega)$ for every q, such that $\mu_q = \mu^n$. By the Berge's maximum theorem, (v_q^*, b_q^*, ℓ_q^*) is an upper-hemicontinuous correspondence and therefore has closed graph. Hence, $\hat{u}(\omega) \in v_q^*(\omega)$ for every q, with $\mu_q = \tilde{\mu}$. We obtain the desired contradiction, since $v_q^*(\omega) = \bar{u}(\omega)$ for such q. □

Proof of Proposition 5. Fix $q \in \mathbb{R}_{++}^\Omega$. Suppose that a FIC mechanism x_q^* is optimal. Then, we have

$$v_q^*(\omega) = u_q^*(\omega) + \sum_{a\in A} t_q^*(a, \omega) x_q^*(a|\omega) \geq u_q^*(\omega),$$

where the inequality follows from $x_q^*(\cdot|\omega) \in CE(\Gamma_\omega)$ for all ω. Since by Lemma 1 we must have $\sum_\omega v_q^*(\omega)q(\omega) = \sum_\omega u_q^*(\omega)q(\omega)$, it follows that $v_q^*(\omega) = u_q^*(\omega)$ for all ω. Finally, since x_q^* is optimal, it must be that $u_q^*(\omega) = \bar{u}(\omega)$ for all ω. Now, note that v_q^* defines a supporting hyperplane of the iso-payoff line of level $U^*(q)$ at q. The intercept of such an hyperplane on each ω-axis is $\hat{q}_\omega(\omega) = \frac{U^*(q)}{\bar{u}(\omega)}$ and $\hat{q}_\omega(\omega') = 0$ for $\omega' \neq \omega$. By definition, each $\hat{q}_\omega$ also belongs to the iso-payoff line of level $U^*(q)$ and therefore $U^*(q) = U^*(\hat{q}_\omega)$ for all ω. In other words, the intercepts of the hyperplane and the iso-payoff line coincide for all ω.

Now consider any $q' \in \mathbb{R}_{++}$, $q' \neq q$, that belongs to the supporting hyperplane of level $U^*(q)$ at q. By definition, we can obtain q' as a convex combination of intercepts $\hat{q}_\omega$ on each axis. Specifically, there exists $\beta \in \Delta(\Omega)$ such that $q'(\omega) = \beta(\omega)\hat{q}_\omega(\omega)$ for all ω. By concavity of $U^*(q)$ (Footnote 18), we must have that

$$U^*(q') = \sum_{\omega\in\Omega} v^*_{q'}(\omega)q'(\omega) \leq U^*(q) = \sum_{\omega\in\Omega} \beta(\omega)U^*(\hat{q}_\omega) = \sum_{\omega\in\Omega} \bar{u}(\omega)q'(\omega).$$

But since $v^*_{q'}(\omega) \geq \bar{u}(\omega)$ for all ω by Lemma 2, we must have $v^*_{q'}(\omega) = \bar{u}(\omega)$ for all ω. Then $v^*_{q''}(\omega) = \bar{u}(\omega)$ for all q'' that belong to the supporting hyperplane of level $U^*(q)$ at q. Finally, since v^*_q is invariant to scaling of q, it follows that $v^*_q(\omega) = \bar{u}(\omega)$ for all ω and all $q \in \mathbb{R}^\Omega_+$. □

Proof of Corollary 4. We have $MRS_q(\omega,\omega') = -\frac{\omega}{\omega'}$ for $\pi > \frac{1}{2}$ and all ω,ω'. Consider now $\pi < \frac{1}{2}$:

$$MRS_q(\omega,\omega') = \begin{cases} -\frac{\omega}{\omega'} & \text{if } \omega,\omega' < a_q \\ -\frac{(1-\pi)\omega}{\pi a_q+(1-\pi)(\omega'-a_q)} & \text{if } \omega < a_q \leq \omega' \\ -\frac{\pi a_q+(1-\pi)(\omega-a_q)}{\pi a_q+(1-\pi)(\omega'-a_q)} & \text{if } \omega,\omega' \geq a_q. \end{cases}$$

Thus, we have

$$\frac{\partial MRS_q(\omega,\omega')}{\partial \pi} = \begin{cases} 0 & \text{if } \omega,\omega' < a_q \\ \frac{\omega a_q}{[\pi a_q+(1-\pi)(\omega'-a_q)]^2} & \text{if } \omega < a_q \leq \omega' \\ -\frac{a_q(\omega'-\omega)}{[\pi a_q+(1-\pi)(\omega'-a_q)]^2} & \text{if } \omega,\omega' \geq a_q. \end{cases}$$

Finally, it is easy to see that $MRS_q(\omega,\omega') < -\frac{\omega}{\omega'}$ for $\omega < a_q \leq \omega'$ and that $MRS_q(\omega,\omega') > -\frac{\omega}{\omega'}$ for $\omega' > \omega \geq a_q$. □

Proof of Corollary 5. Fix ω, q, and a refinement σ_ω. Since $u_i(a,\omega) = \mathbb{E}_{\sigma_\omega}[u_i(a,\omega')|\omega]$ for all i, by (4) we have

$$\begin{aligned} v^*_q(\omega) &= \max_{a\in A} \sum_{\omega'\in\Omega} [u_0(a,\omega') + t^*_q(a,\omega')]\sigma_\omega(\omega') \\ &\leq \sum_{\omega'\in\Omega} \max_{a\in A}[u_0(a,\omega') + t^*_q(a,\omega')]\sigma_\omega(\omega') = \sum_{\omega'\in\Omega} v^*_q(\omega')\sigma_\omega(\omega'). \end{aligned} \tag{17}$$

Thus, if refining $\alpha q(\omega)$ of the original records of type ω according to σ_ω does not change the value of any record, then (17) implies the desired inequality. Now suppose that acquiring better data changes the value of datapoints: There exists a share $\alpha > 0$ such that refining $\alpha q(\omega)$ of the current ω-datapoints according to σ_ω leads to a new database q_α such that $v^*_{q_\alpha}(\omega') \neq v^*_q(\omega')$ for some $\omega' \in \operatorname{supp}\sigma_\omega$ or $\omega' = \omega$. Since the total quantity of datapoints does not change, we have that $\mu_{q_\alpha}(\omega) < \mu_q(\omega)$ and $\mu_{q_\alpha}(\omega') > \mu_q(\omega')$ for all $\omega' \in \operatorname{supp} \sigma_\omega$. By Proposition 4, it follows that $v^*_{q_\alpha}(\omega) \geq v^*_q(\omega)$ and $v^*_{q_\alpha}(\omega') \leq v^*_q(\omega')$ for all $\omega' \in \operatorname{supp} \sigma_\omega$. Now, note that for all α,

$$\sum_{\omega' \in \Omega} v^*_{q_\alpha}(\omega')\sigma_\omega(\omega') \geq v^*_{q_\alpha}(\omega) \geq v^*_q(\omega), \tag{18}$$

where the first inequality follows from (17). This implies that the value of acquiring better data is always non-negative.

Now, suppose that there exists a common $\tilde{a} \in \operatorname{supp} x^*_q(\cdot|\omega)$ that satisfies $x^*_q(\tilde{a}|\omega'') > 0$ for all $\omega'' \in \operatorname{supp} \sigma_\omega$. By complementary slackness, it follows that for all $\omega'' \in \operatorname{supp} \sigma_\omega$, we have $v^*_q(\omega'') = u_0(\tilde{a}, \omega'') + t^*_q(\tilde{a}, \omega'')$. Therefore, by the scarcity principle,

$$\sum_{\omega'' \in \Omega} v^*_{q_\alpha}(\omega'')\sigma_\omega(\omega'') \leq \sum_{\omega'' \in \Omega} v^*_q(\omega'')\sigma_\omega(\omega'') = v^*_q(\omega) \leq v^*_{q_\alpha}(\omega),$$

which, combined with (18), implies the desired equality.

Conversely, suppose that for every $\hat{a} \in \operatorname{supp} x^*_q(\cdot|\omega)$ there exists $\omega' \in \operatorname{supp} \sigma_\omega$ that satisfies $x^*_q(\hat{a}|\omega') = 0$. If the solution to the data-value problem is unique for database q, then $x^*_q(\hat{a}|\omega') = 0$ implies $v^*_q(\omega') > u_0(\hat{a}, \omega') + t^*_q(\hat{a}, \omega')$ by strict complementary slackness. The desired strict inequality is then obtained.

Proof of Proposition 6. The directional derivative of U^* at q along the linear path from q to q_α is equal to

$$q(\omega)\left[\sum_{\omega' \in \Omega} v^*_q(\omega')\sigma_\omega(\omega') - v^*_q(\omega)\right].$$

The linear path from q to q_α can be parametrized as follows: for $t \in [0, 1]$, define $q_t(\omega) = q(\omega) - t\alpha q(\omega)$, $q_t(\omega') = q(\omega') + t\alpha\sigma_\omega(\omega')q(\omega)$ for $\omega' \in \operatorname{supp} \sigma_\omega$, and $q_t(\omega'') = q(\omega'')$ for remaining ω''.

Note that $\sum_{\omega' \in \Omega} v^*_{q_t}(\omega')\sigma_\omega(\omega') - v^*_{q_t}(\omega)$ is non-negative by (17) and decreasing in t by the scarcity principle.

Finally, by the gradient theorem,

$$U^*(q_\alpha) - U^*(q) = \int_0^1 v^*_{q_t} \cdot \nabla q_t \, dt = \alpha q(\omega) \int_0^1 \left[\sum_{\omega' \in \Omega} v^*_{q_t}(\omega')\sigma_\omega(\omega') - v^*_{q_t}(\omega)\right] dt \geq 0,$$

where ∇q_t is the gradient of q_t with respect to t. □

Proof of Proposition 7 . We provide a proof for the general case where the principal can choose $a_0 \in A_0$ and each agent i can privately observe some own data $\omega_i \in \Omega_i$ about the interaction he is in. Fix $(a^*, \omega^*) \in \mathbf{X}$ and introduce $\mathbf{1}_{a^*,\omega^*}$ as a vector of size $|\mathbf{X}|$ with $\varepsilon > 0$ in the position indexed by (a^*, ω^*) and 0 in all other positions. Constitute a matrix $\mathbf{W}$ such that its rows are indexed by $(a, \omega) \in \mathbf{X}$, its columns are indexed by (i, a_i', a_i, ω_i), $i \in I$, and its entries are as follows:

$$\mathbf{W}\left((\tilde{a}, \tilde{\omega}), (i, a_i', a_i, \omega_i)\right) = 1\left\{a_i = \tilde{a}_i, \omega_i = \tilde{\omega}_i\right\}\left(u_i(a_i, \tilde{a}_{-i}, \omega_i, \tilde{\omega}_{-i}) - u_i(a_i', \tilde{a}_{-i}, \omega_i, \tilde{\omega}_{-i})\right).$$

By a variant of the Farkas' lemma, either there exists $\lambda \geq 0$, such that $\mathbf{W}\lambda \leq -\mathbf{1}_{a^*,\omega^*}$, or else there exists $\chi \geq 0$, such that $\mathbf{W}^T\chi \geq 0$, with $\chi^T\mathbf{1}_{a^*,\omega^*} > 0$. Now we show that the latter is true. Indeed, we can pick $\chi(a, \omega) = q(\omega)x(a|\omega)$, where x is obedient and satisfies $x(a^*|\omega^*) > 0$. We can find such x, since $(a^*, \omega^*) \in \mathbf{X}$. Then $\chi \geq 0$ and $\chi^T\mathbf{1}_{a^*,\omega^*} > 0$ are satisfied automatically. Finally, $\mathbf{W}^T\chi \geq 0$ corresponds exactly to the set of obedience constraints in $\mathcal{U}_q$ restricted to the subdomain $\mathbf{X}$.

Since any λ can be decomposed as $\lambda_i(a_i'|a_i, \omega_i) = b_i(a_i, \omega_i)\ell_i(a_i'|a_i, \omega_i)$, we conclude that there is no $(b, \ell) \in G(\mathbf{X})$ that satisfies $t(a, \omega) \leq 0$ for every $(a, \omega) \in \mathbf{X}$ and $t(a^*, \omega^*) < -\varepsilon$. The result then follows, since the choice of $(a^*, \omega^*) \in \mathbf{X}$ and $\varepsilon > 0$ was arbitrary. □

Proof of Proposition 8. We will argue by contradiction. Suppose $q \in \mathbb{R}^{\Omega}_{++}$ and $\mathcal{U}_q$ admits an FIC solution x_q^{**} and hence $x_q^{**}(\cdot|\tilde{\omega}) \in CE(\Gamma_{\tilde{\omega}})$ and $u_q^{**}(\tilde{\omega}) = \bar{u}(\tilde{\omega})$ for all $\tilde{\omega} \in \Omega$. Then $v_q^{**}(\tilde{\omega}) = u_q^{**}(\tilde{\omega}) = \bar{u}(\tilde{\omega})$ for all $\tilde{\omega} \in \Omega$ by Proposition 5.

Now suppose that (a, ω) satisfies both conditions in the statement of the proposition. For $(v_q^{**}, b_q^{**}, \ell_q^{**})$ to be feasible for $\mathcal{V}_q$, we must have for all $\tilde{\omega} \in \Omega$,

$$v_q^{**}(\tilde{\omega}) \geq u_0(a, \tilde{\omega}) + t_q^{**}(a, \tilde{\omega}).$$

Since $u_0(a, \omega) > \bar{u}(\omega) = v^{**}(\omega)$, we must have $t_q^{**}(a, \omega) < 0$. Therefore, there exists a pair $(i, \hat{a}_i)$ that satisfies $u_i(a_i, a_{-i}, \omega) < u_i(\hat{a}_i, a_{-i}, \omega)$ and $\ell_{q,i}^{**}(\hat{a}_i|a_i, \omega_i) > 0$. For such a pair $(i, \hat{a}_i)$, there exists $x(\cdot|\omega') \in CE(\Gamma_{\omega'})$ with the properties listed in the proposition. Then, since $b_q^{**} > 0$ and $\ell_{q,i}^{**}(\hat{a}_i|a_i, \omega_i) > 0$,

$$\begin{aligned}
& \sum_{\tilde{a} \in A} u_0(\tilde{a}, \omega')x(\tilde{a}|\omega') + \sum_{\tilde{a} \in A} t_q^{**}(\tilde{a}, \omega')x(\tilde{a}|\omega') \\
\geq \ & \sum_{\tilde{a} \in A} u_0(\tilde{a}, \omega')x(\tilde{a}|\omega') \\
& + b_{q,i}^{**}(a_i, \omega_i)\ell_{q,i}^{**}(\hat{a}_i|a_i, \omega_i)\left\{\sum_{\tilde{a}_{-i} \in A_{-i}} \left(u_i(a_i, \tilde{a}_{-i}, \omega') - u_i(\hat{a}_i, \tilde{a}_{-i}, \omega')\right)x(a_i, \tilde{a}_{-i}|\omega')\right\} \\
> \ & \sum_{\tilde{a} \in A} u_0(\tilde{a}, \omega')x(\tilde{a}|\omega') = v_q^{**}(\omega'),
\end{aligned}$$

where the first inequality follows because $x(\cdot|\omega') \in CE(\Gamma_{\omega'})$. The strict inequality is incompatible with constraint (4) and delivers the desired contradiction. □

C A Sufficient Condition for Suboptimality of Full Disclosure

We provide a sufficient condition on Γ for suboptimality of full disclosure for the general case where the principal can choose $a_0 \in A_0$ and each agent i can privately observe some own data $\omega_i \in \Omega_i$ about the interaction he is in. Recall that if the principal fully reveals all ω, then she must be implementing a correlated equilibrium of the complete-information game Γ_ω for all ω, i.e., $x_q^*(\cdot|\omega) \in CE(\Gamma_\omega)$. The definition of CE in terms of inequalities can be adjusted to incorporate the principal's a_0.

Proposition 8. *Fix Γ. Suppose there exists (a, ω) that satisfies:*

(1) $u_0(a, \omega) > \bar{u}(\omega)$,

(2) for every agent i and action $\hat{a}_i$, such that $u_i(a_i, a_{-i}, \omega) < u_i(\hat{a}_i, a_{-i}, \omega)$, there exists an $x(\cdot|\omega') \in CE(\Gamma_{\omega'})$ for some ω', with $\omega'_i = \omega_i$, that satisfies

$$\sum_{a \in A} u_0(a, \omega') x(a|\omega') = \bar{u}(\omega'),$$

$$\sum_{a_{-i} \in A_{-i}} \left(u_i(a_i, a_{-i}, \omega') - u_i(\hat{a}_i, a_{-i}, \omega')\right) x(a_i, a_{-i}|\omega') > 0.$$

Then $\mathcal{U}_q$ does not admit an FIC solution for any $q \in \mathbb{R}^{\Omega}_{++}$.

Condition (1) is clearly necessary: If for every datapoint ω every action profile a cannot deliver a payoff higher than the full-information payoff $\bar{u}(\omega)$, then it is clearly optimal for the principal to fully reveal every ω. Given an outcome (a, ω) with $u_0(a, \omega) > \bar{u}(\omega)$, there must be an agent who would have a profitable deviation from a_i to $\hat{a}_i$ if he knew (a_{-i}, ω_{-i}). Otherwise, given a_0, the profile a_{-0} is a Nash Equilibrium of Γ_ω and hence $a_{-0} \in CE(\Gamma_\omega)$, which would imply $u_0(a, \omega) \leq \bar{u}(\omega)$. Then condition (2) requires that agent i's data ω_i is consistent with another datapoint ω'—so that he cannot tell ω and ω' apart based on his own data only—which admits a principal-preferred correlated equilibrium that also recommends i to play a_i and renders the deviation to $\hat{a}_i$ strictly suboptimal.

D Analysis of the Leading Example

This section presents the calculations that back up our statements regarding the leading example. We can ignore the buyers and build their decisions into the utility functions of the surplus-maximizing platform ($i = 0$) and the seller ($i = 1$). There are three types of datapoints, labeled by $\omega \in \{\omega_L, \omega_H, \omega^\circ\}$, where $\omega_H > \omega_L > 0$, and corresponding to whether the buyer's revealed valuation is ω_L, ω_H, or unknown to the platform. Suppose ω° turns into ω_H with probability h and ω_L with probability $1-h$. The prices the seller can charge are $a \in \{\omega_L, \omega_H\}$. The payoffs are $u_0(a,\omega) = \max\{\omega - a, 0\}$, $u_1(a,\omega) = a$ if $a \leq \omega$, and $u_1(a,\omega) = 0$ if $a > \omega$. Given this, we have $u_i(a,\omega^\circ) = hu_i(a,\omega_H) + (1-h)u_i(a,\omega_L)$ for $i = 0,1$. For completeness, we solve both the information-design problem and the data-value problem separately. Since our goal here is to only find the optimizers in both problems, we can work in the space of databases that satisfy $q(\omega_L) + q(\omega_H) + q(\omega^\circ) = 1$, so that $q(\omega) = \mu_q(\omega)$.

D.1 Information-Design Problem

The objective of the platform is

$$(\omega_H - \omega_L)x(\omega_L|\omega_H)\mu_q(\omega_H) + h(\omega_H - \omega_L)x(\omega_L|\omega^\circ)\mu_q(\omega^\circ).$$

The obedience constraints are

$$-\omega_L x(\omega_H|\omega_L)\mu_q(\omega_L) + (\omega_H - \omega_L)x(\omega_H|\omega_H)\mu_q(\omega_H) + (h\omega_H - \omega_L)x(\omega_H|\omega^\circ)\mu_q(\omega^\circ) \geq 0,$$
$$\omega_L x(\omega_L|\omega_L)\mu_q(\omega_L) - (\omega_H - \omega_L)x(\omega_L|\omega_H)\mu_q(\omega_H) - (h\omega_H - \omega_L)x(\omega_L|\omega^\circ)\mu_q(\omega^\circ) \geq 0.$$

We consider two cases depending on whether $h\omega_H - \omega_L > 0$, or $h\omega_H - \omega_L \leq 0$. If $h\omega_H - \omega_L > 0$, or $h > \frac{\omega_L}{\omega_H}$, then from the second obedience constraint $x_q^*(\omega_L|\omega_L) = 1$. The first obedience constraint is then automatically satisfied. Since $h \in (0,1)$, it is always true that $\frac{h\omega_H - \omega_L}{\omega_H - \omega_L} < h$. Then the solution satisfies $x_q^*(\omega_L|\omega_H) = 0$ and $x_q^*(\omega_L|\omega^\circ) = \frac{\omega_L}{h\omega_H - \omega_L}\frac{\mu_q(\omega_L)}{\mu_q(\omega^\circ)}$, as long as $\frac{\omega_L}{h\omega_H - \omega_L}\frac{\mu_q(\omega_L)}{\mu_q(\omega^\circ)} \leq 1$. We conclude that the solution is as follows:

1. If $\mu_q(\omega_L) \leq \frac{h\omega_H - \omega_L}{\omega_L}\mu_q(\omega^\circ)$, then

$$x_q^*(\omega_L|\omega_L) = 1, x_q^*(\omega_L|\omega_H) = 0, \text{ and } x_q^*(\omega_L|\omega^\circ) = \frac{\omega_L}{h\omega_H - \omega_L}\frac{\mu_q(\omega_L)}{\mu_q(\omega^\circ)};$$

2. If $\frac{\omega_H - \omega_L}{\omega_H} - \mu_q(\omega^\circ)(1-h) \geq \mu_q(\omega_L) \geq \frac{h\omega_H - \omega_L}{\omega_L}\mu_q(\omega^\circ)$, then

$$x_q^*(\omega_L|\omega_L) = 1, x_q^*(\omega_L|\omega_H) = \frac{\omega_L\mu_q(\omega_L) - (h\omega_H - \omega_L)\mu_q(\omega^\circ)}{(\omega_H - \omega_L)\mu_q(\omega_H)}, \text{ and } x_q^*(\omega_L|\omega^\circ) = 1;$$

3. If $\mu_q(\omega_L) \geq \frac{\omega_H - \omega_L}{\omega_H} - \mu_q(\omega^\circ)(1-h)$, then

$$x_q^*(\omega_L|\omega_L) = 1, x_q^*(\omega_L|\omega_H) = 1, \text{ and } x_q^*(\omega_L|\omega^\circ) = 1.$$

Now suppose that $h\omega_H - \omega_L \le 0$, or $h \le \frac{\omega_L}{\omega_H}$. Combining obedience constraints in the standard manner for communication problems with binary action, we get

$$\omega_L x(\omega_L|\omega_L)\mu_q(\omega_L) - (\omega_H - \omega_L)x(\omega_L|\omega_H)\mu_q(\omega_H) - (h\omega_H - \omega_L)x(\omega_L|\omega^\circ)\mu_q(\omega^\circ) \ge$$
$$\max\left\{\omega_H\mu_q(\omega_L) + (1-h)\omega_H\mu_q(\omega^\circ) - (\omega_H - \omega_L), 0\right\}.$$

It is immediate that $x_q^*(\omega_L|\omega^\circ) = x_q^*(\omega_L|\omega_L) = 1$, since this choice relaxes the platform's problem as much as possible. The obedience constraint then becomes

$$\omega_L\mu_q(\omega_L) - (h\omega_H - \omega_L)\mu_q(\omega^\circ) - \max\left\{\omega_H\mu_q(\omega_L) + (1-h)\omega_H\mu_q(\omega^\circ) - (\omega_H - \omega_L), 0\right\} \ge$$
$$(\omega_H - \omega_L)x(\omega_L|\omega_H)\mu_q(\omega_H)$$

We then conclude that the solution is as follows:

1. If $\mu_q(\omega_L) \le \frac{\omega_H - \omega_L}{\omega_H} - \mu_q(\omega^\circ)(1-h)$, then

$$x_q^*(\omega_L|\omega_L) = 1, x_q^*(\omega_L|\omega_H) = \frac{\omega_L\mu_q(\omega_L) - (h\omega_H - \omega_L)\mu_q(\omega^\circ)}{(\omega_H - \omega_L)\mu_q(\omega_H)}, \text{ and } x_q^*(\omega_L|\omega^\circ) = 1;$$

2. If $\mu_q(\omega_L) \ge \frac{\omega_H - \omega_L}{\omega_H} - \mu_q(\omega^\circ)(1-h)$, then

$$x_q^*(\omega_L|\omega_L) = 1, x_q^*(\omega_L|\omega_H) = 1, \text{ and } x_q^*(\omega_L|\omega^\circ) = 1.$$

D.2 Data-Value Problem

Let $\lambda(a_1'|a_1) = b(a_1)l(a_1'|a_1)$. The data-value problem is then

$$\min_{v,\lambda} \mu_q(\omega_L)v(\omega_L) + \mu_q(\omega_H)v(\omega_H) + \mu_q(\omega^\circ)v(\omega^\circ),$$

subject to $\lambda(\omega_H|\omega_L), \lambda(\omega_L|\omega_H) \ge 0$,

$$\begin{aligned}
v(\omega_L) &= \max\{\omega_L\lambda(\omega_H|\omega_L), -\omega_L\lambda(\omega_L|\omega_H)\} = \omega_L\lambda(\omega_H|\omega_L),\\
v(\omega_H) &= \max\{\omega_H - \omega_L - (\omega_H - \omega_L)\lambda(\omega_H|\omega_L), (\omega_H - \omega_L)\lambda(\omega_L|\omega_H)\} =\\
&\quad (\omega_H - \omega_L)\max\{1 - \lambda(\omega_H|\omega_L), \lambda(\omega_L|\omega_H)\},\\
v(\omega^\circ) &= \max\{h(\omega_H - \omega_L) + (\omega_L - h\omega_H)\lambda(\omega_H|\omega_L), (h\omega_H - \omega_L)\lambda(\omega_L|\omega_H)\} =\\
&\quad h(\omega_H - \omega_L)\max\left\{1 - \frac{h\omega_H - \omega_L}{h(\omega_H - \omega_L)}\lambda(\omega_H|\omega_L), \frac{h\omega_H - \omega_L}{h(\omega_H - \omega_L)}\lambda(\omega_L|\omega_H)\right\}.
\end{aligned}$$

As we noted before, $\frac{h\omega_H-\omega_L}{h(\omega_H-\omega_L)} < 1$. Suppose that $h > \frac{\omega_L}{\omega_H}$. Then it is optimal to set $\lambda^*_q(\omega_L|\omega_H) = 0$ to relax the problem as much as possible. We then have

$$\begin{aligned}
v(\omega_L) &= \omega_L \lambda(\omega_H|\omega_L), \\
v(\omega_H) &= (\omega_H - \omega_L) \max\{1 - \lambda(\omega_H|\omega_L), 0\}, \\
v(\omega^\circ) &= h(\omega_H - \omega_L) \max\left\{1 - \frac{h\omega_H - \omega_L}{h(\omega_H - \omega_L)}\lambda(\omega_H|\omega_L), 0\right\}.
\end{aligned}$$

There are three candidates for optimal $\lambda(\omega_H|\omega_L)$, specifically, 0 and two kinks of the maxima in the expressions above, 1 and $\frac{h(\omega_H-\omega_L)}{h\omega_H-\omega_L} > 1$.

When $\lambda(\omega_H|\omega_L) = 0$, the objective is $S_0 := (1 - \mu_q(\omega_L) - \mu_q(\omega^\circ)(1-h))(\omega_H - \omega_L)$.

When $\lambda(\omega_H|\omega_L) = 1$, the objective is $S_1 := \mu_q(\omega_L)\omega_L + \mu_q(\omega^\circ)(1-h)\omega_L$.

When $\lambda(\omega_H|\omega_L) = \frac{h(\omega_H-\omega_L)}{h\omega_H-\omega_L}$, the objective is $S_f := \mu_q(\omega_L)\frac{h(\omega_H-\omega_L)}{h\omega_H-\omega_L}\omega_L$.

The following claims are true:

- $S_0 \leq S_1$ if and only if $\mu_q(\omega_L) + \mu_q(\omega^\circ)(1-h) \geq \frac{\omega_H-\omega_L}{\omega_H}$;
- $S_0 \leq S_f$ if and only if $\mu_q(\omega_L)\left(1 + \frac{h\omega_L}{h\omega_H-\omega_L}\right) + \mu_q(\omega^\circ)(1-h) \geq 1$;
- $S_1 \leq S_f$ if and only if $\mu_q(\omega_L)\frac{\omega_L}{h\omega_H-\omega_L} \geq \mu_q(\omega^\circ)$.

Figure 2 captures the resulting regions of $\mu_q(\omega_L)$ and $\mu_q(\omega^\circ)$ that correspond to the value of the problem being equal to one of S_0, S_1, and S_f.

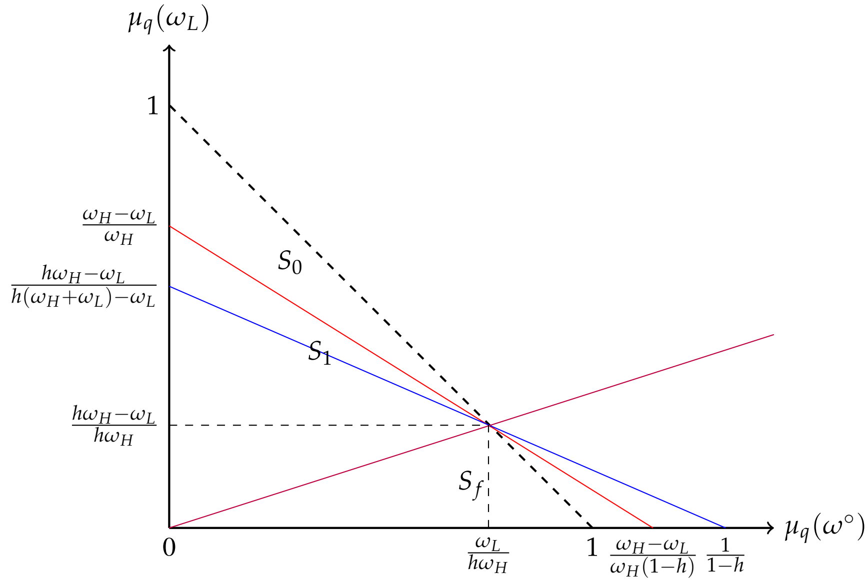


Figure 2: This figure pins down the minimum value of the data-value problem depending on μ_q when $h > \frac{\omega_L}{\omega_H}$. The red line corresponds to $S_0 = S_1$, the blue line corresponds to $S_0 = S_f$, and the purple line corresponds to $S_1 = S_f$.

We then conclude that the solution is as follows:

1. If $\mu_q(\omega_L) \geq \frac{\omega_H-\omega_L}{\omega_H} - \mu_q(\omega^\circ)(1-h)$, then $\lambda_q^*(\omega_H|\omega_L) = 0$ is optimal. The resulting values are
$$v_q^*(\omega_L) = 0, v_q^*(\omega_H) = \omega_H - \omega_L, \text{ and } v_q^*(\omega^\circ) = h(\omega_H - \omega_L);$$
2. If $\frac{h\omega_H-\omega_L}{\omega_L}\mu_q(\omega^\circ) \leq \mu_q(\omega_L) \leq \frac{\omega_H-\omega_L}{\omega_H} - \mu_q(\omega^\circ)(1-h)$, then $\lambda_q^*(\omega_H|\omega_L) = 1$ is optimal. The resulting values are
$$v_q^*(\omega_L) = \omega_L, v_q^*(\omega_H) = 0, \text{ and } v_q^*(\omega^\circ) = (1-h)\omega_L;$$
3. If $\mu_q(\omega_L) \leq \frac{h\omega_H-\omega_L}{\omega_L}\mu_q(\omega^\circ)$, then $\lambda_q^*(\omega_H|\omega_L) = \frac{h(\omega_H-\omega_L)}{h\omega_H-\omega_L}$ is optimal. The resulting values are
$$v_q^*(\omega_L) = \frac{h(\omega_H-\omega_L)}{h\omega_H-\omega_L}\omega_L, v_q^*(\omega_H) = 0, \text{ and } v_q^*(\omega^\circ) = 0.$$

Suppose now that $h \leq \frac{\omega_L}{\omega_H}$. Then immediately
$$v(\omega^\circ) = h(\omega_H - \omega_L) - (h\omega_H - \omega_L)\lambda(\omega_H|\omega_L).$$

$\lambda_q^*(\omega_L|\omega_H) = 0$ is optimal again. There are only two candidates for optimal $\lambda(\omega_H|\omega_L)$, specifically, 0 and 1. The solution then is as follows:

1. If $\mu_q(\omega_L) \geq \frac{\omega_H - \omega_L}{\omega_H} - \mu_q(\omega^\circ)(1-h)$, then $\lambda_q^*(\omega_H|\omega_L) = 0$ is optimal. The resulting values are

$$v_q^*(\omega_L) = 0, v_q^*(\omega_H) = \omega_H - \omega_L, \text{ and } v_q^*(\omega^\circ) = h(\omega_H - \omega_L);$$

2. If $\mu_q(\omega_L) \leq \frac{\omega_H - \omega_L}{\omega_H} - \mu_q(\omega^\circ)(1-h)$, then $\lambda_q^*(\omega_H|\omega_L) = 1$ is optimal. The resulting values are

$$v_q^*(\omega_L) = \omega_L, v_q^*(\omega_H) = 0, \text{ and } v_q^*(\omega^\circ) = (1-h)\omega_L.$$

D.3 Summary

All the cases considered can be grouped into three scenarios based on $\mu_q(\omega_L)$ and $\mu_q(\omega^\circ)$.

Scenario 1. Suppose that $\mu_q(\omega_L) \leq \frac{h\omega_H - \omega_L}{\omega_L}\mu_q(\omega^\circ)$. Note that this scenario appears only if $h > \frac{\omega_L}{\omega_H}$. The solution to the information-design problem is presented in Table 2.

$x_q^*(a\|\omega)$		ω		
		ω_L	ω_H	ω°
a	ω_L	1	0	$\frac{\omega_L}{h\omega_H - \omega_L}\frac{\mu_q(\omega_L)}{\mu_q(\omega^\circ)}$
	ω_H	0	1	$1 - \frac{\omega_L}{h\omega_H - \omega_L}\frac{\mu_q(\omega_L)}{\mu_q(\omega^\circ)}$

Table 2: Platform Example, x_q^* for Scenario 1.

The solution to the data-value problem is $\lambda_q^*(\omega_L|\omega_H) = 0$, $\lambda_q^*(\omega_H|\omega_L) = \frac{h(\omega_H - \omega_L)}{h\omega_H - \omega_L}$ and the unit values of datapoints are $v_q^*(\omega_L) = \frac{h(\omega_H - \omega_L)}{h\omega_H - \omega_L}\omega_L$, $v_q^*(\omega_H) = 0$, and $v_q^*(\omega^\circ) = 0$.

Scenario 2. Suppose that $\frac{h\omega_H - \omega_L}{\omega_L}\mu_q(\omega^\circ) \leq \mu_q(\omega_L) \leq \frac{\omega_H - \omega_L}{\omega_H} - \mu_q(\omega^\circ)(1-h)$. Note that the lower bound on $\mu_q(\omega_L)$ is meaningful only if $h > \frac{\omega_L}{\omega_H}$. The solution to the information-design problem is presented in Table 3.

$x_q^*(a\|\omega)$		ω		
		ω_L	ω_H	ω°
a	ω_L	1	$\frac{\omega_L\mu_q(\omega_L) - (h\omega_H - \omega_L)\mu_q(\omega^\circ)}{(\omega_H - \omega_L)\mu_q(\omega_H)}$	1
	ω_H	0	$\frac{\omega_H - \omega_L - \mu_q(\omega_L)\omega_H - \mu_q(\omega^\circ)(1-h)\omega_H}{(\omega_H - \omega_L)\mu_q(\omega_H)}$	0

Table 3: Platform Example, x_q^* for Scenario 2.

The solution to the data-value problem is $\lambda_q^*(\omega_L|\omega_H) = 0$, $\lambda_q^*(\omega_H|\omega_L) = 1$, and the unit values of datapoints are $v_q^*(\omega_L) = \omega_L$, $v_q^*(\omega_H) = 0$, and $v_q^*(\omega^\circ) = (1-h)\omega_L$.

Scenario 3. Suppose that $\mu_q(\omega_L) \geq \frac{\omega_H - \omega_L}{\omega_H} - \mu_q(\omega^\circ)(1-h)$. The solution to the information-design problem is presented in Table 4.

$x_q^*(a\|\omega)$		ω		
		ω_L	ω_H	ω°
a	ω_L	1	1	1
	ω_H	0	0	0

Table 4: Platform Example, x_q^* for Scenario 3.

The solution to the data-value problem is $\lambda_q^*(\omega_L|\omega_H) = \lambda_q^*(\omega_H|\omega_L) = 0$ and the unit values of datapoints are $v_q^*(\omega_L) = 0$, $v_q^*(\omega_H) = \omega_H - \omega_L$, and $v_q^*(\omega^\circ) = h(\omega_H - \omega_L)$. $\triangle$