THE DESIGN AND PRICE OF INFORMATION

By

Dirk Bergemann, Alessandro Bonatti, and Alex Smolin

July 2016
Revised July 2017

COWLES FOUNDATION DISCUSSION PAPER NO. 2049R

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

The Design and Price of Information*

Dirk Bergemann† Alessandro Bonatti‡ Alex Smolin§

June 3, 2017

Abstract

A data buyer faces a decision problem under uncertainty. He can augment his initial private information with supplemental data from a data seller. His willingness to pay for supplemental data is determined by the quality of his initial private information. The data seller optimally offers a menu of statistical experiments. We establish the properties that any revenue-maximizing menu of experiments must satisfy. Every experiment is a non-dispersed stochastic matrix, and every menu contains a fully informative experiment. In the cases of binary states and actions, or binary types, we provide an explicit construction of the optimal menu of experiments.

Keywords: information design, price of information, statistical experiments, mechanism design, price discrimination, hypothesis testing.

JEL Codes: D42, D82, D83.

*We thank the co-editor, Jeff Ely, and three anonymous referees for their productive suggestions. We are grateful for conversations with Ben Brooks, Giacomo Calzolari, Gabriel Carroll, Gonzalo Cisternas, Jacques Crémer, Andrei Hagiu, Teck Ho, Bruno Jullien, Emir Kamenica, Alessandro Lizzeri, Alessandro Pavan, Mike Powell, Phil Reny, Mike Riordan, Maher Said, Jean Tirole, Juuso Toikka, Alex Wolitzky, and to seminar participants at Berkeley, Bocconi, Bologna, Carnegie Mellon, Chicago, Eief, Harvard, Mannheim, Northwestern, NYU, Oxford, QMUL, Toulouse, UCL, Vienna, Warwick, Yale, and the 2015 World Congress of the Econometric Society. We also thank Ian Ball for excellent research assistance.

†Yale University, 30 Hillhouse Ave., New Haven, CT 06520, USA, dirk.bergemann@yale.edu.

‡MIT Sloan School of Management, 100 Main Street, Cambridge, MA 02142, USA bonatti@mit.edu.

§University of Bonn, Lennéstraße 37, 53113, Bonn, Germany alexey.v.smolin@gmail.com.

1 Introduction

1.1 Motivation

The mechanisms by which information is traded can shape the creation and the distribution of surplus in many important markets. Information about individual borrowers guides banks' lending decisions, information about consumers' characteristics facilitates targeted online advertising, and information about a patient's genome enhances health care delivery. In all these settings, information buyers (i.e., lenders, advertisers, and health care providers) have private knowledge relevant to their decision problem at the time of contracting (e.g., independent knowledge of a borrower, prior interactions with specific consumers, access to a patient's family history). Thus, potential data buyers seek to acquire *supplemental* information to improve the quality of their decision making.

In this paper, we develop a framework to analyze the sale of supplemental information. We consider a data buyer who faces a decision problem under uncertainty. A monopolist data seller owns a database containing information about a "state" variable that is relevant to the buyer's decision. Initially, the data buyer has only partial information about the state. This information is private to the data buyer and unknown to the data seller. The precision of the buyer's private information determines his willingness to pay for any supplemental information. Thus, from the perspective of the data seller, there are many possible types of the data buyer. We investigate the revenue-maximizing information policy, i.e., how much information the data seller should provide and how she should price access to the data.

In order to screen the heterogeneous data buyer types, the seller offers a menu of information products. In our context, these products are statistical experiments—signals that reveal information about the payoff-relevant state. Only the information product itself is assumed to be contractible. By contrast, payments *cannot* be made contingent on either the buyer's action or the realized state and signal. Consequently, the value of an experiment to a buyer is determined by the buyer's private belief and can be computed independently of the price of the experiment. We can recast the resulting screening problem as a nonlinear pricing framework wherein the buyer's type is given by his prior belief. In other words, the seller's problem is to *design* and *price* different versions of experiments, that is, different information products from the same underlying database. Because the design of information can be rephrased in terms of hypothesis testing, the present analysis can also be interpreted as a pricing model for statistical tests.

A large body of literature studies the problem of versioning information goods, emphasizing that digital production allows sellers to easily customize (or degrade) the attributes of such products (Shapiro and Varian, 1999). This argument applies even more forcefully

to information products (i.e., statistical experiments). In a nutshell, the data seller's problem consists of degrading the quality of the information sold to some buyers in order to charge higher prices to higher-value buyers. We show that the very nature of information products enriches the scope of price discrimination. Because information is valuable to the extent to which it affects decision making, buyers with different beliefs do not simply value experiments differently: they may even disagree on their ranking. In this sense, the value of information naturally has both a vertical element (the quality of the information), and a horizontal element (the position of the information).

We show that the optimal menu contains, in general, both the fully informative experiment and partially informative, "distorted" experiments. The distorted information products are not simply noisy versions of the same data. Instead, optimality imposes considerable structure on the distortions in the information provided. In particular, every experiment offered as part of the optimal menu is *non-dispersed*, i.e., it contains a signal realization that rules out one of the states. Moreover, if the buyer's decision problem is to match his action with a state, every experiment is *concentrated*, i.e., it induces the buyer to take the correct action with probability one, conditional on at least one realized state.

We provide a full characterization of the optimal menu in the case of binary states and actions. This setting yields sharp insights into the profitability of discriminatory pricing for selling information. In the binary-state environment, the buyer's types are one dimensional and the utilities are piecewise linear with a kink at the belief at which the buyer would switch his optimal action. If all buyer types are congruent, i.e., they take identical actions without additional information, an intuition analogous to the "no-haggling" result for monopoly pricing (Myerson, 1981; Riley and Zeckhauser, 1983) applies: the seller simply offers the fully informative experiment at a fixed price. In general, however, the seller's problem consists of screening types both *within* and *across* classes of congruent types.

The beneficial use of partial information can be seen with two types that are ranked according to their valuation of the fully informative experiment. The "high" type is ex ante less informed, while the "low" type is ex ante more informed. Suppose that types would pursue distinct actions in the absence of additional information. A feasible policy for the seller is to offer the high type the fully informative experiment and the low type a partially informative experiment that generates one of two signals: with small but positive probability, the signal informs the low type *without noise* about the state that he considers less likely ex ante; with the remaining probability, it sends a second noisy signal. This allows the low type to improve the quality of his decision making; thus, he would be willing to pay a positive amount for the experiment. By contrast, the high type would not attach a positive value to this partial information. After all, he would have chosen the action suggested by the

noiseless signal under his prior anyway, and given his prior, the noisy signal is too weak to modify his action.

The optimal menu for two types exploits the horizontal element of information to extract value from the low type without conceding any rents to the high type. Such profitable screening by providing partial information is the novel element that distinguishes the pricing of information from other monopoly problems, such as designing insurance or goods of differentiated quality. With a continuum of types, the optimal menu still contains at most two experiments: one is fully informative, and the other contains two signals, one of which perfectly reveals the true state. In particular, the optimal menu involves discriminatory pricing (i.e., two different experiments are offered) only if "ironing" is required (Myerson, 1981). Intuitively, the second experiment intends to serve buyers in one group, while charging higher prices to the other group.

Our findings have concrete implications for the sale of information. In Section 5, we illustrate our results in the context of the information being sold by online data brokers, focusing on a broad class of products ("data appends") that are used for marketing and risk-mitigation purposes. We use the language of hypothesis testing and statistical errors to demonstrate how the design of data products can be informed by the structural properties of the optimal experiments we identified in our analysis. In particular, we argue that no experiment in an optimal menu should add unbiased noise to the seller's information, and we discuss whether enabling the buyer to access only a portion of the seller's data is equivalent to introducing noise.

In Section 6, we take the first step toward more general results. We fully characterize the menu with two types that face a matching decision problem. In an optimal menu, the high type purchases the fully informative experiment, while the low type purchases an experiment that is at least partially informative. This can occur even if the types are congruent. The latter experiment provides "directional" information about the states the low type perceives to be relatively more likely and induces him to take the corresponding actions more often. In this way, the seller optimally reduces the information rents of the less informed type, possibly to zero, without needing to exclude the more informed type.

1.2 Related Literature

Our paper is part of the body of literature on selling information to imperfectly informed decision makers. In seminal papers, Admati and Pfleiderer (1986, 1990) analyze the sale of information to a continuum of *ex ante homogeneous* agents, all with the same prior information. After the purchase of supplemental information, the agents trade an asset

with a common value. They show that it is optimal to provide noisy, idiosyncratic and, hence, heterogeneous information. This idiosyncratic information guarantees the traders a local monopoly, which preserves the value of acquiring information even in an informative, rational-expectations equilibrium. Thus, Admati and Pfleiderer (1986, 1990) explicitly consider interactions among data buyers that we do not pursue here. By contrast, we focus on *ex ante heterogeneous* types of a single buyer who value information differently due to their different prior beliefs. The data seller in our setting offers noisy versions of the data to screen the buyer's initial information and to extract more surplus, leading to profound differences in the optimal experiments. A second contribution of this paper relative to Admati and Pfleiderer (1986, 1990) is that we consider *all* feasible statistical experiments, whereas they restrict their attention to normally distributed priors and signals. We shall see that the optimal experiment is outside of the normal class, even if the priors are normally distributed.

In recent work, Babaioff, Kleinberg, and Paes Leme (2012) also analyze the optimal mechanisms for selling information. While we consider the same general question, the details of the model, the contracting environment, and the nature of the results differ substantially. In their model, the ex post payoff function of the data buyer depends on *two state variables*. The seller has private information about one state variable, and the buyer has private information about the other. Their contracting environment differs from ours in that the seller is allowed to make the information disclosure and the price dependent on his privately observed signal. By contrast, we ask the data seller to commit to a selling mechanism before the realization of any state variable. The central results of their paper are statements of the revelation principle and algorithms for the optimal mechanism using surplus extraction arguments, as in Cremer and McLean (1988).

Within the mechanism design literature, our approach is related to, yet conceptually distinct from, models of discriminatory information disclosure in which the seller of a good discloses match-value information and sets a price. Several papers, including Lizzeri (1999), Ottaviani and Prat (2001), Johnson and Myatt (2006), Bergemann and Pesendorfer (2007), Eső and Szentes (2007a), Krähmer and Strausz (2015), and Li and Shi (2015), analyze this problem from an ex ante perspective, where the seller commits (simultaneously or sequentially) to a disclosure rule and a pricing policy.[1] Eső and Szentes (2007b) consider a related model of selling advice. Their model is distinct from our analysis in two dimensions. First, the private information of the agent is the expected value difference between two possible actions. Thus, the private information is one dimensional rather than multidimensional.

[1] In addition, a number of more recent papers, including Balestrieri and Izmalkov (2014), Celik (2014), Koessler and Skreta (2016), and Mylovanov and Tröger (2014), analyze this question from an informed principal perspective.

Second, the seller can make the payment contingent on both the statistical experiment and the buyer's action. By contrast, our seller can price the information but not the action itself. Commitment to a disclosure policy is present in the literature on Bayesian persuasion, e.g., Rayo and Segal (2010), Kamenica and Gentzkow (2011), and Kolotilin, Li, Mylovanov, and Zapechelnyuk (2015). In contrast to this line of work, our model admits monetary transfers and rules out any direct effect of the buyer's ex post action on the seller's utility.

Our previous work (Bergemann and Bonatti, 2015) considered the information-acquisition policy of a data buyer who then decided on the placement of display advertising. This earlier model was simpler in many respects. First, the price of information was given or determined by a competitive market. Second, the data buyer did not have any private information. Third, despite allowing for a continuum of matching values (states) and advertising levels (actions), the available information structures were restricted to simple queries that perfectly revealed individual state realizations. The analysis focused on the nature of the buyer's optimal queries given the distribution of match values and the cost of advertising.

Hörner and Skrzypacz (2016) share a similar title but consider a very different setting. They consider a dynamic hold-up game, except that information rather than a physical object is sold. At the beginning of the game, the buyer has no private information and wants to hire a competent data seller. The data seller knows whether she is competent and can prove her competence by sequentially undertaking tests within a fixed subclass of statistical experiments. Hörner and Skrzypacz (2016) allow for sequential monetary transfers and characterize an equilibrium that is most favorable to the competent seller.

Finally, our seller's problem bears some resemblance to a bundling problem. With more than two states, the buyer types are multidimensional and it is well-known—see, for example, Pavlov (2011b)—that the single-price result of Myerson (1981) and Riley and Zeckhauser (1983) does not hold. Indeed, the optimal menu involves stochastic bundling quite generally, and the structure of the bundles offered can be quite rich.[2] Stochastic bundles are analogous to the partially informative experiments in our model. To further distinguish from these classic multidimensional problems, stochastic bundling can arise in our setting even when buyer types are one dimensional.

[2] For example, see Manelli and Vincent (2006), Pycia (2006), Pavlov (2011a), and Rochet and Thanassoulis (2015). In particular, Daskalakis, Deckelbaum, and Tzamos (2017) construct an example wherein the types follow a Beta distribution, and the optimal menu contains a continuum of stochastic allocations.

2 Model

A single decision maker, the data buyer, faces a decision problem under uncertainty. The state of nature ω is drawn from a finite set $\Omega = \{\omega_1, ..., \omega_i, ..., \omega_I\}$. The data buyer chooses an action a from a finite set $A = \{a_1, ..., a_j, ..., a_J\}$. The ex post utility is denoted by

$$u(\omega_i, a_j) \triangleq u_{ij} \in \mathbb{R}_+. \tag{1}$$

The ex post payoffs can thus be represented by an $I \times J$ matrix:

u	a_1	$\cdots$	a_J
ω_1	u_{11}	$\cdots$	u_{1J}
$\vdots$	$\vdots$		$\vdots$
ω_I	u_{I1}	$\cdots$	u_{IJ}

We impose the following weak assumptions on the matrix: *(i)* $I \leq J$, and *(ii)* $u_{ii} > u_{ij}$ for all $j \neq i$. These two assumptions capture the idea that the action space is at least as rich as the state space and that for every state ω_i, there is a unique action (labeled a_i) that maximizes the decision maker's utility in that state.

Matching Utility A useful special case is one in which the data buyer faces binary ex post payoffs in each state, i.e., he seeks to match the state and the action. In that case, we frequently drop the second subscript for the ex post utility on the diagonal. The utility function $u(\omega_i, a_j)$ is then given by

$$u(\omega_i, a_j) \triangleq \mathbb{I}_{[i=j]} \cdot u_i, \text{ and } u_i \triangleq u_{ii}; \tag{2}$$

or in matrix form,

u	a_1	$\cdots$	a_I
ω_1	u_1		0
$\vdots$		$\ddots$	
ω_I	0		u_I

This formulation assumes that, in each state, the data buyer assigns the same value (normalized to zero) to each *wrong* action. This assumption has no bite across states because adding a state-dependent translation to the utility function does not affect preferences over actions. Under this assumption, it is without loss of generality to assume that the sets of actions and states have the same cardinality: $|A| = |\Omega| = I = J$.

Prior Information The *interim belief* θ about the state is the type of the data buyer

$$\theta \in \Theta \triangleq \Delta\Omega,$$

where θ_i denotes the interim probability that type θ assigns to state ω_i, with $i = 1, ..., I$. The interim beliefs of the data buyer are his private information. From the perspective of the data seller, these beliefs are distributed according to a distribution

$$F \in \Delta\Theta,$$

which we take as a primitive of our model.[3]

In line with the interpretation of selling supplemental information, however, we note that the beliefs $\theta \in \Theta$ can be generated from a common prior and privately observed signals. Thus, suppose there is a common prior $\mu \in \Delta\Omega$. The decision maker privately observes a signal $r \in R$ according to a commonly known experiment

$$\lambda : \Omega \to \Delta R.$$

The decision maker then forms his interim belief via Bayes' rule

$$\theta(\omega \mid r) \triangleq \frac{\lambda(r \mid \omega)\,\mu(\omega)}{\sum_{\omega' \in \Omega} \lambda(r \mid \omega')\,\mu(\omega')}.$$

The interim beliefs $\theta(\omega \mid r)$, simply denoted by θ, are thus the private information of the data buyer. From the data seller's perspective, the common prior $\mu \in \Delta\Omega$ and the distribution of signals $\lambda : \Omega \to \Delta R$ induce a distribution $F \in \Delta\Theta$ of interim beliefs.

Supplemental Information The data buyer seeks to augment his initial private information by obtaining additional information from the data seller in order to improve the quality of his decision making. A statistical *experiment* (equivalently, an information structure) $E = (S, \pi)$ consists of a set S of signals s and a likelihood function:

$$\pi : \Omega \to \Delta S.$$

We assume throughout that the realization of the buyer's private signal $r \in R$ and that of the signal $s \in S$ from any experiment E are independent, conditional on the state ω. In other words, the buyer and the seller draw their information from independent sources.

For a given experiment $E = (S, \pi)$, let S denote the (finite) set of signals that are in the

[3] As usual, the model allows for the alternative interpretations of a single buyer and a continuum of buyers.

support of the experiment and π_{ik} the conditional probability of signal $s_k \in S$ in state ω_i. Letting $K \triangleq |S|$, we have

$$\pi_{ik} \triangleq \Pr\left(s_k \mid \omega_i\right),$$

where $\pi_{ik} \geq 0$ and $\sum_{k=1}^{K} \pi_{ik} = 1$ for all i. We then obtain the stochastic matrix

$$\begin{array}{c|ccccc} E & s_1 & \cdots & s_k & \cdots & s_K \\ \hline \omega_1 & \pi_{11} & \cdots & \pi_{1k} & \cdots & \pi_{1K} \\ \vdots & \vdots & & & & \vdots \\ \omega_i & \pi_{i1} & & \pi_{ik} & & \pi_{iK} \\ \vdots & \vdots & & & & \vdots \\ \omega_I & \pi_{I1} & \cdots & \pi_{Ik} & \cdots & \pi_{IK} \end{array}. \tag{3}$$

The following experiments are of particular interest:

1. a *non-dispersed* experiment that contains at least one nil entry $\pi_{ik} = 0$ for some i, k;
2. a *concentrated* experiment that contains a standard basis vector $\pi_{ii} = 1$ for some i;
3. the *fully informative* experiment $\overline{E}$, with $\pi_{ii} = 1$ for all i.

In non-dispersed experiment, one signal s_k allows the decision maker to rule out *some* state ω_i. In a concentrated experiment, there exists a state ω_i that is ruled out by *all* signals $s_k \neq s_i$. The fully informative experiment perfectly reveals the true state.

A *menu of experiments* $\mathcal{M} = \{\mathcal{E}, t\}$ (or an information policy) consists of a collection $\mathcal{E}$ of experiments E and an associated tariff function

$$t : \mathcal{E} \to \mathbb{R}_+.$$

Our goal is to characterize the revenue-maximizing menu for the seller. The timing of the game is as follows:

1. the seller posts a menu $\mathcal{M}$;
2. the true state ω and the buyer's type θ are realized;
3. the buyer chooses an experiment $E \in \mathcal{E}$ and pays the corresponding price $t\left(E\right)$;
4. the buyer observes a signal s from experiment E and chooses an action a.

We emphasize that the data seller is unrestricted in her choice of statistical experiment (i.e., the seller can improve upon the buyer's original information with arbitrarily accurate

signals), and that the marginal cost of providing the information is nil. These assumptions capture settings in which sellers hold very precise data and distribution costs are negligible.

We deliberately focus on the pure problem of the design and pricing of statistical experiments. We thus assume that the seller commits to a menu of experiments before the realization of the state ω and the type θ. We further assume that none of the following are contractible: the buyer's action a, the realized state ω, or the signal s. Thus, scoring rules and other belief-elicitation schemes that compare the elicited beliefs with the realization of some random variable are not available to the seller. Finally, we consider only static mechanisms and do not investigate sequential protocols. We expect that the sequential sale of experiments would allow the data seller to extract additional surplus relative to static mechanisms, as this would allow the seller to correlate individual payments with the realized states. However, the nature of the incentive constraints would not be affected (at least in the last round of communication), and we expect the qualitative results to remain unchanged.

3 Information Design

3.1 Value of Information

We first describe the value of the buyer's initial information and then determine the incremental value of an experiment $E = (S, \pi)$. The value of the data buyer's problem under prior information *only* is given by choosing the action a_j that maximizes the expected utility given the interim belief θ:

$$a(\theta) \triangleq \arg\max_{a_j \in A} \left\{ \sum\nolimits_{i=1}^{I} \theta_i u_{ij} \right\}.$$

The expected utility of type θ is therefore given by

$$u(\theta) \triangleq \max_{a_j \in A} \left\{ \sum\nolimits_{i=1}^{I} \theta_i u_{ij} \right\}.$$

By contrast, if the data buyer has access to an experiment $E = (S, \pi)$, he first observes the signal realization $s_k \in S$, updates his beliefs and then chooses an appropriate action. The marginal distribution of signals s_k from the perspective of type θ is given by

$$\Pr[s_k \mid \theta] = \sum\nolimits_{i=1}^{I} \theta_i \pi_{ik}.$$

Consequently, for any signal s_k that occurs with strictly positive probability $\Pr[s_k \mid \theta]$, the

action that maximizes the expected utility of type θ is given by

$$a\left(s_k \mid \theta\right) \triangleq \underset{a_j \in A}{\arg\max} \left\{ \sum\nolimits_{i=1}^{I} \left(\frac{\theta_i \pi_{ik}}{\sum_{i'=1}^{I} \theta_{i'} \pi_{i'k}} \right) u_{ij} \right\}, \tag{4}$$

which leads to the following conditional expected utility:

$$u\left(s_k \mid \theta\right) \triangleq \max_{a_j \in A} \left\{ \sum\nolimits_{i=1}^{I} \left(\frac{\theta_i \pi_{ik}}{\sum_{i'=1}^{I} \theta_{i'} \pi_{i'k}} \right) u_{ij} \right\}. \tag{5}$$

Integrating over all signal realizations s_k and subtracting the value of prior information, the (net) value of an experiment E for type θ is given by

$$V\left(E, \theta\right) \triangleq \mathbb{E}\left[u\left(s \mid \theta\right)\right] - u\left(\theta\right) = \sum\nolimits_{k=1}^{K} \max_{j} \left\{ \sum\nolimits_{i=1}^{I} \theta_i \pi_{ik} u_{ij} \right\} - u\left(\theta\right). \tag{6}$$

In the case of the previously defined matching utility (2), the value of information takes the simpler form

$$V\left(E, \theta\right) = \sum\nolimits_{k=1}^{K} \max_{i} \left\{ \theta_i \pi_{ik} u_i \right\} - \max_{i} \left\{ \theta_i u_i \right\}. \tag{7}$$

The value of the prior information, given by $\max_i \theta_i u_i$, is generated by the action that has the highest value-weighted probability of matching the state. The value of an experiment E is generated by choosing an action on the basis of the posterior belief $\theta_i \pi_{ik}$ induced by each signal s_k. Under matching utility, the value of information is given by the (value-weighted) *incremental probability* of choosing the correct action.[4]

In Figure 1, we illustrate the value of an experiment in a model with three actions and three states $\omega_i \in \{\omega_1, \omega_2, \omega_3\}$. The prior belief of each agent is therefore an element of the two-dimensional simplex, $\theta = (\theta_1, \theta_2, 1 - \theta_1 - \theta_2)$. The utility function is given by state-action matching with uniform weights, i.e., $u_i \triangleq 1$. We display the value of the fully informative experiment $\overline{E}$ and of a partially informative experiment E as a function of the buyer's prior.[5]

[4] The value of information for the data buyer differs from a consumer's value for multiple goods or bundles of characteristics. In particular, the first max operator in (6) and (7) corresponds to the optimality condition for the buyer's action given the available information. The second max operator reflects the type-dependent nature of participation constraints. Both elements are missing from the multiproduct monopolist's problem (Pavlov, 2011a,b; Daskalakis, Deckelbaum, and Tzamos, 2017).

[5] The information structure in the right panel is given by the following stochastic matrix:

E	s_1	s_2	s_3
ω_1	1/2	1/4	1/4
ω_2	0	3/4	1/4
ω_3	0	1/4	3/4

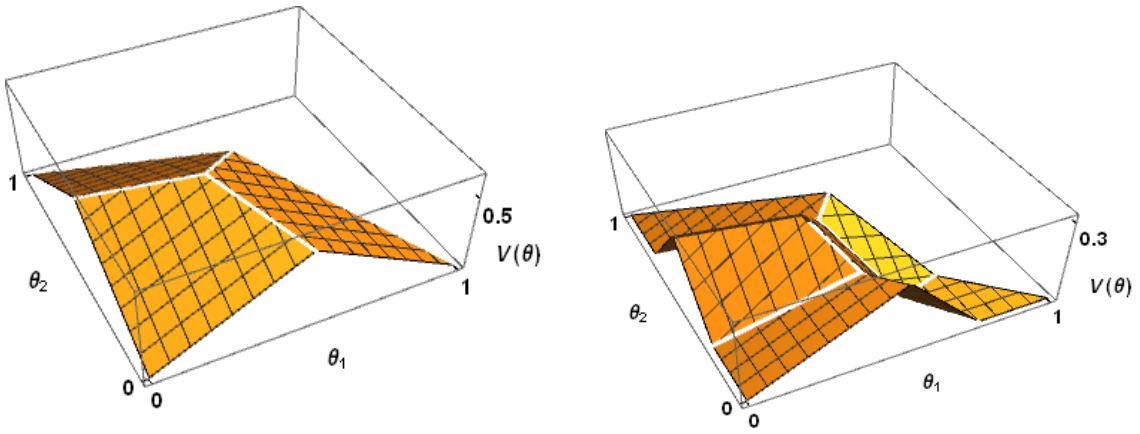

Figure 1: Value of Full and Partial Information, $I = J = 3$

Viewed as a function of the types, the value of an experiment $V(E, \theta)$ is piecewise linear in θ with a finite number of linear components. The linearity of the value function is a consequence of the Bayesian nature of our problem, where types are prior probabilities of states. The downward kinks are due to the max operator in the buyer's reservation utility $u(\theta)$. They correspond to changes in the buyer's action without supplemental information. The upward kinks are generated by the max operator in (5) and reflect changes in the buyer's preferred action upon observing a signal. Finally, the experiment E is only valuable if at least two signals lead to different actions. If the buyer chooses a constant action following every signal s_k, then $V(E, \theta) = 0$, as can be seen in formulations (6) and (7), as well as in Figure 1 (right).

3.2 The Seller's Problem

The seller's choice of a profit-maximizing menu of experiments may involve, in principle, designing one experiment per buyer type. In turn, each experiment entails a type-dependent mapping from signals into actions. The seller's problem can, however, be simplified by reducing the set of menus of experiments to a smaller and very tractable class.

First, by the revelation principle, we can restrict our attention to direct mechanisms $\mathcal{M} = \{E(\theta), t(\theta)\}$ that assign an experiment $E(\theta) = (S(\theta), \pi(\theta))$ and a price $t(\theta)$ to each type θ of data buyer. We denote the indirect (net) utility for the truth-telling agent by

$$V(\theta) \triangleq V(E(\theta), \theta) - t(\theta).$$

The seller's problem consists of maximizing the expected transfers

$$\max_{\{E(\theta),t(\theta)\}} \int_{\theta\in\Theta} t(\theta)\,\mathrm{d}F(\theta) \tag{8}$$

subject to incentive-compatibility constraints

$$V(\theta) \geq V(E(\theta'),\theta) - t(\theta'), \quad \forall\,\theta,\theta' \in \Theta, \tag{9}$$

and individual-rationality constraints

$$V(\theta) \geq 0, \quad \forall\theta \in \Theta.$$

Second, given any direct mechanism $\mathcal{M} = \{E(\theta), t(\theta)\}$, we say that experiment $E(\theta)$ is *responsive* if every signal $s \in S(\theta)$ leads type θ to a different optimal choice of action and, in particular,

$$a(s_k \mid \theta) = a_k \text{ for all } s_k \in S(\theta). \tag{10}$$

Importantly, condition (10) is only required for every experiment $E(\theta)$ and for the corresponding type θ. In other words, we do not require this condition to hold if signals $s_k \in S(\theta)$ are evaluated by a different type $\theta' \neq \theta$. Finally, we define an *outcome* of a menu as the joint distribution of states, actions, and monetary transfers resulting from every type's optimal choice of experiment and subsequent choice of action.

Proposition 1 establishes that it is without loss of generality to restrict attention to *responsive menus*—direct revelation mechanisms in which every experiment $E(\theta)$ is responsive.

Proposition 1 (Responsive Menus)
The outcome of every menu $\mathcal{M}$ can be attained by a responsive menu.

Our proof closely follows the argument of the revelation principle for Bayesian games of communication established by Myerson (1982). We show that we can always reduce the size of the signal space to the size of the action space recommended in equilibrium. The intuition is straightforward. Consider an incentive-compatible menu that contains an experiment $E(\theta)$ with more signals than actions. We can combine all signals in $E(\theta)$ that lead to the same choice of action for type θ. The value of this experiment remains constant for type θ, who does not modify his behavior. However, because the new experiment is (weakly) less informative than the experiment we started with, $V(E(\theta),\theta')$ decreases (weakly) for all $\theta' \neq \theta$, relaxing the incentive constraints. Finally, because every signal sent with positive probability under experiment $E(\theta)$ leads type θ to a different action, we can order the signals such that each $s_k \in S(\theta)$ recommends the corresponding action a_k.

The language of the Bayesian games of communication, as suggested by Myerson (1982), is helpful for understanding the nature of the seller's problem more generally. The solution to the data seller's problem has to satisfy two different constraints, the truth-telling (or honesty) constraint given by (9) and the obedience constraint given by (4). Thus, the buyer must be jointly honest and obedient. In particular, double deviations (lying and disobeying) must not be profitable for the buyer.

An immediate implication of Proposition 1 is that, without loss of generality, we can restrict our attention to experiments in which the signal space has the cardinality of the action space, i.e., $K = J$. This insight allows us to write the likelihood function of every experiment (3) as a matrix with the same dimensions as the payoff matrix, i.e.,

$E(\theta)$	$s_1(=a_1)$	$\cdots$	$s_i(=a_j)$	$\cdots$	$s_J(=a_J)$
ω_1	π_{11}	$\cdots$	π_{1j}	$\cdots$	π_{1J}
$\vdots$	$\vdots$				$\vdots$
ω_i	π_{i1}		π_{ij}		π_{iJ}
$\vdots$	$\vdots$				$\vdots$
ω_I	π_{I1}	$\cdots$	π_{Ij}	$\cdots$	π_{IJ}

Thus, we can replace the signal s_j with the action recommendation a_j. This property does not require that every action a_j is recommended with strictly positive probability. For instance, some signals s_j may never be sent, corresponding to a column vector of zeros at the j-th position. The resulting value of experiment $E(\theta)$ for type θ can be written as

$$V(E(\theta),\theta) = \sum\nolimits_{j=1}^{J} \sum\nolimits_{i=1}^{I} \theta_i \pi_{ij} u_{ij} - \max_j \left\{ \sum\nolimits_{i=1}^{I} \theta_i u_{ij} \right\}. \tag{11}$$

Under the restriction to responsive experiments, formulation (11) removes the first max operator from the value of experiment $E(\theta)$ for the truth-telling type θ. Because a misreporting type need not always take the recommended action, we must still use the original formulation (6) when computing the value of experiment $E(\theta)$ for type θ'.

When payoffs are given by state-action matching as in (2), the value of experiment $E(\theta)$ for the truth-telling type θ simplifies to

$$V(E(\theta),\theta) = \sum\nolimits_{i=1}^{I} \theta_i \pi_{ii} u_i - \max_i \{\theta_i u_i\}. \tag{12}$$

In formulation (12), the value of experiment $E(\theta)$ for type θ is fully determined by its diagonal entries π_{ii}. By contrast, the off-diagonal entries π_{ij} may enter the value of experiment $E(\theta)$ for different types θ', as in (7).

3.3 Structural Properties

We now leverage the nature of the value of information to impose additional structure on the experiments that are part of an optimal menu.

Proposition 2 (Optimal Experiments)

1. *The fully informative experiment* $\overline{E}$ *is part of an optimal menu.*
2. *Every experiment in any optimal menu is non-dispersed, i.e.,* $\pi_{ij} = 0$ *for some* $i \neq j$*.*
3. *In the matching case, every experiment in any optimal menu is concentrated, i.e.,* $\pi_{ii} = 1$ *for some* i*.*

The first part of this result can be established via contradiction. Every type θ values the fully informative experiment $\overline{E}$ the most among all the experiments. Suppose, then, that $\overline{E}$ is not part of the menu, and denote the most expensive item currently on the menu by E'. The seller can replace experiment E' with the complete experiment $\overline{E}$, keeping all other prices constant and charging a higher price for $\overline{E}$ than for E'. The new menu weakly increases the seller's revenue without lowering the net utility of any buyer type.[6]

The second and third parts imply that every optimal experiment eliminates the buyer's uncertainty along some dimension. They are also established by an improvement argument. Fix an experiment E and suppose all entries π_{ij} are strictly positive. In each state (row) i, the seller increases the probability π_{ii} of the signal s_i that yields the highest payoff u_{ii} for an obedient type in that state. Concurrently, the seller reduces the probability π_{ij} of the signal s_j that yields the lowest payoff $\min_j u_{ij}$ for an obedient type. The seller shifts a probability mass *inversely proportional* to the difference in payoffs $u_{ii} - u_{ij}$, and hence, the resulting payoff gain is uniform across states ω_i. This procedure is applied until the first entry π_{ij} reaches 0. Because the beliefs of each type θ sum to one, this shift is valued uniformly by all obedient types and weakly less by any other type. A commensurate increase in the price of the experiment offsets the value of the additional information provided and, hence, strictly increases profits while (weakly) relaxing the truth-telling and participation constraints.

Thus, with arbitrary payoffs, every experiment E contains a signal s_j that allows the buyer to rule out some state ω_i. The limits of this result are related to the possibility of double deviations evoked earlier. A misreporting type θ' may not necessarily choose the action recommended by every signal in the experiment $E(\theta)$ intended for type θ. For

[6] The fully informative experiment is part of every optimal menu if all types assing positive probability to all states. It need not be part of *every* optimal menu if some types assign zero probability to some states. In that case, an optimal menu may contain a partially informative experiment that, combined with some type's prior, fully reveals all states that have positive probability, and so remains fully informative in this weaker sense.

the above improvement argument to discourage double deviations, it is critical that the probability mass is shifted away from the worst action in each state ω_i, thus yielding the largest possible marginal benefit $(\max_j u_{ij} - \min_j u_{ij})$ for every obedient type in state ω_i. Any other shift may yield a strictly higher benefit to type θ' than to type θ and lead to the violation of the truth-telling constraints.

In the case of matching utility, the seller can shift the probability mass to the diagonal from *any* off-diagonal entry until the first entry π_{ii} reaches 1. Indeed, when the probability is shifted to the diagonal, any type that does not follow the signal's recommendation obtains a strictly lower benefit relative to an obedient type. As a result, any optimal experiment is concentrated, i.e., there exists at least one state ω_i under which signal s_i is sent with probability one, and the buyer takes the correct action a_i. Conversely, the buyer is able to rule out (at least) one state ω_i after observing any signal $s_k \neq s_i$.

In Sections 4 and 6, we show that with binary states and actions or binary types, respectively, it is sufficient to consider truth-telling and obedience separately. By contrast, Example 3 in the Supplemental Appendix demonstrates that with three (or more) actions and types, double deviations typically impose additional restrictions on the optimal menu.

4 Optimal Menu with Binary Actions

In this section, we consider an environment with two actions $a \in \{a_1, a_2\}$ and two states $\omega \in \{\omega_1, \omega_2\}$. In this setting, the restriction to matching utility functions entails no loss of generality relative to a general payoff matrix. To wit, for every state ω_i, we can always subtract the (state-dependent) constant u_{ij} with $i \neq j$. This linear transformation normalizes the payoffs of the data buyer by setting $u_{12} = u_{21} = 0$ without affecting the optimality conditions of the data buyer's decision problem.

We thus obtain a diagonal payoff matrix as in (2) with positive entries given by

$$u_1 \triangleq u_{11}, \ u_2 \triangleq u_{22}.$$

With binary states, the interim belief of the data buyer (his type) is one dimensional. We identify each type with the interim probability of state ω_1:

$$\theta \triangleq \Pr\left[\omega = \omega_1\right] \in [0, 1].$$

We denote the interim belief type θ that is indifferent between action a_1 and a_2 by θ^* as

$$\theta^* u_1 = (1-\theta^*)\, u_2 \Leftrightarrow \theta^* = \frac{u_2}{u_1 + u_2}. \tag{13}$$

4.1 Binary Experiments

With binary actions, Proposition 1 implies that it is sufficient to consider for every type θ experiments $E(\theta)$ that generate (at most) two signals:

$$\begin{array}{c|cc} E(\theta) & s_1 & s_2 \\ \hline \omega_1 & \pi_{11}(\theta) & 1-\pi_{11}(\theta) \\ \omega_2 & 1-\pi_{22}(\theta) & \pi_{22}(\theta) \end{array}. \tag{14}$$

With binary signals, we can simplify the notation by dropping the second subscript for diagonal entries, as we did for the payoff function:

$$\pi_1(\theta) \triangleq \pi_{11}(\theta),\ \pi_2(\theta) \triangleq \pi_{22}(\theta).$$

Without loss of generality, we assume that $\pi_1(\theta) + \pi_2(\theta) \geq 1$. In other words, signal s_1 is relatively more likely to occur than signal s_2 under state ω_1 than under state ω_2, or

$$\frac{\pi_1(\theta)}{1-\pi_1(\theta)} \geq \frac{1-\pi_2(\theta)}{\pi_2(\theta)}.$$

With binary states and actions, the general payoff environment conforms to the matching utility environment. We can therefore write the value for an arbitrary experiment E as follows by modifying expression (12):

$$V(E,\theta) = \max\left\{\theta\pi_1 u_1 + (1-\theta)\,\pi_2 u_2 - \max\{\theta u_1, (1-\theta)\, u_2\}, 0\right\}. \tag{15}$$

As in the earlier formulation (12), the diagonal entries of the matrix $\pi(\theta)$ generate the probability that experiment $E(\theta)$ allows type θ to match the realized state with his action. Conversely, the first max operator accounts for the possibility that a misreporting type θ' does not follow the recommendation implicit in one of the signals of $E(\theta)$ and, hence, derives no value from the supplemental information.

In Figure 2, we illustrate how the value of information changes as a function of the type θ for the case $u_1 = u_2 = 1$. We compare two experiments with binary signals, namely, the fully informative experiment $E' = (\pi'_1, \pi'_2) = (1,1)$ and a partially informative experiment $E'' = (\pi''_1, \pi''_2) = (1/2, 1)$.

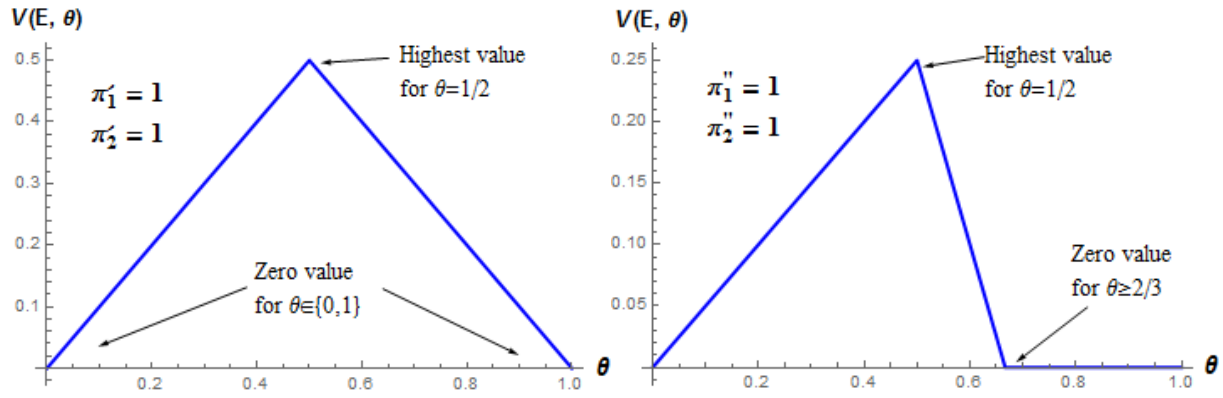

Figure 2: Value of Full and Partial Information ($I = J = 2, \quad u_1 = u_2 = 1$)

The value of information as a function of the buyer's interim beliefs θ reflects many intuitive properties that we formally establish in the next subsection:

1. The most valuable type for the seller is the ex ante least informed. In the examples in Figure 2, this is type $\theta^* = 1/2$. Conversely, the most informed types $\theta \in \{0, 1\}$ have zero value of information. The linear decline in each direction away from θ^* follows from the linearity of the value of information in the interim probability.

2. When we consider any asymmetric experiment, such as the one displayed in the right panel of Figure 2, the distance from the least informed type $|\theta - \theta^*|$ is *not* a sufficient statistic for the value of information. The different slopes on each side of θ^* indicate different marginal benefits for matching state ω_1 versus state ω_2 on the basis of differences in the interim beliefs θ.

Thus, even in an environment where types are clearly one dimensional, information products are inherently multidimensional (in this case, two dimensional). In particular, information always has both vertical (quality) and horizontal (positioning) dimensions. Unlike in models of nonlinear pricing (with respect to either quantity or quality) where all the types *agree* on the relative ranking of all the products, in the current environment, the types *disagree* on the very ranking of all partially informative experiments.

4.2 Binary Types

We derive the optimal menu with two data buyer types, $\theta \in \Theta = \left\{\theta^L, \theta^H\right\}$. The *high (value) type* θ^H assigns a higher value to receiving the fully informative experiment, i.e.,

$$V\left(\overline{E}, \theta^H\right) > V\left(\overline{E}, \theta^L\right). \tag{16}$$

With uniform weights $u_1 = u_2 > 0$, this simply means the *high (value) type* is less well-informed ex ante, i.e.,

$$\left|\theta^H - 1/2\right| \leq \left|\theta^L - 1/2\right|.$$

We refer to the distance between the interim belief θ and the indifference belief θ^* as the precision of the type θ. We denote the frequency of a high type as

$$\gamma \triangleq \Pr(\theta = \theta^H).$$

The following distinction proves helpful. The interim beliefs of the two types are said to be *congruent* if both types would choose the same action without additional information. If we adopt the convention that the high type chooses action a_1 under his prior information (i.e., $\theta^H > \theta^*$), then beliefs (and corresponding types) are *congruent* if $\theta^* < \theta^H < \theta^L$ and *noncongruent* if $\theta^L < \theta^* < \theta^H$.

We first establish two familiar properties: "no distortion at the top," i.e., for the high type and "no rent at the bottom," i.e., for the low type.

Proposition 3 (Binding Constraints)
In an optimal menu:

1. *type θ^H purchases the fully informative experiment $\overline{E}$;*
2. *the participation constraint of type θ^L binds;*
3. *the incentive-compatibility constraint of type θ^H binds.*

We note that these properties of the two-type environment hold for any number of states and actions under arbitrary payoffs, and we shall revisit them in Section 6. The description of the optimal menu is then completed by characterizing the experiment $E(\theta^L)$ purchased by the low type. Here, it is productive to distinguish between congruent and noncongruent beliefs.

Congruent Beliefs. In the case of congruent beliefs $\left(\theta^* < \theta^H < \theta^L\right)$, the argument is related to the classic monopoly pricing problem. By Proposition 2, we know that the optimal experiment $E(\theta^L)$ is concentrated. With congruent beliefs, both types would choose action a_1 absent any additional information. The data seller does not want to reduce the information relative to the outside option and, thus, sets $\pi_1(\theta^L) = 1$. The issue is then how much information to provide about state ω_2, i.e.,

$E(\theta^L)$	s_1	s_2
ω_1	1	0
ω_2	$1 - \pi_2(\theta^L)$	$\pi_2(\theta^L)$

.

Now, any partially informative experiment with $\pi_2(\theta^L) \in (0,1)$ geared towards the low type is also valuable to the high type. Using the expression for the value of information given by (15), we can write the incentive constraint for the high type as follows:

$$\theta^H u_1 + \left(1-\theta^H\right) u_2 - t(\theta^H) \geq \theta^H u_1 + \left(1-\theta^H\right) \pi_2(\theta^L) u_2 - t(\theta^L),$$

since $\pi_1(\theta^H) = \pi_2(\theta^H) = 1$ and $\pi_1(\theta^L) = 1$. The incentive constraint for the high type thus reduces to:

$$\underbrace{\left(1-\theta^H\right)}_{\Pr(\omega_2)} \cdot \underbrace{\left(1-\pi_2(\theta^L)\right)}_{\text{additional precision}} \cdot u_2 \;\geq t(\theta^H) - t(\theta^L). \tag{17}$$

Hence, we observe that both the objective and the constraints in the seller's problem are linear in the choice variable π_2. We can therefore appeal to the no-haggling result of Riley and Zeckhauser (1983) that establishes the optimality of an extremal policy. Such a policy consists of either allocating the object (here, the information) with probability one or not allocating it at all, hence, $\pi_2 \in \{0,1\}$. As in the single-good monopolist's problem, the optimal policy depends on the distribution of buyer types. In particular, the low type receives the fully informative experiment if and only if the probability γ of the high type is sufficiently small, or

$$\left(1-\theta^L\right) u_2 \geq \gamma \left(1-\theta^H\right) u_2 \Longleftrightarrow \gamma \leq \frac{1-\theta^L}{1-\theta^H}.$$

Noncongruent Beliefs. In the case of noncongruent beliefs $\left(\theta^L < \theta^* < \theta^H\right)$, both the argument and the result are distinct from those of the classic monopoly problem. In the absence of additional information, the two types choose different actions. The seller can then provide information in a format that has positive value to one type but zero value to the other type. For example, suppose that $\pi_2(\theta^L) = 1$ and $\pi_1(\theta^L)$ is chosen such that after receiving signal s_2, the high type θ^H is indifferent between actions a_1 and a_2 (i.e., his posterior belief is θ^*):

$$\theta^H \left(1-\pi_1'\right) u_1 = \left(1-\theta^H\right) u_2 \Longleftrightarrow \pi_1' = \frac{u_1\theta^H - u_2\left(1-\theta^H\right)}{u_1\theta^H} \in (0,1)\,. \tag{18}$$

Because type θ^H chooses action a_1 without additional information, the experiment

E'	s_1	s_2
ω_1	π_1'	$1-\pi_1'$
ω_2	0	1

(19)

does not lead to a strict improvement in the decision making (or utility) of the high type θ^H. By contrast, the low type θ^L assigns a positive value to experiment E'. After all, signal s_1 would lead him to match his action to state ω_1, which he would never achieve without additional information. Thus, the seller can offer partial information to the low type without incurring any implicit cost in terms of surplus extraction vis-à-vis the high type.[7]

The argument in support of a partially informative experiment is illustrated for the case $u_1 = u_2 = 1$ in Figure 3, which depicts the value of two experiments *net of the price* as a function of the buyer's type $\theta \in [0, 1]$. In this example, we set $\theta^L = 1/5 < 1/2 < 7/10 = \theta^H$.

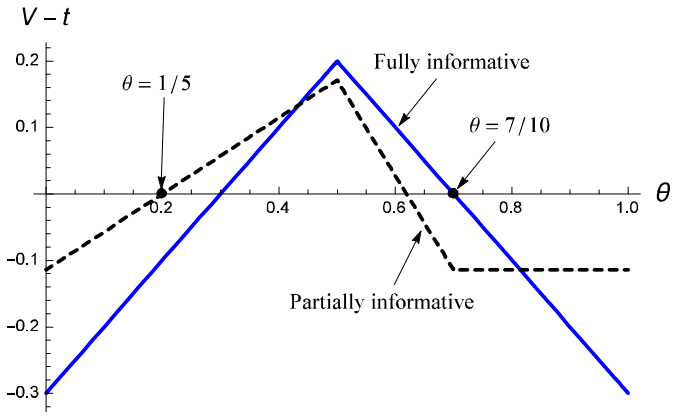

Figure 3: Suboptimal Menu: $(\pi_1, \pi_2) \in \{(1, 1), (4/7, 1)\}$

The net value of the fully informative experiment $\overline{E}$ is depicted by the solid line at a price $t(\theta^H) = 1 - \theta^H$ that leaves the type θ^H indifferent between buying and not buying the fully informative experiment. The dashed line depicts the partial information experiment given by π_1', as described by (18). The associated partially informative experiment E' leaves the low type θ^L indifferent between buying and not buying. As for the high type θ^H, this experiment offers zero value at a positive price, leaving his incentive constraint slack. Correspondingly, for the high type, the net value of the partially informative experiment E' is *strictly* below the net value of the fully informative experiment.

As the incentive-compatibility constraints are slack given this specific menu of experiments for both types, the seller can offer a more informative experiment E'' to the low type θ^L while still satisfying the incentive constraint for the high type θ^H. As the partially informative experiment E'' has a positive price tailored to type θ^L, i.e., $t(\theta^L) = \theta^L \pi_1'' u_1$, the high type can be made indifferent between experiments $\overline{E}$ and E'' by making sure that incentive

[7] Discriminatory menus that do not shut down the low-value type are optimal in nonlinear screening models, such as the monopolistic insurance market studied in Stiglitz (1977). The novel element of selling information to buyers with noncongruent priors is that the seller can more easily extract (possibly all) the information rents.

constraint (17) is binding:

$$\underbrace{\theta^H}_{\Pr(\omega_1)} \cdot \underbrace{\left(1-\pi_1''\right)}_{\text{additional precision}} \cdot u_1 \quad = \quad \underbrace{\left(1-\theta^H\right)u_2}_{t(\theta^H)} - \underbrace{\theta^L \pi_1'' u_1}_{t(\theta^L)} \iff \pi_1'' = \frac{u_1\theta^H - u_2\left(1-\theta^H\right)}{u_1\left(\theta^H-\theta^L\right)}. \tag{20}$$

The experiment E'' characterized in (20) is as informative as possible while satisfying the high type's incentive compatibility constraint and *both* participation constraints with equality. Experiments $\overline{E}$ and E'' are illustrated in Figure 4, again for the case $u_1 = u_2 = 1$.

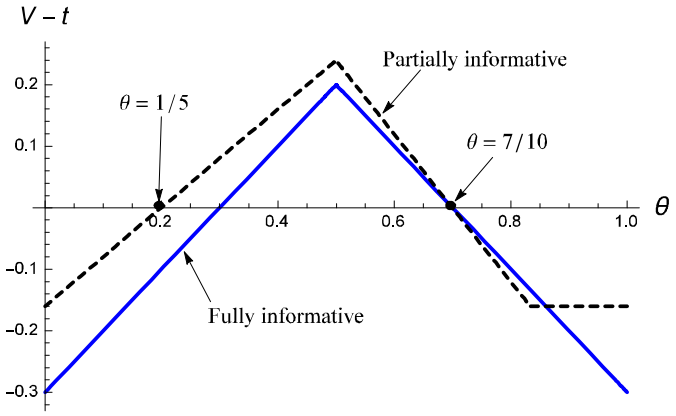

Figure 4: Optimal Menu: $(\pi_1, \pi_2) \in \{(1,1), (4/5, 1)\}$

This example highlights the horizontal aspect of selling information that increases the scope of screening: the high type θ^H buys the perfectly informative experiment; the low type θ^L buys a partially informative experiment; and the seller extracts the entire surplus (which, however, falls short of the socially efficient surplus). The relative frequency of each buyer type determines the shape of the optimal menu. In particular, the partially informative experiment E'' is replaced by the fully informative experiment when there is a high proportion of low types, i.e., when $\gamma < \theta^L/\theta^H$ for all payoffs (u_1, u_2).

We have thus shown the following results for the fully binary model.

Proposition 4 (Partial Information)

1. *With congruent priors, the low type receives either zero or complete information.*
2. *With noncongruent priors, the low type receives either partial or complete information.*

With congruent priors, both types receive complete information if they are sufficiently similar or if the low type is relatively frequent. With noncongruent priors, the optimal menu offers the fully informative experiment to both types if they are sufficiently similar in their prior information, or if the low type is sufficiently frequent. Offering two experiments

becomes optimal if the two types are sufficiently *different* in their level of informativeness. Importantly, the low type always receives some information in that case and is not excluded.[8] Furthermore, the high type receives positive rents only if he is pooled with the low type. Otherwise, the seller extracts the entire surplus that is generated.

We now examine the comparative statics of the optimal menu. Because the high type buys the fully informative experiment $\overline{E}$ and the low type's experiment is concentrated with $\pi_2 = 1$, the optimal experiment $E(\theta^L)$ can be described by its first diagonal entry π_1.

Proposition 5 (Comparative Statics)
The informativeness π_1 of the optimal experiment $E(\theta^L)$ is

1. *decreasing in the probability γ of the high type;*
2. *decreasing in the precision of the low type's prior belief $\left|\theta^L - \theta^*\right|$;*
3. *increasing in the precision of the high type's prior belief $\left|\theta^H - \theta^*\right|$ when priors are congruent or the menu is discriminatory.*

Thus, even though the shape of the optimal menu depends on whether types have congruent or noncongruent priors, the comparative statics of the optimal experiment are robust across the different scenarios. The rent extraction vs. efficiency trade-off is resolved at the expense of the low type as *(i)* the fraction of high types increases, or *(ii)* the low type's willingness to pay for the complete experiment decreases.

Finally, as the high type's willingness to pay for information decreases (his prior becomes more precise), the optimal menu may switch from offering full to partial information to the low type. This occurs because separating the two types becomes profitable whenever the gap in their prior beliefs is sufficiently wide. Once partial information is offered, however, the optimal distortions decrease with the precision of the high type. This accounts for the qualifying statement in Proposition 5.3.

4.3 Continuum of Types

We now complete the analysis of the binary-action environment. We denote the interim belief of the data buyer by $\theta \triangleq \Pr(\omega = \omega_1)$ and consider a continuum of types $\theta \in [0, 1]$ on the unit interval, with a distribution $F(\theta)$ and associated density $f(\theta)$. We shall show that

[8] In Section 6, we establish a stronger result: with many actions and states, there always exists a distribution γ such that partial information is provided to the low type unless the two types agree on both the most likely state *and* the relative likelihood of any other two states. In the binary-state model, the latter condition is vacuously satisfied. However, with more than two states, the latter condition fails generically.

many qualitative properties of the two-type case—including the cardinality of the optimal menu—hold in this setting.

Recall the value of information was described in (15):

$$V(E,\theta) = \max\left\{\theta\pi_1 u_1 + (1-\theta)\,\pi_2 u_2 - \max\{\theta u_1, (1-\theta)\,u_2\}, 0\right\}.$$

We can rewrite the value of information as

$$V(E,\theta) = \max\left\{\theta\left(\pi_1 u_1 - \pi_2 u_2\right) + \pi_2 u_2 - \max\{\theta u_1, (1-\theta)\,u_2\}, 0\right\},$$

and we capture the value of experiment E for type θ via a one-dimensional variable

$$q(\theta) \triangleq \pi_1(\theta)u_1 - \pi_2(\theta)u_2 \in [-u_2, u_1],$$

which describes the *differential informativeness* of the experiment. The endpoints of the interval $[-u_2, u_1]$ identify two extreme experiments $q \in \{-u_2, u_1\}$ that are attained when either one of the two signals occurs with probability one in both states. In either case, the resulting experiment reveals no information to the data buyer. Conversely, the fully informative experiment is given by $q = u_1 - u_2$. Because Proposition 2 establishes that either $\pi_1(\theta) = 1$ or $\pi_2(\theta) = 1$ (or both) for each type θ, we know that $q(\theta) > u_1 - u_2$ implies $\pi_1(\theta) = 1$, and that $q(\theta) < u_1 - u_2$ implies $\pi_2(\theta) = 1$. Collecting terms, we can rewrite the value of an experiment in terms of the one-dimensional variable q as follows:

$$V(q,\theta) = \max\left\{\theta q + u_2 + \min\left\{u_1 - u_2 - q, 0\right\} - \max\{\theta u_1, (1-\theta)\,u_2\}, 0\right\}. \tag{21}$$

The value of information in (21) illustrates the main properties of our screening problem: *(i)* the buyer has a type-dependent participation constraint; *(ii)* the ex ante indifferent type θ^* has the highest willingness to pay for any experiment q; *(iii)* the experiment $q = u_1 - u_2$ is the most valuable for all types θ; *(iv)* different types θ rank partially informative experiments differently; and *(v)* the utility function $V(q,\theta)$ has the single-crossing property in (θ, q).

The single-crossing property indicates that types with a higher θ, those who believe that state ω_1 is more likely, assign a higher value to experiments with a higher q. In turn, these experiments contain a signal that delivers stronger evidence regarding state ω_2, which they deem less likely ex ante. As in the binary-type case, the vertical dimension (quality of the information) and the horizontal dimension (position of the information) cannot be chosen separately by the seller. In particular, it is not possible to change the differential informativeness of the experiment (i.e., to choose a very high or a very low q) without

reducing its overall informativeness.

We know from Proposition 1 that we can focus on *responsive* menus. In the case of binary actions, an experiment is responsive if and only if the value of following both signals' recommendations is non-negative. Therefore, equation (21) implies that experiment $q(\theta)$ offered to type θ is responsive if and only if

$$q(\theta) \in Q(\theta) \triangleq \begin{cases} \left[-u_2, \frac{u_1}{1-\theta} - u_2\right] & \text{for} \quad \theta \leq \theta^*, \\ \left[u_1 - \frac{u_2}{\theta}, u_1\right] & \text{for} \quad \theta \geq \theta^*. \end{cases} \tag{22}$$

In other words, the restriction $q(\theta) \in Q(\theta)$ allows us to eliminate the first max operator from (21) when computing the value $V(q(\theta), \theta)$. We thus obtain a characterization of *implementable* and *responsive* menus of experiments.

Lemma 1 (Implementable and Responsive Menus)
A menu $\{q(\theta)\}_{\theta \in \Theta}$ *is* implementable *and* responsive *if and only if*

$$q(\theta) \in [-u_2, u_1] \text{ is non-decreasing} \tag{23}$$

and

$$\int_0^1 q(\theta)\, d\theta = u_1 - u_2. \tag{24}$$

To obtain some intuition for constraint (24), observe that the interior type θ^* assigns the highest value to any experiment q, and that the utility function $V(q, \theta)$ has a downward kink at $\theta = \theta^*$. In order to compute the buyer's rent function V, we apply the envelope theorem to the two intervals on $[0, \theta^*]$ and $[\theta^*, 1]$ separately. Because information has zero value for types $\theta \in \{0, 1\}$, we know that $V(0) = V(1) = 0$. We thus obtain two (potentially different) expressions for $V(\theta^*)$. Finally, the buyer's gross value of an experiment (21) is jointly continuous in (q, θ), and hence, the rent function V is continuous by the Maximum Theorem.[9] Incentive compatibility then imposes the following restriction on responsive allocations:

$$V(\theta^*) = \int_0^{\theta^*} V_\theta(q, \theta) \mathrm{d}\theta = -\int_{\theta^*}^1 V_\theta(q, \theta) \mathrm{d}\theta.$$

Computing the buyer's marginal rent from (21) and using the definition of θ^* to simplify (**??**) yields the integral condition in the Lemma.[10] Notably, condition (22) does not appear

[9] Intuitively, if the rent function had a downward jump at θ^*, some nearby type $\theta^* + \varepsilon$ could purchase the experiment intended for type $\theta^* - \varepsilon$. This yields a payoff that is arbitrarily close to $V(\theta^* - \varepsilon)$, and hence, gives $\theta^* + \varepsilon$ an incentive to misreport his type.

[10] Our environment with type-dependent participation constraints is an instance of the "high convexity"

in the statement of Lemma 1 because it is implied by the monotonicity condition (23) and by the integral constraint (24).

With this result in place, the transfer $t(\theta)$ associated with every experiment $q(\theta)$ can be computed from the envelope formula on the $[0, \theta^*]$ and $[\theta^*, 1]$ separately. Using integral constraint (24) to simplify further, the seller's problem can be written as

$$\max_{q(\cdot)} \int_0^1 \left[\left(\theta f(\theta) + F(\theta) \right) q(\theta) + \min \left\{ \left(u_1 - u_2 - q(\theta) \right) f(\theta), 0 \right\} \right] \mathrm{d}\theta, \tag{25}$$

subject to constraints (23) and (24).

Observe that the seller's objective (25) is piecewise linear (and concave) in the experiment $q(\theta)$. Thus, absent the integral constraint, the optimal experiments take values at the kinks, i.e., $q^*(\theta) \in \{-u_2, u_1 - u_2, u_1\}$ for all θ. This corresponds to a menu containing the fully informative experiment only. While such a menu is, in fact, optimal under some conditions, the seller can sometimes do better by offering (at most) one additional experiment.

Proposition 6 (Cardinality of the Optimal Menu)
An optimal menu consists of at most two experiments.

To establish this property of optimal menus, we reduce the seller's problem (25) to a linear program with equality and non-negativity constraints. An application of the Fundamental Theorem of Linear Programming (e.g., Theorem 8.4 in Chvatal (1983)) implies that the solution is an increasing step function with at most three jumps.[11] Because it is optimal to set $q(0) = -u_2$ and $q(1) = u_1$, this means that the optimal menu contains at most two informative experiments $q \in (-u_2, u_1)$. When offered, the second experiment contains a signal that perfectly reveals one state. However, the linearity of the environment prevents the seller from offering more than one distorted experiment, i.e., no further *versioning* is optimal. We formalize this intuition in the following subsection.

4.4 Single-Item vs. Discriminatory Pricing

To obtain further insights into the properties of the optimal menu, we refine our approach to the seller's problem. We combine Lagrange methods, as in the type-dependent participation

case in Jullien (2000). As such, the integral condition for implementability (24) differs from budget, capacity, or enforceability constraints because the distribution $F(\theta)$ does not appear in the integrand. A similar condition, for different reasons and with different implications, appears as a persuasion budget constraint in Kolotilin, Li, Mylovanov, and Zapechelnyuk (2015).

[11] This result is related to linear mechanism-design problems with budget constraints, such as Samuelson (1984), Brusco and Hopenhayn (2007), and Fuchs and Skrzypacz (2015). In our setting, the reduction to a linear program is obtained through a change of variable from $q(\theta)$ to its increments, and by imposing the additional constraint that some type $\theta^* \in (0, 1)$ purchases the fully informative experiment.

constraints model of Jullien (2000), with the *ironing* procedure developed by Toikka (2011) extending that in Myerson (1981). This approach allows us to overcome two difficulties posed by our problem: *(i)* integral constraint (24) and objective (25) have generically different weights, $\mathrm{d}\theta$ and $\mathrm{d}F(\theta)$; hence, *(ii)* the problem is non-separable in the type θ and the experiment $q(\theta)$, which interact in two different terms. In particular, the "virtual values" $\phi(\theta, q)$—defined as the partial derivative of the integrand in (25) with respect to q—are a non-constant function of the experiment.

We derive the solution to the seller's problem by maximizing the virtual values. Because the integrand in objective (25) is piecewise linear in q, its partial derivative $\phi(\theta, q)$ takes on only two values. We then define the following two functions

$$\begin{aligned} \phi^{-}(\theta) &\triangleq \theta f(\theta) + F(\theta), \\ \phi^{+}(\theta) &\triangleq (\theta - 1) f(\theta) + F(\theta) \end{aligned}$$

that describe the virtual value $\phi(\theta, q)$ for $q < u_1 - u_2$ and $q \geq u_1 - u_2$, respectively. Heuristically, the two virtual values represent the marginal benefit to the seller of increasing each type's experiment q from $-u_2$ to $u_1 - u_2$ and from $u_1 - u_2$ to u_1. If ironing à la Myerson is required, we denote the *ironed* virtual values as $\bar{\phi}^{-}(\theta)$ and $\bar{\phi}^{+}(\theta)$.

Finally, we say that a menu satisfies the *pooling property* if it is constant on any interval where the relevant (ironed) virtual value is constant.

Proposition 7 (Optimal Menu)
The menu $\{q^*(\theta)\}_{\theta \in \Theta}$ is optimal if and only if the following conditions hold:

1. *there exists $\lambda^* > 0$ such that, for all θ,*

$$q^*(\theta) = \arg \max_{q \in [-u_2, u_1]} \left[\bar{\phi}^{-}(\theta) \min\{q, u_1 - u_2\} + \bar{\phi}^{+}(\theta) \max\{0, q - (u_1 - u_2)\} - \lambda^* q \right];$$

2. *$\{q^*(\theta)\}_{\theta \in \Theta}$ has the pooling property and satisfies integral constraint (24).*

We now illustrate the optimal menu under flat and discriminatory pricing separately.

Corollary 1 (Single-Item Menu)
The optimal menu contains a single item whenever any of the following hold:

1. *almost all types have congruent priors, i.e., $F(\theta^*) \in \{0, 1\}$;*
2. *the monopoly price for experiment $\overline{E}$ is equal on $[0, \theta^*]$ and $[\theta^*, 1]$;*
3. *both virtual values $\phi^{-}(\theta)$ and $\phi^{+}(\theta)$ are strictly increasing.*

To grasp the intuition behind these results, consider a relaxed problem where the seller contracts separately with two groups of buyers, $\theta < \theta^*$ and $\theta \geq \theta^*$. Part (1) essentially addresses this case. Because of the linearity of the problem, the optimal mechanism is a cutoff mechanism: the seller offers only the fully informative experiment to each group, at generically different prices. Part (2) states that, if the cutoff types in the two subproblems have the same willingness to pay, then those cutoffs also solve the unrestricted problem. For instance, when $u_1 = u_2$, any distribution that is symmetric around $\theta = 1/2$ satisfies this condition. Part (3) identifies regularity conditions on the distribution of types that rule out, for example, most bimodal distributions. Under these conditions, the seller prefers to offer the fully informative experiment to all buyers at an intermediate price.

We note that only condition (3) can be stated independently of the matching values (u_1, u_2). For example, an optimal menu contains a single item whenever the values are uniformly distributed, irrespective of the payoffs. Figure 5 describes the optimal menu for the case of constant match values, $u_1 = u_2 = 1$, and uniformly distributed types. Because both virtual values are strictly increasing, they cross the threshold level λ^* only once. Thus, the optimal menu is a step function $q(\theta) \in \{-1, 0, 1\}$ assigning the fully informative experiment $q = 0$ to types $\theta \in [1/4, 3/4]$ at a price $t = 1/4$, and no information to all other types.

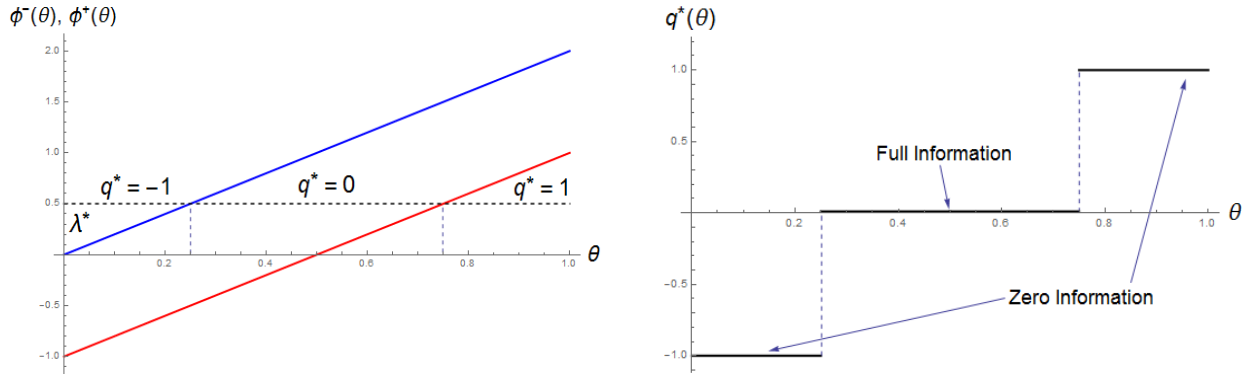

Figure 5: Optimal Allocation with Uniform Distribution

If the conditions of Corollary 1 fail, however, the seller may choose to offer a second (distorted) experiment to one group in order to maintain a high price for the fully informative experiment. Furthermore, Corollary 1.3 implies that the seller offers a second experiment *only if* the distribution of types requires *ironing* of the virtual values. When types correspond to interim beliefs, it is natural to consider bimodal densities that fail the strong regularity conditions and, therefore, introduce the need for ironing. This is the case, for example, if most buyers are well informed ex ante (with most types close to 0 or 1). In other words, ironing is not a technical curiosity in our case but rather a technique that becomes unavoidable because of the properties of the information environment. Figure 6 (left) illustrates the ironed virtual

values for a bimodal probability density function (drawn on a different scale).[12] Figure 6 (right) illustrates the resulting optimal menu, again for the case $u_1 = u_2$.

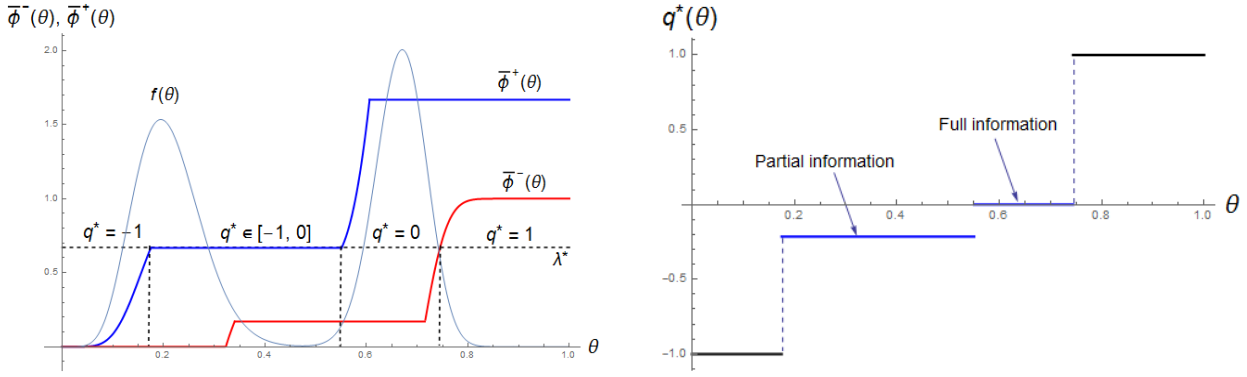

Figure 6: Probability Density Function and Optimal Allocation

As in the optimal menu in the binary-type setting, the partially informative experiment $q \approx -0.21$ contains one signal that perfectly reveals state ω_1. This experiment is relatively unattractive for higher types, and it allows the monopolist to increase the price for the large mass of types located around $\theta \approx 0.7$. Note that the ex ante least informed type θ^* need not purchase the fully informative experiment, despite having the highest value of information and obtaining the highest rent $V(\theta^*)$. Indeed, the type that is indifferent between the two experiments in Figure 6 is $\theta \approx 0.55$. From the seller's perspective, incentivizing type $\theta = 1/2$ to purchase the fully informative experiment would require further distortions (and, hence, a lower price) for the second experiment. This leads to a loss of revenue from the types around $\theta \approx 0.2$. Because such types are quite frequent, this loss more than offsets the gain in revenue from the types around $\theta = 1/2$.

Finally, a sufficient condition for the optimality of two-item menus can be obtained by continuity with the binary-type case. In that setting, when the two types are equally likely, symmetric about $1/2$ and sufficiently well informed (i.e., $\theta^H > 2/3$), the optimal menu is discriminatory for all $u_1 \neq u_2$.

5 Implications for Data Pricing

We discuss how to bring our model's results to bear on the design of real-world information products. We focus on offline and online brokers of big data—firms such as Acxiom, Nielsen, and Oracle—that sell information about individual consumers to business customers.

[12] The distribution in Figure 6 (left) represents a "perturbation" of the two-type example in Section 4.2. It is a mixture (with equal weights) of two Beta distributions with parameters $(8, 30)$ and $(60, 30)$.

Data obtained from brokers is typically used to facilitate marketing efforts and to mitigate risks.[13] Information used for marketing purposes is typically sold through *data appends* and *marketing lists*. Data appends reveal supplemental information about a firm's existing or potential customers, allowing the firm to place them into more precise market segments. For instance, all major data brokers offer data management platforms (DMPs)—customized software that enables websites to track their users and integrate their own data with 3rd-party data. Most risk-mitigation data products offered by credit rating agencies are also of this kind. Conversely, marketing lists facilitate targeted advertising to new consumers. Advertisers can choose whether to acquire standard lists of potential consumers with pre-specified sets of characteristics, or to customize their list of desired consumer attributes.

Our model of selling supplemental information is best suited to analyzing data appends.[14] We therefore describe these information products in greater detail, and we comment on marketing lists and other data services at the end of this section. For concreteness, consider the following examples of data appends.

- Oracle ID-Graph tracks firms' customers across several devices, augmenting the data collected on the firms' websites with behavioral observations from different sources.[15]
- Email Intelligence by TowerData attaches demographic, income, intent, and purchase information to a merchant's own list of email addresses.[16]
- The credit reporting agency Equifax offers its business customers (e.g., banks and credit card companies) a risk-mitigation product called Undisclosed Debt Monitoring. This product tracks an individual borrower to identify new negative information that arrives between the original loan approval and the closing date.[17]

Each of these products is available in several versions, which differ mainly in terms of the number of informative variables the seller discloses to the buyer. For instance, ID-Graph and Email Intelligence allow buyers to customize their queries to the database (e.g., a consumer's age group, income level, interests, and intent). Similarly, the versions of Undisclosed Debt

[13] Marketing and risk-mitigation products generate 46% and 41% of data brokers' revenues, respectively (Federal Trade Commission, 2014).

[14] Many brokers offer both kinds of products. For example, Nielsen Total Audience identifies two ways of packaging its data: a DMP to "expand, optimize, segment and activate your customer data across all marketing channels and platforms" and Data as a Service (DaaS) to "find your target audience segment, or customize your own based on the characteristics that are most important to you." See http://www.nielsen.com/us/en/solutions/capabilities/total-audience.html for more details.

[15] https://www.oracle.com/marketingcloud/products/data-management-platform/id-graph.html

[16] http://www.towerdata.com/email-intelligence/pricing

[17] https://www.equifax.com/business/undisclosed-debt-monitoring

Monitoring differ in terms of the number of potentially negative events (e.g., late payments, credit inquiries, bankruptcy filings) that are monitored and disclosed.

The most common buyers of data products are marketing and advertising firms, lenders and financial services firms, and retail companies (Federal Trade Commission, 2014). The data buyers differ along several dimensions, including their ability to process data and the richness of their action space.[18] Any of these dimensions of heterogeneity can lead to interesting sorting patterns of buyers into product versions. Here we focus on the differences in the strength of the buyers' priors, i.e., in the availability of initial information.

Our main structural results (Propositions 1 and 2) derive the properties of the distribution of states and signals that are associated with monopolistic screening. In particular, our results impose restrictions on the support of the conditional distribution of signals in a given state. In order to leverage these insights to evaluate and inform the design of data appends, it is useful to rephrase the design of a statistical experiment in terms of hypothesis testing. Our structural results identify the types of *statistical errors* incurred by data buyers as a consequence of market power in the sale of information.[19]

For the present purpose, consider a data buyer, such as an advertiser or a lender. The data buyer wishes to test the null hypothesis "target with an ad" or "grant a loan" for a specific consumer, i.e., to distinguish a null hypothesis H_0 from an alternative hypothesis H_1. The data buyer is a Bayesian decision maker with a prior distribution over the hypotheses. He can take one of two actions, each of which is optimal under the respective hypothesis.

The data seller has a *test statistic* whose distribution, conditional on the true state of the world, is given by H_0 or H_1 as in Figure 7. We assume that H_0 and H_1 satisfy the monotone likelihood ratio property. While the data seller could potentially disclose arbitrary information about the distribution of the test statistic, suppose she chooses to inform the data buyer if the statistic is above or below a certain threshold. The data buyer then chooses the corresponding action. We can represent this experiment as a statistical test of the null hypothesis $H_0 = \{\omega_2\}$,

E	s_1	s_2
ω_1	$1-\beta$	β
ω_2	α	$1-\alpha$

[18] For instance, a local bank deciding whether to give a mortgage at the prevailing rate has a coarser action space than a major credit card company deciding on a new account's credit limit and interest rate.

[19] Any data broker enjoys some degree of market power to the extent that its data sources are not perfectly correlated with those of other brokers or it has a superior ability to process information. Furthermore, it can be shown that the structural properties of optimal menus (Proposition 2) hold even in an imperfectly competitive setting with differentiated products (e.g., competition in nonlinear prices among sellers with partially correlated databases). In that case, non-dispersed and concentrated experiments are the most profitable instruments through which to provide any given utility level to the data buyer.

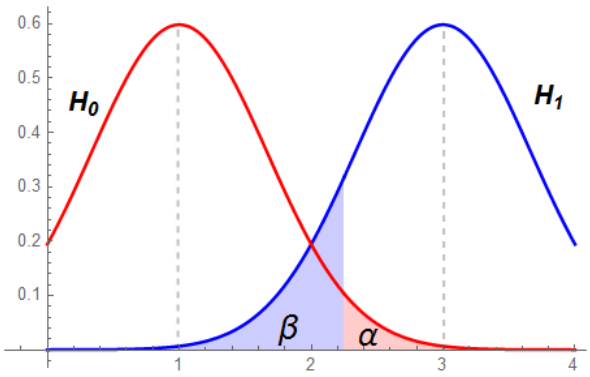

Figure 7: Conditional Distributions of the Test Statistic

where α and β denote the probability of a type I and type II statistical error, respectively.[20]

Given the information contained in the database of the seller in Figure 7, the set of feasible statistical tests is described by the area between the blue curve and the red line in the left panel of Figure 8. The blue curve identifies the loci of the least type I and type II statistical errors given the data *available*. As the data seller can always introduce noise into the test statistic, the set of feasible statistical tests is given by the entire area. In our model, we assumed that the data seller has complete information, and thus, the boundary is given by the blues lines in the right panel of Figure 8, which coincide with the α and β axes.[21]

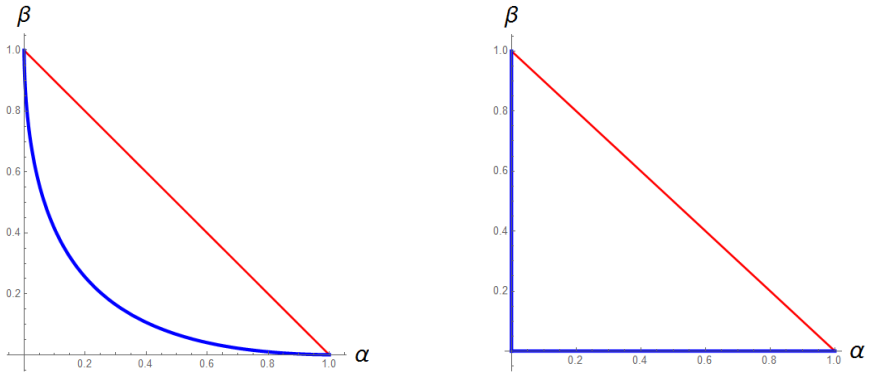

Figure 8: Feasible Information Structures

The central issue for the data seller is that she does not know the data buyer's prior beliefs and, hence, the buyer's willingness to pay for this information. She must therefore employ a richer mechanism to screen heterogeneous buyers. In particular, the seller offers the buyer a menu of binary ("pass/fail") tests. Each test reports the outcome "pass" when

[20] A type I error leads the decision maker to reject the null hypothesis even though it is true. By contrast, a type II error leads the decision maker to accept the null hypothesis even though it is false.

[21] Because distributions H_0 and H_1 satisfy the monotone likelihood ratio property, the feasibility frontier is spanned by threshold tests. As the informativeness of the statistic increases, the feasible set expands and approaches our unrestricted setting.

the test statistic is below a particular threshold. Each test yields a different combination of type I and type II errors (α, β).

Our main structural result (Proposition 2) identifies systematic patterns in the optimal design of partially informative statistical tests. With binary states and actions, and no constraints on (α, β), the seller induces some buyers to make *either* type I *or* type II errors, but not both. The logic underlying this result extends to the case of a partially informed seller: an identical argument establishes the stronger result that all optimal tests lie on the lower boundary of the feasible set in Figure 8 (left). Separation in the optimal menu is then supported by the differences in the error structure of each test and by the buyers' heterogeneous preferences over statistical errors.

A concrete implication of the optimality of non-dispersed experiments is that no product in an optimal menu should add unbiased noise to the seller's information. Adding idiosyncratic noise can be useful when multiple buyers compete in a downstream market, as in Admati and Pfleiderer (1986), or when it is important to preserve the anonymity of the data, as in the differential privacy literature (Dwork, 2008). In the absence of these concerns, our results suggest that this practice reduces the seller's revenues. Instead, optimal experiments should minimize the type-II error for any level of type-I error (as in Figure 8). For example, if the data buyer faces a binary advertising decision, the optimal menu should lead to excessively broad *or* to excessively narrow campaigns, i.e., advertising to a subset of high-value consumers but not to low-value consumers, or to all high-value and to some low-value consumers.

Interestingly, in practice, none of the data products described earlier appear to introduce noise into the data. For example, credit rating agencies do not offer both precise and noisy versions of the same information (e.g., computing a consumer's credit score on the basis of more or less detailed data). Instead, information is degraded by revealing only a portion of the available data to the buyer. We now discuss how omitting explanatory variables can be seen as implementing our optimal mechanism under specific conditions on the buyer's decision problem and on the seller's data.

Consider the following stylized description of the above-mentioned Equifax product Undisclosed Debt Monitoring. A credit rating agency collects K binary characteristics $x_k \in \{0, 1\}$ for each potential borrower. Each realization $x_k = 1$ corresponds to a "red flag" on the borrower's record. Suppose, for simplicity, that the data buyer's (the lender's) payoff-relevant state is $\omega = \mathbf{1}\left[\Sigma_{k=1}^{K} x_k = 0\right]$, i.e., it is optimal to grant a loan if and only if there are no red flags. The data seller knows the borrower's vector of characteristics $(x_1, \ldots, x_K)$, while the data buyer privately knows the (identical) probability distribution of each binary characteristic x_k. In this case, hiding the realized value of any characteristic x_k induces the data

buyer to grant a loan to an unqualified borrower with positive probability, but no qualified borrower is ever turned down.

In this example, the data buyer incurs only one type of statistical error. Thus, omitting variables can be part of an optimal design if subsets of the seller's data contain conclusive evidence *against* at least one state. In other settings, withholding data may introduce full-support noise into the experiment. This is the case any time a hidden variable could have overturned the signals based on the revealed data and changed the buyer's action *in any direction*. For example, it is not difficult to construct examples in which a broker (e.g., Oracle or Nielsen) provides demographic, income and interest information about consumers, but using demographic and income data as proxies for interest leads the data buyer to incur both types of statistical errors.

Data brokers offer several other information products in addition to data appends. For instance, marketing lists are queries to a broker's database of individual consumer records. Such queries allow advertisers to identify potential customers with a set of desired characteristics.[22] Thus, a marketing list can be viewed as a statistical experiment where each signal indicates (possibly with error) a pre-specified consumer segment. From a modeling standpoint, however, selling a list is very different from appending data to the buyer's existing information. In particular, when selling a list, the seller is able to charge a price *contingent* on the *signal realization*.[23] Marketing lists can also be sold contextually to advertising space. This occurs when a data provider (e.g., Bluekai) partners with a publisher of space (Doubleclick) and charges the advertiser a price per impression in addition to the cost of the advertising space. Thus the price of the data augments the cost of the ads. In other words, the buyer's advertising decision is contractible.[24]

6 Optimal Menu with Many Actions

We now extend the analysis of the optimal menu to environments with many actions and many states. In order to make progress in this richer environment, we restrict our attention to the case of binary types *and* matching payoffs defined in (2). We then provide a suitable generalization of the optimal menu derived in Section 4. We continue to define the low

[22] For example, the online invitation website Evite sells lists of attendees of events at specific locations; the share-buttons provider AddThis sells lists of internet users' tastes for news; and Mailways sells lists of physical mailing addresses at the mail carrier route level (Anderson and Simester, 2013).

[23] Babaioff, Kleinberg, and Paes Leme (2012) provide further insight into the role of realization-contingent pricing. In our model, contracting on signal realization leads to the first-best profits for the seller.

[24] In our setting, contracting on the buyer's action leads to the first-best profits when the payoff of matching state and action is constant, though the same is not true in more general environments. Data provider-publisher partnerships can be analyzed more comprehensively in the setting of Eső and Szentes (2007b).

and the high type $\theta \in \left\{\theta^L, \theta^H\right\}$ such that the high type values the completely informative experiment more than the low type, as in (16).

The construction of the optimal menu now proceeds differently than in Section 4. We first solve a *relaxed* problem wherein we require the high type θ^H to take action a_i upon observing signal s_i even when buying the experiment destined for the low type θ^L. We then show that the solution to the relaxed problem—which disregards the obedience constraints—indeed satisfies the original constraints; hence, it also solves the full problem. In other words, we first guess and then verify that the optimal mechanism is *obedient* on and off the equilibrium path. In Example 3, we show that this relaxed approach fails to deliver a valid solution with more than two types. Thus, a different approach is required if we want to consider problems with arbitrary cardinality in *both* the action/state space *and* the type space.

In the relaxed problem, we replace the high type's incentive-compatibility constraint

$$V(\theta^H) \geq V\left(E(\theta^L), \theta^H\right) - t(\theta^L) = \sum\nolimits_{j=1}^{I} \max_i \left\{\theta_i^H u_i \pi_{ij}(\theta^L)\right\} - \max_i \left\{\theta_i^H u_i\right\} - t(\theta^L),$$

with the weaker constraint

$$V(\theta^H) \geq \sum\nolimits_{i=1}^{I} \theta_i^H u_i \pi_{ii}(\theta^L) - \max_i \left\{\theta_i^H u_i\right\} - t(\theta^L). \tag{26}$$

The relaxed version of the constraint drops the max operator and simply asks type θ^H to accept the recommendation a_i implicit in signal s_i.

Proposition 3 established that both the low type's participation constraint and the high type's incentive constraint bind. This reduces the seller's problem to choosing the diagonal entries of the low-value type's experiment $\pi_{ii}(\theta^L) \in [0,1]$. Substituting $t(\theta^L) = V(E(\theta^L), \theta^L)$ in the right-hand side of (26) above, the seller maximizes

$$(1-\gamma) \underbrace{\left(\sum\nolimits_{i=1}^{I} \theta_i^L u_i \pi_{ii}(\theta^L) - \max_i \theta_i^L u_i\right)}_{t(\theta^L) = V(E(\theta^L), \theta^L)} + \gamma \underbrace{\left(\sum\nolimits_{i=1}^{I} \theta_i^H u_i - \max_i \theta_i^H u_i - V(\theta^H)\right)}_{t(\theta^H) = V(\overline{E}, \theta^H) - V(\theta^H)} \tag{27}$$

subject to the high type's participation constraint

$$V(\theta^H) = \sum\nolimits_{i=1}^{I} (\theta_i^H - \theta_i^L) u_i \pi_{ii}(\theta^L) - \max_i \left\{\theta_i^H u_i\right\} + \max_i \left\{\theta_i^L u_i\right\} \geq 0. \tag{28}$$

The relaxation of the incentive constraints—fixing a mapping from signals to actions for both types and experiments—turns the seller's problem into a linear program.

Intuitively, the seller's choice of a partially informative experiment $E(\theta^L)$ is guided by the *disagreement* in the beliefs over states of the two types. The seller is most willing to

introduce noise into signals about states that the low type considers relatively less likely than the high type. The resulting distortions in the decisions facilitate screening without sacrificing too much of the surplus. To formalize this intuition, we re-order the states ω_j by the likelihood ratios of the two types' beliefs. In particular, let

$$\frac{\theta_1^L}{\theta_1^H} \leq \cdots \leq \frac{\theta_i^L}{\theta_i^H} \leq \cdots \leq \frac{\theta_I^L}{\theta_I^H}. \tag{29}$$

A basic application of the Lagrange multiplier method yields the following characterization result, which generalizes the optimal experiment in Section 4.2.

Proposition 8 (Optimal Menu with Two Types)
There exists a critical state $i^* < I$ such that an optimal experiment $E(\theta^L)$ has $\pi_{ii} = 0$ for all $i < i^*$ and $\pi_{ii} = 1$ for all $i > i^*$.

Because, without loss of generality, the low type chooses action a_i when observing signal s_i, a nil diagonal entry $\pi_{ii} = 0$ means signal s_i is never sent by experiment $E(\theta^L)$. Unlike in the binary action case, the partially informative experiment $E(\theta^L)$ may thus contain fewer signals than the available actions, as the seller drops some signals to reduce the information rent of the high type. Furthermore, Proposition 8 implies that the optimal experiment $E(\theta^L)$ has a lower triangular shape, with at most one strictly interior diagonal entry $\pi_{i^*i^*} \in (0, 1)$:

$$\begin{array}{c|ccccccc} E(\theta^L) & s_1 & \cdots & & s_{i^*} & \cdots & & s_I \\ \hline \omega_1 & 0 & \cdots & 0 & \pi_{1i^*} & \cdots & & \pi_{1I} \\ \vdots & & & & \vdots & & & \vdots \\ \omega_{i^*} & \vdots & & \vdots & \pi_{i^*i^*} & \cdots & & \pi_{i^*I} \\ & & & & 0 & 1 & & 0 \\ \vdots & & & & \vdots & & \ddots & \\ \omega_I & 0 & \cdots & 0 & 0 & 0 & & 1 \end{array} . \tag{30}$$

In other words, the seller chooses a subset of "targeted" states $i \in \{i^*, ..., I\}$ in which the buyer takes the correct action with positive probability. Conversely, when a "residual" state $i \in \{1, ..., i^* - 1\}$ is realized, the buyer never takes the correct action. Residual states are used to degrade the information revealed about the targeted states. The seller chooses how to partition states based on two factors: the two types' relative beliefs; and the relative frequency of each type.

In order to construct the optimal experiment $E(\theta^L)$, we substitute the expression for the information rent $V(\theta^H)$ from (28) into the seller's objective (27). Up to an additive constant,

the seller's profits are given by

$$\sum\nolimits_{i=1}^{I}(\theta_i^L - \gamma\theta_i^H)u_i\pi_{ii}(\theta^L). \tag{31}$$

The shape of the optimal experiment then depends on whether the participation constraint of the high type binds. In particular, if (28) is slack, the seller's profits (31) are maximized by assigning only extreme values to the diagonal, $\pi_{ii} \in \{0, 1\}$. Because the likelihood ratios θ_i^L/θ_i^H are increasing, we have $\pi_{ii} = 1$ for all $i \geq i_s$, where

$$i_s \triangleq \min\{i : \gamma \leq \theta_i^L/\theta_i^H\}. \tag{32}$$

To determine whether the participation constraint of the high type is satisfied by this solution, define the following function

$$R(j) \triangleq \sum\nolimits_{i=j}^{I}(\theta_i^H - \theta_i^L)u_i - \max_i\left\{\theta_i^H u_i\right\} + \max_i\left\{\theta_i^L u_i\right\}.$$

This expression corresponds to the rent of the high type if the seller drops all signals s_i (and, hence, recommended actions a_i) with $i < j$ and sets $\pi_{ii} = 1$ for all $i \geq j$. In particular, $R(1)$ is the difference in the value of full information between the high and the low type. By construction, it is positive. We further note that $R(j)$ is strictly decreasing in j if $\theta_j^H > \theta_j^L$ and increasing otherwise. Furthermore, because of the likelihood ratio order on states i, it attains its minimum at $i_c \triangleq \min\{i : 1 \leq \theta_i^L/\theta_i^H\}$, where $i_c \geq i_s$. In the proof of Proposition 9, we show that $R(i_c) < 0$. Therefore, if $R(i_s) \geq 0$, the constraint (28) is slack at the optimum.

Conversely, if $R(i_s) < 0$, then the solution to unconstrained problem (31) violates the high type's participation constraint. Therefore, constraint (28) must bind at the optimum. In particular, the seller chooses a state i_b and a diagonal entry $\pi_{i_b i_b} \in [0, 1]$ to satisfy (28) with equality when, in addition, $\pi_{ii} = 1$ for all states $i > i_b$ and $\pi_{ii} = 0$ for all $i < i_b$. By the properties of $R(j)$, this critical state i_b is given by the unique solution to

$$R(i_b) > 0 > R(i_b + 1).$$

Intuitively, state i_b corresponds to the minimum number of signals (and corresponding action recommendations) that must be eliminated in order to satisfy the high type's incentive constraint while extracting all the rent. Proposition 9 establishes that the participation constraint of the high type binds if and only if $i_b < i_s$.

Proposition 9 (Information Rents)

1. *The critical state ω_{i^*} in experiment (30) is given by $i^* = \min\{i_s, i_b\}$.*
2. *If $i_b \geq i_s$, then $E(\theta^L)$ has $\pi_{i^*i^*} = 1$ and $V(\theta^H) > 0$.*
3. *If $i_b < i_s$, then $E(\theta^L)$ has $\pi_{i^*i^*} = \pi^B_{i^*}$ and $V(\theta^H) = 0$.*

To grasp the intuition, note that the definition of state i_b does not depend on the distribution of types, while state i_s is increasing in the fraction of high types γ. The fraction of high types represents the shadow cost of providing information to the low types. When this opportunity cost is low, the monopolist prefers to limit distortions and concede rents to the high type. As γ increases, the informativeness of the low type's experiment decreases. For example, if $\gamma < \theta^L_1/\theta^H_1$ (i.e., $i_s = 1$), Proposition 9 implies that both types purchase the fully informative experiment $\overline{E}$, and that the high type obtains positive rents. Conversely, if $\gamma > \theta^L_i/\theta^H_i$ for all i such that $\theta^L_i/\theta^H_i < 1$, then $i_s > i_b$ and the high type obtains no rent.

To complete the description of the optimal experiment $E(\theta^L)$, we need to specify the distribution of signals π_{ij} for $i \leq i^*$. We construct an off-diagonal assignment procedure that induces both types to follow the recommendation of every signal. In particular, for every state $i \leq i^*$, we begin with the last signal s_I and assign the off-diagonal probabilities π_{iI} such that the high type is indifferent between actions a_i and a_I. We then proceed backward to signal s_{I-1} preserving indifference and placing the residual probability, if any, on π_{ii^*}. We show that the high type prefers action a_{i^*} to any action a_i with $i < i^*$. Therefore, the solution to the relaxed problem satisfies the original constraints. We illustrate the construction of the optimal experiment and the implications for information rents in Example 1 in the Supplemental Appendix.

Earlier, we defined any two types θ^L and θ^H as congruent if they shared the same optimal action a_i for some i given their interim beliefs:

$$\arg\max_{a_i \in A} \left\{ \sum\nolimits_{j=1}^{J} \theta^L_j u_{ij} \right\} = \arg\max_{a_i \in A} \left\{ \sum\nolimits_{j=1}^{J} \theta^H_j u_{ij} \right\} = a_i. \tag{33}$$

We now define two types to be *strongly congruent* if the posterior belief of θ^L can also be represented as a convex combination of θ^H and the vertex of the probability simplex identifying state ω_i, i.e., if (33) holds and there exists $\lambda \in [0, 1]$ such that

$$\theta^L = \lambda \theta^H + (1 - \lambda)\,(0, ..., 0, 1, 0, ..., 0)\,.$$

In other words, if θ^L and θ^H are strongly congruent, then θ^L lies on a ray that goes through the vertex $(0, ..., 0, 1, 0, ..., 0)$ and θ^H. An implication of this geometric condition is that the types θ^L and θ^H have a constant likelihood ratio $\theta^L_j/\theta^H_j = \lambda$ across all states $\omega_j \neq \omega_i$.

With more than two states, discriminatory pricing can now be profitable even if the two types are congruent, as long as they are not strongly congruent.

Proposition 10 (Partial Information)

1. *With strongly congruent priors, type θ^L receives zero or full information.*
2. *Without strongly congruent priors, type θ^L receives zero, partial or full information.*

Proposition 10 thus strengthens and generalizes Proposition 4. With more than two states, the seller can exploit disagreement along *any dimension* and extract all the surplus through discriminatory pricing. Example 2 in the Supplemental Appendix illustrates how congruent, but not strongly congruent beliefs, allow for surplus extraction.

With more than two types (and more than two actions and states), our relaxed approach is not always valid, i.e., the optimal menu leads different types to choose different actions in response to the same signal realizations. In particular, type θ need not follow the recommendation of every signal in all experiments $E(\theta')$, for $\theta \neq \theta'$. We provide an instance of this additional issue in Example 3 of the Supplemental Appendix.

7 Conclusion

We have studied a monopolist who sells supplemental information to privately informed buyers. The resulting screening problem reflects several key properties of information goods that set it apart from traditional models of price discrimination.

First, the Bayesian nature of the buyers' decision making is fundamental to the seller's problem. Differences in the buyers' private beliefs introduce a novel aspect of horizontal differentiation that widens the seller's scope for price discrimination.

Second, information is inherently rich and can be modified in many ways. Even one-dimensional data, such as a consumer's credit score, can be turned into a rich set of information products, as some data buyers may want to identify consumers with excellent scores, while other data buyers may be concerned with avoiding consumers with very low scores.

Third, ultimately, instrumental information is useful as long as it guides the decision of the data buyer. Information therefore enters as an input into the buyer's decision problem. Thus, buyers with different private beliefs may act differently upon receiving the same supplemental information. We have shown that the buyer's ability to adjust his behavior in response to new information complicates the seller's screening problem by introducing the possibility of double deviations. In this respect, selling information is akin to selling multiple inputs

to heterogeneous buyers who can combine them in different ways, according to a privately known production technology.

In this paper, we have deliberately relied solely on belief heterogeneity to motivate sales of supplemental information. In practice, however, buyers of information may differ along several alternative or additional dimensions, including their cost of choosing specific actions, their ability to process data, or their preferences for timely information.[25] Each of these extensions can be implemented within the framework we have outlined. Combining different sources of heterogeneity appears more challenging but promises to yield additional insights. Finally, we have focused on the packaging or versioning problem of a seller who is free to acquire and degrade information. Thus, our results represent only a first pass at understanding the trade-offs involved in selling information products. A richer model would distinguish the fixed cost of acquiring the information (e.g., building a database) from the variable cost of duplicating, distributing, and potentially degrading the available information.

[25] For example, the Consumer Sentiment Index (released by Thomson-Reuters and the University of Michigan) screens buyers of time-sensitive data by offering information products that differ only in the timing of their availability.

A Appendix

Proof of Proposition 1. Consider any type θ and experiment $E = (S, \pi)$. Without loss of generality, let the type choose a single action after each signal.[26] Let S^a denote the set of signals in experiment E that induces type θ to choose action a. Thus, $\cup_{a\in A} S^a = S$. Construct experiment $E' = (S', \pi')$ as a recommendation for type θ based on experiment E, with signal space $S' = \{s_a\}_{a\in A}$ and

$$\pi'(s_a|\omega) = \int_{S^a} \pi(s \mid \omega)\, \mathrm{d}s \quad \omega \in \Omega, a \in A.$$

By construction, E' and E induce the same outcome distribution for type θ; hence, $V(E', \theta) = V(E, \theta)$. Moreover, E' is a garbling of E. By Blackwell's theorem, we have $V(E', \theta') \leq V(E, \theta')$ for all θ'. Therefore, for any incentive-compatible and individually rational direct mechanism $\{E(\theta), t(\theta)\}$, we can construct another direct mechanism $\{E'(\theta), t(\theta)\}$ whose experiments lead type θ to take action a after observing signal $s_a \in S'(\theta)$ that is also incentive compatible and individually rational, thus yielding weakly larger profits. ■

Proof of Proposition 2. (1.) The argument is given in the text.
(2.) Let $\mathcal{M} = \{E(\theta), t(\theta)\}$ be an individually rational and incentive-compatible direct mechanism. Fix an experiment $E \in \mathcal{M}$, let π_{ij} denote the conditional probability of signal s_j in state ω_i, and suppose $\pi_{ij} > 0$ for all i and j. We argue the seller can improve her profits by replacing E with a non-dispersed experiment. By Proposition 1, we can restrict attention to responsive experiments with J signals. Hence, the value of experiment E for any obedient type θ is given by

$$V_{\text{ob}}(E, \theta) = \sum\nolimits_{i=1}^{I} \theta_i \sum\nolimits_{j=1}^{J} \pi_{ij} u_{ij} - u(\theta). \tag{34}$$

For each state ω_i, define the *worst* action $a_{j(i)}$ and the corresponding signal $s_{j(i)}$, where $j(i) \in \arg\min_j u_{ij}$. Now let

$$\varepsilon_i \triangleq \frac{\eta}{u_{ii} - u_{ij(i)}} \quad \text{with} \quad \eta \triangleq \min_i \left\{ \left(u_{ii} - u_{ij(i)} \right) \pi_{ij(i)} \right\},$$

and construct a new experiment E' where $\pi'_{ii} = \pi_{ii} + \varepsilon_i$ and $\pi'_{ij(i)} = \pi_{ij(i)} - \varepsilon_i$ for all i.

[26] Any signal inducing a mixed action could first be split (independently of ω) into subsignals that each induce one of the pure actions in the support the mixed action. The resulting experiment still satisfies obedience and truth-telling and induces the same outcome.

Experiment E' is *non-dispersed* by construction, i.e., $\pi_{i^*j(i^*)} = 0$, where

$$i^* \in \arg\min_i \{(u_{ii} - u_{ij(i)})\pi_{ij(i)}\}.$$

We now argue that the seller can improve his profits and relax the incentive constraints. Using (34), we write the incremental value of experiment E' for an obedient type θ as follows:

$$V_{\text{ob}}(E', \theta) - V_{\text{ob}}(E, \theta) = \sum\nolimits_{i=1}^{I} \theta_i \left(u_{ii} - u_{ij(i)}\right) \varepsilon_i = \sum\nolimits_{i=1}^{I} \theta_i \eta = \eta.$$

The seller can therefore increase the price of experiment E by exactly η and leave the net utility of any truth-telling type θ unchanged.

Now consider any other type θ' who chooses a different action $a_j \neq a_{j(i)}$ after signal $s_{j(i)}$ from experiment E. The marginal benefit to type θ' in state ω_i from the shift in probability from $\pi_{ij(i)}$ to π_{ii} is given by

$$u_{ii} - u_{ij} \leq u_{ii} - u_{ij(i)}$$

by definition of action $j(i)$. Furthermore, if the discrete shift in probabilities causes type θ' to change his action in response to a given signal, the average benefit of such a shift will be a convex combination of $u_{ii} - u_{ij}$ and $u_{ii} - u_{ij(i)}$. Therefore, the value of experiment E' for type θ' exceeds the value of E by at most by η.

Finally, suppose the original experiment was intended for type θ, i.e., $E = E(\theta)$. The direct mechanism $\mathcal{M}'$, where $E'(\theta)$ replaces $E(\theta)$ and $t(\theta) + \eta$ replaces $t(\theta)$, is individually rational and incentive compatible. Moreover, experiment $E'(\theta)$ is non-dispersed by construction, and all transfers are weakly greater than in the original mechanism $\mathcal{M}$.
(3.) For some utility functions u_{ij}, multiple actions can be critical for any given state. The uniform-improvement procedure can then be applied to an experiment as long as it assigns positive probabilities to critical actions in every state. As a result, every optimal experiment will contain one row with a number of zero entries greater than or equal to $\min_i |\arg\min_j u_{ij}|$. In the case of matching utility, $|\arg\min_j u_{ij}| = J-1$ for all i; hence, every optimal experiment is *concentrated*. ■

Proof of Proposition 3. (1.) We know from Proposition 2 that at least one type must buy the fully informative experiment $\overline{E}$. Suppose only type θ^L buys $\overline{E}$ as part of the optimal menu. Then the price of $\overline{E}$ is at most $V(\overline{E}, \theta^L)$. By incentive compatibility, if the high type θ^H purchases $E \neq \overline{E}$, it must be that $t(\theta^H) < V(\overline{E}, \theta^L)$. Therefore, eliminating experiment $E(\theta^H)$ from the menu strictly improves the seller's profits, yielding a contradiction.
(2.) The participation constraint of θ^L must bind. Indeed, some participation constraint must bind, otherwise the seller could increase both prices. Suppose the constraint of θ^H is

binding and that of θ^L is not. Since θ^H is served by $\overline{E}$, then $t(\theta^H) = V\left(\overline{E}, \theta^H\right) \geq V\left(\overline{E}, \theta^L\right)$. Hence, we can increase the price $t(\theta^L)$ without violating incentive compatibility.
(3.) The incentive constraint of type θ^H must bind. Suppose not and then consider two cases: if the participation constraint of type θ^H does not bind, we can increase $t(\theta^H)$; if the participation constraint of type θ^H does bind, then it must be that $E(\theta^L)$ is not equal to $\overline{E}$. Since payoffs are continuous in π_{ij} we can increase both the informativeness of $E(\theta^L)$ and the price $t(\theta^L)$. ■

Proof of Proposition 4. (1.) Recall the definition $\theta^* = u_2/(u_1 + u_2)$, and consider the case of congruent priors $\theta^L > \theta^H > \theta^*$. It follows from Proposition 9 that an optimal menu contains only the fully informative experiment $\overline{E}$. The value of experiment $\overline{E}$ is given by $(1-\theta^H)u_2$ and $(1-\theta^L)u_2$ for the high type and the low type, respectively. The profits from selling to one or both types are given by $\gamma\left(1-\theta^H\right)u_2$ and $(1-\theta^L)u_2$, respectively. Thus, it is optimal to serve both types if and only if $\gamma \leq (1-\theta^L)/(1-\theta^H)$.
(2.) Consider the case of noncongruent priors, $\theta^L < \theta^* < \theta^H$. Let $q \triangleq \pi_{11}u_1 - \pi_{22}u_2$ and denote $q^L = q(\theta^L), q^H = q(\theta^H)$. It follows from Proposition 9 that in an optimal menu, we have $q^L \leq q^H = u_1 - u_2$. If $\gamma < \theta^L/\theta^H$, we wish to show that flat pricing is optimal, i.e., $q^L = q^H = u_1 - u_2$ and $t_1 = t_2 = (1-\theta^L)u_1$. Fix an incentive-compatible menu $\left(q^H, q^L, t^H, t^L\right)$ with $q^L < u_1 - u_2$ and define the following modification:

$$\left(q^{H\prime}, t^{H\prime}, q^{L\prime}, t^{L\prime}\right) = \left(q^H, t^H - \varepsilon\left(\theta^H - \theta^L\right), q^L + \varepsilon, t^L + \varepsilon\theta^L\right). \tag{35}$$

If $q^L < u_1 - u_2$, modification (35) with $\varepsilon \in (0, u_1 - u_2 - q_L)$ preserves incentive compatibility. Furthermore, because $\gamma < \theta^L/\theta^H$, this improves profits by at least

$$\varepsilon\theta^L(1-\gamma) - \varepsilon\left(\theta^H - \theta^L\right)\gamma > 0,$$

which yields a contradiction. Finally, we argue that if $\gamma > \theta^L/\theta^H$, then discriminatory pricing is optimal. First, notice that the individual rationality of the high type must bind, i.e., $q^H = u_1 - u_2$ and $t^H = (1-\theta^H)u_2$; otherwise, modification (35) would be profitable for some $\varepsilon < 0$. Second, q^L and t^L maximize the payment of the low type, subject to his individual-rationality constraint and to the high type's incentive-compatibility constraint. At the optimum, both constraints bind, and the solution is given by

$$q^L = \frac{\theta^H u_1 - (1-\theta^L)u_2}{\theta^H - \theta^L}, t^L = \frac{\left(\theta^H u_1 - (1-\theta^H)u_2\right)\theta^L}{\theta^H - \theta^L}.$$

Substituting the definition of q yields expression (20) in the text. ■

Proof of Proposition 5. If priors are noncongruent and the optimal menu is discriminatory, experiment $E(\theta^L)$ has

$$\hat{\pi}_1 = \frac{u_1\theta^H - u_2\left(1-\theta^H\right)}{u_1\left(\theta^H - \theta^L\right)} \leq 1.$$

The informativeness π_1 is increasing in types: $\partial\hat{\pi}_1/\partial\theta^H \geq 0$ and $\partial\hat{\pi}_1/\partial\theta^L \geq 0$. The proposition then follows directly from the optimal menu characterization presented in the text.
(1.) If priors are congruent then $\pi_1 = 1$ for $\gamma \leq (1-\theta^L)/(1-\theta^H)$ and $\pi_1 = 0$ otherwise. If priors are noncongruent then $\pi_1 = 1$ for $\gamma \leq \theta^L/\theta^H$ and $\pi_1 = \hat{\pi}_1$ otherwise.
(2.) If priors are congruent then $\left|\theta^L - \theta^*\right| = \theta^L - \theta^*$. The optimal menu has $\pi_1 = 1$ if $\theta^L \leq 1-\gamma\left(1-\theta^H\right)$ and $\pi_1 = 0$ otherwise. If priors are noncongruent then $\left|\theta^L - \theta^*\right| = \theta^* - \theta^L$. The optimal menu has $\pi_1 = 0$ if $\theta^L < \gamma\theta^H$ and $\pi_1 = \hat{\pi}_1$, which is increasing in θ^L, otherwise.
(3.) By the normalization of a high type $\left|\theta^H - \theta^*\right| = \theta^H - \theta^*$. If priors are congruent then the optimal menu is $\pi_1 = 0$ for $\theta^H < 1 - \left(1-\theta^L\right)/\gamma$ and $\pi_1 = 1$ otherwise. If priors are noncongruent and the menu is discriminatory, then $\pi_1 = \hat{\pi}_1$, which is increasing in θ^H. ■

Proof of Lemma 1. We begin with necessity. Consider a responsive menu $\{q(\theta)\}_{\theta\in\Theta}$ and any two types $\theta_2 > \theta_1$ who follow the recommendations of signals in both experiments q_1 and q_2. We can then rewrite the net utility of experiment q in (21) as

$$V(q,\theta) = \theta q + u_2 + \min\{u_1 - u_2 - q, 0\} - \max\{\theta u_1, (1-\theta)u_2\}. \tag{36}$$

Because the menu is implementable, the incentive-compatibility constraints imply

$$\begin{aligned} V(q_1,\theta_1) - t_1 &\geq V(q_2,\theta_1) - t_2, \\ V(q_2,\theta_2) - t_2 &\geq V(q_1,\theta_2) - t_1, \\ V(q_2,\theta_2) - V(q_1,\theta_2) \geq t_2 - t_1 &\geq V(q_2,\theta_1) - V(q_1,\theta_1). \end{aligned}$$

The strict single-crossing property of $V(q,\theta)$ in (36) implies $q_2 \geq q_1$; hence, $q(\theta)$ is increasing. Because the buyer's rent is differentiable with respect to θ on $[0,\theta^*]$ and $[\theta^*,1]$ respectively, we can compute the function $V(\theta)$ on these two intervals separately. By inspection of (36), the rent is increasing on the former interval and decreasing on the latter. Furthermore, because the utility function $V(q,\theta)$ is continuous in θ, the rent function V is continuous by the Maximum Theorem. We thus obtain the expression in the text:

$$V(\theta^*) = V(0) + \int_0^{\theta^*} V_\theta(q,\theta)\,\mathrm{d}\theta = V(1) - \int_{\theta^*}^1 V_\theta(q,\theta)\,\mathrm{d}\theta.$$

By the envelope theorem, $V_\theta(q,\theta) = q + u_2$ for $\theta < \theta^*$, and $V_\theta(q,\theta) = q - u_1$ for $\theta > \theta^*$. Finally, because $V(0) = V(1)$, we can simplify the equation above and obtain

$$\int_0^1 q(\theta)\,\mathrm{d}\theta = u_1 - u_2.$$

We now turn to sufficiency. Suppose the menu $\{q(\theta)\}_{\theta\in\Theta}$ is increasing and satisfies the integral constraint (24). Then, construct the following transfers:

$$t(\theta) = \begin{cases} \theta q(\theta) + u_2 + \min\{u_1 - u_2 - q(\theta), 0\} - \int_0^\theta (q(x) + u_2)\,\mathrm{d}x - (1-\theta)u_2 & \text{if} \quad \theta < \theta^*, \\ \theta q(\theta) + u_2 + \min\{u_1 - u_2 - q(\theta), 0\} + \int_\theta^1 (q(x) - u_1)\,\mathrm{d}x - \theta u_1 & \text{if} \quad \theta \geq \theta^*. \end{cases} \tag{37}$$

Because the menu satisfies the integral constraint, we have

$$\int_0^\theta q(x)\,\mathrm{d}x = u_1 - u_2 - \int_\theta^1 q(x)\,\mathrm{d}x,$$

and we can express all transfers $t(\theta)$ in (37) as

$$t(\theta) = \theta q(\theta) + \min\{u_1 - u_2 - q(\theta), 0\} - \int_0^\theta q(x)\,\mathrm{d}x. \tag{38}$$

Under these transfers, the expected utility of type θ from reporting type θ' is given by

$$V(q(\theta'),\theta) - t(\theta') = (\theta - \theta')\,q(\theta') + \int_0^{\theta'} q(x)\,\mathrm{d}x + u_2 - \max\{\theta u_1, (1-\theta)u_2\}.$$

Because q is monotone, the expression on the right-hand side is maximized at $\theta' = \theta$, and hence, the incentive constraints are satisfied. Because the rent $V(\theta) \triangleq V(q(\theta),\theta) - t(\theta)$ is non-negative for all $\theta \in [0,1]$, the participation constraints are also satisfied.

Finally, note that the integral constraint (24) and the monotonicity condition imply the responsiveness condition (22). The set $Q(\theta)$ of responsive experiments is described in Figure 9 below. Suppose to the contrary, that $q(\theta) \notin Q(\theta)$ and, in particular, that $q(\theta) = \hat{q} < u_1 - u_2/\theta$ for some $\theta > \theta^*$. Then, by monotonicity, we would have

$$\int_0^1 q(x)\,\mathrm{d}x \leq \hat{q}\theta + u_1(1-\theta) < u_1 - u_2,$$

which yields a contradiction. ■

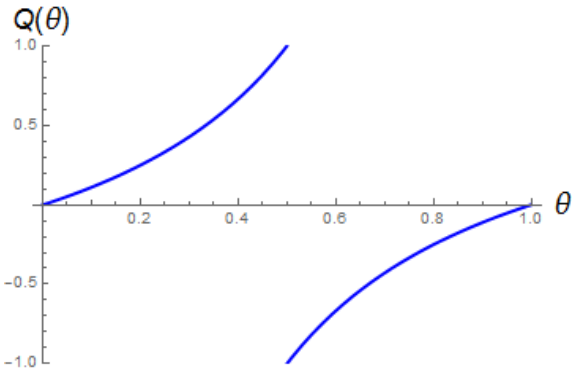

Figure 9: Bounds on Responsive Experiments $Q(\theta)$

Proof of Proposition 6. The Fundamental Theorem of Linear Programming applies to settings with $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$, with $m < n$. Let A be an $m \times n$ matrix with the m rows being linearly independent. If the linear problem

$$\begin{aligned} &\max_{x \in \mathbb{R}^n} \quad c \cdot x \\ &\text{s.t. } Ax = b \text{ and } x \geq 0 \end{aligned}$$

has a solution, then it has a solution with all but m entries being zero.

In this form, the Fundamental Theorem cannot be directly applied to the monopolist's problem (25): in this problem, *(i)* the number of choice variables is a continuum; *(ii)* the objective function is nonlinear; and *(iii)* monotonicity constraints are absent in the canonical representation of the theorem. To address *(i)*, we discretize the state space into a fine grid with a radius ε, $[0, \varepsilon, 2\varepsilon, \ldots, 1]$. To address *(ii)*, we recall from Proposition 2 that an optimal menu contains the fully informative experiment. Hence, for any optimal menu $\{q(\theta)\}_{\theta \in \Theta}$, there is some type θ^I such that $q\left(\theta^I\right) = u_1 - u_2$. Consequently, any optimal q must solve the problem with the additional constraint $q\left(\theta^I\right) = u_1 - u_2$ for an appropriately chosen θ^I. Because the menu is monotone, the objective can be written linearly with $\min\{u_1 - u_2 - q(\theta), 0\} = 0$ for all $\theta < \theta^I$ and $\min\{u_1 - u_2 - q(\theta), 0\} = u_1 - u_2 - q(\theta)$ for all $\theta > \theta^I$. Finally, to address *(iii)*, we change variables from $q(\theta)$ to its increment $\hat{q}(\theta) \triangleq q(\theta) - q(\theta - \varepsilon)$, where $\hat{q}(0) \triangleq 0$. Finally, we substitute $q(\theta) = -u_2 + \sum_{x=0}^{\theta} \hat{q}(x)$ and rewrite the monotonicity constraint as $\hat{q}(\theta) \geq 0$.

The monopolist's problem (25) can be restated as

$$\begin{aligned}
\max_{\hat{q}(\theta)} \ & \sum_{\theta=0}^{1} \hat{q}(\theta) \left(\sum_{x=\theta}^{1} (x f(x) + F(x)) - \sum_{x=\theta^I}^{1} f(x) \right) \\
\text{s.t.} \ & \sum_{\theta=0}^{\theta^I} \hat{q}(\theta) = u_1, \\
& \sum_{\theta=\theta^I}^{1} \hat{q}(\theta) = u_2, \\
& \sum_{\theta=0}^{1} \left(-u_2 + \sum_{x=0}^{\theta} \hat{q}(x) \right) = u_1 - u_2, \\
& \hat{q}(\theta) \geq 0.
\end{aligned}$$

This is a canonical linear programming problem with three linearly independent equality constraints. The first two correspond to the total change from the experiment $q(0) = -u_2$ to the fully informative experiment $q\left(\theta^I\right) = u_1 - u_2$ to $q(1) = u_1$. The last equality is integral constraint (24). Hence, by the Fundamental Theorem, for any θ^I there is an optimal menu of experiments with at most three positive entries. Since $q(1) = u_1$ corresponds to an uninformative experiment, this implies there are at most two informative experiments in any optimal menu.

Finally, consider an arbitrary discrete distribution that converges in distribution to a continuous distribution. The profits converge for any fixed menu, and hence, optimal profits converge as well. Furthermore, the set of profits that can be achieved by two-item menus is compact. Therefore, there exists an optimal menu with continuous types that has at most two items. ■

Proof of Proposition 7. We first derive the seller's objective in the usual way. Using (38) to write the expected transfers and integrating by parts, we obtain

$$\int_0^1 t(\theta) \, \mathrm{d}F(\theta) = \int_0^1 \left(\theta q(\theta) + \min\{u_1 - u_2 - q(\theta), 0\} - \frac{1 - F(\theta)}{f(\theta)} q(\theta) \right) \mathrm{d}F(\theta).$$

Using the integral constraint, we obtain (up to an additive constant)

$$\int_0^1 t(\theta) \, \mathrm{d}F(\theta) = \int_0^1 \left[(\theta f(\theta) + F(\theta)) q(\theta) + \min\{u_1 - u_2 - q(\theta), 0\} f(\theta) \right] \mathrm{d}\theta.$$

We now establish that the solution to the seller's problem (25) can be characterized through Lagrangian methods. For necessity, note that the objective is concave in the experiment; the set of non-decreasing functions is convex, and the integral constraint can be

weakened to the real-valued inequality constraint

$$\int_0^1 q(\theta)\mathrm{d}\theta \leq u_1 - u_2. \tag{39}$$

Necessity of the Lagrangian then follows from Theorem 8.3.1 in Luenberger (1969). Sufficiency follows from Theorem 8.4.1 in Luenberger (1969). In particular, any solution maximizer of the Lagrangian $q(\theta)$ with

$$\int_0^1 q(\theta)\mathrm{d}\theta = \bar{q}$$

maximizes the original objective subject to the inequality constraint

$$\int_0^1 q(\theta)\mathrm{d}\theta \leq \bar{q}.$$

Thus, any solution to the Lagrangian that satisfies the constraint solves the original problem.

Because the Lagrangian approach is valid, we apply the results of Toikka (2011) to solve the seller's problem for a given value of the multiplier λ on the integral constraint. Write the Lagrangian as

$$\int_0^1 \left[\left(\theta f\left(\theta\right) + F\left(\theta\right)\right) q\left(\theta\right) + \min\left\{u_1 - u_2 - q\left(\theta\right), 0\right\} f\left(\theta\right) + \lambda\left(u_1 - u_2 - q\left(\theta\right)\right)\right] \mathrm{d}\theta.$$

In order to maximize the Lagrangian subject to the monotonicity constraint, consider the *generalized virtual surplus*

$$\Phi(\theta, q) := \int_{-u_2}^{q} \left(\bar{\phi}\left(\theta, x\right) - \lambda^*\right) \mathrm{d}x,$$

where $\bar{\phi}\left(\theta, x\right)$ denotes the ironed virtual value for experiment x. (Up to a constant, $\Phi(\theta, q)$ is a general formulation of the virtual value in the statement of the proposition.) Note that the virtual surplus $\Phi(\theta, q)$ is weakly concave in q. Because the multiplier λ shifts all virtual values by a constant, the result in Proposition 7 follows from Theorem 4.4 in Toikka (2011). Finally, note that the optimal value of λ^* is strictly positive. If not, the pointwise maximizer of $\bar{\phi}\left(\theta, q\right)$ would be weakly above $u_1 - u_2$ and strictly so for some θ, leading to a violation of the integral constraint. Therefore, the inequality constraint (39) must bind. ■

Proof of Proposition 9. We know from Proposition 3 that the high type θ^H purchases the fully informative experiment. We now derive the optimal experiment $E(\theta^L)$. Suppose (as we later verify) that both types θ^H and θ^L choose action a_i after observing signal s_i from

experiment $E(\theta^L)$. The seller's relaxed problem can be written as

$$\max_{0\leq\pi_{ii}\leq 1,\ V(\theta^H)} (1-\gamma)\sum\nolimits_{i=1}^{I}\theta_i^L u_i\pi_{ii} - \gamma V(\theta^H),$$
$$\text{s.t. } V(\theta^H) \geq \sum\nolimits_{i=1}^{I}\pi_{ii}u_i\left(\theta_i^H-\theta_i^L\right) - \max_i \theta_i^H u_i + \max_i \theta_i^L u_i \geq 0.$$

where the latter inequality ensures that the high type θ^H achieves a non-negative payoff when misreporting his type and following every signal's recommendation.

We now simplify the problem as follows. By Proposition 3, the incentive-compatibility constraint of type θ^H binds and the relaxed problem can be rewritten as

$$\max_{0\leq\pi_{ii}\leq 1} \sum\nolimits_{i=1}^{I}\pi_{ii}u_i(\theta_i^L-\gamma\theta_i^H), \tag{40}$$
$$\text{s.t. } \sum\nolimits_{i=1}^{I}\pi_{ii}u_i(\theta_i^H-\theta_i^L) - \max_i\left\{\theta_i^H u_i\right\} + \max_i\left\{\theta_i^L u_i\right\} \geq 0. \tag{41}$$

Now, arrange the states such that

$$\frac{\theta_1^L}{\theta_1^H} \leq \cdots \leq \frac{\theta_I^L}{\theta_I^H},$$

and consider the function $R(i)$ defined in the text

$$R(j) \triangleq \sum\nolimits_{i=j}^{I} u_i(\theta_i^H-\theta_i^L) - \max_i\left\{\theta_i^H u_i\right\} + \max_i\left\{\theta_i^L u_i\right\}.$$

The function $R(j)$ is decreasing in j for $j<i_c$ and increasing in j for $j>i_c$, where

$$i_c \triangleq \min\left\{i : 1 \leq \theta_i^L/\theta_i^H\right\}. \tag{42}$$

We now show that $\min_j R(j) = R(i_c) < 0$. This is immediate if $\max_i\left\{\theta_i^H u_i\right\} \geq \max_i\left\{\theta_i^L u_i\right\}$, as we have $\sum_{i=i_c}^{I} u_i(\theta_i^H-\theta_i^L) < 0$ by construction. Conversely, if $\max_i\left\{\theta_i^H u_i\right\} < \max_i\left\{\theta_i^L u_i\right\}$, then we let $i_L \triangleq \arg\max_i\left\{\theta_i^L u_i\right\}$, and we derive the following bound:

$$R(i_c) \leq R(i_L) \leq \sum\nolimits_{i=i_L}^{I} u_i(\theta_i^H-\theta_i^L) - \theta_{i_L}^H u_{i_L} + \theta_{i_L}^L u_{i_L} = \sum\nolimits_{i=i_L+1}^{I} u_i(\theta_i^H-\theta_i^L) < 0$$

where the latter inequality follows from the fact that $\max_i\left\{\theta_i^H u_i\right\} < \max_i\left\{\theta_i^L u_i\right\}$ implies $\theta_{i_L}^L > \theta_{i_L}^H$, which means $i_L \geq i_c$.

We now turn to the solution to the seller's problem. Notice that (40)-(41) is a linear problem that attains a monotone solution. In particular, any solution must be a monotone sequence: $\pi_{ii}=0$ for $i<\hat{\imath}$ and $\pi_{ii}=1$ for $i>\hat{\imath}$, for some $\hat{\imath}\in\{1,\ldots,I\}$. (This structure can

be seen immediately from the Lagrange multiplier method, as the solution must maximize some linear combination of the objective function and the constraint.)

Suppose that the constraint does not bind at the optimum. Then, the solution assigns $\pi_{ii} = 1$ if and only if $i \geq i_s \triangleq \min\left\{i : \gamma \leq \theta_i^L/\theta_i^H\right\}$. Note that $i_s \leq i_c$, which was defined in (42). Therefore, if $R(i_s) \geq 0$ the constraint is satisfied at the unconstrained optimum. If, instead, we have $R(i_s) < 0$, then by the definition of the critical state i_b,

$$R(i_b) > 0 > R(i_b + 1),$$

we have $i_b < i_s$. In this case, the solution is given by $\pi_{ii} = 0$ for $i < i_b$, $\pi_{ii} = 1$ for $i > i_b$ and by setting $\pi_{i_b i_b}$ to satisfy (41) with equality, i.e.,

$$\pi_{i_b i_b} = \frac{\sum_{j=i_b+1}^{I}(\theta_j^L - \theta_j^H)u_j - \max_j \theta_j^L u_j + \max_j \theta_j^H u_j}{(\theta_{i_b}^H - \theta_{i_b}^L)u_{i_b}}. \tag{43}$$

Combining these two arguments, we conclude that the optimal cutoff state $\hat{\imath}$ is given by $i^* = \min\{i_b, i_s\}$ and that $\pi_{i^*i^*} \in \{\pi_{i_b i_b}, 1\}$ depending on whether $i_b \lesseqgtr i_s$.

To complete the menu, we now need to specify the off-diagonal entries π_{ij} to ensure both types θ^H and θ^L choose action a_i when observing signal s_i. This requires

$$\pi_{ii}\theta_i u_i \geq \pi_{ji}\theta_j u_j$$

for both types and for all $j < i$, because the signal matrix can be taken to be lower triangular. In particular, we need to ensure that, for all $j < i^*$,

$$\pi_{i^*i^*}u_{i^*}\theta_{i^*}^H \geq \pi_{ji^*}u_j\theta_j^H,$$
$$\pi_{i^*i^*}u_{i^*}\theta_{i^*}^L \geq \pi_{ji^*}u_j\theta_j^L.$$

Because $\theta_j^L/\theta_j^H \leq \theta_i^L/\theta_i^H$ for $j < i$, it suffices to satisfy the constraint of type θ^H.

We proceed as follows. Fix an alternative action a_j with $j < i^*$. For any $i > i^*$, we make type θ^H indifferent between following the recommendation of signal i and choosing action a_j; we do so beginning with π_{jI} and proceeding backward as long as required. If this procedure assigns positive weight to π_{ji^*}, then it must be that

$$\pi_{ji^*} = 1 - \sum_{i=i^*+1}^{I} \frac{\theta_i^H u_i}{\theta_j^H u_j}. \tag{44}$$

We argue that type θ^H has strict incentives to follow the recommendation of signal i^* when

$\pi_{i^*i^*} = \pi^B_{i^*}$. (A fortiori, type θ^H has strict incentives to choose action a_i following any signal s_i with $i \geq i^*$ if $\pi_{ii} = 1$.) Recall the definition

$$\pi_{i^*i^*}\theta^H_{i^*}u_{i^*} = \frac{\sum_{i=i^*+1}^{I} u_i(\theta^L_i - \theta^H_i) - \max_i \theta^L_i u_i + \max_i \theta^H_i u_i}{\theta^H_{i^*} - \theta^L_{i^*}}\theta^H_{i^*}.$$

Let $i_L = \arg\max_i \theta^L_i u_i$ and $i_H = \arg\max_i \theta^H_i u_i$, and consider the following two cases.

(1.) If $i_L > i^*$, then we know that

$$\sum\nolimits_{i=i^*+1}^{I} \theta^L_i u_i > \max_i \theta^L_i u_i,$$

and we bound $\pi_{i^*i^*}\theta^H_{i^*}u_{i^*}$ by

$$\pi_{i^*i^*}\theta^H_{i^*}u_{i^*} > \frac{\theta^H_{i^*}}{\theta^H_{i^*} - \theta^L_{i^*}}\left(\theta^H_j u_j - \Sigma^I_{i=i^*+1}\theta^H_i u_i\right) > \theta^H_j u_j - \Sigma^I_{i=i^*+1}\theta^H_i u_i = \pi_{ji^*}\theta^H_j u_j.$$

(2.) If $i_L \leq i^*$, then from (43) and (44), we know that the difference $\pi_{i^*i^*}\theta^H_{i^*}u_{i^*} - \pi_{ji^*}\theta^H_j u_j$ is proportional to

$$\begin{aligned} & \sum\nolimits_{i=i^*+1}^{I} u_i(\theta^L_i - \theta^H_i) - \max_i \theta^L_i u_i + \max_i \theta^H_i u_i - \left(1 - \frac{\theta^L_{i^*}}{\theta^H_{i^*}}\right)\left(\theta^H_j u_j - \Sigma^I_{i=i^*+1}\theta^H_i u_i\right) \\ = & \sum\nolimits_{i=i^*+1}^{I} u_i\left(\theta^L_i - \frac{\theta^L_{i^*}}{\theta^H_{i^*}}\theta^H_i\right) - \max_i \theta^L_i u_i + \max_i \theta^H_i u_i - \left(1 - \frac{\theta^L_{i^*}}{\theta^H_{i^*}}\right)\theta^H_j u_j. \end{aligned}$$

Notice that every term in the sum is positive because the likelihood ratio is increasing in i. The remaining terms can be written as

$$\begin{aligned} -\max_i \theta^L_i u_i + \max_i \theta^H_i u_i - \left(1 - \frac{\theta^L_{i^*}}{\theta^H_{i^*}}\right)\theta^H_j u_j &= -\frac{\theta^L_{i_L}}{\theta^H_{i_L}}\theta^H_{i_L}u_{i_L} + \theta^H_{i_H}u_{i_H} - \left(1 - \frac{\theta^L_{i^*}}{\theta^H_{i^*}}\right)\theta^H_j u_j \\ &\geq \theta^H_{i_H}u_{i_H}\left(1 - \frac{\theta^L_{i_L}}{\theta^H_{i_L}}\right) - \left(1 - \frac{\theta^L_{i^*}}{\theta^H_{i^*}}\right)\theta^H_j u_j. \end{aligned}$$

Here, the likelihood ratios are ranked (because $i_L \leq i^*$ in this case), and $\theta^H_{i_H}u_{i_H} \geq \theta^H_j u_j$ by the definition of i_H. ■

B Supplemental (Online) Appendix

B.1 Many Actions and States

We illustrate the construction of the optimal experiment as determined by Proposition 8 and the implications for information rents in the following example.

Example 1 (Noncongruent Types) *Consider uniform match values ($u_i = 1$ for all $i = 1,2,3$) and two types, $\theta^L = (1/10, 1/10, 8/10)$ and $\theta^H = (4/10, 3/10, 3/10)$. These types are noncongruent: without additional information, θ^H would choose action a_1 and θ^L would choose a_3. The likelihood ratios θ_i^L/θ_i^H are $(1/4, 1/3, 8/3)$. This implies $i_b = 2$, whereas $i_s \in \{1,2\}$ depending on the prior probability γ of the high type. For $\gamma \in [0, 1/4]$ and $\gamma \in [1/4, 1/3]$, Proposition 9.2 applies, and the high type obtains positive rents. Furthermore, for $\gamma \geq 1/4$, the partially informative experiment $E(\theta^L)$ involves dropping signal s_1. The optimal experiment $E(\theta^L)$ as a function of γ is given by*

$E(\theta^L)$	s_1	s_2	s_3
ω_1	1	0	0
ω_2	0	1	0
ω_3	0	0	1
	if $\gamma < 1/4$,		

$E(\theta^L)$	s_1	s_2	s_3
ω_1	0	1/4	3/4
ω_2	0	1	0
ω_3	0	0	1
	if $\gamma \in [1/4, 1/3]$,		

$E(\theta^L)$	s_1	s_2	s_3
ω_1	0	1/4	3/4
ω_2	0	1/2	1/2
ω_3	0	0	1
	if $\gamma > 1/3$.		

Example 2 illustrates how congruent, but not strongly congruent beliefs, allow for surplus extraction. In the example, the two types deem state ω_2 the most likely. Thus the types are congruent but not strongly congruent, as they disagree on the relative likelihood of states ω_1 and ω_3.

Example 2 (Congruent Priors) *Consider uniform match values ($u_i = 1$ for all $i = 1,2,3$) and two types, $\theta^L = (5/10, 1/10, 4/10)$ and $\theta^H = (4/10, 3/10, 3/10)$. Because $i_b = 1$, the optimal experiment $E(\theta^L)$ as a function of γ is given by*

$E(\theta^L)$	s_1	s_2	s_3
ω_1	1	0	0
ω_2	0	1	0
ω_3	0	0	1
	if $\gamma \leq 1/3$,		

$E(\theta^L)$	s_1	s_2	s_3
ω_1	1	0	0
ω_2	0	1/2	1/2
ω_3	0	0	1
	if $\gamma > 1/3$,		

and the high type obtains positive rents only if $\gamma < 1/3$.

In Example 3 we illustrate that the relaxed approach is not valid with many types. In the example below with three types, no experiment $E(\theta^1)$ can lead both types θ^2 and θ^3 to follow the action recommended by every signal. Thus, the profits in the relaxed problem are strictly greater than those in original problem.

Example 3 (Many Types and Actions) *Consider uniform match values ($u_i = 1$ for all $i = 1, 2, 3$) and three types, $\theta^1 = (1/6, 1/6, 4/6)$, $\theta^2 = (1/2, 1/2, 0)$, and $\theta^3 = (1/2, 0, 1/2)$, which are all equally likely. In the* relaxed problem*, the monopolist sells the fully informative experiment to types θ^2 and θ^3. Type θ^1 is offered the partially informative experiment*

$E(\theta^1)$	s_1	s_2	s_3
ω_1	1/2	0	1/2
ω_2	0	1	0
ω_3	0	0	1

,

and the seller's revenues are equal to $5/12$*. However, if type θ^2 purchased experiment $E(\theta^1)$, he would choose action a_1 when observing signal s_3. In the solution to the* full problem*, which we can construct by a guess-and-verify approach, the optimal experiment $E(\theta^1)$ consists of*

$E(\theta^1)$	s_1	s_2	s_3
ω_1	1/2	0	1/2
ω_2	1/2	1/2	0
ω_3	0	0	1

, (45)

which yields revenues of $1/3$*, i.e., revenues are strictly lower in the relaxed program.*

B.2 More Actions than States

We consider a setting with two types, and we relax the assumption of matching state-action payoffs. In particular, we consider the following example with two types, two states, and three actions. The data buyer's payoff is given by

$u(\omega, a)$	a_1	a_2	a_3
ω_1	1	0	4/5
ω_2	0	1	4/5

.

Thus, action a_i is the optimal action in state ω_i, but action a_3 provides a lower bound on the payoffs that is uniform across states—an insurance action. Let the two types be given by $\theta^L = (1/10, 9/10)$ and $\theta^H = (6/10, 4/10)$. As the results of Proposition 3 do not rely

on matching payoffs, we know type θ^H receives full information, $E(\theta^H) = \overline{E}$, his incentive constraint binds, and the participation constraint of type θ^L binds.

For the case $\gamma \triangleq \Pr\left(\theta = \theta^H\right) = 3/4$, an optimal menu contains the experiment

$E(\theta^L)$	s_1	s_2	s_3
ω_1	1/3	2/15	8/15
ω_2	0	4/5	1/5

(46)

at a price $t(\theta^L) = 1/25$ and the experiment $\overline{E}$ at a price $t(\theta^H) = 1/5$. The optimal menu is illustrated in Figure 10.

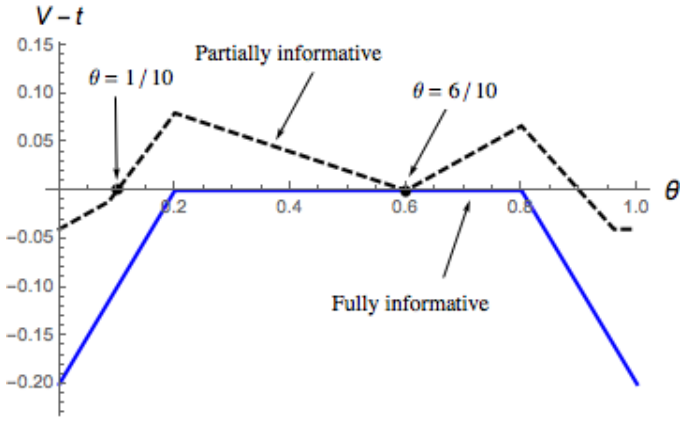

Figure 10: Optimal Menu

This menu has the following notable properties: *(i)* the seller extracts all the surplus from both types; *(ii)* type θ^L follows the recommendation of every signal in $E(\theta^L)$; *(iii)* type θ^H, if purchasing experiment $E(\theta^L)$, is indifferent between action a_2 and action a_3 when observing signal s_2 as well as between a_1 and a_3 when observing signal s_3; and *(iv)* the optimal profits are strictly lower than those in the relaxed problem.

Indeed, ignoring the off-path obedience constraints, the optimal menu is discriminatory, and the seller extracts all the rents by offering the experiment

$E(\theta^L)$	s_1	s_2
ω_1	2/5	3/5
ω_2	0	1

(47)

at a price $t(\theta^L) = 1/25$ and the experiment $\overline{E}$ at a price $t(\theta^H) = 2/5$.

In the full problem, the seller cannot turn experiment $E(\theta^L)$ in (46) into the *more informative* one in (47). If she did, buyer type θ^H would choose action a_3 after deviating

and observing signal s_2.[27] In other words, the seller extracts the surplus, but at the cost of additional distortion—notably, there is no "1" entry in (46).

Interestingly, action a_3 may not be induced in an optimal menu yet still restrict the seller. Indeed, if we modify the above example by setting $\gamma = 2/3$, an optimal menu can be calculated to contain the experiment

$E(\theta^L)$	s_1	s_2
ω_1	5/6	1/6
ω_2	0	1

(48)

at a price $t(\theta^L) = 1/12$ and the experiment $\overline{E}$ at a price of $t(\theta^H) = 11/60 < 1/5 = V\left(\theta^H, \overline{E}\right)$. As before, type θ^H is indifferent between a_2 and a_3 after deviating to $E(\theta^L)$ and observing s_2. Contrary to our earlier examples, the high type makes positive rents despite the seller's discriminatory menu offering.

To emphasize, action a_3 is not chosen in an optimal menu by either type, yet it looms large and prevents the seller from extracting the full surplus. In particular, the seller would like to reduce π_{11} in order to relax the high type's incentive constraint and increase t^H. However, by doing so she would induce type θ^H to choose action a_3 after s_2. This means that the high type's *marginal benefit* from a probability shift from π_{11} to π_{12} is $\theta_1^H(-1 + 4/5) = -3/25$, while the corresponding *marginal* change in price t^L is $-\theta_1^L = -1/10$. Therefore, t^H can only increase at rate $1/50$, which is not profitable for the seller when the fraction of high types is $\gamma = 2/3$. If both types were instead required to follow the signals' recommendations, the high type's misreporting value would change at rate $-\theta_1^H = -3/5$, allowing the seller to increase t^H at the profitable rate of $1/2$. The kink in the "exchange rate" when the high type is indifferent among several actions prevents the seller from making the modification.

To reinforce the point, maintain the assumption $\gamma = 2/3$ but exclude action a_3 from the set of available actions. The optimal menu is again given by the experiment in (47), which is now *less informative* than (48), but allows the seller to extract all the rents.

B.3 Sequential Design

We show that sequential design of experiments can increase the seller's revenues in our leading binary-type example. We focus on the simplest instance of a dynamic protocol, whereby the seller first releases a free informative experiment to the buyer and then, without observing

[27] Similarly, one can show that type θ^H must be indifferent after signal s_3 in experiment (46). The argument here is by contradiction: if he strictly preferred action a_3, the seller could make the experiment more valuable for θ^L without changing its value for θ^H; and if he strictly preferred a_1, the seller could rearrange all signals and relax the incentive-compatibility constraint.

the realized signal, offers a menu of (experiment, price) pairs from which to choose.

Let $\Omega = \{\omega_1, \omega_2\}$, $A = \{a_1, a_2\}$, and assume uniform match values, i.e.,

$$u(\omega_i, a_j) = \mathbb{I}_{[i=j]}.$$

Consider two equally likely types with interim beliefs $\theta^L = 1/8$ and $\theta^H = 1/4$,respectively, where $\theta \triangleq \Pr[\omega_1]$.

Because the two types are *congruent*, an optimal static mechanism (Proposition 4) contains only the fully informative experiment. In the current example, the seller is indifferent between charging prices $t = 1/8$ and $t = 1/4$. In either case, the monopoly profits are

$$\pi^*_{\text{static}} = 1/8.$$

Consider the following sequential scheme. First, the seller reveals an outcome of the following experiment E_0 at no cost to the buyer

E_0	s_1	s_2
ω_1	1	0
ω_2	1/3	2/3

After observing signal s_2, the buyer is convinced that the state is $\omega = \omega_2$, which confirms his prior, and does not buy further information. After realization s_1, however, the buyer's beliefs are updated to

$$\theta^L(s_1) = 3/10, \quad \theta^H(s_1) = 1/2.$$

At this point, the seller offers the fully informative experiment at a price $\bar{t} = 3/10$.

The key observation is that signal s_1 under experiment E_0 is more likely be realized for the high type $\theta^H = 1/4$ than for the low type $\theta^L = 1/8$. In particular, the signal distribution is given by

$$\Pr\left[s_1 \mid \theta^L\right] = \frac{5}{12}, \ \Pr\left[s_1 \mid \theta^H\right] = \frac{1}{2}.$$

As a consequence, the monopolist's profit is given by

$$\pi^*_{\text{dyn}} \triangleq \bar{t}\left(\gamma \Pr\left[s_1 \mid \theta^H\right] + (1-\gamma)\Pr\left[s_1 \mid \theta^L\right]\right) = 11/80.$$

Thus, the sequential sale outperforms the static sale in this example, i.e.,

$$\pi^*_{\text{dyn}} = 11/80 > 1/8 = \pi^*_{\text{static}}.$$

Taking a step back, it is clear that the seller would ideally like to condition payments on the realized states. In this case, she could charge a payment of 1 upon realization of state ω_1, which is the state less likely for either type. Both types would accept such a contract, and the seller achieves the first-best profits. As we do not allow for the payments to be made contingent on the realization of the state, a sequential mechanism essential represents a costly instrument to (partially) circumvent this restriction.

In essence, the proposed sequential scheme charges a constant price $\bar{t} = 3/10$ upon realization of signal s_1. Because the signal is correlated with the state under experiment E_0, it occurs more frequently for the higher type, allowing the seller to effectively price discriminate without ever giving the buyer a choice of experiment.

Finally, note that the seller could do better within the simple class of mechanisms that initially release a free experiment, followed by a menu.

Intuitively, as the correlation between state and signal s_1 becomes more precise (i.e., as s_1 becomes more informative), the seller's ability to condition payments on states improves. Ultimately, however, the seller must balance the ability to correlate payments with the willingness to pay for supplemental information after observing signal s_1 (e.g., the signal cannot be arbitrarily precise).

To formalize the intuition, consider offering free experiments of the following form

$E(x)$	s_1	s_2
ω_1	1	0
ω_2	$1-x$	x

.

These experiments lead to posterior beliefs

$$\begin{aligned} \theta^L(x) &\triangleq \Pr\left[s_1 \mid \theta^L\right] = \frac{1}{8-7x}, \\ \theta^H(x) &\triangleq \Pr\left[s_1 \mid \theta^H\right] = \frac{1}{4-3x}. \end{aligned}$$

These beliefs satisfy the condition $1/2 = \gamma \geq \theta^L(x)/\theta^H(x)$ for all x. Therefore, after releasing experiment $E(x)$ the seller optimally offers the fully informative experiment $\bar{E}$ at a price

$$\bar{t}(x) = \min\left\{\theta^L(x), 1-\theta^H(x)\right\}.$$

Finally, a straightforward calculation reveals that the seller's profits are maximized by choosing x such that $\theta^L(x) < 1/2 < \theta^H(x)$. In particular, it is optimal for the seller to induce

the two types to have identical willingness to pay for the full information, i.e.,

$$\theta^L(x^*) = 1 - \theta^H(x^*).$$

The optimal experiment has

$$x^* = 1 - 1/\sqrt{21} \approx 0.781,$$

which is larger than $x = 2/3$, as used in the initial example, and yields profits $\pi^* = (7 + 2\sqrt{21})/112 \approx 0.144$ that exceed $11/80$, as computed above.

References

Admati, A. R., and P. Pfleiderer (1986): "A monopolistic market for information," *Journal of Economic Theory*, 39(2), 400–438.

——— (1990): "Direct and indirect sale of information," *Econometrica*, 58(4), 901–928.

Anderson, E. T., and D. Simester (2013): "Advertising in a Competitive Market: The Role of Product Standards, Customer Learning and Switching Costs," *Journal of Marketing Research*, 50(4), 489–504.

Babaioff, M., R. Kleinberg, and R. Paes Leme (2012): "Optimal Mechanisms for Selling Information," in *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pp. 92–109.

Balestrieri, F., and S. Izmalkov (2014): "Informed seller in a Hotelling market," Discussion paper, HP Labs and New Economic School.

Bergemann, D., and A. Bonatti (2015): "Selling Cookies," *American Economic Journal: Microeconomics*, 7(3), 259–294.

Bergemann, D., and M. Pesendorfer (2007): "Information Structures in Optimal Auctions," *Journal of Economic Theory*, 137(1), 580–609.

Brusco, S., and H. Hopenhayn (2007): "Deregulation with consensus," *Economic Theory*, 32(1), 223–250.

Celik, L. (2014): "Information unraveling revisited: disclosure of horizontal attributes," *Journal of Industrial Economics*, 62(1), 113–136.

Chvatal, V. (1983): *Linear Programming*. Freeman and Co.

Cremer, J., and R. McLean (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247–1258.

Daskalakis, C., A. Deckelbaum, and C. Tzamos (2017): "Strong Duality for a Multiple-Good Monopolist," *Econometrica*, 85(3), 735–767.

Dwork, C. (2008): "Differential Privacy: A Survey of Results," in *International Conference on Theory and Applications of Models of Computation*, pp. 1–19. Springer.

Eső, P., and B. Szentes (2007a): "Optimal information disclosure in auctions and the handicap auction," *Review of Economic Studies*, 74(3), 705–731.

——— (2007b): "The price of advice," *Rand Journal of Economics*, 38(4), 863–880.

Federal Trade Commission (2014): *Data Brokers: a Call for Transparency and Accountability.*

Fuchs, W., and A. Skrzypacz (2015): "Government interventions in a dynamic market with adverse selection," *Journal of Economic Theory*, 158, 371–406.

Hörner, J., and A. Skrzypacz (2016): "Selling information," *Journal of Political Economy*, 124(6), 1515–1562.

Johnson, J. P., and D. P. Myatt (2006): "On the Simple Economics of Advertising, Marketing, and Product Design," *American Economic Review*, 96(3), 756–784.

Jullien, B. (2000): "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 93(1), 1–47.

Kamenica, E., and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101(6), 2590–2615.

Koessler, F., and V. Skreta (2016): "Informed seller with taste heterogeneity," *Journal of Economic Theory*, 165, 456–471.

Kolotilin, A., M. Li, T. Mylovanov, and A. Zapechelnyuk (2015): "Persuasion of a Privately Informed Receiver," Discussion paper, University of New South Wales.

Krähmer, D., and R. Strausz (2015): "Ex post information rents in sequential screening," *Games and Economic Behavior*, 90, 257–273.

Li, H., and X. Shi (2015): "Discriminatory Information Disclosure," Discussion paper, University of British Columbia and University of Toronto.

Lizzeri, A. (1999): "Information revelation and certification intermediaries," *Rand Journal of Economics*, 30(2), 214–231.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods.* John Wiley & Sons.

Manelli, A. M., and D. R. Vincent (2006): "Bundling as an optimal selling mechanism for a multiple-good monopolist," *Journal of Economic Theory*, 127(1), 1–35.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

——— (1982): "Optimal Coordination Mechanism in Generalized Principal-Agent Problems," *Journal of Mathematical Economics*, 10, 67–81.

Mylovanov, T., and T. Tröger (2014): "Mechanism Design by an Informed Principal: Private Values with Transferable Utility," *Review of Economic Studies*, 81(4), 1668–1707.

Ottaviani, M., and A. Prat (2001): "The value of public information in monopoly," *Econometrica*, 69(6), 1673–1683.

Pavlov, G. (2011a): "Optimal Mechanism for Selling Two Goods," *The BE Journal of Theoretical Economics*, 11(1), 1–33.

——— (2011b): "A Property of Solutions to Linear Monopoly Problems," *The B.E. Journal of Theoretical Economics*, 11(1), 1–16.

Pycia, M. (2006): "Stochastic vs Deterministic Mechanisms in Multidimensional Screening," Discussion paper, UCLA.

Rayo, L., and I. Segal (2010): "Optimal Information Disclosure," *Journal of Political Economy*, 118(5), 949–987.

Riley, J., and R. Zeckhauser (1983): "Optimal Selling Strategies: When to Haggle, When to Hold Firm," *Quarterly Journal of Economics*, 98, 267–290.

Rochet, J., and J. Thanassoulis (2015): "Stochastic Bundling," Discussion paper, University of Zürich.

Samuelson, W. (1984): "Bargaining under Asymmetric Information," *Econometrica*, 54(4), 995–1005.

Shapiro, C., and H. R. Varian (1999): *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press.

Stiglitz, J. E. (1977): "Monopoly, Non-Linear Pricing and Imperfect Information: The Insurance Market," *Review of Economic Studies*, 44(3), 407–430.

Toikka, J. (2011): "Ironing without Control," *Journal of Economic Theory*, 146(6), 2510–2526.

Disclosure Statement

Dirk Bergemann

I gratefully acknowledge the financial support of the National Science Foundation through grant ICES 1215808.

Alessandro Bonatti
I disclose that I have not received financial support related to this research.

Alex Smolin
I disclose that I have not received financial support related to this research.