THE DESIGN AND PRICE OF INFORMATION

By

Dirk Bergemann, Alessandro Bonatti, and Alex Smolin

July 2016

COWLES FOUNDATION DISCUSSION PAPER NO. 2049

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§

July 2016

Abstract

This paper analyzes the trade of information between a data buyer and a data seller. The data buyer faces a decision problem under uncertainty and seeks to augment his initial private information with supplemental data. The data seller is uncertain about the willingness-to-pay of the data buyer due to this private information. The data seller optimally offers a menu of (Blackwell) experiments as statistical tests to the data buyer. The seller exploits differences in the beliefs of the buyer's types to reduce information rents while limiting the surplus that must be sacrificed to provide incentives.

Keywords: selling information, experiments, mechanism design, price discrimination, product differentiation.

JEL Codes: D42, D82, D83.

*We thank Ben Brooks, Giacomo Calzolari, Gabriel Carroll, Gonzalo Cisternas, Jeff Ely, Teck Ho, Emir Kamenica, Alessandro Lizzeri, Alessandro Pavan, Mike Powell, Phil Reny, Mike Riordan, Maher Said, Juuso Toikka, Alex Wolitzky and seminar participants at Berkeley, Bocconi, Bologna, Carnegie Mellon, Chicago, Harvard, Mannheim, Northwestern, NYU, Toulouse, Vienna, Yale and the 2015 World Congress of the Econometric Society for their helpful comments.

†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.

§Yale University, 30 Hillhouse Ave., New Haven, CT 06520, USA, alexey.smolin@yale.edu.

1 Introduction

The mechanisms by which information is traded help 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 or informal 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 canonical 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 relevant to the buyer's decision. Initially, the data buyer has private and partial information about the state. The precision of this private information affects the buyer's willingness to pay for any supplemental information. Thus, from the point of view of the data seller, there are many possible types of the data buyer. We investigate the revenue-maximizing policy for the data seller in terms of how much information the seller should provide and how the seller should price access to her data.

In order to screen the heterogeneous types of the data buyer, the seller offers a menu of products. In our context, these products are Blackwell experiments, i.e., signals that reveal information about the buyer's 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 for a given buyer can be computed independently of its price. Finally, even though the buyer is partially informed by his initial private beliefs, the analysis differs considerably from a belief-elicitation problem. Instead, we cast the problem into a nonlinear pricing framework where the buyer's type is determined by his prior belief. The seller's problem then reduces to *designing* and *pricing* information products in different versions.

The design of information can be phrased in terms of a hypothesis test. Indeed, the entire analysis can be viewed as a pricing model for statistical tests. For concreteness, consider a medical expert (the "data buyer") who seeks to distinguish a null hypothesis H_0 from a mutually exclusive alternative hypothesis H_1. The medical expert is a Bayesian decision maker with a prior distribution over the true hypothesis. The expert can take one of two actions, and each one is optimal under the respective hypothesis.

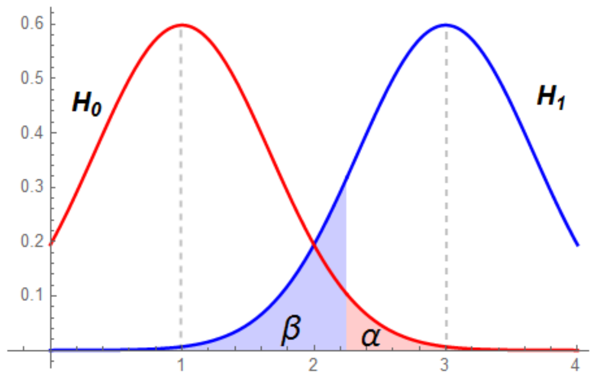

Figure 1: Conditional Distributions of the Test Statistic

The genetic testing company (the "data seller") has access to a test statistic that is distributed conditional on the true state of the world, H_0 or H_1, as in Figure 1. We study how the data seller should reveal (possibly partial) information about the test statistic and how she should price this information. For instance, the seller can reveal the exact value of the test statistic to the buyer. The buyer then chooses the action corresponding to H_1 if the test statistic is above a certain threshold. This information structure induces type-I and type-II errors (α, β) as in Figure 1. It also improves the buyer's payoff relative to acting on his prior information only. The novel aspect in our analysis is that the seller does not know the buyer's prior beliefs and, hence, the buyer's willingness to pay for this information. She therefore must employ a richer mechanism to screen the heterogeneous buyers.

Among other options, the seller can offer the buyer a menu of binary ("Pass/Fail") tests. Each test reports the outcome "Pass" when, say, the actual test statistic is below a different threshold and leads the buyer to take the optimal action under H_0. Each one of these tests yields a different combination of type-I and type-II errors (α, β). Given the information available to the seller in Figure 1, the set of feasible statistical tests is described by the area between the red line and the blue curve in the left panel of Figure 2.

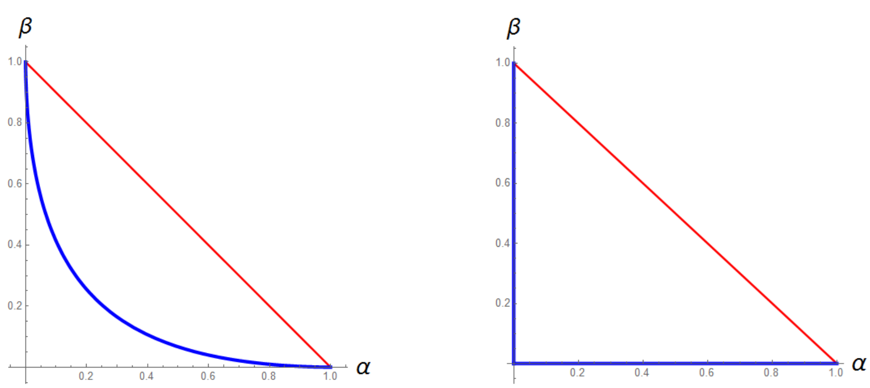

Figure 2: Feasible Information Structures

In this paper, we study a canonical version of this problem. We assume that providing information (e.g., running tests) is costless and that all information structures are available to the seller. In the right panel of Figure 2, this is illustrated by the horizontal and vertical lines which describe statistical tests that minimize one type of error while holding fixed the other type of error at a given level. We then characterize the revenue-maximizing menu of experiments, i.e., statistical tests.

The very nature of information products enriches the scope of price discrimination and leads to new insights relative to the classic nonlinear pricing problem of Mussa and Rosen (1978) and Maskin and Riley (1984). Because information is valuable only if it affects the decision maker's action, buyers with heterogeneous beliefs do not simply value informative signals differently—they disagree on their ranking. In this sense, the value of information naturally has both a vertical quality element and a horizontal positioning element akin to the trade-off between type-I and type-II errors in Figure 2. The data seller can therefore profitably exploit the incompleteness of Blackwell's order and design a menu of information structures that appeal to different buyer types.

As expected, the optimal menu contains, in general, both the fully informative experiment and partially informative, "distorted" experiments. However, 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., the seller induces each buyer to take the correct action in at least one state with probability one. In other words, with two states and actions, the seller induces some buyers to make *either* type-I *or* type-II errors, but not both.

We first provide a full characterization of the optimal menu for the case of two types. We rank the types according to the probability that they make the correct decision in the absence of additional information. Indeed, the seller can provide a larger incremental value to a buyer with a smaller probability of making the correct decision. Thus, the "high" type is the ex ante less informed type, while the "low" type is the ex ante more informed type.

In the optimal mechanism, the high type purchases the fully informative experiment, while the low type buys a distorted experiment. The seller's screening problem is facilitated by the possibility of providing "directional" information: the optimally distorted experiment introduces a type-II error in a subset of signals that the low type considers relatively less likely than the high type, allowing the seller to optimally reduce the information rents of the less-informed type without the need to exclude the more-informed type. As a result, discriminatory pricing can be profitable even if all buyer types have *congruent* prior beliefs, i.e., they would take identical actions in the absence of new information.

Then, we restrict attention to binary states and actions but allow more than two types. This setting yields sharper 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 linear. An intuition analogous to the monopoly pricing problems of Myerson (1981) and Riley and Zeckhauser (1983) suggests that the seller should simply post a price for the fully informative experiment. Instead, we show that the seller's problem consists of screening both *within* and *across* the two subsets of buyer types with congruent priors.

If the seller could perfectly discriminate across groups, the intuition from linear monopoly problems would apply, and the optimal menu would involve two (generically different) posted prices for the fully informative experiment. Flat pricing of the complete information structure is still optimal under strong regularity conditions on the distribution of types. More generally, the optimal menu for a continuum of types 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) when "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 6, we illustrate our results in the context of the markets discussed above, i.e., medical and genetic testing, credit reports, and targeted advertising.

Our paper joins the literature on selling information and monopolistic screening. A large body of literature studies the problem of versioning goods, emphasizing how 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., experiments) and suggests that *versioning* should be an attractive price-discrimination technique (Sarvary, 2012). In this paper, we investigate the validity of these claims in a simple contracting environment à la Mussa and Rosen (1978).

Admati and Pfleiderer (1986, 1990) provide the classic treatment of selling information to a continuum of agents with *ex-ante identical* information who trade in a rational expectations equilibrium. They show that the provision of noisy and hence heterogeneous information turns traders into local monopolists over their idiosyncratic signals, thus preserving the value of acquiring information even in the presence of competition. In contrast, our paper focuses on *ex ante heterogeneous* buyer whose value of information differs due their different prior beliefs. Consequently, 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 information structures.

Eső and Szentes (2007b) and Hörner and Skrzypacz (2015) make more recent contributions to the problem of selling information, focusing on specific, distinct aspects of the problem (i.e., bundling products and information or dynamic information disclosure) in very different contracting environments.

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 both discloses horizontal match-value information and sets a price. Several papers, including 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 to a pricing policy.[1] In related contributions, Lizzeri (1999) considers vertical information acquisition and disclosure by a monopoly intermediary, and Abraham, Athey, Babaioff, and Grubb (2014) study vertical information disclosure in auctions. In a setting similar to ours, Babaioff, Kleinberg, and Paes Leme (2012) derive the optimal dynamic mechanism for selling information sequentially.

Commitment to a disclosure policy is also 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.

2 Model

A single decision maker faces a decision problem under uncertainty. The state of nature ω is drawn from a finite set Ω. The decision maker must choose an action a from a finite set A. We assume without loss of generality that the sets of actions and states have the same cardinality $|A| = |\Omega| = N$.

Payoffs The ex post utility function of the decision maker is denoted by $u(a, \omega)$. We specialize the utility function by assuming that the decision maker seeks to match the state and the action. His ex post utility function $u(a, \omega)$ is therefore given by

$$u(a, \omega) = \mathbf{1}_{[a=\omega]}. \tag{1}$$

This formulation assumes that the decision maker assigns uniform gains and losses across states. More general formulations are equally tractable but complicate the exposition.

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

Prior Information The type of the decision maker, $\theta \in \Theta = \Delta\Omega$, consists of his interim beliefs about the states; these beliefs $\theta \in \Theta$ are private information and can be generated from a common prior and privately observed signals. In particular, suppose there is a common prior $\mu \in \Delta\Omega$. The decision maker privately observes an informative signal $r \in R$ according to a commonly known information structure

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

The decision maker then forms his interim beliefs $\theta \in \Delta\Omega$ via Bayes' rule:

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

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

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

which we take as a primitive of our model.

Incremental Information The data buyer seeks to augment his initial private information with additional information – or experiments – from the data seller in order to improve the quality of his decision making. An *experiment* (or an information structure) $I = \{\pi, S\}$ consists of a set of signals $s \in S$ and a likelihood function:

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

We assume throughout that the realizations of the buyer's private signal $r \in R$ and of the signal $s \in S$ from any experiment I are independent, conditional on the state ω. In other words, the buyer and the seller draw their information from independent sources.[2]

Experiment (or Information Structure) For a given information structure I, we denote by π_{ij} the conditional probability of signal s_j in state ω_i,

$$\pi_{ij} = \Pr(s_j \mid \omega_i),$$

[2] As usual, the model allows for the alternative interpretations of a single buyer and a continuum of buyers. With a continuum of buyers, we will assume that states ω are identically and independently distributed across buyers and that different buyers' signals are conditionally independent.

where $\pi_{ij} \geq 0$ and $\sum_j \pi_{ij} = 1$ for all i. We then obtain the stochastic matrix

$$\begin{array}{c|ccccc} I & s_1 & \cdots & s_j & \cdots & s_J \\ \hline \omega_1 & \pi_{11} & \cdots & \pi_{1j} & \cdots & \pi_{1J} \\ \vdots & \vdots & \ddots & & & \\ \omega_i & \pi_{i1} & & \pi_{ij} & & \\ \vdots & \vdots & & & \ddots & \\ \omega_N & \pi_{N1} & & & & \pi_{NJ} \end{array} \tag{2}$$

The following experiments are of particular interest:

1. the *fully informative* experiment with $\pi_{ii} = 1$ for all i and $|N| = |J|$;
2. a *non-dispersed* experiment that contains a unit row vector, $\pi_{ii} = 1$ for some i;
3. a *noise-free* experiment that contains a column s_j with only one positive entry, $\pi_{jj} > 0$ for some j.

A non-dispersed experiment conveys the information about *some* state i with a single signal s_i. Conversely, any signal other than s_i rules out the state ω_i. A noise-free experiment reveals *some* state ω_i without noise, i.e., the posterior belief after observing s_i is concentrated at ω_i. The fully informative experiment is non-dispersed and noise-free everywhere.

An *information policy* or simply a *menu* $\mathcal{M} = \{\mathcal{I}, t\}$ consists of a collection of information structures and an associated tariff function, i.e.,

$$\mathcal{I} = \{I\}, \qquad t : \mathcal{I} \rightarrow \mathbb{R}^+.$$

Our goal is to characterize the revenue-maximizing menu for the seller.

We deliberately focus on the pure problem of designing and selling information structures. We therefore assume that the seller commits to a menu before the realization of the state ω and the type θ, and that neither the buyer's action a, the realized state ω nor the signal s are contractible. Thus, while the data buyer holds private information, scoring rules and other belief-elicitation schemes are not available to the seller to the extent that the conditioning event is not verifiable. 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 $I \in \mathcal{I}$ and pays the corresponding price t;
4. the buyer observes a signal s from the experiment I and chooses an action a.

The data seller is unconstrained in her choice of information structures (e.g., the seller can improve upon the buyer's original information with arbitrarily accurate signals), and the marginal cost of providing the information is nil. These assumptions capture settings in which sellers own very precise datasets and distribution costs are negligible.

3 Information Design

We now describe the value of the buyer's initial information and the incremental value of an information structure $I = \{\pi, S\}$. Let θ_i denote the interim probability that type θ assigns to state ω_i, with $i = 1, ..., N$. For an arbitrary utility function $u(a, \omega)$, the value of the buyer's problem under prior information is given by

$$u(\theta) \triangleq \max_a \left\{ \sum\nolimits_{i=1}^{N} \theta_i u(a, \omega_i) \right\}.$$

Now, suppose that the buyer has access to an incremental information structure $I = \{\pi, S\}$ that generates signals $\pi : \Theta \to \Delta S$, where the number of signal realizations is $|S| = J$. The buyer chooses an action after updating her beliefs, leading to the expected (gross) utility

$$u(\theta, s_j) \triangleq \max_a \left\{ \sum\nolimits_{i=1}^{N} \frac{\theta_i \pi_{ij}}{\sum\nolimits_{i=1}^{N} \theta_i \pi_{ij}} u(a, \omega_i) \right\}. \tag{3}$$

The marginal distribution of signals s_j from the perspective of type θ is given by

$$\Pr[s_j \mid \theta] = \sum\nolimits_{i=1}^{N} \theta_i \pi_{ij}.$$

Integrating over all signal realizations s_j and subtracting the value of prior information, the (net) value of an information structure I for type θ is given by

$$V(I, \theta) \triangleq \mathbb{E}_s[u(\theta, s)] - u(\theta) = \sum_{j=1}^{J} \max_a \left\{ \sum\nolimits_{i=1}^{N} u(a, \omega_i) \theta_i \pi_{ij} \right\} - u(\theta).$$

With the restriction to utility functions that match action to state, as defined earlier in (1), the value of information takes the simpler form:

$$V(I, \theta) = \sum_{j=1}^{J} \max_i \{\theta_i \pi_{ij}\} - \max_i \{\theta_i\}. \tag{4}$$

The value of the prior information, given by $\max_i \theta_i$, is generated by choosing the action that has the highest prior probability. The value of the information structure is generated by choosing an action on the basis of the posterior belief $\theta_i \pi_{ij}$ induced by every signal s_j. Quite simply, the value of an information structure is given by the *incremental probability* of the buyer choosing the correct action given the state.

An information structure I is valuable only if different signals s_j lead to different actions a_i. In particular, if $\arg\max_i \theta_i \pi_{ij}$ is constant across all the signals s_j, then (4) immediately implies that $V(I, \theta) = 0$. Conversely, the fully informative experiment I^* guarantees that the buyer takes the correct action in every realization of state ω_i. The value of the fully informative experiment is therefore given by

$$V(I^*, \theta) = 1 - \max_i \theta_i. \tag{5}$$

In Figure 3, we illustrate the value of an information structure in a model with 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, $\theta = (\theta_1, \theta_2, 1 - \theta_1 - \theta_2)$. We display below the fully informative experiment I^* and a partial information experiment I as a function of the buyer's prior.[3]

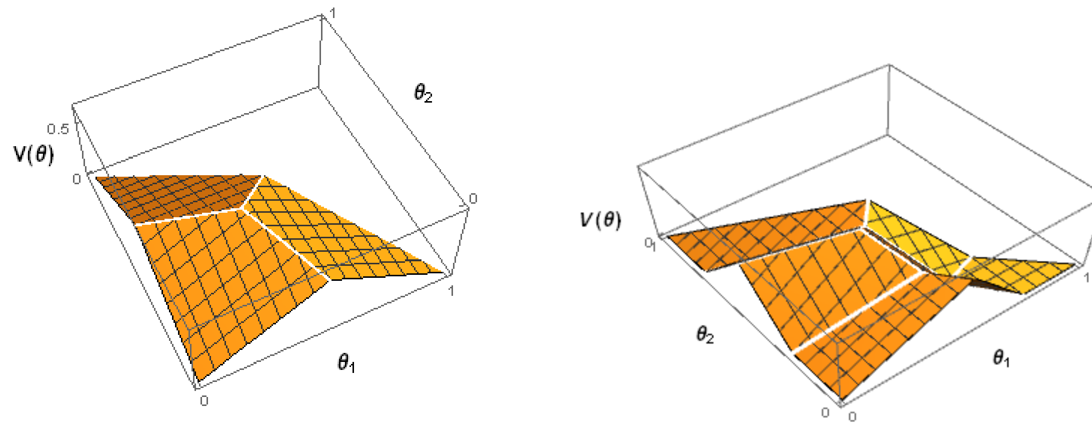

Figure 3: Value of Fully and Partially Informative Experiments, $N = 3$

Viewed as a function of the types, the value of an experiment $V(I, \theta)$ is piecewise linear in θ with a finite number of kinks. The linearity of the value function is a consequence of

[3] The information structure in the right panel is given by the likelihood matrix:

I	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

the Bayesian nature of our problem, where types are prior probabilities over 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 incremental information. The upward kinks are generated by the max operator in (3) and reflect changes in the buyer's preferred action upon observing a signal.

We now characterize the menu of experiments that maximizes the seller's profits. By the revelation principle, we may state the seller's problem as designing a direct mechanism $\mathcal{M} = \{I(\theta), t(\theta)\}$that assigns an information structure I denoted by $I(\theta)$ to each type θ of the buyer at a price $t(\theta)$. The seller's problem consists of maximizing the expected transfers

$$\max_{\{I(\theta),t(\theta)\}} \int_{\theta\in\Theta} t(\theta)\,\mathrm{d}F(\theta) \tag{6}$$

subject to incentive-compatibility constraints

$$V(I(\theta),\theta) - t(\theta) \geq V(I(\theta'),\theta) - t(\theta'), \quad \forall\,\theta,\theta' \in \Theta,$$

and individual rationality constraints

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

The seller's problem can be simplified by reducing the set of optimal information structures: a very tractable class of experiments can generate every outcome (i.e., distribution of buyer actions and seller profits) attainable with an incentive-compatible menu.

Proposition 1 (Responsive Information Structures)
Every incentive-compatible menu $\mathcal{M}$ can be represented as a collection of information structures $I(\theta)$ in which every type θ takes a different action after observing each signal $s \in S(\theta)$.

This result follows from the insight of Myerson (1986) for multi-stage games, where signals consist of recommendations over actions. The intuition is straightforward: consider an incentive-compatible information policy that contains an experiment $I(\theta)$ with more signals than actions; the seller could combine all signals in $I(\theta)$ that lead to the same choice of action for type θ; clearly, the value of this experiment remains constant for type θ (who does not modify his behavior). In addition, because the original signal is strictly less informative than the new one, $V(I(\theta),\theta')$ decreases (weakly) for all $\theta' \neq \theta$, relaxing the incentive constraints. We refer to incentive-compatible information structures in which the truthtelling agent chooses a distinct action for every distinct signal in a *responsive* information structure.

An immediate corollary of Proposition 1 is that, without loss of generality, we can restrict attention to experiment in which the signal space has at most the cardinality of the action space. This process allows us to write the likelihood function of every information structure, described earlier in (2), as a square matrix with $|J| = |N|$

$$\begin{array}{c|ccccc} I(\theta) & s_1 & & s_i & \cdots & s_N \\ \hline \omega_1 & \pi_{11} & & \pi_{1i} & \cdots & \pi_{1N} \\ & & \ddots & & & \\ \omega_i & \pi_{i1} & & \pi_{ii} & & \\ \vdots & \vdots & & & \ddots & \\ \omega_N & \pi_{N1} & & & & \pi_{NN} \end{array} . \tag{7}$$

Moreover, because every signal sent with positive probability under experiment $I(\theta)$ leads to a different action, we can order signals so that each $s_i \in S(\theta)$ recommends the corresponding action a_i to type θ. The resulting value of experiment $I(\theta)$ for type θ can be written as

$$V(I(\theta), \theta) = \sum_{i=1}^{N} \theta_i \pi_{ii} - \max_i \{\theta_i\}. \tag{8}$$

Thus, the diagonal entries of the matrix generate the value of the experiment $I(\theta)$ for type θ. By contrast, the off-diagonal entries serve as instruments to guarantee the incentive-compatibility constraints across types.[4]

We emphasize that the square matrix property does not require that every action a_i is recommended with strictly positive probability. In particular, some signals s_i may never be sent, thus corresponding to a column vector of zeros at the i-th position. Finally, the nature of the value of information imposes additional structural properties on an optimal menu.

Proposition 2 (Non-Dispersed Information)
In any optimal menu of experiments:

1. *the fully informative experiment I^* is part of the menu;*
2. *every experiment is non-dispersed, 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 I^* the most among all experiments. Suppose, then, that I^* is not part of the menu and denote the most expensive item currently on the menu by $\overline{I}$.

[4] Given responsive information structures, we can discard the first max operator in the value of experiment $I(\theta)$ for any truthtelling type θ. We still need to use the original formulation (4) when computing type θ's value of misreporting and buying experiment $I(\theta')$.

The seller can replace experiment $\bar{I}$ with the complete information structure I^*, keeping all other prices constant, and charging a strictly higher price for I^* than for $\bar{I}$. The new menu increases the revenue of the seller without lowering the net utility of any buyer type.

The second part implies that, for every experiment I, there exists a state ω_i that leads buyer θ to take the correct action with probability one. In other words, any time the realized state is ω_i, signal s_i is sent, and the buyer takes action a_i. Conversely, the buyer can rule out the state ω_i after observing any other signal $s_k \neq s_i$. This result can also be established by an improvement argument. Consider an experiment I and suppose there was no diagonal entry $\pi_{ii} = 1$. The seller can profitably increase all diagonal elements π_{ii} by the same amount until the first diagonal entry reaches 1. Concurrently, the seller can raise the price of the experiment by the same amount. Since the beliefs of each type θ over the states sum to one, the uniform increase in the probability of matching the state is valued uniformly across all types $\theta \in \Theta$. Thus, a commensurate increase in the price completely offsets the value of the additional information provided and, hence, leaves all incentive-compatibility and participation constraints unaffected.

Having established the general structural properties of the optimal information policy, we now provide a complete characterization of the optimal menu in two important environments. In Section 6, we discuss the implications of these properties for the observable characteristics of real-world information products.

4 Optimal Menu with Binary Types

In this section, we allow for any finite number N of states and actions but restrict the private information to consist of two types only. A type is thus a N-dimensional probability vector. By contrast, in the next section, we restrict the state and action space to be binary but allow the private information to consist of a continuum of (one-dimensional) private beliefs. These two settings provide complementary results regarding the nature of the optimal pricing and design of information. We discuss towards the end of this section the difficulties that arise in generalizing the results beyond two types and two states (and actions) simultaneously. This will also give us the chance to discuss the current model's relationship with the multi-item bundling problem.

We thus consider a setting with N states and actions and two possible buyer types, $\theta \in \{\theta', \theta''\}$. In the absence of additional information, each type θ would choose the action that matches the state that has the highest prior probability in the vector $\theta = (\theta_1, ..., \theta_i, ..., \theta_N)$. The value of a perfectly informative experiment (5) is therefore larger for the type $\theta \in \{\theta', \theta''\}$ who displays a lower prior probability on the state he deems most likely. This suggests that the data seller can provide the largest value to the type θ whose largest prior probability is smallest. Equivalently, the lower of the two types is willing to pay more to resolve the uncertainty completely. We therefore identify the high type as

$$\theta^H \triangleq \arg\min_{\theta\in\{\theta',\theta''\}} \max_{i\in N} \theta_i,$$

and the low type as

$$\theta^L \triangleq \arg\max_{\theta\in\{\theta',\theta''\}} \max_{i\in N} \theta_i.$$

We denote the probability that the high type appears in the population by

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

To facilitate the description of the seller's problem, we establish two familiar properties: "no distortion at the top," confirming the intuition that the least-informed type is the "high type" who receives an efficient allocation; and "no rent at the bottom," i.e., for the low type, which coincides with the ex ante more-informed type.

Proposition 3 (Binding Constraints)
In any optimal menu, type θ^H purchases the fully informative experiment I^*; the incentive-compatibility constraint of type θ^H binds; and the participation constraint of type θ^L binds.

With this result in place, the remaining question is what kind of information the low and ex ante more-informed type receives in the optimal menu and what prices the seller can charge. It is useful to develop some intuition for the case of two actions and two states.

4.1 Binary Actions and States

For the moment, consider an environment with two actions $a \in \{a_1, a_2\}$ and two states $\omega \in \{\omega_1, \omega_2\}$. The buyer type is a two-dimensional vector $\theta = (\theta_1, \theta_2)$. However, with minor abuse of notation, we identify each type simply with the prior probability of the state ω_1:

$$\theta \triangleq \Pr[\omega = \omega_1]. \tag{9}$$

The definition of high and low types therefore implies

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

In light of Proposition 1, it is sufficient to consider for every type θ information structures $I(\theta)$ that generate (at most) two signals:

$$
\begin{array}{c|cc}
I(\theta) & s_1 & s_2 \\
\hline
\omega_1 & \pi_{11}(\theta) & 1-\pi_{11}(\theta) \\
\omega_2 & 1-\pi_{22}(\theta) & \pi_{22}(\theta)
\end{array} . \tag{10}
$$

Without loss of generality, we assume $\pi_{11}(\theta) + \pi_{22}(\theta) \geq 1$. 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

$$
\frac{\pi_{11}(\theta)}{1-\pi_{22}(\theta)} \geq \frac{1-\pi_{11}(\theta)}{\pi_{22}(\theta)}.
$$

As explained in (8) above, the diagonal entries of the matrix π generate the probability that information structure $I(\theta)$ allows type θ to match the realized state ω_i with his action a_i. Therefore, we can write type θ's value for an arbitrary information structure I as

$$
V(I,\theta) = \max\left\{\pi_{11}\theta + \pi_{22}(1-\theta) - \max\{\theta, 1-\theta\}, 0\right\}, \tag{11}
$$

where the first max operator accounts for the possibility that type θ may not follow the recommendation of one of the signals and, hence, derive no value from the incremental information. (For each type θ and experiment $I(\theta)$, Proposition 1 also implies that we can ignore the max operator.) In Figure 4, we illustrate how the value of information changes as a function of the type θ for two different experiments, namely, the fully informative experiment $I' = (\pi'_{11}, \pi'_{22}) = (1,1)$ and a partially informative experiment $I'' = (\pi''_{11}, \pi''_{22}) = (1/2, 1)$.

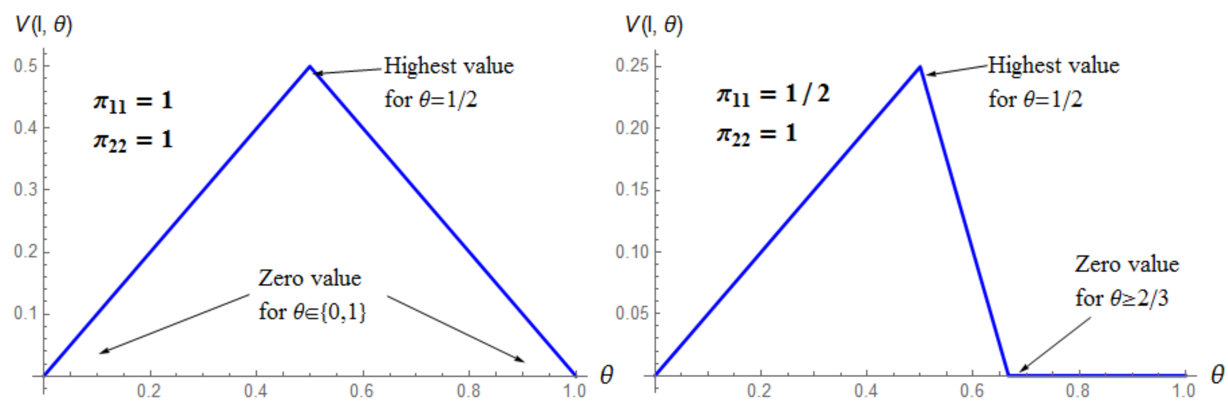

Figure 4: Value of Fully and Partially Informative Experiments, $N = 2$

The behavior of the value of information as a function of the buyer's interim beliefs θ reflects many properties that we have already formally established. First, the most valuable type for the seller is the ex ante least-informed type, namely, $\theta = 1/2$. Conversely, the most-informed types $\theta \in \{0, 1\}$ have zero value of information. The linear decline in each direction away from $\theta = 1/2$ follows from the linearity of the value of information in the interim probability. Second, when we consider any asymmetric experiment, such as the one displayed in the right panel of Figure 4, the distance from the least-informed type $|\theta - 1/2|$ is *not* a sufficient statistic for the value of information. The different slopes on each side of $\theta = 1/2$ 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 high-dimensional (in this case, two dimensional). In particular, information always has both a vertical (quality) and a horizontal (positioning) dimension. 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 ranking of all partial information structures, and agree only on the top ranking of the complete information structure.

Towards more general results, we offer the following distinction: the interim beliefs of the two types are said to be *congruent* if both types would choose the same action in the absence of the additional information. If we adopt the convention that the high type chooses action a_1 under prior information (i.e., $\theta^H \geq 1/2$), then priors are *congruent* if $1/2 \leq \theta^H \leq \theta^L$ and are *noncongruent* if $\theta^L \leq 1/2 \leq \theta^H$. In Proposition 4, we complete the characterization of the optimal menu in the setting with binary states.

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.*

Congruent Priors. In the case of congruent priors, the argument is related to the classic monopoly pricing problem. By Proposition 2, we know that the optimal information structure has to be non-dispersed. With congruent priors, both types would choose action a_1 absent any additional information. As we establish formally in Proposition 5, the seller sets $\pi_{11}(\theta^L) = 1$. The issue is therefore how much information to provide about state ω_2, i.e.,

$I(\theta^L)$	s_1	s_2
ω_1	1	0
ω_2	$1 - \pi_{22}(\theta^L)$	$\pi_{22}(\theta^L)$

Any partially informative experiment with $\pi_{22}(\theta^L) \in (0,1)$ is also valuable for the high type, whose incentive constraint in this case reduces to

$$t\left(\theta^H\right) - t\left(\theta^L\right) \leq \underbrace{\left(1-\theta^H\right)}_{\Pr(\omega_2)} \cdot \underbrace{\left(1-\pi_{22}\left(\theta^L\right)\right)}_{\text{additional precision}}.$$

Hence, we observe that the utility of the buyer type and both the objective and the constraints in the seller's problem are linear in the choice variable π_{22}. We can therefore appeal to the result of Riley and Zeckhauser (1983) on monopoly pricing 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 to allocate it all, hence $\pi_{22} \in \{0,1\}$. As in the monopoly pricing problem, which policy is optimal depends on the size of each segment. Thus, the low type receives the fully informative experiment if the high type is sufficiently infrequent relative to the willingness to pay for the information, or

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

Noncongruent Priors. In the case of noncongruent priors, or $\theta^L < 1/2 < \theta^H$, both the argument and the result are rather distinct from those of the classic monopoly problem. With noncongruent priors, in the absence of any information, the two types would choose different actions. Therefore, each type has a willingness to pay to learn more about the state that the other type believes to be more likely. Thus, the seller can provide information that is valuable to one type but has zero value for the other type. For example, suppose that $\pi_{22}(\theta^L) = 1$ and $\pi_{11}(\theta^L)$ is chosen such that, after receiving signal s_2, the high type θ^H places equal posterior probability on ω_1 and ω_2:

$$\theta^H\left(1-\pi'_{11}\right) = \left(1-\theta^H\right) \Longleftrightarrow \pi'_{11} = \frac{2\theta^H-1}{\theta^H} \in (0,1). \tag{12}$$

Type θ^H is thus indifferent between action a_1 and a_2. Because he would choose action a_1 absent any information, the information structure I'

$$\begin{array}{c|cc} I' & s_1 & s_2 \\ \hline \omega_1 & \pi'_{11} & 1-\pi'_{11} \\ \omega_2 & 0 & 1 \end{array} \tag{13}$$

does not lead to a strict improvement in his decision making (and, hence, his utility). By contrast, the low type assigns positive value to I' since signal s_1 would lead type θ^L to match his action to state ω_1, which type θ^L would never have achieved absent any 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.

In fact, since the seller could charge the low type a positive price for the above information structure I', whereas the high type would assign zero value to it, the incentive constraint of the high type would not bind. Hence, the seller can offer an even more informative experiment I'' to the low type until the incentive constraint of the high type eventually becomes binding. As the partially informative signal I'' has a positive price tailored to type θ^L, i.e., $t\left(\theta^L\right) = \theta^L \pi''_{11}$, the high type can be made indifferent between buying the information structure or not by setting

$$\underbrace{1-\theta^H}_{t(\theta^H)} - \underbrace{\theta^L \pi''_{11}}_{t(\theta^L)} = \underbrace{\theta^H}_{\Pr(\omega_1)} \cdot \underbrace{(1-\pi''_{11})}_{\text{additional precision}} \iff \pi''_{11} = \frac{2\theta^H - 1}{\theta^H - \theta^L}. \tag{14}$$

Experiment I'' is as informative as possible while satisfying the high type's incentive compatibility constraint and *both* participation constraints with equality. The argument in support of a partially informative experiment is illustrated below in Figure 5, which depicts the value of two experiments *net of the price* as a function of the buyer's type $\theta \in [0, 1]$. We consider the case on noncongruent preferences with $\theta^L = 1/5 < 1/2 < 7/10 = \theta^H$. Both types are equally likely.

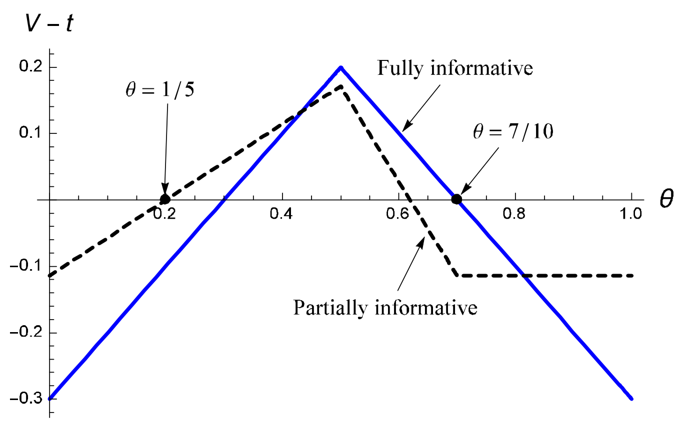

Figure 5: Suboptimal Menu: $(\pi_{11}, \pi_{22}) \in \{(1, 1), (4/7, 1)\}$

The net value of the fully informative experiment I^* is depicted by the straight line at a price $t\left(\theta^H\right) = 1 - \theta^H$ 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'_{11} , as described by (12). The associated partial information experiment I' 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, the net value of the partial information experiment I' is *strictly* below the net value of the fully informative experiment for the high type.

In fact, the incentive compatibility constraints are slack with this menu of experiments for both types. Therefore, the seller can increase the quality of the experiment sold to the low type θ^L while satisfying the incentive constraint for the high type θ^H. The corresponding experiment I'' with the associated π''_{11} as given by (14) is illustrated in Figure 6.

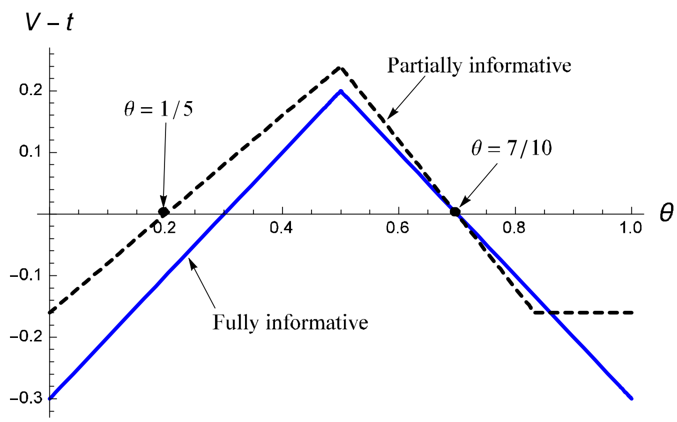

Figure 6: Optimal Menu: $(\pi_{11}, \pi_{22}) \in \{(1,1),(4/5,1)\}$

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). In the optimal menu, the partial informative experiment I'' is eventually replaced by the fully informative experiment if the proportion of low types becomes large.

To summarize the results in the binary model, (i) with congruent priors, both types receive complete information if they are sufficiently similar or if the low type is relatively frequent, while (ii) with noncongruent priors, the optimal menu offers the perfect information structure 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, as we would expect from a conventional screening model.[5] Furthermore, the high type receives positive rents only if he is pooled with the low type. Otherwise, the seller extracts the entire surplus generated.

[5]In Section 4.2, we show that exclusion requires not only that types agree on the most likely state but also that their relative beliefs about any two other states coincide. This requirement is vacuous with two states only but non-generic with $N > 2$.

Relationship to Type I/II Errors In a binary action environment, any information structure can be readily interpreted as a statistical test offered to the buyer. The test may be subject to type-I and type-II statistical errors.[6] In line with the standard notation in statistics, we denote the probability of type-I and type-II errors as α and β, respectively. We can then represent the information structure as a hypothesis test, say for the low type who has a null hypothesis $H_0 = \{\omega_2\}$,

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

In the case of noncongruent priors, in view of Proposition 4, the optimal test offered to the low type has the following structure:

I	s_1	s_2
ω_1	$1-\beta$	β .
ω_2	0	1

Proposition 4 thus asserts that the optimal hypothesis test eliminates the type-I error ($\alpha = 0$) and allows for some type-II error ($\beta > 0$). In fact, the optimal hypothesis test minimizes the type-II error subject to the incentive compatibility constraint. Conversely, we can ask what would happen if the high type who has a null hypothesis $H_0 = \{\omega_1\}$ (as $\theta^H > 1/2$) were to misreport his type and attempt to purchase the hypothesis test meant for the low type, $\theta^L < 1/2$. Then, he would purchase a hypothesis test with strikingly different properties, namely, one that induces no type-II errors but substantial type-I errors. The difference in the error structure of the test across types supports the separation in the optimal menu.

We now examine the comparative statics of the optimal menu. Because the high type buys the fully informative experiment and the low type's experiment always has zero type-I error, the optimal information structure is determined by the level of type II error.

Corollary 1 (Comparative Statics of Type II Error)
The type II error β in the optimal experiment $I(\theta^L)$ is

1. *increasing in the frequency γ of the high type;*
2. *increasing in the precision of the prior of the high type $\left|\theta^H - 1/2\right|$;*
3. *decreasing in the precision of the prior of the low type $\left|\theta^L - 1/2\right|$.*

[6] 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.

Thus, even though the shape of the optimal menu depends on whether types have congruent or noncongruent priors, the comparative statics of the optimal information structure are robust across the different scenarios. The rent extraction vs. efficiency trade-off is resolved at the expense of the low type whenever *(i)* the fraction of high types is larger, *(ii)* the high type has a higher willingness to pay for information, or *(iii)* the low type has a lower willingness to pay for the complete information structure.

4.2 Many States and Actions

With more than two states and actions, the seller's problem involves all the difficulties associated with multidimensional screening. Nevertheless, the structural properties of the optimal experiments extend to multidimensional settings. We first characterize the optimal menu within a binary type environment and then discuss the issues that arise with an arbitrary finite number of types, as well as the relationship with monopoly bundling problems.

We allow for N states and actions but momentarily only two types of buyers, $\theta \in \left\{\theta^L, \theta^H\right\}$. We recall that the high-value type satisfies $\max_i \theta_i^H < \max_i \theta_i^L$ and that Proposition 3 reduces the seller's problem to finding the low type's experiment $I\left(\theta^L\right)$ and the two payments t^L, t^H. Our approach is to solve a *relaxed* problem where we require the high-value type to take action a_i upon observing signal s_i *even when buying the low-value type's experiment*. We then show that the solution to the relaxed problem satisfies the original constraints; hence, it also solves the full problem. Formally, we replace the incentive compatibility constraint

$$U\left(\theta^H\right) \geq \sum_{j=1}^{N} \max_i \left\{\theta_i^H \pi_{ij}(\theta^L)\right\} - \max_i \left\{\theta_i^H\right\} - t(\theta^L),$$

with the weaker constraint that

$$U\left(\theta^H\right) \geq \sum_{j=1}^{N} \theta_j^H \pi_{jj}(\theta^L) - \max_i \left\{\theta_i^H\right\} - t(\theta^L).$$

Without loss of generality, we arrange the signals in the experiment $I(\theta^L)$ so that the low type chooses action a_i when observing signal s_i. Under these assumptions and ignoring additive constant terms, the seller's problem consists of choosing the information rent for the high type, $U(\theta^H) \geq 0$, and the diagonal entries $\pi_{ii} \in [0,1]$ of the low-value type's information structure $I(\theta^L)$ to maximize

$$(1-\gamma)\underbrace{\left(\sum_{i=1}^{n}\theta_i^L\pi_{ii}(\theta^L)-\max_i\theta_i^L\right)}_{t(\theta^L)=V(I(\theta^L),\theta^L)} + \underbrace{\gamma\Big(1-\max_i\theta_i^H-U(\theta^H)\Big)}_{t(\theta^H)=V\left(I^*,\theta^H\right)-U(\theta^H)}, \tag{15}$$

subject to the high type's incentive-compatibility constraint, which can be rewritten as

$$\sum_{i=1}^{n}(\theta_i^L-\theta_i^H)\pi_{ii} \geq \max_i\theta_i^L-\max_i\theta_i^H-U(\theta^H). \tag{16}$$

By inspecting (15) and (16), it is immediate that a higher π_{ii} increases profits and relaxes the constraint for all ω_i such that $\theta_i^L > \theta_i^H$. In other words, if the low type believes a given state ω_i is more likely to occur than does the high type, the seller should send signal s_i with probability one in that state. In particular, the optimal menu has $\pi_{ii} = 1$ for the state corresponding to the low type's default action.

With two states only (e.g., in example of Proposition 4), this implies the remaining entry in row i is $\pi_{ij} = 0$. With $N > 2$ states, the seller has considerably more instruments to distort the low type's allocation. In particular, the partially informative experiment may contain fewer signals than available actions–and, thus, the seller may "drop" some signals from $I(\theta^L)$ to reduce the information rents. The logic above then suggests that the seller should distort the signal distribution in states the high type deems very likely but the low type does not.

To formalize this intuition, we re-order the states ω_i by the likelihood ratios of the two types' beliefs. In particular, let

$$\frac{\theta_1^L}{\theta_1^H} \leq \cdots \leq \frac{\theta_i^L}{\theta_i^H} \leq \cdots \leq \frac{\theta_N^L}{\theta_N^H}. \tag{17}$$

We then define two particular states. The first state is defined as

$$i^* = \min\left\{i : \frac{\theta_i^L}{\theta_i^H} \geq \gamma\right\}, \tag{18}$$

which is well-defined because the likelihood ratio must exceed one for some i.

To define the second state, consider the following quantity:

$$k_j \triangleq \frac{\sum_{i=j+1}^{n}\left(\theta_i^L-\theta_i^H\right)-\max_i\theta_i^L+\max_i\theta_i^H}{\theta_j^H-\theta_j^L}. \tag{19}$$

In particular, k_j corresponds to the value of π_{jj} that satisfies (16) with equality when the high type's rent is nil, $\pi_{ii} = 1$ for all states $i > j$, and $\pi_{ii} = 0$ for all $i < j$. Let

$$j^* \triangleq \min j : k_j \geq 0, \tag{20}$$

which is also well-defined since the definition of the low type implies that at least $k_N \geq 0$.

Proposition 5 establishes several properties of any optimal menu.

Proposition 5 (Optimal Menu with Two Types)

1. *Experiment* $I(\theta^L)$ *has* $\pi_{ii} = 0$ *for all* $i < \min\{i^*, j^*\}$ *and* $\pi_{ii} = 1$ *for all* $i > \min\{i^*, j^*\}$.
2. *If* $j^* < i^*$, *then* $I(\theta^L)$ *has* $\pi_{j^*j^*} = k_{j^*}$ *and* $U(\theta^H) = 0$.
3. *If* $j^* \geq i^*$, *then* $I(\theta^L)$ *has* $\pi_{i^*i^*} = 1$ *and* $U(\theta^H) > 0$.

To complete the description of the optimal experiment $I(\theta^L)$, we need to specify the distribution of signals π_{ij} for $i < \min\{i^*, j^*\}$. In the Appendix, we construct an algorithm that assigns the entries π_{ij} in such a way that both types always follow the recommendation implied by each signal. Therefore, the solution to the relaxed problem solves our original problem. The resulting optimal experiment $I(\theta^L)$ has a lower triangular shape, with $\pi_{ii} \in \{k_i, 1\}$ depending on whether $i^* \lesseqgtr j^*$, as described in Proposition 5:

$$\begin{matrix} 0 & \cdots & 0 & \pi_{1i} & \cdots & \cdots & \pi_{1n} \\ \vdots & & \vdots & \vdots & & & \vdots \\ \vdots & & \vdots & \pi_{ii} & \cdots & \cdots & \pi_{in} \\ \vdots & & \vdots & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{matrix} \tag{21}$$

Intuitively, the seller's choice of experiments in the optimal menu depends on two factors: (a) the two types' relative beliefs, and (b) the distribution of types in the population. Proposition 5 shows that the problem is separable in this respect. In particular, the optimal experiment $I(\theta^L)$ is chosen from a discrete set that depends only on the two types (θ^H, θ^L). Each element of this set distorts progressively more signals, as in (21). The distribution of types γ then determines the optimal element of the set, i.e., the number of null columns and value of the diagonal entry in the critical state $\min\{i^*, j^*\}$.

More specifically, the extent of the two types' belief *disagreement* guides the seller's choice of a partially informative experiment. The seller is most willing to introduce noise in signals that the low type considers relatively less likely than the high type because these distortions facilitate screening without sacrificing surplus. That is, distortions are more likely to occur in states that maximize relative disagreement. The extent of disagreement is captured by the critical state j^* that represents the minimum number of signals that must be eliminated in order to satisfy incentive compatibility, while holding the high type to his reservation utility.

Finally, the profitability of distorting experiment $I(\theta^L)$ depends on the distribution of types. In particular, the measure of high types (captured by the critical state i^*) represents the shadow cost of providing information to the low types. For high shadow cost, the monopolist is willing to distort the allocation by removing as many signals as necessary to satisfy (16) and hold the high type to his participation constraint. For low opportunity cost, the monopolist prefers to limit distortions and concede rents to the high type. Thus, the informativeness of the low type's experiment is decreasing in the fraction of high types γ.

The following examples with three states and actions illustrate Proposition 5.

Example 1 (Noncongruent Priors) *Let* $\theta^L = (1/10, 1/10, 4/5)$ *and* $\theta^H = (2/5, 3/10, 3/10)$. *Because* $k_1 < 0 < k_2$, *the partially informative experiment* $I(\theta^L)$ *can involve dropping signal* s_1, *i.e., setting* $\pi_{11} = 0$. *In particular, the optimal experiment* $I(\theta^L)$ *is given by*

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

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

$I(\theta^L)$	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 $\gamma > 1/3$.		

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

Example 2 shows that discriminatory pricing can be profitable even if the two types have congruent beliefs, as long as the likelihood ratio θ_i^L/θ_i^H takes more than two distinct values.

Example 2 (Congruent Priors) *Let* $\theta^H = (3/10, 2/5, 3/10)$ *and* $\theta^L = (1/10, 1/2, 2/5)$. *Because* $k_1 > 0$, *an optimal experiment* $I(\theta^L)$ *is given by*

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

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

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

The key property of Example 2 is that, even though both types initially deem state ω_2 the most likely, they disagree on the relative likelihood of states ω_1 and ω_3, i.e., $\theta_1^H/\theta_3^H \neq \theta_1^L/\theta_3^L$. If $\gamma > 1/3$, then in the optimal menu, the seller offers type θ^L an experiment $I(\theta^L)$ that reveals state ω_1 with a conditional probability of 1/2. Because type θ^L assigns prior probability 1/10 to state ω_1 only, this distortion does not reduce his willingness to pay considerably. Conversely, the prior beliefs of type θ^H assign a much larger weight to state ω_1, which means he perceives the quality of experiment $I(\theta^L)$ as significantly lower, enabling the seller to charge a high price for the full information structure $I(\theta^H)$.

Thus, with more than two states, the seller can exploit disagreement along *any dimension* and extract all the surplus through discriminatory pricing. Nonetheless, it is not always profitable to offer two distinct experiments. In Example 2, if high-value types are sufficiently scarce ($\gamma \leq 1/3$), the seller prefers to pool the two types. Corollary 2 identifies two polar cases in which the distribution of types pins down the rent of the high type.

Corollary 2 (Information Rents)

1. *If $\gamma < \min_i \theta_i^L/\theta_i^H$, both types purchase the fully informative experiment I^* and the high type obtains positive rent $U(\theta^H) > 0$.*
2. *If $\gamma > \theta_i^L/\theta_i^H$ for all i such that $\theta_i^L/\theta_i^H < 1$, the high type obtains no rent.*

Finally, Corollary 3 deals with the case of exclusion. It generalizes the intuition from the fully binary model in which noncongruent priors remove the need to exclude the low type.

Corollary 3 (Exclusion)

1. *If priors are noncongruent or if they are congruent but the two types do not lie on a ray in the simplex, it is never optimal to exclude type θ^L.*
2. *In both cases, the experiment $I(\theta^L)$ is partially informative if γ is sufficiently large.*

4.3 More than Two Types

With more than two types and $N > 2$, our relaxed approach is not always valid—in the optimal menu, different types would choose different actions in response to the same signal realizations. More formally, it is always without loss of generality – see Proposition 1 – to assume that type θ follows the recommendation of each signal in his own experiment $I(\theta)$. However, type θ need not follow the recommendation of every signal in other experiments $I(\theta')$. The full incentive-compatibility constraints can then be written as

$$\sum_{j=1}^{N} \theta_j \pi_{jj}(\theta) - t(\theta) \geq \sum_{j=1}^{N} \max_i \theta_i \pi_{ij}(\theta') - t(\theta'). \tag{22}$$

In this case, computing the optimal menu is more difficult. Nevertheless, the max operator in (22) can be captured by a set of additional linear inequality constraints. Thus, for a discrete type space and at the cost of additional complexity, the monopolist's problem can still be solved via (numerical) linear programming, as illustrated by the following example.

Example 3 (Signals and Actions) *Let types* $\theta^1 = (1/6, 1/6, 2/3)$*,* $\theta^2 = (1/2, 1/2, 0)$*, and* $\theta^3 = (1/2, 0, 1/2)$ *be uniformly distributed. In the* relaxed problem*, the monopolist sells the fully informative experiment to types* θ^2 *and* θ^3*. Type* θ^1 *buys*

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

.

However, when contemplating buying $I\left(\theta^1\right)$*, type* θ^2 *would choose action* a_1 *when observing signal* s_3*. In the solution to the* full problem*,* $I\left(\theta^1\right)$ *is given by*

$I(\theta^1)$	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

. (23)

It is easy to see that, in Example 3, no experiment $I(\theta^1)$ can lead both types θ^2 and θ^3 to follow the action recommended by every signal. Thus, the solutions to the relaxed and general problems must differ.

Relationship with the Monopoly Bundling Problem At a first glance, the problem of the data buyer resembles a consumer's demand for multiple goods or bundles of characteristics. We recall that the value of an information structure is given by:

$$V(I, \theta) = \sum_j \max_i \{\pi_{ij}\theta_i\} - \max_i \{\theta_i\}. \tag{24}$$

We can interpret the buyer's type θ as a vector of tastes, states ω_i as products, and the likelihood function π as the quantity of each product available to the consumer. With this interpretation, an experiment I generates a random allocation $s \in S$. Observing signal s_j is then analogous to being offered the set of options $(\pi_{1j}, \ldots, \pi_{Nj})$. Finally, the consumer chooses a single product from the set according to his tastes. Consequently, the consumer's utility is convex in the probability distribution over options, generating the upward kinks in (24). In addition, type-dependent reservation utilities generate the downward kinks. In contrast, in bundling problems, buyers' utilities are typically linear in the goods they acquire, and their outside options are nil.

Now, as the data buyer must ultimately choose an action, acquiring information introduces an additional choice variable (and optimality condition) represented by the first max operator in (24). Hence, the data buyer's problem does not reduce to selecting a bundle of goods from a menu. Indeed, a buyer of incremental information behaves like a consumer with unit demand who faces a stochastic choice problem. This property of buying information brings additional incentive constraints to the seller's problem differentiating it from a multiproduct monopolist's problem. Furthermore, the outside option for the buyer is given by his initial information. This is reflected in the second max operator in (24). Thus, the participation constraint is naturally type dependent.

Despite these important differences, our seller's problem bears some resemblance to a bundling problem. In particular, when buyers' types are multidimensional (i.e., $N > 2$ in our case), 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.[7] Partially informative experiments are the analog of stochastic bundles in our model. To further mark the difference to these classic multidimensional problems, stochastic bundling can arise in our setting even when buyer types are one-dimensional. This aspect emerged in the fully binary setting of Proposition 4, which we generalize in the next section.

5 Optimal Menu with a Continuum of Types

Finally, we return to the environment with binary actions and states and allow for a continuum of types. Letting $\theta \triangleq \Pr[\omega_1]$, the value of information is given in (11), i.e.,

$$V(I,\theta) \triangleq \max\{\pi_{11}\theta + \pi_{22}(1-\theta) - \max\{\theta, 1-\theta\}, 0\}.$$

Because types are one dimensional, we can define for each experiment $I(\theta)$ the following measure of *relative informativeness*,

$$q(\theta) \triangleq \pi_{11}(\theta) - \pi_{22}(\theta) \in [-1,1].$$

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

For each θ, Proposition 2 implies that either $\pi_{11}(\theta) = 1$ or $\pi_{22}(\theta) = 1$ (or both). Therefore, we adopt the convention that $q(\theta) > 0$ implies $\pi_{11}(\theta) = 1$ and $q(\theta) < 0$ implies $\pi_{22}(\theta) = 1$. We can then more succinctly rewrite the value of an experiment q as

$$V(q, \theta) = \max\{\theta q - \max\{q, 0\} + \min\{\theta, 1 - \theta\}, 0\}. \tag{25}$$

With this notation, the two information structures $q \in \{-1, 1\}$ correspond to releasing no information to the buyer. (These experiments show the same signal with probability one.) Conversely, the fully informative experiment is given by $q = 0$.

The value of information in (25) reflects the more general properties of our screening problem: (a) buyers have type-dependent participation constraints;[8] (b) buyer type $\theta = 1/2$ has the highest willingness to pay for any experiment q; (c) the experiment $q = 0$ is the most valuable for all types θ; (d) different types θ rank partially informative experiments differently; and (e) the utility function $V(q, \theta)$ has the single-crossing property in (θ, q).

The single-crossing property indicates that buyers with a higher θ, who are relatively more optimistic about state ω_1, assign a relatively higher value to experiments with a higher q—such experiments contain a signal that is more revealing of state ω_2, which they deem less likely. However, the "vertical quality" and "horizontal position" of an information product cannot be chosen separately by the seller. In particular, it is not possible to change the relative informativeness of a product (i.e., choose a very high or very low q) without reducing its overall informativeness.

Recall that Proposition 1 allows us to focus on *responsive* allocations. Therefore, we simplify the problem by eliminating the first max operator from (25). Next, we derive a characterization of all implementable responsive allocations q. Observe that the buyer's utility function has a downward kink in θ. This is a consequence of having an interior type ($\theta = 1/2$) assign the highest value to any allocation and of the linearity of the buyer's problem. We thus compute the buyer's rents $U(\theta)$ on $[0, 1/2]$ and $[1/2, 1]$ by applying the envelope theorem to each interval separately. We then obtain two possibly different expressions for the rent of type $\theta = 1/2$. However, continuity of the rent function then requires that

$$U(1/2) = U(0) + \int_0^{1/2} V_\theta(q, \theta) \mathrm{d}\theta = U(1) - \int_{1/2}^{1} V_\theta(q, \theta) \mathrm{d}\theta.$$

Because types $\theta = 0$ and $\theta = 1$ assign zero value to any experiment and transfers are nonnegative, incentive compatibility requires $U(0) = U(1) = 0$. Therefore, while any type's

[8] In general, our setting involves both bunching and exclusion, making it difficult to directly apply existing approaches, such as the one in Jullien (2000).

utility can always be written in the above form, the novel element of our model is that no further endogenous variables appear.[9] Differentiating (25) with respect to θ and simplifying, we obtain the following restriction on an incentive-compatible allocation

$$\int_0^{1/2} (q(\theta) + 1)\mathrm{d}\theta = -\int_{1/2}^{1} (q(\theta) - 1)\mathrm{d}\theta. \tag{26}$$

The constraint (26) is a new condition that sets our framework apart from most other screening problems. In particular, incentive constraints impose an aggregate (integral) constraint on the allocation.[10] We formalize this in the following result.

Lemma 1 (Implementable Allocations)
Any responsive allocation q is implementable if and only if the following two conditions hold:

$$\begin{aligned} &q(\theta) \in [-1, 1] \ \textit{ is non-decreasing;} \\ &\textit{and } \int_0^1 q(\theta)\, d\theta = 0. \end{aligned} \tag{27}$$

The transfers $t(\theta)$ associated with the allocation $q(\theta)$ can be computed in the usual way on the two intervals $[0, 1/2]$ and $[1/2, 1]$ separately. With the addition of the integral constraint (27) for implementability, we can state the seller's problem as follows:

$$\max_{q(\cdot)} \int_0^1 \left[\left(\theta - \frac{F(1/2) - F(\theta)}{f(\theta)} \right) q(\theta) - \max\{q(\theta), 0\} \right] \mathrm{d}F(\theta), \tag{28}$$

$$\text{s.t. } q(\theta) \in [-1, 1] \text{ non-decreasing,}$$

$$\int_0^1 q(\theta)\, \mathrm{d}\theta = 0.$$

5.1 Optimal Menu

In order to solve the seller's problem (28) and characterize the optimal menu, we rewrite the objective with the density $f(\theta)$ explicitly in each term:

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

[9] For instance, in Mussa and Rosen (1978) and in Myerson (1981), the rent of the highest type $U(1)$ depends on the entire allocation q.

[10] The integral constraint thus differs from other instances of screening under integral constraints (e.g., constraints on transfers due to budget or enforceability, or capacity constraints). In the model of persuasion with private information of Kolotilin, Li, Mylovanov, and Zapechelnyuk (2015) a similar integral constraint appears as a persuasion budget constraint. However, because theirs is a model without transfers, the constraint operates differently and has very different implications.

This minor modification highlights two important features of our problem: (i) the constraint and the objective have generically different weights, $\mathrm{d}\theta$ and $\mathrm{d}F(\theta)$; and, hence, (ii) the problem is non-separable in the type and the allocation, which interact in two different terms. In particular, consider the "virtual values" $\phi(\theta, q)$, defined as the partial derivative of the integrand with respect to q. Unlike in standard separable models, the virtual values are a function of the allocation. Because our objective is piecewise linear, the function $\phi(\theta, q)$ takes on two values only, i.e.,

$$\phi(\theta, q) := \begin{cases} \theta f(\theta) + F(\theta) & \text{for} \quad q < 0, \\ (\theta - 1) f(\theta) + F(\theta) & \text{for} \quad q > 0. \end{cases}$$

Heuristically, the two virtual values represent the marginal benefit to the seller (ignoring the constraint) of increasing each type's allocation from -1 to 0, and from 0 to 1, respectively.

If ironing à la Myerson is required, we denote the *ironed* virtual value for experiment q as $\bar{\phi}(\theta, q)$. Finally, we say that the allocation satisfies the *pooling property* if it is constant on any interval where the relevant (ironed) virtual value is constant. The solution to the seller's problem is then obtained by combining standard Lagrange methods with the ironing procedure developed by Toikka (2011), which extends the approach of Myerson (1981).

Proposition 6 (Optimal Allocation Rule)
The allocation $q^*(\theta)$ is optimal if and only if:

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

$$q^*(\theta) = \arg\max_q \left[\int_{-1}^{q} \left(\bar{\phi}(\theta, x) - \lambda^* \right) dx \right];$$

2. *q^* has the pooling property and satisfies the integral constraint (27).*

To gain some intuition, observe that the problem is piecewise-linear (but concave) in the allocation. Thus, absent the integral constraint, the seller would choose an allocation that takes values at the kinks, i.e. $q^*(\theta) \in \{-1, 0, 1\}$ for all θ. In other words, the seller would offer a menu containing the fully informative experiment only. Flat pricing turns out to be optimal for the seller in a number of circumstances. The main novel result of this section is that the seller can sometimes do better by offering one additional experiment.

Proposition 7 (Optimal Menu)
An optimal menu consists of at most two experiments.

1. *The first experiment is fully informative.*
2. *The second experiment contains a signal that perfectly reveals one state.*

To obtain some intuition for this result, consider a relaxed problem where the seller contracts separately with buyers $\theta < 1/2$ and $\theta > 1/2$. Because of the linearity of the problem, the optimal menu for each group is degenerate: the seller offers the perfectly revealing experiment at a flat price. Now, consider the solution to the full problem. If the optimal prices for each separate group are similar, the seller offers the fully informative experiment at an intermediate price. If they are quite different, the seller prefers to distort the information sold to one group to maintain a high price for the fully informative experiment. However, the linearity of the environment prevents the seller from offering more than one distorted experiment, i.e., no further *versioning* is optimal.

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

5.2 Flat vs. Discriminatory Pricing

Let types be uniformly distributed, $F(\theta) = \theta$, and consider the virtual values $\phi(\theta, q)$ for $q < 0$ and $q \geq 0$ separately. These values are constant in q; hence, we refer to $\phi(\theta, -1)$ and $\phi(\theta, 1)$, respectively. For a given value of the multiplier λ, the allocation that maximizes the expected virtual surplus in Proposition 6 assigns $q^*(\theta) = -1$ to all types θ for which $\phi(\theta, -1) < \lambda$; it assigns $q^*(\theta) = 0$ to all types θ for which $\phi(\theta, -1) > \lambda > \phi(\theta, 1)$; and $q^*(\theta) = 1$ for all types θ for which $\phi(\theta, 1) > \lambda$.

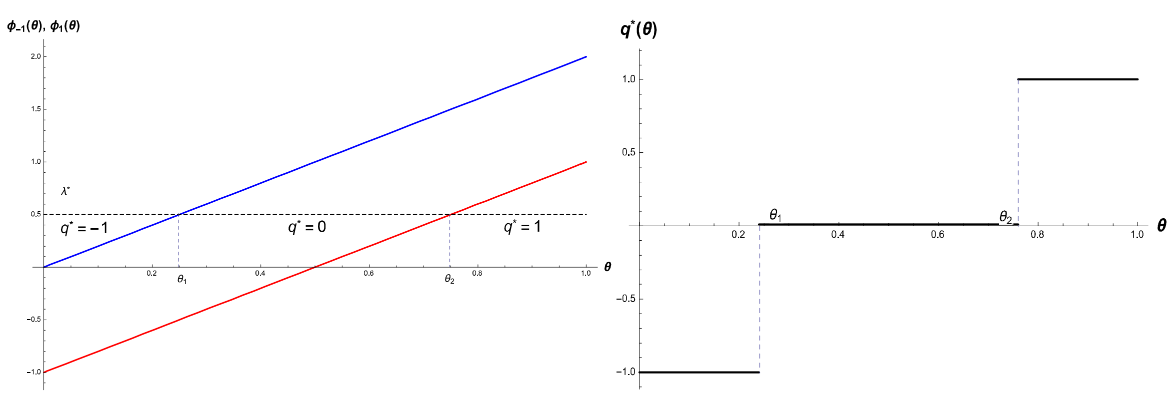

Figure 7: Uniform Distribution: Virtual Values and Optimal Allocation

Figure 7 illustrates the resulting allocation rule. In order to satisfy the integral constraint, the optimal value of the multiplier λ^* must identify two symmetric threshold types (θ_1, θ_2)

that separate types receiving the efficient allocation $q = 0$ from those receiving no information at all, $q = -1$ or $q = 1$. The allocation then clearly satisfies the integral constraint (27). More generally, if both virtual values $\phi(\theta, -1)$ and $\phi(\theta, 1)$ are strictly increasing in θ, the optimal menu consists of charging the monopoly price for the fully informative experiment.

Flat pricing is optimal under weaker conditions than strictly increasing virtual values. We now summarize the sufficient conditions for this result.

Proposition 8 (Flat Pricing) *Suppose any of the following conditions hold:*

1. $F(\theta) + \theta f(\theta)$ *and* $F(\theta) + (\theta - 1) f(\theta)$ *are strictly increasing;*
2. *the density* $f(\theta) = 0$ *for all* $\theta > 1/2$ *or* $\theta < 1/2$*;*
3. *or the density* $f(\theta)$ *is symmetric around* $\theta = 1/2$*.*

The optimal menu contains only the fully informative experiment.

An implication of Proposition 8 is that the seller offers a second experiment *only if* ironing is required. At the same time, there exist examples with non-monotone virtual values and one-item menus. Distributions that are symmetric around $1/2$ are one such instance: for any distribution $F(\theta)$, the solution to the restricted problem on $[0, 1/2]$ or $[1/2, 1]$ is a cutoff policy. Because the cutoffs under symmetric distribution are themselves symmetric, the solutions to the two subproblems satisfy the integral constraint and, hence, provide a tight upper bound to the seller's profits.

The monotonicity conditions of Proposition 8 that guarantee increasing virtual values are quite demanding. When types correspond to interim beliefs, it is natural to consider bimodal densities (e.g., a well-informed population in a binary model) that fail the regularity conditions and, therefore, introduce the need for ironing. For example, starting from the common prior, if buyers observe binary signals, a bimodal distribution of beliefs would result with types holding beliefs above and below the mean of the common prior μ. Thus, non-monotone densities and distributions violating the standard monotonicity conditions are a natural benchmark. Therefore, ironing is not a technical curiosity in our case, but rather a technique that becomes unavoidable because of the features of the information environment.

We now illustrate the ironing procedure when virtual values are not monotone and how it leads to a richer (two-item) optimal menu. Figure 8 considers a bimodal distribution of types, illustrating the probability density function and the associated virtual values.[11]

[11] The distribution in the left panel represents a "perturbation" of the two-type example in Section 4.1. In particular, it is a mixture (with equal weights) of two Beta distributions with parameters $(8, 30)$ and $(60, 30)$.

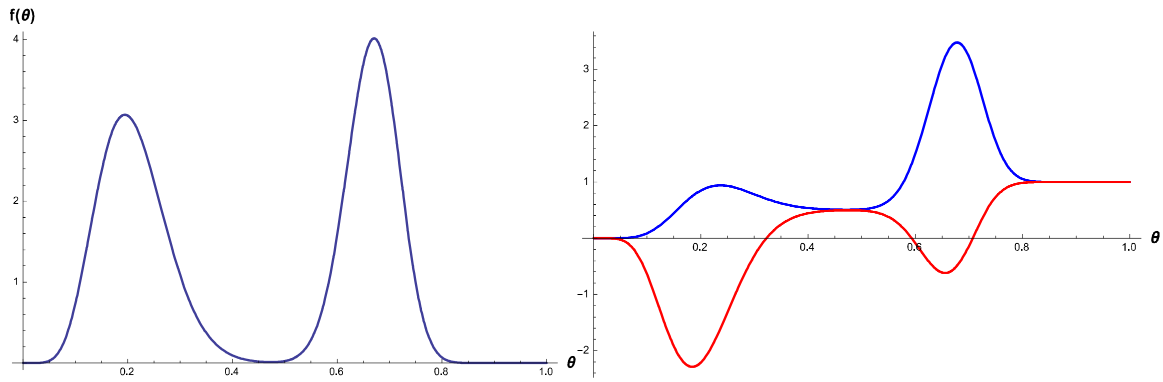

Figure 8: Probability Density Function and Virtual Values

We therefore consider the "ironed" versions of each virtual value and identify the equilibrium value of the multiplier λ^*. In this example, the value of the multiplier must equal the constant level of one of the virtual values.

Figure 9 illustrates the optimal two-item menu. For types in the "pooling" region $\theta \in [0.17, 0.55]$, the seller is indifferent among all values of $q \in [-1, 0]$. The optimal allocation ($q^* \approx -0.22$) is pinned down by (the pooling property and) the integral constraint.

In both examples, extreme types with a low value of information are excluded from the purchase of informative signals. In the latter example, the monopolist offers a second information structure that is tailored towards relatively lower types. This structure (with $q < 0$) contains one signal that perfectly reveals the high state. 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 $\theta \approx 0.7$.

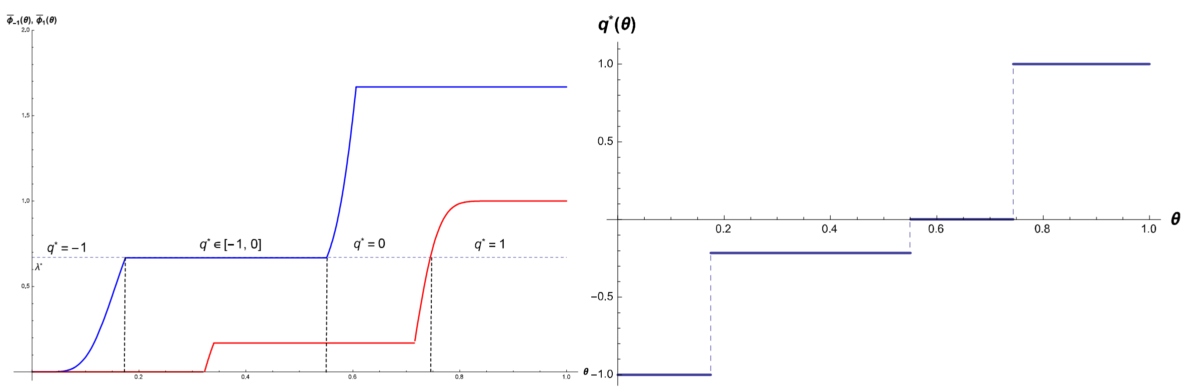

Figure 9: Ironed Virtual Values and Optimal Allocation

The properties of the optimal discriminatory pricing scheme reflect the fact that the type structure is quite different from the standard screening environment. While information rents $U(\theta)$ peak at $\theta = 1/2$, the ex ante least informed type $\theta = 1/2$ need not purchase the fully informative experiment, despite having the highest value of information.[12] In the above example, inducing the types around $\theta = 1/2$ to purchase the fully informative experiment would require further distortions (hence, a lower price) for the second experiment. This leads to a loss of revenue from types around $\theta \approx 0.2$. Because such types are quite frequent, this loss more than offsets the gain in revenue from types around 1/2.

6 Implications for Data Pricing

Menu pricing of information products is widely used by many data sellers, including genetic testing companies, credit score agencies, and big data brokers. The former two offer menus of products consisting of tests or inquiries that reveal, with varying degrees of confidence, a patient's predisposition to develop a disease or a borrower's risk of default. The latter offer both standardized and customized data services that facilitate targeting advertisements to specific consumers.[13]

In this Section, we relate the solution to our screening problem to the practical design of information products. In a nutshell, our seller's problem consists of degrading the quality of the information sold to some buyers in order to charge higher prices to "premium" buyers.[14] Our results contribute to identifying systematic patterns in the optimal design of "distorted" partially informative experiments.

Remark 1 (No White Noise) *No product in an optimal menu adds unbiased noise to the seller's information. Instead, each product contains "directionally informative" signals.*

Proposition 2 shows that every experiment is non-dispersed and, therefore, the buyer always takes the correct action in at least one state. In reality, sellers hardly have perfect information about the state. Nonetheless, in the special case of binary states and actions, every optimal experiment minimizes the type-II error for any level of type-I error (as in Figure 2). This result has several practical implications.

Genetic testing companies such as Ambry Genetics exploit advances in sequencing technology to identify gene mutations associated with various diseases. The tests on its menu

[12] This feature of the optimal menu already emerges in three type examples. Type $\theta = 1/2$ nevertheless obtains the highest information rent.

[13] Brokers such as Acxiom and Oracle offer Data Management Platforms—customized software that automates the integration of 1st- and 3rd-party data, enabling websites to track their users and place them into more precise market segments—as well as less flexible data packages (see, for example, http://www.acxiom.com/data-packages/). Similarly, AddThis offers a choice between using a customized consumer audience and a standard audience (http://www.addthis.com/custom-audiences).

[14] Our setting differs from the "damaged goods" model of Deneckere and McAfee (1996) because degrading quality is costless.

vary mainly in the number of genes tested for each medical condition. Importantly, the low-end products always test a strict subset of the genes tested by the higher-end products.[15] Through the lens of our model, all such tests eliminate the type-I error, and buyers are screened on the basis of their relative distaste for different types of statistical errors.

Similarly, credit reporting agencies such as Equifax 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, Equifax offers its business customers products such as "Prescreen" and "Undisclosed Debt Monitoring." The former identifies consumers who are likely to open and maintain an account. Among other criteria, it uses a flexible threshold on the credit score to construct the *target list*. Varying the threshold then corresponds to varying the information provided in an optimal way, balancing the trade-off between type-I and type-II errors. The latter identifies default risk based on new information arriving between the time of the original credit inquiry and the closing of the loan. This product's versions differ in the number of events (e.g., late payments or inquiries made to competing agencies Experian and Transunion) that are monitored for potentially "bad news."

This approach to selling consumer-level information is in sharp contrast to earlier methods. For instance, mailing lists are typically sold at the level of mail-carrier routes.[16] From the information buyer's point of view, using approximate information (e.g., one's home address as a proxy for income) inevitably allows for both types of statistical errors.

Remark 2 (Targeted and Residual Categories) *With two buyer types, all states (except one at most) are partitioned into targeted states that send non-dispersed information and residual states that send dispersed signals. Only actions corresponding to targeted states are chosen with positive probability.*

Proposition 5 characterizes the optimal menu for the binary type setting. With two buyer types, the partially informative experiment is represented by the triangular matrix (21). This means that the seller chooses a subset of "targeted" realized states in which the buyer always takes the correct action. Conversely, when a "residual" state is realized, the buyer never takes the correct action. In other words, some states are used to "degrade" the information revealed about the targeted states. This property of the optimal menu resonates with the design of the information products offered by online data brokers.

In the online advertising market, data brokers record several attributes for each individual consumer. They then offer "data packages" to advertisers. Such packages can be viewed as

[15] For example, breast and ovarian cancer tests examine between 2 and 17 genes. Their prices range from \$400 to \$3900 (http://www.ambrygen.com/sites/default/files/Master_Pricelist_testCode_2.pdf).

[16] See http://mailways.net or the paper by Anderson and Simester (2013) on direct marketing.

experiments where each signal "flags" a consumer segment. If we view a state realization as a specific consumer profile (i.e., a combination of attributes), every *optimal* signal will flag a segment consisting of *all* consumers fitting a single profile and will also include additional, heterogeneous consumers.

Data brokers choose which segments to include in each package on the basis of the relative demand for each individual consumer profile. A very coarse data package would, for example, partition the consumer population into segments corresponding to an "Advertise" and a "Do Not Advertise" recommendation. Depending on which consumer profiles advertisers wish to include and exclude—both kinds of targeting strategies are empirically relevant—data packages may lead to excessively *broad* or excessively *narrow* campaigns.[17]

Remark 3 (Simple Menus for Simple Problems) *Regardless of the heterogeneity in the consumer population, the optimal menu for binary choice problems contains two items—a "full" version and a "light" version of the seller's information.*

Proposition 7 establishes the optimality of two-item menus for binary state and action problems. While there are other reasons to offer relatively coarse menus, our results shed light on how the complexity of the buyer's decision problem is related to the complexity of the optimal menu—not just to the cardinality of the signal space by Proposition 1.

For example, in the case of genetic testing, the main choice of action consists of whether to order additional exams or increase the frequency with which a patient is monitored. In line with our results, Ambry Genetics offers a rich product line targeting several diseases. However, within each condition (e.g., a certain type of cancer or cardiac disease), very few options are typically available—one that covers all the genes the company is able to test and another that induces type-II errors.[18]

Two-item menus are also used by LinkedIn to sell member profiles to potential employers who, essentially, must decide whom to interview: a Lite version that allows one to condition searches on a limited number of criteria and a Premium version that grants access to the entire database. Conversely, in the case of online advertising, the action space is much larger (e.g., how much to invest and which product to advertise). Customization of the data packages then leads to a menu with a far greater number of options.

[17] In our earlier work (Bergemann and Bonatti, 2015), we examined a linear pricing problem for individual consumer profiles, i.e., "cookies."

[18] Passing the genetic test does not guarantee that the patient will not develop the disease. In that sense, the seller has only partial information (whose accuracy depends on the disease). For our purposes, we can, however, identify the fully informative experiment with the revelation of all the seller's information.

7 Conclusions

We have studied the problem of a monopolist who sells incremental information to privately informed buyers. Fundamental to the seller's incentives to degrade information is the Bayesian nature of the buyer's problem. In particular, selling information introduces a novel aspect of horizontal differentiation across buyer types that widens the scope for price discrimination through directional information.

We have deliberately focused on the "packaging" or versioning problem of a seller who is (a) free to acquire and degrade information, and (b) uninterested in the buyer's actions. Thus, our results provide only a first pass at understanding the trade-offs involved in selling information products. A richer model would distinguish the cost of acquiring information (i.e., building the database) from the cost of duplicating and distributing the information. Furthermore, in many applications, it is costly to introduce noise in the data . This occurs, for instance, when the seller is concerned with preserving the anonymity of her information.

Selling pre-packaged data is just one mechanism for trading information. For instance, large providers of online advertising space, such as Google and Microsoft, do not just sell information—they also internalize advertisers' spending decisions. Thus, they effectively bundle data and advertising products by offering a menu of targeting opportunities. In this sense, our approach is better suited to analyze a data broker's pricing problem when merchants wish to buy targeted "advertising products" from a different seller. At the same time, this application suggests an interesting model with a richer contract space.

Finally, we have relied solely on belief heterogeneity to motivate the sale of incremental information. However, buyers of information may differ along several alternative or additional dimensions. For example, buyers of time-sensitive data can be screened according to how much they value timely information.[19] Each of these extensions can be implemented in the framework we have outlined. Combining different sources of heterogeneity (e.g., beliefs and discount rates) appears more challenging but promises additional insights.

[19] Essentially homogeneous information products that differ in availability timing include: the Consumer Sentiment Index released by Thomson-Reuters and the University of Michigan, as well as the Exome tests for prenatal genetic diagnoses offered by Ambry Genetics.

A Appendix

Proof of Proposition 1. Consider any type θ and experiment I. Let S_I^a denote the sets of the signals in experiment I that induce type θ to choose action a. Thus, $\cup_{a\in A} S_I^a = S_I$. Construct the experiment I' as a recommendation for type θ based on the experiment I, $S_{I'} = \{s_a\}_{a\in A}$ and

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

By construction, I' induces the same outcome distribution for type θ as I so $V(I', \theta) = V(I, \theta)$. At the same time, I' is a garbling of I so by Blackwell's theorem $V(I', \theta') \leq V(I, \theta')\ \forall \theta'$. Thus, for any incentive compatible and individually rational direct mechanism $\{I(\theta), t(\theta)\}$ we can construct another direct mechanism $\{I'(\theta), t(\theta)\}$ with its experiments leading type θ to take a different action after observing each signal $s \in S(\theta)$ that is also incentive compatible and individually rational, and yields weakly larger profits. ■

Proof of Proposition 2. The argument for part (1.) is given in the text. For part (2.), consider any individually rational and incentive compatible direct mechanism $\mathcal{M} = \{I(\theta), t(\theta)\}$. Fix a type θ with the associated experiment $I(\theta)$ and let $\pi_{ij}(\theta)$ denote the conditional probability of signal s_j in state ω_i under experiment $I(\theta)$. By Proposition 1, each type θ has a different optimal action for each signal in $I(\theta)$. Without loss of generality, we then arrange the signals in $I(\theta)$ so that type θ takes action a_i when observing s_i. If type θ never takes action a_i, we drop signal s_i from $I(\theta)$, i.e. we set the i-th column of $\pi(\theta)$ to zero. Because beliefs $\Sigma_i \theta_i = 1$, we can write the value of information (4) as

$$\begin{aligned} V(I(\theta), \theta) &= \left[\sum_{i=1}^{N} \theta_i \pi_{ii}(\theta) - \max_i \theta_i\right]^+ \\ &= \left[\sum_{i=1}^{N-1} \theta_i (\pi_{ii}(\theta) - \pi_{NN}(\theta)) + \pi_{NN}(\theta) - \max_i \theta_i\right]^+. \end{aligned} \tag{29}$$

Now define $\varepsilon(\theta) := 1 - \max \pi_{ii}(\theta)$, and construct a new experiment $I'(\theta)$ where $\pi'_{ii}(\theta) = \pi_{ii}(\theta) + \varepsilon(\theta)$ for all i and for all θ. For each state ω_i, the off-diagonal entries $\pi_{ij}(\theta)$ with $j \neq i$, are correspondingly reduced by $\varepsilon(\theta)$, without further restrictions on how this is operation is performed. It then follows from (29) that

$$[V(I'(\theta), \theta) - \varepsilon(\theta)]^+ = V(I(\theta), \theta).$$

Furthermore, for all types $\theta' \neq \theta$, the value of mimicking type θ increases by less than $\varepsilon(\theta)$

(strictly so, if type $\theta' \neq \theta$ responds to the signals in $I(\theta)$ differently from type θ). Suppose type θ' chooses action $a_{i(j)}$ upon observing signal s_j from experiment $I(\theta)$. We then have

$$V(I(\theta), \theta') = \left[\sum_{j=1}^{N} \theta'_{i(j)} \pi_{i(j)j}(\theta) - \max_i \theta'_i\right]^+ .$$

Furthermore, because $i(j)$ need not coincide with j, the entries $\pi_{i(j)j}$ in $V(I(\theta), \theta')$ increase at most by $\varepsilon(\theta)$. Therefore, we have

$$[V(I'(\theta), \theta') - \varepsilon(\theta)]^+ \leq V(I(\theta), \theta').$$

Consequently, the direct mechanism $\mathcal{M}' = \{I'(\theta), t(\theta) + \varepsilon(\theta)\}$ is also individually rational and incentive compatible. Moreover, all experiments $I(\theta)$ are non-dispersed by construction, and all transfers are weakly greater than in the original mechanism $\mathcal{M}$. ■

Proof of Proposition 3. (i) We know from Proposition 2 that at least one type must buy the fully informative experiment I^*. Suppose only type θ^L buys I^* as part of the optimal menu. Then the price of I^* is at most $V(I^*, \theta^L)$. By incentive compatibility, if the high type θ^H purchases $I \neq I^*$, it must be that $t(\theta^H) < V(I^*, \theta^L)$. Therefore, eliminating the experiment $I(\theta^H)$ from the menu strictly improves the seller's profits. Contradiction. (iii) The participation constraint of θ^L must bind. Indeed, some participation constraint must bind, otherwise increase both prices. To the contrary, let the constraint of θ^H bind and not of θ^L. Since θ^H is served by I^* then $t(\theta^H) = V(I^*, \theta^H) \geq V(I^*, \theta^L)$. Hence can increase $t(\theta^L)$ without violating incentive compatibility. (ii) The incentive constraint of θ^H must bind. To the contrary, assume that it does not bind. Then there are two cases. If θ^H participation constraint doesn't bind then can increase $t(\theta^H)$. If θ^H participation constraint binds then it must be that $I(\theta^L)$ is not equal to I^*. Since payoffs are continuous in π_{ij} we can increase both informativeness of $I(\theta^L)$ and $t(\theta^L)$. ■

Proof of Proposition 4. (1.) Consider the case of congruent priors. It follows from Proposition 5 that an optimal menu contains the fully informative experiment only. The value for this experiment are $1 - \theta^H$ and $1 - \theta^L$ for the high and the low type, respectively. The profits from non-exclusive and exclusive pricing are $\gamma(1 - \theta^H)$ and $1 - \theta^L$. Thus, it is optimal for both types to participate if and only if $\gamma \leq (1 - \theta^L)/(1 - \theta^H)$.

(2.) Consider the case of noncongruent priors. Let $q \triangleq \pi_{11} - \pi_{22}$ and denote $q^L = q(\theta^L), q^H = q(\theta^H)$. It follows again from Proposition 5 that in, an optimal menu, we have $q^L \leq 0$ and $q^H = 0$. If $\gamma < \theta^L/\theta^H$, we wish to show that flat pricing is optimal, i.e., $q^L = q^H = 0$ and $t_1 = t_2 = 1 - \theta^L$. Now, for an arbitrary incentive compatible menu

$\left(q^H, q^L, t^H, t^L\right)$ define the following modification:

$$\left(q^{H\prime}, t^{H\prime}, q^{L\prime}, t^{L\prime}\right) = \left(q^H, t^H - \varepsilon\left(\theta^H - \theta^L\right), q^L + \varepsilon, t^L + \varepsilon\left(1 - \theta^L\right)\right). \tag{30}$$

If $q^L < 0$, the modification (30) with $\varepsilon \in (0, -q_L)$ preserves incentive compatibility and improves profits by $\varepsilon\theta^H(\theta^L) > 0$, which yields a contradiction.

Finally, we wish to show that if $\gamma > \theta^L/\theta^H$, then discriminatory pricing is optimal. Notice first that the individual rationality of the high type must bind, $q^H = 0, t^H = 1-\theta^H$, otherwise modification (30) would be profitable for some $\varepsilon < 0$. Second, the 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

$$q^L = \frac{\theta_H + \theta_L - 1}{\theta_H - \theta_L}, t^L = \frac{\left(2\theta^H - 1\right)\theta^L}{\theta^H - \theta^L}.$$

Substituting the definition of q yields the expression (14) in the statement. ■

Proof of Proposition 5. We know from Proposition 3 that the high type θ^H purchases the fully informative experiment. We now derive the optimal experiment $I(\theta^L)$. Suppose (as we later verify) that both types θ^H and θ^L would choose action a_i after observing signal s_i from the optimal experiment $I(\theta^L)$. The seller then chooses $\pi_{ii} \in [0,1]$, $i = 1, \ldots n$ to solve the linear program (15) subject to (16). We can write the Lagrangian as

$$L = (1-\gamma)\left[\sum_{i=1}^{n} \pi_{ii}\left(\theta_i^L + \lambda\left(\theta_i^L - \theta_i^H\right)\right) - \lambda \max_i \theta_i^L + \lambda \max_i \theta_i^H + \lambda U(\theta^H)\right] - \gamma U(\theta^H),$$

where $\lambda \geq 0$ represents the shadow value of increasing type the informational rent $U(\theta^H)$. It follows immediately that

$$\pi_{ii} = \begin{cases} 0 & \text{if} \quad \theta_i^L + \lambda\left(\theta_i^L - \theta_i^H\right) < 0, \\ [0,1] & \text{if} \quad \theta_i^L + \lambda\left(\theta_i^L - \theta_i^H\right) = 0, \\ 1 & \text{if} \quad \theta_i^L + \lambda\left(\theta_i^L - \theta_i^H\right) > 0. \end{cases} \tag{31}$$

and that

$$U(\theta^H) = \begin{cases} \sum_{i=1}^{n}(\theta^H)\pi_{ii} + \max_i \theta_i^L - \max_i \theta_i^H & \text{if} \quad \gamma < \lambda\left(1-\gamma\right), \\ 0 & \text{if} \quad \gamma \geq \lambda\left(1-\gamma\right). \end{cases} \tag{32}$$

To characterize the solution, we need only identify the value of the multiplier λ^*. We proceed in the following steps.

1. Recall states ω_i are ordered by increasing likelihood ratio of beliefs θ_i^L/θ_i^H, and consider state ω_1. If ω_1 is the only state i for which $\pi_{ii} < 1$, (31) implies that the multiplier is given by

$$\lambda^* = \min_i \frac{1}{\frac{\theta_i^H}{\theta_i^L} + 1} = \frac{\theta_1^L}{\theta_1^L + \theta_1^H}. \tag{33}$$

Therefore, π_{11} solves (16) with equality, i.e., the candidate optimal value is $\pi_{11} = k_1$ as defined in (19). If $k_1 \geq 0$ and $\theta_1^L/\theta_1^H < \gamma$, then (32) implies $U(\theta^H) = 0$ and (31) implies we must set $\pi_{ii} = 1$ for all $i \geq 2$. If instead $\gamma < \theta_1^L/\theta_1^H$ then then $\pi_{11} = 1$ and $U(\theta^H) > 0$ solves (16), i.e., $U(\theta^H) = \max_i \theta_i^L - \max_i \theta_i^H$. Finally, if $k_1 < 0$ and $\theta_1^L/\theta_1^H < \gamma$, there is no solution with the value of λ^* given in (33).

2. Suppose the above candidate solution is not, in fact, the optimum. The value of the multiplier must then satisfy $\lambda^* > \theta_1^L/(\theta_1^L + \theta_1^H)$, which implies the non-negativity constraint binds, i.e., $\pi_{11} = 0$. This means the optimal partially informative experiment has less than full rank. The next candidate value for the multiplier λ^* is $\theta_2^L/(\theta_2^L + \theta_2^H)$. Again, if $k_2 \geq 0$ and $\theta_2^L/\theta_2^H < \gamma$, then (32) implies $U(\theta^H) = 0$ and (31) implies $\pi_{ii} = 1$ for all $i \geq 3$. If instead $\gamma < \theta_1^L/\theta_1^H$ then then $\pi_{11} = 1$ and $U(\theta^H) > 0$ solves (16) with equality. Finally, if $k_2 < 0$ and $\theta_2^L/\theta_2^H < \gamma$, there is no solution with the value of $\lambda^* = \theta_2^L/(\theta_2^L + \theta_2^H)$ and the next candidate is $\lambda^* = \theta_3^L/(\theta_3^L + \theta_3^H)$, which implies $\pi_{11} = \pi_{22} = 0$.

3. The procedure iterates until either (a) state i^* is reached, i.e., $\theta_i^L/\theta_i^H > \gamma$, or (b) state j^* is reached, i.e., $k_j \geq 0$.

4. We must then verify that the off-diagonal entries π_{ij}, $i \neq j$ can be set to ensure both types θ^H and θ^L choose action a_i when observing signal s_i. This requires

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

for both θ^L and θ^H and for all $j < i$, because the signal matrix can be taken to be lower triangular. Fix a signal i and an alternative action a_j. We need

$$\frac{\theta_i}{\theta_j} \geq \frac{\pi_{ji}}{\pi_{ii}}.$$

By construction we have $\theta_i^H/\theta_j^H < \theta_i^L/\theta_j^L$ for all $i > j$, hence, we need only worry about the incentives of type θ^H. We can then assign the probabilities π_{ji} following the procedure described in Algorithm 1: for any $i' > i$, we make type θ^H indifferent between following the recommendation of signal i' and choosing actions a_i; we do so beginning with π_{jn} and proceeding backward as long as required; if the procedure assigns positive weight to π_{ji} then it must be that $\pi_{ji} = 1 - \Sigma_{s=i+1}^{n} \theta_s^H / \theta_j^H$. It follows that type θ^H has strict incentives to follow the recommendation of signal i. Note that $\pi_{ii} \in \{k_i, 1\}$, hence, if $\pi_{ii} = k_i$, we have by definition

$$k_i \theta_i^H = \frac{\sum_{s=i+1}^{n} \left(\theta_s^L - \theta_s^H\right) - \max_j \theta_j^L + \max_j \theta_j^H}{\theta_i^H - \theta_i^L} \theta_i^H.$$

Now, because $\arg\max_j \theta_j^L > i$ we can bound the above expression by

$$k_i \theta_i^H > \frac{\theta_i^H}{\theta_i^H - \theta_i^L} \left(\theta_j^H - \Sigma_{s=i+1}^{n} \theta_s^H\right) > \theta_j^H - \Sigma_{s=i+1}^{n} \theta_s^H = \theta_j^H \pi_{ji}.$$

A fortiori, type θ^H has strict incentives to choose a_i if $\pi_{ii} = 1$.

This ends the proof. ■

Algorithm 1 (Optimal Menu with Binary Type) We construct the two optimal experiments as a function of the distribution of types. This algorithm contains three subroutines, beginning with the construction of the allocation described in part (1.) of Proposition 5. We refer to this allocation as the maximally informative zero rent experiment.

Maximally Informative Zero Rent Experiment. Order states i in increasing order of the likelihood ratios θ_i^L / θ_i^H and set $U(\theta^H) = 0$. Let $\pi_{ii} = 1$ for $i = 2, \ldots, n$ and solve (16) with equality with respect to π_{11}. The solution is given by k_1 in (19). If $k_1 \geq 0$, stop. If $k_1 < 0$, set $\pi_{11} = 0$ and $\pi_{ii} = 1$ for $i = 3, \ldots, n$. Solve (16) with equality with respect to π_{22}, which yields k_2. If $k_2 \geq 0$, stop, otherwise iterate the procedure. The procedure terminates at state j^* defined in (20).

We now use the distribution of types to identify which step of the construction yields the profit-maximizing experiment.

Optimal Experiment. Begin at state j^*. If $\gamma > \theta_{j^*}^L / \theta_{j^*}^H$ the maximally informative zero rent experiment is part of the optimal menu. If $\gamma < \theta_{j^*}^L / \theta_{j^*}^H$, set $\pi_{j^* j^*} = 1$ and consider $j^* - 1$. If $\gamma > \theta_{j^*-1}^L / \theta_{j^*-1}^H$, stop, and choose $U(\theta^H) > 0$ to satisfy (16) with equality. Otherwise set $\pi_{j^*-1, j^*-1} = 1$ and consider $j^* - 2$, iterating until reaching state i^* defined in (18) and adjusting the rent $U(\theta^H)$ to satisfy the incentive constraint.

Thus, as the fraction of high types γ increases, the optimal menu involves potentially steeper distortions. Conversely, when $i^* = 1$, the menu involves bunching both types at the fully informative experiment. Finally, we illustrate a simple procedure to assign the off-diagonal probabilities to the partially informative experiment.

Off-Diagonal Entries. Suppose the optimal menu sets $\pi_{ii} = 0$ for all $i < \hat{\imath}$, for some $\hat{\imath} > 1$. To assign the off diagonal entries $\pi_{i,j}$ with $i < \hat{\imath}$ and $j \geq \hat{\imath}$, fix a state $i < \hat{\imath}$ and begin with signal n. Let $\pi_{in} = \min\{\theta_n^H/\theta_i^H, 1\}$. If $\theta_n^H/\theta_i^H > 1$, stop. If $\theta_n^H/\theta_i^H < 1$, set $\pi_{i,n-1} = \min\{\theta_{n-1}^H/\theta_i^H, 1-\theta_n^H/\theta_i^H\}$, and proceed backwards until reaching $\pi_{i,i+1}$. The entries so constructed satisfy the recommendation constraint because, by the definition of k_j in (19), we have $\theta_{\hat{\imath}}^H > k_{\hat{\imath}}\theta_{\hat{\imath}}^H > \theta_i^H \left(1 - \Sigma_{j=\hat{\imath}+1}^n \theta_j^H/\theta_i^H\right)$.

Proof of Lemma 1. We begin with necessity. Consider an incentive compatible allocation q. For any two types $\theta_2 > \theta_1$ we have

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

It follows from the single-crossing property of $V(q, \theta)$ that $q_2 \geq q_1$ hence $q(\theta)$ is increasing. Because the buyer's rent is differentiable with respect to θ on $[0, 1/2]$ and $[1/2, 1]$ respectively, we can compute the function $U(\theta)$ on these two intervals separately. We obtain the expression in the text,

$$U(1/2) = U(0) + \int_0^{1/2} V_\theta(q, \theta)\, \mathrm{d}\theta = U(1) - \int_{1/2}^1 V_\theta(q, \theta)\, \mathrm{d}\theta.$$

By the envelope theorem $V_\theta(q, \theta) = q + 1$ for $\theta < 1/2$ and $= q - 1$ for $\theta > 1/2$. Finally, because $U(0) = U(1)$ we obtain

$$\int_0^1 q(\theta)\, \mathrm{d}\theta = 0.$$

We now turn to sufficiency. Suppose the allocation q is increasing and satisfies the integral constraint. Then construct the following transfers

$$t(\theta) = \begin{cases} \theta q(\theta) - \max\{q(\theta), 0\} - \int_0^\theta q(x)\, \mathrm{d}x & \text{if} \quad \theta < 1/2, \\ \theta q(\theta) - \max\{q(\theta), 0\} + \int_\theta^1 q(x)\, \mathrm{d}x & \text{if} \quad \theta \geq 1/2. \end{cases} \tag{34}$$

Because the allocation satisfies the integral constraint, we have

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

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

$$t(\theta) = q(\theta)\theta - \max\{q(\theta), 0\} - \int_0^{\theta} q(x)\,\mathrm{d}x.$$

Finally, the expected utility of type θ from reporting type θ' is given by

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

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

Proof of Proposition 6. We first derive the seller's objective in the usual way. Using (34), we write the expected transfers as

$$\begin{aligned}\int_0^1 t(\theta)\,\mathrm{d}F(\theta) &= \int_0^1 (\theta q(\theta) - \max\{q(\theta), 0\})\,\mathrm{d}F(\theta) \\ &\quad - \int_0^{1/2}\int_0^{\theta} q(x)\,\mathrm{d}x\mathrm{d}F(\theta) + \int_{1/2}^1\int_{\theta}^1 q(x)\,\mathrm{d}x\mathrm{d}F(\theta).\end{aligned}$$

Integrating the last two terms by parts, we obtain

$$-F(1/2)\int_0^{1/2} q(x)\,\mathrm{d}x + \int_0^{1/2} q(x)\,F(x)\,\mathrm{d}x - F(1/2)\int_{1/2}^1 q(x)\,\mathrm{d}x + \int_{1/2}^1 q(x)\,F(x)\,\mathrm{d}x,$$

and hence

$$\int_0^1 t(\theta)\,\mathrm{d}F(\theta) = \int_0^1 \left[(\theta q(\theta) - \max\{q(\theta), 0\})\,f(\theta) - (F(1/2) - F(\theta))\,q(\theta)\right]\mathrm{d}\theta.$$

We now establish that the solution to the seller's problem (28) can be characterized through Lagrangian methods. For necessity, note that the objective is concave in the allocation rule; the set of non-decreasing functions is convex; and the integral constraint can be weakened to the real-valued inequality constraint

$$\int_0^1 q(\theta)\mathrm{d}\theta \leq 0. \tag{35}$$

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

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

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

$$\bar{J}(\theta, q) := \int_{-1}^{q} \left(\bar{\phi}\left(\theta, x\right) - \lambda^* \right) \mathrm{d}x,$$

where $\bar{\phi}\left(\theta, x\right)$ denotes the ironed virtual value for allocation x. Note that $\bar{J}(\theta, q)$ is weakly concave in q. Because the multiplier λ shifts all virtual values by a constant, the result in Proposition 6 then follows from Theorem 4.4 in Toikka (2011). Finally, note that $\bar{\phi}(\theta, q) \geq 0$ for all θ implies the value λ^* is strictly positive (otherwise the solution q^* would have a strictly positive integral). Therefore, the integral constraint (35) binds. ■

Proof of Proposition 7. From the Lagrangian maximization, we have the following necessary conditions

$$q^*(\theta) = \begin{cases} -1 & \text{if} \quad \bar{\phi}(\theta,-1) < \lambda^*, \\ \bar{q} \in [-1,0] & \text{if} \quad \bar{\phi}(\theta,-1) = \lambda^*, \\ 0 & \text{if} \quad \bar{\phi}(\theta,-1) > \lambda^* > \bar{\phi}(\theta,1), \\ \bar{q}' \in [0,1] & \text{if} \quad \bar{\phi}(\theta,1) = \lambda^*, \\ 1 & \text{if} \quad \bar{\phi}(\theta,-1) > \lambda^*, \end{cases}$$

and

$$\int_0^1 q^*(\theta)d\theta = 0.$$

If λ^* coincides with the flat portion of one virtual value, then by the pooling property of Myerson (1981), the optimal allocation rule must be constant over that interval, and the level of the allocation is uniquely determined by the integral constraint. Finally, suppose λ^* equals the value of $\bar{\phi}(\theta, q^*(\theta))$ over more than one flat portion of the virtual values $\bar{\phi}(\theta,-1)$ and $\bar{\phi}(\theta,1)$. Without loss of generality, we can select an allocation q^* that assigns experiment $q = 0$ or $q \in \{-1,1\}$ to all types in one of the two intervals. ■

Proof of Proposition 8. (1.) If $F(\theta) + \theta f(\theta)$ and $F(\theta) + (\theta - 1) f(\theta)$ are strictly increasing, then ironing is not required and it follows from the analysis in the text that the optimal solution has $q \in \{-1,0,1\}$ for all θ.

(2.) If all types are located at one side from $1/2$ then the integral constraint has no bite since the allocation rule $q(\theta)$ can always be adjusted on the other side to satisfy it. The solution on each side of $1/2$ involves a cutoff type and $q \in \{-1,0\}$ or $q \in \{0,1\}$, both of which result in flat pricing.

(3.) If types are symmetrically distributed, then the separately optimal menus for types $\theta < 1/2$ and $\theta > 1/2$ are identical. Therefore, the union of the two solutions satisfies the integral constraint, and hence solves the original problem. ■

References

Abraham, I., S. Athey, M. Babaioff, and M. Grubb (2014): "Peaches, Lemons, and Cookies: Designing Auction Markets with Dispersed Information," Discussion paper, Microsoft Research, Stanford, and Boston College.

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.

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

Daskalakis, C., A. Deckelbaum, and C. Tzamos (2015): "Strong duality for a multiple-good monopolist," in *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pp. 449–450. ACM.

Deneckere, R., and R. McAfee (1996): "Damaged Goods," *Journal of Economics and Management Strategy*, 5, 149–174.

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.

Hörner, J., and A. Skrzypacz (2015): "Selling Information," *Journal of Political Economy*, forthcoming.

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 (2014): "Sales Talk," Discussion paper, PSE and UCL.

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.

Maskin, E., and J. Riley (1984): "Monopoly with Incomplete Information," *Rand Journal of Economics*, 15(2), 171–196.

Mussa, M., and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18(2), 301–317.

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

——— (1986): "Multistage Games with Communication," *Econometrica*, 54(2), 323–358.

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.

Sarvary, M. (2012): *Gurus and Oracles: The Marketing of Information*. MIT Press.

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

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