Screening Costly Information*

João Thereze[†]

December 31, 2023

Abstract

We study screening with endogenous information acquisition. A monopolist offers a menu of quality-differentiated products. After observing the offer, a consumer can costly learn about which product is right for them. We characterize the optimal menu and how it is distorted: typically, all types receive lower-than-efficient quality, and distortions are more intense than under standard screening, even when no information is acquired in equilibrium. The additional distortion is due to the threat from the buyer to obtain information that is not optimal for the seller. This threat interacts with the division of surplus: profits are U-shaped in the level of information costs and, when such costs are low, the consumer may be better off than when information is free.

*I am indebted to my advisor, Pietro Ortoleva, for his continued guidance throughout this project. I am grateful to Sylvain Chassang, Ana Luiza Dutra, Alessandro Lizzeri, Xiaosheng Mu, Franz Ostrizek, Rafael Parente, Wolfgang Pesendorfer, Pëllumb Reshidi, Denis Shishkin, Can Urgun, Leeat Yariv and Mu Zhang for very helpful suggestions and feedback. I also thank audiences at Princeton University, the University of California at Davis and the Young Economist Symposium for their helpful comments and discussions.

†Fuqua School of Business, Duke University joao.thereze@duke.edu

1 Introduction

Consumers are often unsure about which product is right for them. In health insurance markets, potential buyers grapple with deciding which plan to contract, incurring large losses for selecting the wrong product (Abaluck and Gruber, 2011; Brown and Jeon, 2020; Handel and Kolstad, 2015). Frequently, additional information that would aid the agent's decision is available, but acquiring and processing it requires effort. In the insurance example, one could look at experience in previous years and research family history to forecast coverage needs, or compare how different plans cover various possible conditions. Whether the buyer expends effort to gauge their own preference or to assess how product characteristics match their personal taste, information acquisition is costly. We study how this costly information acquisition in demand affects product supply and, therefore, equilibrium outcomes.

This paper considers a model of vertical product differentiation in which consumers need to pay a cost to discover their type. Understanding how supply responds to information frictions is important for several reasons. First, extensive evidence shows costly information acquisition significantly affects consumer behavior.[1] Second, supply responses shed light on producers' incentives to aid or hinder consumer learning. Indeed, network provider and insurance company websites alike offer tools that help consumers compare plans, suggesting that sellers act to influence information acquisition. Finally, information frictions justify public interventions; for example, a health insurance policy may include instruments that help customers choose the right product for them (Brown, 2019). However, the consequences of these interventions on consumer welfare cannot be ascertained without considering equilibrium effects. In fact, we show that helping buyers choose often reduces consumer welfare.

The Model As in Mussa and Rosen (1978), a monopolist (she) offers goods of different quality to screen different types of the agent (he). In standard screening, the buyer knows his taste for quality, but the seller does not. Here, we assume the agent's valuation is unknown to both players at the beginning of the game. Valuations may assume one of two values. After observing the menu of available goods, but before making a purchase decision, the buyer can costly acquire information about how much he values quality; that is, the agent is rationally inattentive about his type. Information acquisition can be interpreted as the buyer learning either about his own preferences, or about product characteristics that affect his personal match value with the good. Because learning takes place after the seller makes her offer, the menu of goods will affect information choices: the monopolist can influence which information is acquired by choosing which contracts to provide. This interaction between contracts and costly learning is central to our analysis.

[1] See Abaluck and Gruber (2011), Brown and Jeon (2020), and Taubinsky and Rees-Jones (2018), for example for evidence on consumer behavior.

Results Information acquisition brings a new dimension to the monopolist's problem: contracts must induce the buyer to acquire the information the seller wants him to learn. This requirement implies two additional constraints to optimal screening. First, the principal must fine-tune goods' quality for the agent to acquire specific information. If quality levels are far apart, selecting the wrong product implies large utility losses, and the buyer has considerable incentives to learn. If quality levels are close together, mistakes are less consequential, and the buyer will not choose to learn much. Thus, which information is acquired depends on the extent of product differentiation. Second, prices must be low enough for the consumer to buy. The more expensive products are, the higher the incentives for the buyer to learn: if he learns his value for quality is low, he does not purchase any goods, and thus he avoids paying too much for quality he does not enjoy. To discourage this behavior, the principal must offer price discounts.

We characterize the solution to the principal's problem and derive two key results. First, we show information acquisition adds to, rather than mitigates, the inefficiencies generated by information asymmetry alone. Quality distortions are both more widespread and larger than in standard screening. They are more widespread because agents with all valuations typically receive below-efficient quality levels — in contrast to the standard screening result, where the agent with the highest valuation is assigned efficient quality. Distortions are larger in the sense that, on average, quality levels are further away from the efficient ones, when compared with the standard screening benchmark. The reason is that the consumer shifts the balance of power in the agency relationship by threatening to learn what the seller does not want him to. The buyer receives rents not only due to the private information he actually bears, but also because of the information he could have obtained.

For an intuition, consider the following simple case. When acquisition costs are high enough, we prove no learning takes place, so information is symmetric in equilibrium. In standard screening, if information is symmetric, the seller offers a single good to the buyer, with efficient quality tailored to his valuation; prices are such that the consumer is indifferent between buying or not. Here, this offer does not work. If the agent is ex-ante indifferent between buying and not buying, then he has incentives to learn whether his real value is just a little smaller than his ex-ante belief. If he finds out his type is low, the agent does not purchase the good. To dissuade the buyer from learning, the seller gives the product a price discount and, to save on costs, degrades its quality. Thus, the contract is inefficient even when information is symmetric. This argument highlights the distinction between this model and standard screening: the key distortion here is the *threat* of obtaining information, rather than the presence of private information.

The second main result is that profits and consumer surplus are non-monotonic in the level of information costs. Profits are U-shaped, whereas consumer surplus often increases for low acquisition costs, and always decreases when costs are high. The reason is that the credibility of the agent's threat varies as costs change. When information costs are small, a fair amount of information is acquired in equilibrium. In that setting, the agent earns rents by threatening the principal not to learn as much as she wants. This threat gets more credible as acquisition costs rise and learning becomes harder. Thus, in that range, an increase in costs benefits the agent at the expenses of the principal. By contrast, when acquisition costs are high, information acquisition is very limited in equilibrium, and the agent extracts rents from the threat of learning more than the principal wants. Then, an increase in costs reduces the credibility of the threat, benefiting the seller and hurting the buyer.

Implications The non-monotonicity result has several implications. First, it sheds light on sellers' incentives to manipulate learning. These incentives vary with the level of information costs. If costs are low, it tends to be optimal for the principal to facilitate learning, by favoring transparency and providing tools that ease information acquisition. If acquisition costs are high, sellers may benefit from hiding product information and dissuading learning. Importantly, the seller may benefit from manipulating learning even if her action does not change information choices in equilibrium. Rather, changing the credibility of the buyer's threat is what drives the redistribution of surplus. This rationale complements insights from the literature on product obfuscation, in which firms dissuade consumer learning in order to profit against competitors.[2] Conversely, in our model, the seller aims to reduce the buyer's strategic advantage and may even help buyers acquire information.

Second, the non-monotonicity of consumer welfare has consequences for the design of policies that facilitate consumer learning. These transparency policies are often used in practice (Brown, 2019; Hackethal et al., 2012). The rationale supporting such interventions is based on how information costs affect consumers by causing choice mistakes: buyers fail to purchase the product that is right for them. Thus, when supply is fixed, consumer welfare increases if acquisition costs decrease. We show this argument may fail when supply is allowed to respond. The seller benefits from more informed consumers, because she is then able to tailor products to consumer types. Hence, she is willing to provide incentives for the agent to learn and make fewer mistakes. When information costs are not too high, incentive provision more than compensates the buyer for his choice mistakes. Thus, transparency policies could have unintended equilibrium consequences, harming consumers, instead of benefiting them.

Related literature This paper contributes to the large body of work applying rational inattention models to different strategic interactions (Bloedel and Segal, 2020; Matějka and Tabellini, 2016; Ravid, 2020; Yang, 2019).[3] In particular, it relates to the literature on product market equilibrium when consumers are inat-

[2]See Ellison (2005), Ellison and Wolitzky (2012), and Petrikaitė (2018).
[3]See Mackowiak et al. (2020) for a thorough review of applications of rational inattention by field.

tentive (Boyacı and Akçay, 2018; Hefti, 2018; Martin, 2017; Matějka and McKay, 2012). In these papers, firms offer a single good to agents who are rationally inattentive with respect to quality, price, or both. Closest to ours are Mensch (2022) and Mensch and Ravid (2022). Mensch (2022) studies the problem of selling a single good of fixed quality in the same information environment as this paper. Our framework differs in that here the principal offers multiple products, with endogenous quality levels, to induce or deter information acquisition. Thus, our model connects incentives for learning with product differentiation, a central contribution of this paper.

In concurrent work, Mensch and Ravid (2022) study a similar model of monopolistic screening with endogenous information acquisition, and also prove that quality distortions are more widespread than under standard screening. The papers differ in two dimensions. First, we provide additional results on the effect of these distortions on surplus. In particular, we show information costs may affect profits and consumers' surplus non-monotonically, with implications for firms' incentives to manipulate learning, as well as for the design of transparency policies. Second, we model information costs distinctly. While the two frameworks are equivalent when the state space is binary, their model allows for an uncountable state space and focuses on information costs that depend on the distribution of the buyer's posterior means.

Our paper complements a literature that studies how information acquisition interacts with mechanism design. Roesler and Szentes (2017) model a bilateral trade setting in which the buyer obtains information before the seller offers a menu. Similarly, Ravid et al. (2022) study the same setting when acquisition decisions and price setting occur simultaneously. These papers differ from ours in the timing of information acquisition: here, the monopolist commits to the menu beforehand, providing incentives for the information she wants to be acquired, which gives rise to the key distortion in our model. This timing distinction is critical: the main takeaway of Ravid et al. (2022) is that the consumer can be significantly better off when information is freely available, versus when it is extremely cheap. Our timing assumption reverses that message, in that the buyer can be better off with costly information than when information is free. In section 5, we discuss this reversal and the role of timing.

Li and Shi (2017) and Guo et al. (2018) study information disclosure by a principal in a screening setting. In contrast to our model, information is controlled by the seller, not by the buyer. Our work is thematically related to an older literature on mechanisms with information acquisition (Bergemann and Välimäki, 2002; Crémer and Khalil, 1992; Shi, 2012; Szalay, 2009). We differ from these papers by focusing on multiple products and allowing for flexible information acquisition. The techniques we apply are connected with works on contracting with information design (Boleslavsky and Kim, 2018; Doval and Skreta, 2022; Georgiadis and Szentes, 2020; Ostrizek, 2020). We differ from these papers in that our problem is not one of information design, but rather of information acquisition. Additionally, none of those papers study screening. Finally, we are indebted to tools and ideas developed in decision problems under rational inattention, for example Caplin et al. (2022) and Matějka and McKay (2015). In particular, our solution to the acquisition problem adapts the method developed by Caplin et al. (2022), based on concavification techniques (Aumann and Maschler, 1995; Kamenica and Gentzkow, 2011).

2 A Model of Menu Design with Costly Information Acquisition

A monopolist (the principal, she) aims to sell indivisible goods to a potential buyer (the agent, he). She produces goods of different quality levels $q \geq 0$ at cost $c(q)$, where the cost function c is increasing, strongly convex and twice continuously differentiable.[4] The buyer's valuation for quality is $\vartheta \in \{\underline{\theta}, \overline{\theta}\}$, with $0 < \underline{\theta} < \overline{\theta}$. If an agent with valuation ϑ purchases a good of quality q at price t, he receives utility $\vartheta q - t$. The monopolist offers the buyer a menu of quality-price pairs. We assume throughout that $c(0) = c'(0) = 0$.

The model departs from traditional screening in two ways. First, information is symmetric at the beginning of the game: originally, ϑ is unknown to both players, who share a prior with mean μ. Second, before making his purchasing decision, but after observing the menu, the agent can acquire information about his valuation. Formally, he can choose an information structure $(\mathcal{S}, P)$, which consists of a set of signal realizations $\mathcal{S}$ and a function $P : \{\underline{\theta}, \overline{\theta}\} \to \Delta(\mathcal{S})$ that assigns a distribution over signals for each state. Although the agent is free to choose any information, learning is costly, as described below.

The timing of the game is as follows. The principal acts first, offering a schedule of quality-price pairs. After observing the offer, the agent decides which information to acquire. Upon observing a signal realization, the agent decides whether to buy one of the options from the menu or none — in which case he obtains a utility of zero. Importantly, the choice of information happens only after the menu is observed and, hence, may depend on the menu of options offered by the monopolist.

Information and Acquisition Costs Because the agent's utility is linear, it depends on information only through the mean of posterior beliefs. Following standard practice, we associate each signal to the posterior mean it generates (Dworczak and Martini, 2019; Gentzkow and Kamenica, 2016). Formally, a signal realization is $\theta \in [\underline{\theta}, \overline{\theta}] \equiv \Theta$, and, upon observing signal θ, the buyer's expected payoff is $\theta q - t$. We refer to the realization of the agent's private information, θ, as his type or, abusing notation, his posterior. Then, any information structure is a distribution over realizations $F \in \Delta(\Theta)$ satisfying a Bayesian consistency constraint (Kamenica and Gentzkow, 2011). Because the state is binary, this constraint can be written as:

$$\mathbb{E}_F[\theta] = \mu. \tag{BC}$$

[4] A function $f \in \mathcal{C}^2$ is strongly convex if $\min_x f''(x) > 0$.

We assume information-acquisition costs are posterior-separable (Caplin et al., 2022). A convex function over posterior means, $H : \Theta \to \mathbb{R}$, exists, with $H(\mu) = 0$, and a scalar $k \geq 0$ such that the cost of an information structure F is defined as:

$$K(F) = k\mathbb{E}_F\left[H(\theta)\right].$$

This class of cost functions generalizes mutual information-based acquisition costs, widely used in works on rational inattention, and encompasses almost all of the costs contained in the flexible information acquisition literature.[5] By obtaining information, the agent moves his posterior away from the prior. Heuristically, H defines this informal notion of distance between prior and posterior. The cost K reflects the expectation of this distance across all possible signal realizations according to information structure F. Note K is monotonic in the Blackwell order, in the sense that more informative structures are costlier. The parameter k scales the cost, and it is later used for comparative statics. We assume $H \in C^3$ and strongly convex.

Two subclasses of posterior-separable cost functions are of particular relevance. We say that a cost function has the unbounded marginal costs (UMC) property if increasing the precision of a signal becomes arbitrarily costly as the signal becomes more precise. Formally, H is UMC if $\lim_{\theta \to \vartheta} |H'(\theta)| = \infty$, for $\vartheta \in \{\underline{\theta}, \overline{\theta}\}$. Mutual-information costs belong to this class. Conversely, we say a function is of bounded marginal costs (BMC) if $|H'|$ is bounded. A prominent example is quadratic costs, $H(\theta) = \frac{(\theta-\mu)^2}{2}$, which measure the expected reduction in prior variance obtained by observing information F.

The Principal's Problem. We study the mechanism that maximizes expected profits for the monopolist. When choosing a mechanism, the seller takes into account that her offer affects the decisions of the buyer in two ways. First, it determines the buyer's choice given his information. In standard screening, the revelation principle guarantees that it suffices for the seller to offer one contract to each type of agent. However, here the agent type is determined by the information he decides to acquire, which is endogenous. As a consequence, the traditional revelation principle does not apply directly in this setting. Nonetheless, the principal can restrict attention to menus of quality-transfer pairs $\mathcal{M} = \{q(\theta), t(\theta)\}_{\theta \in \Theta}$, in which one contract is offered to each possible posterior of the agent, including those that have zero probability in equilibrium.[6]

[5] The standard formulation for posterior-separable information costs is in terms of posterior distributions. Because the state is binary, we can write it as a function of posterior means only. To embed mutual information, $H(\theta) = \sum_{\vartheta \in \{\underline{\theta}, \overline{\theta}\}} \left\{ \frac{|\vartheta-\theta|}{\overline{\theta}-\underline{\theta}} \log \frac{|\vartheta-\theta|}{\overline{\theta}-\underline{\theta}} - \frac{|\vartheta-\mu|}{\overline{\theta}-\underline{\theta}} \log \frac{|\vartheta-\mu|}{\overline{\theta}-\underline{\theta}} \right\}$. Part of the literature has been devoted to finding costs of acquisition that are alternative to mutual information, for example Hébert and Woodford (2020) and Pomatto et al. (2023). The cost specifications in these papers satisfy posterior-separability. Applications have either used Shannon entropy or other posterior-separable costs. Examples are Matějka and McKay (2015), Hébert and La'O (2020) and Yang (2019).

[6] We prove this revelation principle in the presence of information acquisition in Theorem OA 1 in Online Appendix B. The results is related to revelation principles for general type spaces as in Skreta (2006) and Hellwig (2010).

The second way in which the menu affects consumers' decisions is through information choices. Because information is acquired after contracts are offered, acquisition depends on the terms of those contracts. By designing the menu, the principal indirectly controls information acquisition. Therefore, the seller maximizes profits knowing both which information will be acquired in equilibrium and how each type selects across contracts. Defining the outside option for the consumer as $C_o \equiv \{q(o), t(o)\} = \{0, 0\}$, we can write the principal's problem:

$$\max_{\mathcal{M}, F \in \Delta(\Theta)} \quad \mathbb{E}_F \left[t(\theta) - c\left(q\left(\theta\right)\right)\right]$$

$$\text{s.t.} \quad \theta \in \arg\max_{\omega \in \Theta \cup \{o\}} \theta q(\omega) - t(\omega), \quad \theta \in \Theta \tag{IC}$$

$$F \in \arg\max_{\mathbb{E}_G[\theta]=\mu} \mathbb{E}_G \left[\max_{\omega \in \Theta \cup \{o\}} \{\theta q(\omega) - t(\omega)\} - kH(\theta) \right]. \tag{IA}$$

The monopolist maximizes expected profits subject to two sets of constraints. The first one, IC, subsumes both traditional incentive compatibility and individual rationality constraints. It says that, given a posterior, $\theta \in \Theta$, the buyer chooses from the contracts in the menu and the outside option to maximize utility, obtaining interim rents $U(\theta) \equiv \max_{\omega \in \Theta \cup \{o\}} \{\theta q(\omega) - t(\omega)\}$. Note IC depends only on the offered menu, and not on which information is acquired. Thus, this constraint can be characterized following standard practice in mechanism design (Myerson, 1981). Maximizing expected profits subject to IC is the classic monopolistic screening problem, given a distribution over types F (Maskin and Riley, 1984; Mussa and Rosen, 1978). Throughout this paper, we denote this optimization as the standard screening problem, and we use it as a benchmark.

The information-acquisition constraint, IA, is the departure from standard screening. It says information must be a solution to the agent's acquisition problem. In that problem, the buyer chooses an information structure — namely, a Bayesian-consistent distribution of types — to maximize the expectation of utility net of acquisition costs, $U - kH$. Costs are posterior-separable, and utility is given by interim rents: the maximum payoff of the contracting choice for each type realization. Finally, if multiple information structures solve the acquisition problem, we allow the principal to select his favorite; that is, we consider the principal-optimal equilibrium of the game.

3 Results

3.1 Simplifying the Principal's Problem

In this section, we turn the problem of the principal into a simple optimization that can be solved by standard techniques. We show two features of the problem that connect contracting with information acquisition. These properties allow for rewriting the principal's optimization into a tractable, finite-dimensional problem.

Information acquisition brings two novel features to the contracting problem, which we call (i) marginal incentives for acquisition and (ii) the threat point. Recall that the principal controls information indirectly, by designing the menu. She must provide incentives for any information she wants to be chosen. (i) and (ii) characterize how the principal can provide such incentives: by controlling the marginal benefits of learning through product differentiation; and by adjusting the level of rents. In the discussion below, we explain these properties and their implications in constraining the space of contracts available to the principal.

Henceforth, we follow the standard practice of denoting menus as rent-quality pairs, $\{U(\theta), q(\theta)\}_{\theta\in\Theta}$. In the following discussion, we assume an optimal information has at most two posteriors. As a consequence of binary states, any such information structure can be identified by the two posterior means in its support: $\operatorname{supp} F = \{\theta_L, \theta_H\}$, $\theta_L \leq \mu \leq \theta_H$. Because only two posteriors are chosen, at most two contracts are signed with positive probability, $\{C_L, C_H\}$, with rents and quality levels $\{U_i, q_i\}_{i\in\{L,H\}}$, where $U_i \equiv U(\theta_i)$ and $q_i \equiv q(\theta_i)$. At the end of this subsection, Proposition 1 shows the optimal information structure is indeed binary. We characterize properties (i) and (ii) in terms of these equilibrium contracts.

Marginal incentives. The first property ties together product differentiation and the endogenously acquired information. Because product quality affects the value of learning, it affects information acquisition. Intuitively, agents benefit from information because they can adopt the contract that better reflects their tastes. As quality levels grow apart, this benefit increases, because purchasing the wrong contract has larger consequences. To induce a specific information strategy, the principal must fine-tune the quality of contracts to the information she wants to be acquired. To see this argument heuristically, recall that the consumer chooses information to maximize the expectation of utility net of costs, $U - kH$. Then, marginal net utilities $U' - kH'$ can be understood as the marginal value of information. But by the classic characterization of Myerson (1981), incentive compatibility implies $U'(\theta) = q(\theta)$, so marginal net utilities are affected by product quality.[7] Thus, quality influences information decisions.

[7] A priori, it is not clear that the rent function U is differentiable at the relevant points in the support of F. This is formally shown in the proof of Lemma 1.

The marginal-incentives property makes this interplay between product differentiation and information acquisition precise. Roughly, it states that *marginal net utilities must be equated in the support of an optimal structure*. When this property fails, the buyer prefers to acquire information that is not predicated by the information structure, either because quality levels are too far apart and incentives for learning are excessive, or vice versa. This argument works only partially if an optimal structure fully reveals a state. In that case, because the agent cannot learn more about the fully revealed state, an inequality of marginal utilities holds instead. The following result formalizes the discussion above and establishes the marginal-incentives constraint formally.

Lemma 1. *Let F and $\{U(\theta), q(\theta)\}_{\theta\in\Theta}$ satisfy IC and IA, with* $\operatorname{supp} F = \{\theta_L, \theta_H\}$*. Then, there is $\psi \in \mathbb{R}$ such that, for $i \in L, H$:*

$$\begin{aligned} q_i - kH'(\theta_i) &= \psi, \quad \text{if } \theta_i \in (\underline{\theta}, \overline{\theta}) \\ q_L - kH'(\theta_L) &\le \psi, \quad \text{if } \theta_L = \underline{\theta} \\ \psi &\le q_H - kH'(\theta_H), \quad \text{if } \theta_H = \overline{\theta} \end{aligned} \tag{M}$$

Figure 1 illustrates this result graphically. The figure plots the net utility obtained at different posteriors, which are depicted in the horizontal axis. For each posterior, the agent chooses the best contract from $\{C_L, C_H\}$ or the outside option C_o, as shown in the bottom of the picture. The standard concavification argument implies that any optimal information structure, F, is represented by a segment tangent to the graph of $U - kH$.[8] We identify this segment with its slope ψ. The support of F are the types at which tangency happens — that is, $\operatorname{supp} F = \{\theta_L, \theta_H\}$ —, and the ex-ante utility of the agent is the height of the line segment at the prior mean. Note ψ plays an important role in our analysis. In particular, as we further emphasize later, ψ adjusts the value of information $U - kH$ to take into account the Bayesian-consistency constraint.

An immediate consequence of Lemma 1 is that, insofar as no posterior in the support of F is fully revealing, the difference between quality levels is pinned down by the marginal-incentives property:

$$q_H - q_L = kH'(\theta_H) - kH'(\theta_L)$$

Threat Point. The second property concerns the level, rather than the margin, of acquisition incentives: it is reminiscent of a participation constraint. Recall that the principal wants the agent to pick a contract from the menu she designs. However, the agent can always choose the outside option C_o after information realizes. He does so if he finds out that both contracts are too expensive given his information, that is, if he gets a signal that his type is too low. To guarantee participation, the principal must discourage the acquisition of any information that induces the agent to opt out of the menu. Figure 2a depicts a situation

[8] For concavification in general, see Aumann and Maschler (1995) and Kamenica and Gentzkow (2011). For the argument applied to information acquisition, see Caplin et al. (2022).

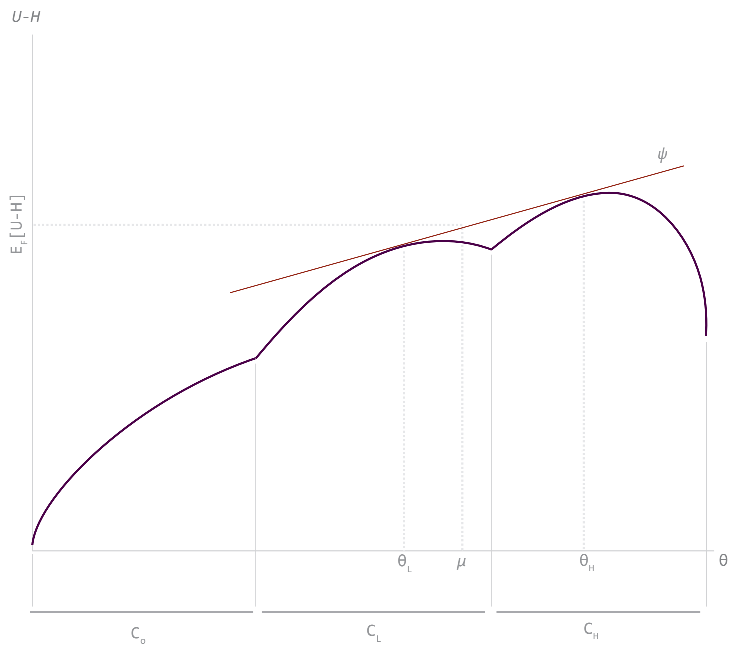
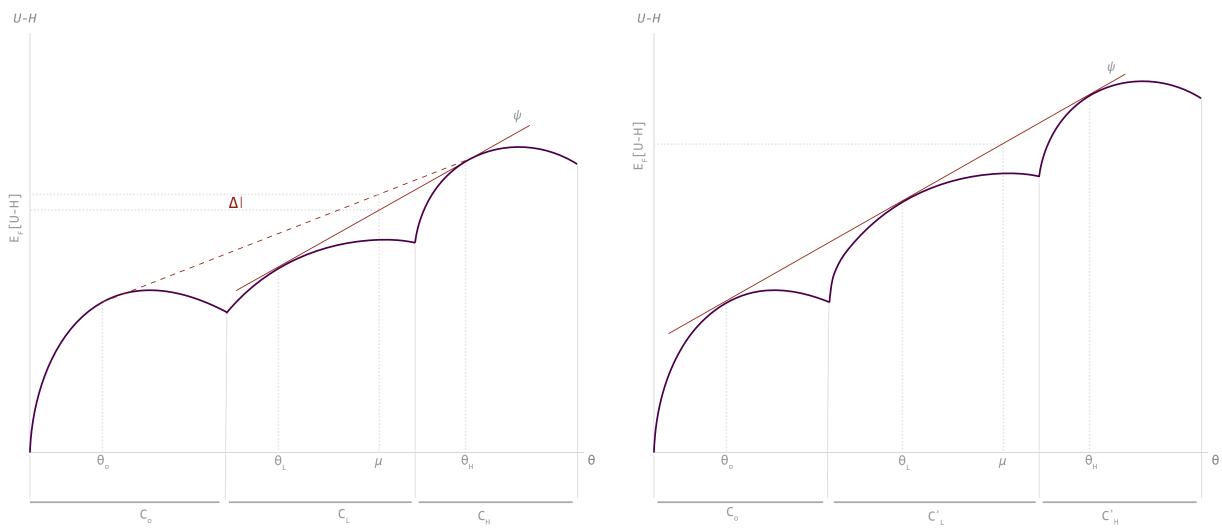


Figure 1: An Optimal Information Structure

of non-participation. In that example, the agent rejects the principal's prescription, F, deviating to the information structure denoted by the dotted segment, with support $\{\theta_o, \theta_H\}$. If signal θ_o realizes, he opts-out of the menu. That deviation generates an expected gain of $\Delta > 0$ for the buyer. In the figure, F satisfies the marginal-incentives property, but the deviation is still attractive to the buyer: the seller must resort to a different set of tools to guarantee participation.

The principal can avoid such deviations by making the menu more attractive. In particular, she can affect the level of net utilities by reducing the prices of C_L and C_H, obtaining new contracts $\{C_L', C_H'\}$, as can be seen in Figure 2b. By reducing transfers in each of these contracts, the principal raises rents — and, thus, net utility — for types that accept those contracts. Additionally, she increases the interval of types that participate in the menu, that is, the interval of posteriors choosing the outside option, C_o, shrinks. This change in contracts renders the prescribed information structure F optimal to the buyer by ruling out the deviation to θ_o.

This intuition generalizes. Given the quality levels and prescribed information, we identify a type that plays a central role in the opting-out deviations: we call it the threat point. The threat point is the posterior that maximizes the ex-ante value of purchasing C_o for the buyer. Recall that the value of information is given by the net utility $U - kH$. By trying to maximize this value, the buyer must take into account



(a) A profitable deviation

(b) No profitable deviation

Figure 2: Net Utilities and the Threat Point

the Bayesian consistency constraint. To do so, the buyer maximizes the adjusted value of information $U - kH - \psi\theta$. The threat point, $\theta_o(\psi)$, is a posterior under which the agent chooses not to buy any contract and that maximizes the adjusted value of information:

$$\theta_o(\psi) \equiv \arg\max_{v \in \Theta}\{-kH(v) - \psi v\}.$$

As in Figure 2b, the threat-point property establishes that, at an optimal information structure, the adjusted value of information in the support is equated with the adjusted value of information at the threat point. The next result states that property formally.

Lemma 2. *Let $F, \{U(\theta), q(\theta)\}$ solve the principal's problem with* $\operatorname{supp} F = \{\theta_L, \theta_H\}$. *Furthermore, assume ψ is as in Lemma 1. Then, the following holds:*

$$U_i - kH(\theta_i) - \psi\theta_i = -kH(\theta_o(\psi)) - \psi\theta_o(\psi) \quad \textit{for all} \;\; i \in \{L, H\} \tag{TP}$$

Lemma 2 shows the relation between contracts in the support and the threat point is tight: the consumer is exactly indifferent between learning $\theta_o(\psi)$ and following F. The expression on the left-hand side is the adjusted value of obtaining posterior θ_i, that is in the support of information structure F, prescribed by the principal. The right-hand side is the adjusted value of learning the threat point. This indifference implies the threat point regulates the level of rents. By determining the level of rents, quality is pinned down because rents are linked to quality by incentive compatibility. A higher threat point reduces quality and rents. Formally, if the principal wants to implement information structure $\operatorname{supp} F = \{\theta_L, \theta_H\}$ such that no type is fully revealing of a state, quality levels must satisfy:

$$q_i = kH'(\theta_i) - kH'(\theta_o(\psi)), \quad i \in \{L, H\}.$$

Together, marginal incentives for acquisition and the threat point tightly constrain the set of contracts the principal can choose from. Marginal incentives for acquisition tie the differences in quality levels for any two contracts to the prescribed information structure, whereas the threat point fixes the quality level. An implication of this discussion is that product differentiation is pinned down by the information structure that the principal wants to induce. Given the choice of information, the only instrument remaining for the principal is the choice of the threat point or, equivalently, the dual variable ψ.

A simpler problem. Lemma 1 and Lemma 2 show M and TP are constraints to the principal's optimization. The following result proves they are the only constraints the principal must satisfy in addition to Bayesian consistency, BC. We say two optimization problems are equivalent if they have the same value and the solution to one can be used to construct a solution to the other.

Proposition 1. *The principal's problem is equivalent to:*

$$\max_{\{\theta_i, U_i, q_i\}_{i\in\{L,H\}}, \psi} \Big\{ \sum_{i\in\{L,H\}} p_i^F \left[\theta_i q_i - c(q_i) - U_i\right] : M, \;\; TP \quad and \quad BC \Big\}, \tag{P}$$

where p_i^F is the probability of θ_i under $\operatorname{supp} F = \{\theta_L, \theta_H\}$.

Furthermore, this problem has a solution.

Proposition 1 shows the principal's problem can be greatly simplified to a finite dimensional optimization with equality and inequality constraints. The objective function is the principal's profits rewritten as surplus minus rents, as standard, and assuming F is at-most-binary. Essentially, the result states that the principal can focus on binary information structures and on equilibrium variables: posteriors and contracts that are signed with positive probability in equilibrium. In the original principal's problem, the seller must take into account the off-equilibrium behavior of the buyer: the information they could have chosen to obtain and the contracts they would pick if they acquired different information. In Proposition 1, instead, the principal only needs to track the slope ψ, which, in addition to equilibrium variables, guarantees that the agent will follow the plan of action she prescribes.

For an intuition, note our discussion thus far has proved that M and TP must hold in a solution to the principal's problem. Thus, provided the restriction to binary distributions, the profits attained in the problem in Proposition 1 are at least as large as the principal's profits. The majority of the proof consists in showing the converse, namely, that any solution to the problem above can be extended to satisfy IC and IA without loss of profits. This result is obtained in several steps. First, we show contracts are incentive compatible in $\{\theta_L, \theta_H\}$, which is guaranteed because M implies quality is monotonic, and TP allows us to compare interim rents. TP also implies individual rationality. In fact, IC holds strictly: the principal needs to provide more incentives for the agent to acquire information, than for him to simply reveal it. Then, we show we can extend contracts to satisfy IC to posteriors that have zero probability in equilibrium. Finally, a concavification argument guarantees IA is also satisfied.

A tractable problem Note Proposition 1 delivers a tractable optimization problem suitable for applications. For an illustration, consider the case in which H is UMC; that is, the marginal cost of learning grows towards infinity as the precision of any signal increases. Thus, the optimal information structure contains no fully revealed state. As a consequence, M holds with equality, allowing us to write quality, q_i, as a function of θ_i and ψ. Similarly, one can use the threat-point property, TP, to solve for interim rents, U_i, and Bayes-consistency can be used to write p_i^F as a function of $\operatorname{supp} F$ and μ.

These three observations allow for substitution of q_i, U_i, and p_i in the seller's optimization. The following remark summarizes this discussion:

Remark. *When H is UMC the problem of the principal can be rewritten as*[9]*:*

$$\max_{\theta_L \le \mu \le \theta_H, \psi} \sum_i \frac{|\theta_j - \mu|}{\theta_H - \theta_L} \Big\{ \theta_i (kH'(\theta_i) - kH'(\theta_o(\psi))) - c(kH'(\theta_i) - kH'(\theta_o(\psi))) \\ - [kH(\theta_i) - kH(\theta_o(\psi)) + \psi(\theta_i - \theta_o(\psi))] \Big\}.$$

The problem above is a simple optimization over three variables. Next, we characterize general properties of the optimal menu and of the surplus, which are the main results of the paper. In the following subsections, we describe these results and discuss how they are guided by the two key forces described above: the marginal-incentives property, M, and the threat point, TP. It is worth noting that, as k converges to zero, the principal's problem converges to the standard screening problem in which the consumer is fully informed.

3.2 Menu Design and Information Acquisition

This section focuses on the efficiency of optimal contracts. We compare these contracts with the first-best allocations and with the ones obtained in standard screening. This comparison is not straightforward. Here, the agent is uninformed to begin with, and all the information is obtained endogenously by acquisition choices. By contrast, in the standard screening model, information is exogenous. To evaluate quality distortions, we keep information constant across models: we first solve for the optimal information structure in our model and then take it as the exogenous information in a standard screening problem.

[9] The simplified optimization problem for the non UMC case can be found in the proof of Proposition 2 in the Appendix.

The efficient (or first-best) quality for type θ_i is the quality that maximizes the production surplus for that type, namely, q_i^{f} such that $c'(q_i^{\mathrm{f}}) = \theta_i$. Under symmetric, exogenous information, each type of buyer is assigned a contract with efficient quality. In the second-best problem, when information is exogenous but asymmetric, the principal wants to dissuade the high-type agent from purchasing contracts aimed toward the low type. The optimal way for her to achieve that goal is to make the low contract less attractive. Thus, in the standard screening model, the principal underprovides quality to the low type, $q_L^s < q_L^{\mathrm{f}}$. On the other hand, the quality assigned to the high type, q_H^s, is efficient. Note the sole driver of this distortion is private information: the principal needs to avoid attracting high types to the low contract, because she is uninformed about consumer's value.

For any information structure F and menu of contracts with quality levels q, we use $\tau(F,q) \equiv \mathbb{E}_F[\theta - c'(q(\theta))]$ to denote the expected wedge: it measures how distorted from efficiency the quality levels in the menu are. In particular, $\tau(F,q^{\mathrm{f}}) = 0$ because the first-best menu maximizes production surplus, and $\tau(F,q^s) > 0$. We now show that when information is endogenously acquired, in contrast to the pure screening case, both types receive below-efficient quality. Additionally, the wedge is larger than in the second-best solution.

Proposition 2 (Distortion Patterns)**.** *Let $F, \{U, q^*\}$ solve the principal's problem. Then:*

> ***Quality is underprovided.*** $c'(q_L^*) < \theta_L$ *and* $c'(q_H^*) \leq \theta_H$
>
> ***Aggregate distortions.*** $\tau(F,q^s) \leq \tau(F,q^*)$

All inequalities are strict if either (i) k is high enough; (ii) $k > 0$ and H is UMC; or (iii) whenever $\operatorname{supp} F = \{\mu\}$.[10]

Proposition 2 shows that in the presence of information acquisition, distortions also happen at the top; moreover, they are larger than under standard screening. To understand the result on the wedge, note that our model adds a new distortion to the screening problem: the threat from the agent to obtain information. The buyer could obtain information on whether his type is low and opt-out of the menu if it is the case. By the threat-point property, the principal must reduce prices to guarantee no such deviation is profitable for the buyer. The efficient way to reduce prices, from the production perspective, is to also degrade quality. As a result, aggregate distortions are higher, on average. For an interpretation of the result, consider the case in which production costs are quadratic, $c(q) = \frac{q^2}{2}$. In that case, the increase in the wedge means that the average quality provided under endogenous information is lower than the average quality provided in the second-best, when information is exogenous: endogeneity increases the misallocation of qualities to types.

[10] Lemma OA1 in Online Appendix A shows that $\bar{k}$ exists such that $\operatorname{supp} F = \{\mu\}$ whenever $k \geq \bar{k}$.

To show that qualities are underprovided, it is helpful to think of the principal's problem P as an information design problem, given the threat point θ_o — or equivalently, the dual variable ψ. The principal wants to maximize the expectation of the profits obtained at posterior θ, $L(\theta) = \theta q(\theta) - c(q(\theta)) - U(\theta)$. Optimal information design requires that the principal's profit is concavified, which in turn implies that marginal profits of changing posteriors are equalized at any point in the support. By the marginal-incentives property, an increase in posteriors leads to an increase in quality, and it's impact in profits is proportional to the wedge. Heuristically:

$$L'(\theta) = \underbrace{(q(\theta) - U'(\theta))}_{0 \text{ by IC}} + (\theta - c'(q(\theta)))\, q'(\theta)$$

Concavification then implies that the wedges have the same sign. Because, on average, the wedge is positive, a distortion at the bottom begets a distortion at the top. As discussed in the Introduction, the inefficiency persists even when information is symmetric: distortions are determined by the threat of information acquisition, rather than by private information itself.

Note the inequalities in Proposition 2 are not always strict. In particular, for low k, acquiring complete information can be optimal. In this case, it is always profit maximizing for the principal to provide the same quality levels as under standard screening — although not the same prices. However, for sufficiently large costs, the inequalities are strict and the difference between this model and standard screening is sharp. Similarly, UMC cost functions guarantee that information cannot be fully acquired and thus that inequalities are always strict. The remainder of this section illustrates these results with a special case in which we completely characterize the optimal solution.

3.2.1 Quadratic Costs

Assume $c(q) = \frac{q^2}{2}$ and $H(\theta) = \frac{(\theta-\mu)^2}{2}$. Under this specification, the cost of acquiring an information structure is proportional to the reduction in prior variance obtained by observing this information.[11] As a function of k, we denote the equilibrium information structure as $F(k)$ and equilibrium quality as $q^*(k) = \{q_L^*(k), q_H^*(k)\}$. We also denote $q^s(k)$ as the second-best quality obtained when the information is exogenously set at $F(k)$. The next proposition characterizes $F(k)$ and compares $q^*(k)$ with $q^s(k)$.

Proposition 3 (Quadratic Costs)**.** *Under quadratic costs, $\underline{k} \leq \overline{k}$ exist such that the optimal structure $F(k)$ satisfies:*

[11] Specifically, $K(F) = \frac{k}{2}\mathbb{V}_F[\mathbb{E}[\vartheta|\theta]]$. Then, by the Variance Decomposition Formula, $K(F) \propto \mathbb{V}[\vartheta] - \mathbb{E}_F[\mathbb{V}[\vartheta|\theta]]$.

$$\operatorname{supp} F(k) = \begin{cases} \{\underline{\theta}, \overline{\theta}\}, & \text{if } k < \underline{k} \\ \{\omega(k), \overline{\theta}\}, & \text{if } \underline{k} \leq k < \overline{k} \\ \{\mu\}, & \text{if } k > \overline{k} \end{cases}$$

where ω is a strictly increasing, continuous function with $\omega(\underline{k}) = \underline{\theta}$. $F(k)$ is unique except, possibly, at $\overline{k}$. Moreover, $q_H^*(k) = q_H^{\mathrm{f}}(k) = \overline{\theta}$, for all $k < \overline{k}$ and $q_L^*(k) \leq q_L^s(k)$.

In words, $F(k)$ has a very simple form: it contains three regions depending on k. For low k, full information is acquired and the state is revealed. For high k, no information is acquired at all. For intermediate levels of costs, the information consists of one posterior that fully reveals the high state, $\overline{\theta}$, and of a low posterior that is partially informative. In that range, as k increases, the low posterior is increasing; that is, the precision of the low signal deteriorates monotonically. Although the specific form of this optimal structure is special, the existence of $\overline{k}$ such that no information is acquired for $k \geq \overline{k}$ holds for any information and production costs. Similarly, the existence of $\underline{k}$ such that the states are perfectly revealed for $k \leq \underline{k}$ is guaranteed by any BMC information costs.[12]

The high state $\overline{\theta}$ is fully revealed because of the marginal-incentives property, M. By M, quality levels must be fine-tuned not to give the agent incentives to make one of his posteriors more precise. However, $\theta_H = \overline{\theta}$ is as precise as a posterior can be, so the principal does not have to worry about that kind of deviation. Thus, the seller is free to provide quality to the high agent that is discontinuously larger than what she could offer if the high posterior were all but perfectly informative. She seizes that opportunity, providing the high type with his efficient quality. As k grows, maintaining full revelation at the top becomes costlier and the seller degrades the precision of the low posterior, increasing the quality of the product sold to the low type.

Because the high type receives efficient quality, all the distortions in this example come from the contract given to the low type. Figure 3 plots $q_L^*(k)$ and compares it with $q_L^s(k)$ and $q_L^{\mathrm{f}}(k)$. Because production costs are quadratic, $q_L^{\mathrm{f}}(k) = \omega(k)$. The distance between q_L^{f} and q_L^s is the standard screening distortion. It stems from asymmetric information exclusively; therefore, it vanishes for $k \geq \overline{k}$, when information is symmetric in equilibrium. The difference between q_L^s and q_L^* is the distortion given by information acquisition: it is a result of the agent's threat of acquiring information that is not prescribed by the principal. This difference does not disappear when information is symmetric: the distortion persists and only vanishes asymptotically, as k approaches infinity and the agent loses his ability to threaten the principal.

[12] These results are proved in Lemma OA1 and Lemma OA2 in Online Appendix A.

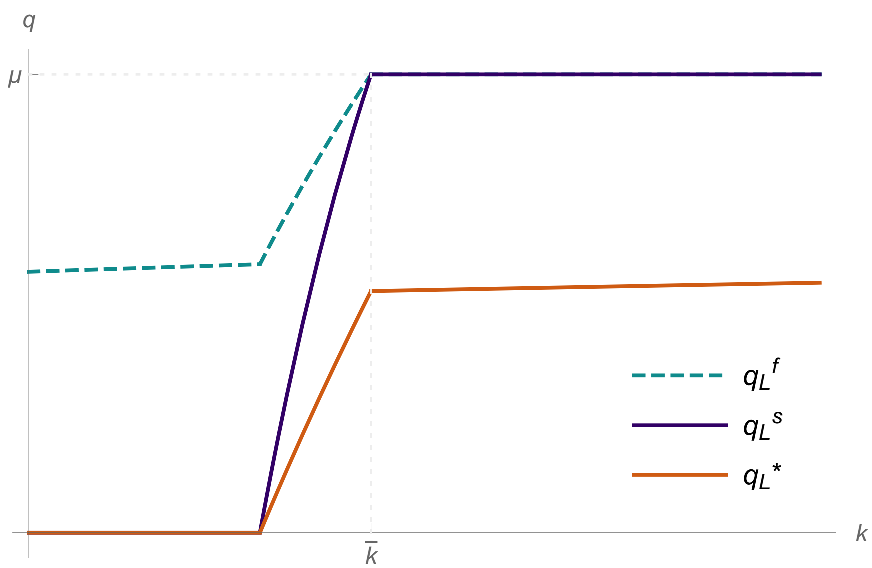


Figure 3: Low Quality Under Quadratic Costs

Notes: This figure plots the quality of the low type under quadratic costs as k varies, q_L^*. For each k, we use the information structure $F(k)$ to solve for the first- and second-best contracts q_L^f and q_L^s. For ease of visualization, we start the horizontal axis from $k > \underline{k}$.

3.3 Nonmonotonic Surpluses

We now turn to discussing how the level of acquisition costs, k, affects profits and consumer surplus. This impact is not clear from the outset, because several mechanisms are at play. First, information has a productive role in this model, helping agents match with contracts of appropriate quality, generating higher surplus. Thus, higher information costs, by constraining the production frontier, could have a negative effect on surpluses. Second, acquired information is costly and must be paid for. To the extent that these expenses vary with the level of information costs, this level affects the surpluses. Finally, costs affect the balance of power in the principal-agent relationship in two ways. One is direct, as k maps into the value to the buyer of deviating from the prescribed information strategy. Because, in equilibrium, he must be compensated for not deviating, k affects the division of surplus. The other is indirect: the level of costs affects which information is acquired and therefore affects information asymmetry. Proposition 4 below describes the end result of all these mechanisms on surpluses. The following assumption is relevant.

Assumption 1. *At least one of the following holds:*

1. *H is BMC;*
2. $(\overline{\theta} - \underline{\theta})^2 \geq (\overline{\theta} - \mu)\overline{\theta}$.

Assumption 1 describes a class of economies. It includes all bounded marginal costs of information. Condition 2 in the assumption does not depend on information costs. It requires, instead, that, under full information, the second-best contract excludes the low type, that is, $q_L^s = 0$. We define consumer surplus as the ex-ante net utility of the agent under the optimal information structure: $\mathbb{E}_F[U(\theta) - kH(\theta)]$.[13]

Proposition 4 (Nonmonotonicity)**.** *The principal's profits are U-shaped in k. The consumer's surplus is decreasing for sufficiently large k. Under Assumption 1, consumer's surplus increases for sufficiently small k.*

Proposition 4 shows that as the level of acquisition costs changes, profits respond non-monotonically. Under Assumption 1, consumer surplus is also non-monotonic and roughly in the opposite way. Profits are U-shaped in that the initial effect of information costs is to make the principal worse off. By contrast, when k is high enough, the principal benefits from an increase in acquisition costs. Under Assumption 1, the opposite holds for consumer surplus. The initial increase in costs improves their expected utility. However, when k becomes too large, a further increase in acquisition costs generates losses to the buyer. This latter effect does not depend on the assumption.

The intuition for this result relies on how the cost of learning affects the balance of power between the players through the threat of the agent. Changing k affects both the value and the credibility of the agent's deviations. When k is small and thus the prescribed information structure is particularly revealing, the most valuable threat for the agent is acquiring too little information. As k increases, that threat becomes more credible: the agent has fewer incentives to learn as learning gets costlier. This increase initially works in his favor and at the expense of the seller. By contrast, when k is very high and the prescribed information structure is extremely opaque, the most valuable threat is learning too much. However, as costs grow, that deviation becomes less credible. Then, further increases benefits the principal at the cost of the buyer.

This explanation oversimplifies the complex dynamics of acquisition costs. Indeed, all mechanisms mentioned earlier are at play. Nonetheless, the interaction between the level of costs and threats is the most prominent. The easiest way to see that is to consider again the case of quadratic costs. Recall that, in that case, the agent acquires a fully informative information structure and contracts have the second-best quality for $k \leq \underline{k}$. Over this interval, production is just as efficient as in the second-best, so the fact that profits are falling — and consumer surplus increasing — over that range is unrelated to the productive role of information. Additionally, even if the principal is reimbursed for acquisition costs, one can show profits are decreasing: the result is not driven by information expenses. A similar rationale works for $k \geq \overline{k}$. In both of these cases — for sufficiently low or high costs — the only force at play is the one emphasized in the previous paragraph. Assumption 1 guarantees that this intuition carries over to more general information cost functions.

In fact, the argument above shows that, under Assumption 1, surplus moves in the opposite direction

[13]The same comparative statics in Proposition 4 holds for gross utility $\mathbb{E}_F[U(\theta)]$.

of consumer surplus for extreme values of k. Consider again $k \leq \overline{k}$ — where contracts are the same as in the second-best, and consumers are fully-informed. In that case, because only information costs are increasing, but production is held fixed, it must be the case that total surplus in the market is reduced. On the other hand, when k is high, information is not acquired at all, but the optimal contract is distorted downwards compared to the second-best. As k increases from then, the only change is that the contract the principal offers approaches efficiency — increasing surplus.

Implications The non-monotonicity result has two main implications. First, it sheds light on firms' incentives to aid or dissuade consumer learning. In many settings, a seller can help or obstruct consumer learning by making testing, experimenting, or having access to valuable information easier or harder. For example, insurance and mobile phone provider websites often help customers find plans that best suit their needs. When should we expect sellers to hinder or help information acquisition? Our nonmonotonicity result can shed light on this issue. If the seller has a limited ability to affect the level of acquisition costs, then she would prefer to hamper consumer learning when those costs are sufficiently large, but to facilitate learning when costs are low. Here, obfuscation is profitable when the agent threatens to learn more than what the monopolist desires. For a given level of costs, preventing this threat would imply distorting production surplus and providing price discounts. On the other hand, by making acquisition harder, the seller discourages learning, achieving the same goal with smaller efficiency losses. When the agent threatens to learn less than the seller wants, facilitating learning has the same effect.

Second, the non-monotonicity of consumer surplus suggests policies that facilitate consumer learning may not benefit consumers. Transparency policies are relatively popular in markets for complex goods.[14] For example, New Hampshire provides the public with the HealthCost website, which is a price information tool for health-care costs (Brown, 2019). Importantly, this website allows consumers to compare out-of-pocket medical procedure costs across providers, taking into account personal information, as their insurance carrier and zip code. Similar tools are also available in many other states (Brown, 2019). The rationale for these policies seems to be that reducing information costs will help consumers make better decisions and, thus, increase their welfare. However, this rationale ignores equilibrium effects, which may reverse the intended effects of such intervention. In particular, under Assumption 1, the nonmonotonicity of consumer surplus implies that some level of opaqueness in the market is optimal for consumers.

[14] See Brancaccio et al. (2017) for an example in finance, Brown (2017, 2019) for health care, and Hackethal et al. (2012) for cellular plans.

4 Discussion

Timing We studied the contracting problem when the agent acquires information after the principal offers the menu. This assumption is realistic: in a variety of settings, the menu of available goods is fixed, and consumers can choose to acquire information after observing it, according to their own timelines. Examples are health care and online shopping, where the available goods and terms of trade are typically readily available and easily observable. In other common environments, however, information can be acquired before the menu is offered. In particular, when observing the terms of trade depends on an action by the buyer — for example, contacting a dealer for a financial-asset quote, or reaching out to a vendor to learn about prices —, information acquisition can happen before the seller's offer is fixed. This possibility can also hold when the producer is able to make exploding or timed offers. Roesler and Szentes (2017) and Ravid et al. (2022) study bilateral trade models in which information acquisition happens before and simultaneously to the design of the mechanism, respectively. Together, their work and this paper shed light on the importance of the timing assumption in determining the outcome of mechanism design under information acquisition.

Our model complements Ravid et al. (2022) by showing our timing assumption reverses their main takeaway. The key message in Ravid et al. (2022) is that, when information is acquired simultaneously to the design of the mechanism, the buyer may be substantially better off having access to free information than being able to purchase the same information at a low cost. When buyer and seller decide at the same time, the corresponding optimal mechanism fails to induce sufficient information acquisition even when costs are arbitrarily small. As the buyer foregoes some amount of information, the authors show the price at the optimal mechanism is higher than it would be if the buyer learned more thoroughly. As a consequence, both the consumer and the producer are worse off than in the full-information equilibrium that could arise when information is free.[15] As previously discussed, we obtain the opposite result: from the point of view of the consumer, low information-acquisition costs can be strictly better than being able to acquire information for free. The reason is that the principal, acting first, chooses to compensate the agent to acquire surplus-enhancing information when doing so is inexpensive. That compensation works in the buyer's favor and may increase consumer surplus.

Multiple States In the Online Appendix C, we extend the model to accommodate a finite number of valuations for the buyer. Under posterior-separability, we show that aggregate distortions are larger than the second-best ones, as is the case with two states. However, with multiple states, it no longer holds that

[15] In fact, they are worse off than under any equilibrium of the costless information economy.

all qualities are underprovided. With two valuations, any posterior-separable cost can be written as a linear function of F, the distribution of posterior means. Doing so when valuations are not binary is not possible. As a consequence, to keep linearity of the acquisition-cost function, we need to write information costs as a function of the whole distribution over posterior beliefs. The relevant Bayes consistency constraint is that the posteriors average out to the prior. The generalized model can be analysed following the same steps as the binary one: we show that conditions analogous to the marginal-incentives property and the threat-point property characterize implementability in that setting. We then transform the principal's problem into an information design problem with the threat point as an extra control variable. An important additional constraint arises: with multiple states, when the principal chooses a value for the threat point, she limits the set of posterior distributions it can generate. This observation, noted in Mensch (2022), implies that the information design problem the seller solves is constrained. By picking a threat point, the seller has to choose posteriors from a lower-dimensional set in the space of distributions.

Aggregate distortions remain below the second-best ones in that case, but underproduction is no longer guaranteed. The intuition for the negative result follows from the information design problem. With binary states, the seller chooses information structures unconstrained, so optimality requires the profit function to be concavified — implying the marginal profit of changing posteriors is equal at all points in the support.[16] Because the marginal profit of changing posteriors is proportional to the production wedge, $\theta - c'(q(\theta))$, concavification implies these wedges have the same sign: underproduction at the bottom begets underproduction at the top. With multiple states, because the information design problem is constrained, concavification is no longer required: optimality does not pin down a relationship between the wedges. Intuitively, the principal would like to distort all types downwards, but she may not be able to do so because she cannot choose information structures freely. One can then ask what restrictions on information costs are required to guarantee widespread downward distortions. Mensch and Ravid (2022) show that if information costs are linear in the distribution of posterior means, underprovision is guaranteed for all types of agents.

Posterior-Separability Finally, we assumed information costs to be posterior-separable (Caplin et al., 2022). This has three related implications. First, under posterior-separability one can restrict attention to binary information structures. This is a consequence of the linearity of posterior-separable costs with respect to the posterior distribution. In the absence of linearity, it is straightforward to construct information costs such that the information the consumer acquires includes more than two posteriors. Second, and more importantly, posterior-separable costs allow for a characterization of implementability as two proper-

[16] To be precise, with two states the information design problem is unconstrained up to a lower bound in the posterior mean.

ties: marginal incentives and the threat point. In our model, we leverage the tools from information design to guarantee that these properties characterize the constraint set of the principal's problem, which cannot be done for general costs. In particular, deviations to non-participation, which for posterior-separable costs are parameterized by a single posterior — the threat point — now depend on the whole deviating distribution of posteriors, compromising the tractability of this approach.

The third implication of posterior-separability is that, conditional on the threat point, the problem of the principal becomes a regular information design optimization. Because our argument for establishing underproduction at every point in the support relies on concavification, it cannot be readily extended for general cost functions. Nevertheless, in a specific case the result of underproduction generalizes: when the principal decides to induce no information acquisition. The key assumption, that is satisfied by posterior-separable costs, is that the consumer can move beliefs from the prior at zero marginal costs. To formalize this result, we restrict attention to binary information structures, and extend the model in the text for general cost functions. Concretely, if $\{\theta_L, \theta_H\} = \operatorname{supp} F$, we let $K(F) = h(\theta_L, \theta_H)$, where h is decreasing in θ_L and increasing in θ_H.

Assumption 2. *h is continuously differentiable and* $\frac{\partial h}{\partial \theta_L}(\mu, \mu) = \frac{\partial h}{\partial \theta_H}(\mu, \mu) = 0$

Proposition 5. *Under Assumption 2, if the principal wants to induce no information, quality is underprovided:* $c'(q^*) < \mu$.

The rationale for this result parallels the one for posterior-separable costs. In standard screening, if information is symmetric, the seller offers a single good to the buyer, with efficient quality for his valuation and prices are such that the consumer is indifferent between buying or not. Here, this offer does not work. If the agent is indifferent between buying and not buying when uninformed, he always has access to a cheap enough experiment that may tell him that his type is slightly lower than the ex-ante type. Therefore, to guarantee that the buyer acquires no information, the principal gives a price discount and optimally distorts quality accordingly. This suggests the force for underproduction is still present even when information costs are not posterior-separable. In particular, the result that distortions exist even under symmetric information does not depend on details of information costs.

Appendix: Auxiliary Results and Proofs

We start proving an auxiliary result. The information acquisition problem can be written in terms of rents as:

$$\begin{aligned} \max_{G\in\Delta(\Theta)} \quad & \mathbb{E}_G\left[U(\theta)-kH(\theta)\right] \\ \text{s.t.} \quad & \mathbb{E}_G[\theta]=\mu \end{aligned} \tag{1}$$

Lemma 3. *Problem 1 has a solution. $F\in\Delta(\Theta)$ solves it if and only if it satisfies BC and there exists $\psi\in\mathbb{R}$ such that:*[17]

$$\operatorname{supp} F \subseteq \arg\max_{v\in\Theta}\{U(v)-kH(v)-\psi v\} \tag{2}$$

Proof of Lemma 3

By strict convexity of c, it is without loss of generality to assume (q,t) is bounded. We start by proving existence of a solution. We then proceed to show necessity and sufficiency of condition 2.

Existence. Because contracts are bounded and Θ is compact, interim rents U are bounded by definition. Thus, $U-kH$ is a bounded function and the objective function is trivially continuous with respect to F in the weak topology. Additionally, because Θ is compact, $\Delta(\Theta)$ inherits compactness in the topology of weak convergence, by an application of Prokhorov's theorem. Finally, the set of F satisfying BC is closed under weak convergence, implying that the constraint set is a closed subset of a compact space, being itself compact. As a consequence, under the topology of weak convergence, problem 1 is one of maximizing a continuous function over a compact set and, therefore, has a solution.

Necessity. That BC is necessary is trivial, as it is a constraint in the problem. For 2, Start by defining the Lagrangian:

$$L(F,\psi)=\mathbb{E}_F\left[U(\theta)-kH(\theta)-\psi\theta\right]$$

Notice that, as $\mu\in(\underline{\theta},\overline{\theta})$, it is an interior point of the set $\{y\in\mathbb{R}:\mathbb{E}_F[\theta]=y \quad \text{for some} \quad F\in\Delta(\Theta)\}$. Given that, Luenberger (1997), Chapter 8, Problem 7 proves that if F solves 1, then there exists ψ such that $F\in\arg\max_{G\in\Delta(\Theta)} L(G,\psi)$. We now prove that this implies 2.

[17] This result is a consequence of the Lagrangian Lemma in Caplin et al. (2022). We adapt it to our framework and provide a short proof.

Define $\chi \equiv \arg\max_{\theta\in\Theta}\{U(\theta) - kH(\theta) - \psi\theta\}$. Assume, so as to find a contradiction, that $v \in \operatorname{supp} F$ exists such that $v \notin \chi$. It is immediate that F cannot put positive weight outside of χ. Then, assume v is a continuity point of F. That implies there is a neighborhood of v, N_1, such that $N_1 \in \operatorname{supp} F$. However, because $U - kH$ is continuous, there is another neighborhood of v, N_2, such that for all $x \in N_2$, $x \notin \chi$. Then, F puts positive weight on $N = N_1 \cap N_2$ with $N \cap \chi = \emptyset$, which is a contradiction. Thus, 2 is necessary.

Sufficiency. Assume F satisfies BC and 2 for ψ. Because it satisfies 2, it clearly maximizes $L(G,\psi) = \mathbb{E}_G[U(\theta) - kH(\theta) - \psi\theta]$. Define the auxiliary Lagrangian $\tilde{L}$ as $\tilde{L}(G,\lambda) = \mathbb{E}_G[U(\theta) - kH(\theta) - \lambda\cdot(1,-1)\theta]$, for $\lambda \in \mathbb{R}^2$.

Because F maximizes L, it must also maximize $\tilde{L}$ when $\lambda\cdot(1,-1) = \psi$. Take $\lambda > 0$ such that this is the case which is, of course, always possible. Luenberger (1997), Chapter 8.4, Theorem 1 shows that if F maximizes $\tilde{L}$ it also solves:

$$\begin{aligned} \max_{G\in\Delta(\Theta)} \quad & \mathbb{E}_G[U(\theta) - kH(\theta)] \\ \text{s.t.} \quad & \mathbb{E}_G[\theta] \leq \mathbb{E}_F[\theta] \\ & -\mathbb{E}_G[\theta] \leq -\mathbb{E}_F[\theta] \end{aligned}$$

Because F satisfies BC, $\mathbb{E}_F[\theta] = \mu$. Therefore, the problem above is equivalent to the acquisition problem.

■

Proof of Lemma 1

By the characterization of IC, U is differentiable almost everywhere, except at discontinuities of q and $U'(\theta) = q(\theta)$. If q is continuous at $\theta \in \operatorname{supp} F \cap (\underline{\theta}, \overline{\theta})$, then first order condition is necessary and implies:

$$q(\theta) - kH'(\theta) = \psi$$

We want to prove that, indeed, q is continuous in $\operatorname{supp} F \cap (\underline{\theta}, \overline{\theta})$, so the equality above holds. Assume, to obtain a contradiction, that this is not the case, so there is θ in that set such that q is discontinuous. By IC, q is monotonic, so we must have:

$$\lim_{z\uparrow\theta} q(z) < \lim_{z\downarrow\theta} q(z)$$

We start proving that θ is not a discontinuity point of F.

F **is not discontinuous at** θ**.** Assume that is the case. Then, pick a small $\varepsilon > 0$. Denote $\theta_+ \equiv \theta + \varepsilon$ and $\theta_- \equiv \theta - \varepsilon$. Define $\tilde{F}$ such that:

$$\tilde{F}(v) = \begin{cases} F(v) & \text{, if } v < \theta_- \\ F(v) + \frac{dF(\theta)}{2}) & \text{, if } v \in [\theta_-, \theta) \\ F(v) - \frac{dF(\theta)}{2}) & \text{, if } v \in [\theta, \theta_+) \\ F(v) & \text{, if } v \geq \theta_+ \end{cases}$$

$\tilde{F}$ clearly satisfies Bayesian consistency. Now consider:

$$\begin{aligned} &\mathbb{E}_{\tilde{F}}[U - kH] - \mathbb{E}_F[U - kH] = \\ \Big(U(\theta_-) - H(\theta_-)\Big)\frac{dF(\theta)}{2} + \Big(U(\theta_+) - kH(\theta_+)\Big)\frac{dF(\theta)}{2} - \Big(U(\theta) - kH(\theta)\Big)dF(\theta) \\ &= \left(\int_{\theta}^{\theta_+}(q(v) - kH'(v))dv - \int_{\theta_-}^{\theta}(q(v) - kH'(v)dv\right)\frac{dF(\theta)}{2} \\ &\geq \left(\lim_{v\downarrow\theta} q(v) - \lim_{v\uparrow\theta} q(v)\right)\varepsilon\frac{dF(\theta)}{2} + k(H'(\theta_-) - H'(\theta_+))\varepsilon\frac{dF(\theta)}{2} \geq 0 \end{aligned}$$

where the last inequality holds for small enough ε, as the term in the first parentheses is bounded away from zero, whereas the term in the second parentheses goes to zero as ε approaches zero. This implies that $\tilde{F}$ increases the value of the objective function of the agent, which is a contradiction with optimality of F, so θ cannot be a point of discontinuity of F.

F **is not continuous at** θ**.** If that was the case, there would be a neighborhood N of θ such that $N \subset \operatorname{supp} F$. Let $\varepsilon > 0$ and $\theta_- = \theta - \varepsilon$, $\theta_+ = \theta + \varepsilon$, such that $\theta_-, \theta_+ \in N$. Notice that ε can be taken so that q is continuous at both θ_- and θ_+ - as q is increasing, it has at most countable discontinuities. Continuity of q at $\theta_-, \theta_+ \in \operatorname{supp} F$ implies, as shown before:

$$q(\theta_+) - kH'(\theta_+) = \psi = q(\theta_-) - kH'(\theta_-)$$

which can be reorganized as:

$$q(\theta_+) - q(\theta_-) = kH'(\theta_+) - kH'(\theta_-)$$

By taking ε sufficiently small, the right hand side can be made as close to zero as one desires - as H' is continuous, whereas the right hand side is bounded below by $\lim_{v\downarrow\theta} q(v) - \lim_{v\uparrow\theta} q(v)$, providing the desired contradiction.

As a consequence of the last two paragraphs, we obtained a contradiction with q discontinuous in

$\operatorname{supp} F \cap (\underline{\theta}, \overline{\theta})$. Then, and if θ is in that set, $q(\theta) - H'(\theta) = \psi$. We finish the proof by showing the result for $\overline{\theta} \in \operatorname{supp} F$. The result for $\underline{\theta}$ is symmetric.

If $\overline{\theta} \in \operatorname{supp} F$. Assume $q(\overline{\theta}) - kH'(\overline{\theta}) < \psi$. Consider $\varepsilon > 0$:

$$\begin{aligned} U(\overline{\theta}) - kH(\overline{\theta}) - \psi\overline{\theta} - (U(\overline{\theta} - \varepsilon) - kH(\overline{\theta} - \varepsilon) - \psi(\overline{\theta} - \varepsilon)) = \\ \int_{\overline{\theta}-\varepsilon}^{\overline{\theta}} (q(z) - kH'(z))dz - \psi\varepsilon \\ \leq (q(\overline{\theta}) - KH'(\overline{\theta} - \epsilon) - \psi)\varepsilon \end{aligned}$$

Where the last inequality comes from monotonicty of q and convexity of H. By continuity of H' and the assumption that $q(\overline{\theta}) - H'(\overline{\theta}) < \psi$, for sufficiently small ε the last expression must become smaller than zero, finding a contradiction with 2. Therefore, $q(\overline{\theta}) - H'(\overline{\theta}) \geq \psi$. A similar argument establishes the result for $\underline{\theta}$, so necessity of M is concluded. ∎

Proof of Lemma 2

Throughout this proof, we say that a menu is feasible for the principal if it satisfies IC and IA. Notice that 2 implies:

$$U(\theta) - kH(\theta) - \psi\theta \geq -kH(\theta_o(\psi)) - \psi\theta_o(\psi) \quad \text{for all} \quad \theta \in \operatorname{supp} F \tag{3}$$

Fix F and let a feasible (U, q) and ψ satisfy 3 with strict inequality. Consider the alternative menu $(\tilde{U}, \tilde{q})$ with $\tilde{U}(\theta) = \max\{U(\theta) - \varepsilon, 0\}$, for some $\varepsilon > 0$ such that $(\tilde{U}, q)$ still satisfies 3, and $\tilde{q}(\theta) = q(\theta)\mathbb{1}_{\tilde{U}(\theta)>0}$. Such an ε exists, because we assumed 3 held with strict inequality. Next, we show this menu is feasible. We finish the proof by showing that it is also more profitable.

Feasibility. First, it satisfies IA. To see that, recall by Lemma 3 that IA is equivalent to 2. Then, because (U, q) is feasible, 2 holds for it:

$$\operatorname{supp} F \subseteq \arg\max_{v \in \Theta} \{U(v) - kH(v) - \psi v\}$$

However, $\tilde{U}(v) - kH(v) - \psi v = \max\{U(v) - kH(v) - \psi v - \epsilon, -kH(v) - \psi v\}$, which is a monotonic transformation of $U(v) - kH(v) - \psi v$. Thus:

$$\arg\max_{v \in \Theta} \{U(v) - kH(v) - \psi v\} = \arg\max_{v \in \Theta} \{\tilde{U}(v) - kH(v) - \psi v\}$$

and IA holds for the new menu $(\tilde{U}, q)$.

For IC, notice that $\tilde{q}$ is increasing and satisfies individual rationality by definition. Finally:

$$\tilde{U}(\theta) = \max\{0, U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} q(v)dv - \varepsilon\} = \max\{0, U(\underline{\theta}) - \varepsilon\} + \int_{\underline{\theta}}^{\theta} q(v)\mathbb{1}_{\tilde{U}(v)>0}dv$$

so the envelope condition also holds and $(\tilde{U}, \tilde{q})$ satisfy IC. Thus, it is feasible.

Profitability. We finally show $(\tilde{U}, \tilde{q})$ is more profitable than (U, q). Notice that $\tilde{U} < U$. We prove that, in the support of F, $q = \tilde{q}$. To see that, take $\theta \in \operatorname{supp} F$. Because $(\tilde{U}, \tilde{q})$ is feasible, we have, by reordering 3:

$$\tilde{U}(\theta) \geq kH(\theta) - kH(\theta_o(\psi)) + \psi(\theta - \theta_o(\psi))$$
$$\geq kH(\theta) - kH(\theta_o(\psi)) - kH'(\theta_o(\psi))(\theta - \theta_o(\psi)) > 0$$

where the first inequality comes from reordering 3, the second from the definition of $\theta_o(\psi)$ and the third from strict convexity of H. Then, $\tilde{U}(\theta) > 0$, implying $\tilde{q}(\theta) = q(\theta)$. Because $\tilde{U} < U$ and, in the support of F, $\tilde{q} = q$, $(\tilde{U}, \tilde{q})$ must be more profitable than (U, q). Thus, (U, q) cannot be optimal, because $(\tilde{U}, \tilde{q})$ is also feasible. Therefore, for any optimal menu, 3 must hold with equality: that is, TP holds, as we wanted to prove.

■

Lemma 4 (Binary is Sufficient). *Assume the principal's problem has a solution. Then, there are $F, (U, q)$ solving it such that $|\operatorname{supp} F| \leq 2$ and $|q(\Theta)/\{0\}| \leq 2$.*

Proof of Lemma 4

By Lemma 3, the problem of the principal can be written as:

$$\max_{G,(U,q),\psi} \{\mathbb{E}_G[\theta q(\theta) - c(q(\theta)) - U(\theta)] : IC,\ BC \quad \text{and} \quad 2\}$$

where we have rewritten profits in the usual surplus minus rents format. Fix any solution to this problem, $(U, q), F, \psi$. First, we show it is sufficient to focus on binary distributions. Then we show we can restrict the menu accordingly.

Binary distribution. For fixed $(U, q), \psi$, we show the principal is at least as well off with a binary distribution. Consider the simplex $\Delta(\operatorname{supp} F)$. By Winkler (1988), the problem of maximizing profits in this simplex subject to BC is solved by an at-most binary distribution, call it $\tilde{F}$. Because $\tilde{F}$ solves this problem, it is at least as profitable as F for the principal. We now show it is feasible under the principal's problem.

Notice that IC does not depend on F, so $(U,q),\psi,\tilde{F}$ satisfy IC. Additionally, because $\operatorname{supp}\tilde{F} \subset \operatorname{supp} F$, then $(U,q),\psi,\tilde{F}$ satisfy 2. Thus, $\tilde{F}$ is feasible. That implies that focusing on at-most-binary distributions is sufficient.

Binary menus. Now start with a solution $(U,q),\psi,F$ with F at-most-binary. Define the following alternative menu:

$$\tilde{U}(\theta) = \max\left\{0, \max_{v \in \operatorname{supp} F}\{U(v) + (\theta - v)q(v)\}\right\}$$

Let $v(\theta) \in \operatorname{supp} F$ be the largest argmax of the maximization problem in the parentheses above. We define $\tilde{q}(\theta) = q(v(\theta))\mathbb{1}_{\tilde{U}(\theta)>0}$. Notice that $(\tilde{U},\tilde{q})$ was constructed to satisfy IC. It is also easy to see it coincides with (U,q) in $\operatorname{supp} F$, so it satisfies 2. Now, this new menu produces exactly the same profits as (U,q), but notice that $q(\Theta) \subset q(\operatorname{supp} F) \cup \{0\}$, proving the result, since $\operatorname{supp} F$ is at-most-binary.

■

Proof of Proposition 1

Recall the problem of the principal can be written as:

$$\max_{G,(U,q)} \{\mathbb{E}_G[\theta q(\theta) - c(q(\theta)) - U(\theta)] : IC \quad \text{and} \quad IA\} \tag{4}$$

where we have rewritten profits in the usual surplus minus rents format. We call the value of this problem L^*. Let the value of problem P be L^P. We start showing that $L^* \leq L^P$. Then, we prove that the converse holds by constructing a transformation that allows us to express the solution to one of the problems as a solution to the other.

First, $L^* \leq L^P$. Take a solution of 4, $\{F,(U,q)\}$. By Lemma 4, assume without loss of generality that F is at-most-binary and $|q(\Theta)/\{0\}| \leq 2$. Because the solution must satisfy IA, by Lemma 3, it also satisfies 2 for some ψ, and F is Bayesian consistent. Additionally, by Lemma 1, IA and IC imply M, and by Lemma 2, optimality in the principal's problem 4 implies TP.

Just as in the text, identify $\operatorname{supp} F = \{\theta_L, \theta_H\}$, with $\theta_L \leq \mu \leq \theta_H$, and then define $q_i \equiv q(\theta_i)$ and $U_i \equiv U(\theta_i)$, $i \in \{L,H\}$. Then, $\left\{F, \{U_i, q_i\}_{i\in\{L,H\}}, \psi\right\}$ is feasible in P, because TP and M depend only on the points in the support of F. Additionally, it is easy to see that profits under $\left\{F, \{U_i, q_i\}_{i\in\{L,H\}}, \psi\right\}$ in P are the same as L^*. Because L^P is the maximum profits obtained at P, $L^* \leq L^P$.

Next, we prove that $L^P \leq L^*$. Take a feasible element of P, $\left\{F, \{U_i, q_i\}_{i\in\{L,H\}}, \psi\right\}$. We show we can construct from that a feasible element of 4, $\{F,(U,q)\}$ that preserves the profits of the principal. The proof proceeds in two steps.

Step 1: Extension to a menu that satisfies IC. For $\theta_i \in \operatorname{supp} F$, Let $q(\theta_i) \equiv q_i$ and $U(\theta_i) \equiv U_i$, so these functions are defined in $\operatorname{supp} F$. We now extend them. Define, for $\theta \notin \operatorname{supp} F$:

$$U(\theta) = \max\{0, \max_{v \in \operatorname{supp} F}\{U(v) + (\theta - v)q(v)\}\}$$

Let $v(\theta)$ be the largest argmax of the maximization problem in the parentheses above. We define $\tilde{q}(\theta) = q(v(\theta))\mathbb{1}_{\tilde{U}(\theta)>0}$. (U,q) is clearly individually rational. We prove that is it also incentive compatible. Start with $\theta, \theta' \in \operatorname{supp} F$. We have:

$$\begin{aligned} U(\theta) - U(\theta') - (\theta - \theta')q(\theta') &= \\ U(\theta') + kH(\theta) - kH(\theta') + \psi(\theta - \theta') - U(\theta') - (\theta - \theta')q(\theta') &\geq \\ kH(\theta) - kH(\theta') + \psi(\theta - \theta') - (\theta - \theta')(\psi + kH'(\theta')) &\geq 0 \end{aligned}$$

where the first equality is a consequence of TP, the first inequality cancels repeated terms and applies M, and the second inequality cancel the new repeated terms and uses convexity of H. This implies incentive compatibility holds in $\operatorname{supp} F$. It is easy to see that U was constructed so that it satisfied incentive compatibility outside of the support, so (U,q) satisfy IC, finishing step 1.

Step 2. $F,(U,q)$ **satisfy IA** We start by proving that $g(\theta) \equiv U(\theta) - kH(\theta) - \psi\theta$ is concave by parts. Notice that, $q(\Theta) \subset \{q_o \equiv 0, q_L, q_H\}$. q is non-decreasing, because (U,q) satisfies IC, so we can define the intervals $I_i = \{\theta \in \Theta : q(\theta) = q_i\}$, for $i \in \{o, L, H\}$. Now, $g|_{I_i}$ is differentiable and we have:

$$g'|_{I_i}(\theta) = q_i - H'(\theta) - \psi$$

which is decreasing, so $g|_{I_i}$ is concave for each $i \in \{o, L, H\}$. Then, by M and the definition of $\theta_o \equiv \theta_o(\psi)$, for each i, $\theta_i \in \arg\max_{v \in I_i}\{g(v)\}$. That implies:

$$\arg\max_{v \in \Theta}\{g(v)\} \subseteq \max_{i \in \{o,L,H\}}\{g(\theta_i)\}$$

But by TP, $g(\theta_i)$ is a constant across $i's$. Because $\operatorname{supp} F = \{\theta_L, \theta_H\}$ we have:

$$\operatorname{supp} F \subseteq \arg\max_{v \in \Theta} g(\theta)$$

which is exactly 2 which, by Lemma 3 implies IA, when summed with the fact that F satisfies BC.

We proved that $\{F,(U,q)\}$ is feasible under 4. But because the menu coincides with $\{U_i,q_i\}_{i\in\{L,H\}}$ in the support of F, it obtains the same profit in 4 as in P. Because this was done for an arbitrary feasible element of P we have: $L^P \leq L^*$.

We have then established $L^P = L^*$ and constructed a mapping between solutions to the problems, proving that they are equivalent.

Existence of Solution. As previously argued, P is an optimization in 7 variables. Recall that we identify F with its support: $\{\theta_L,\theta_H\}$, $\theta_L \leq \mu \leq \theta_H$. Notice that $p_L^F = 1-p_H^F$, and $p_H^F = \frac{\mu-\theta_L}{\theta_H-\theta_L}$, for $\theta_H > \theta_L$. Given M, For $\theta_H = \theta_L = \mu$, we have $q_H = q_L$ and $U_H = U_L$. Because of that, p_H^F is immaterial in that case, so we define $p_H^F = 0$. By inspection, the restriction of the objective function to the constraint set is continuous in $\left\{\{\theta_i,U_i,q_i\}_{i\in\{L,H\}},\psi\right\}$. We proceed to prove the constraint set is compact.

M, TP and BC are clearly closed. As previously argued, q_i,U_i are bounded, and $\theta_i \in \Theta$ are also bounded, for $i\in\{L,H\}$. Then, we just need to prove that ψ is bounded. By M, we have, for $\theta_L < \theta_H$:

$$q_H - q_L \geq kH'(\theta_H) - kH'(\theta_L)$$

Notice that this implies $H'(\theta_i)$, $i\in\{L,H\}$ is uniformly bounded. To see that, assume there is a sequence of feasible choices, with θ_i^n, and $H'(\theta_H^n) \geq n$ unbounded. Then, because q_i are bounded, we have that $H'(\theta_L^n)$ also grows unboundedly, so that the right hand side of the inequality above remains bounded. But this is impossible because $\theta_L^n \leq \mu$ implies $kH'(\theta_L^n) \leq kH'(\mu) < \infty$. A similar argument holds if we assume θ_L^n is unbounded below. Then, using M again we obtain:

$$q_L - kH'(\theta_L) \leq \psi \leq q_H - kH'(\theta_H)$$

which guarantees that ψ is bounded. Thus, P is a problem of maximizing a continuous function over a compact set and, therefore, it has a solution.

■

Proof of Proposition 2

We start by constructing a Lagrangian for the principal's problem. For that, we add two auxiliary variables to the problem. Write

$$q_i = kH'(\theta_i) + \psi + a_i$$

By M, $a_L \leq 0$, with equality for all $\theta_L > \underline{\theta}$ and $a_H \geq 0$, with equality for all $\theta_H < \overline{\theta}$. With that definition, we can eliminate M and plug q in the profit function. Additionally, recall that $\theta_o(\psi) = \arg\max_{v\in\Theta}\{-kH(v) - \psi v\} \leq \theta_L$, as $q_L \geq 0$. Using that, we solve TP for U_i to obtain:

$$U_i = kH(\theta_i) + \psi\theta + \max_{v\in\Theta}\{-kH(v) - \psi v\}$$

This then allows us to eliminate the constraint TP and to, again, rewrite the profit function plugging this equation for U_i. In order for the problems to be equivalent, then, we need to impose two types of constraints on a_i. In particular, $a_L \leq 0$, $a_H \geq 0$, to which we associate multipliers α_i, and $a_L(\theta_L - \underline{\theta}) = 0$ and $a_H(\overline{\theta} - \theta_H) = 0$, to which we associate β_i. For shortness, we encode $\vartheta_L = \underline{\theta}$ and $\vartheta_H = \overline{\theta}$. Finally, recall that $q_i \geq 0$ is also a constraint, that we associate with multiplier y_i. We can then write the following Lagrangian for P, plugging in the equation above for U_i:

$$\begin{aligned}\mathcal{L}(\{\theta_i, a_i, \alpha_i, \beta_i, y_i\}_{i\in\{L,H\}}, \psi) = \sum_{i\in\{L,H\}} p_i^F \Big\{ &\theta_i\,(kH'(\theta_i) + \psi + a_i) - c\,(kH'(\theta_i) + \psi + a_i) \\ &- \Big(kH(\theta_i) + \psi\theta_i + \max_{v\in\Theta}\{-kH(v) - \psi v\}\Big) \\ &- \alpha_i a_i - \beta_i a_i(\theta_i - \vartheta_i) - y_i\,(kH'(\theta_i) + \psi + a_i)\Big\}\end{aligned} \tag{5}$$

Notice that, by M, $q_H \geq q_L$, so $y_H = 0$, and we omit it whenever convenient. Start with first order conditions for ψ:

$$[\psi]: \quad \mu - \theta_o(\psi) = \sum_{i\in\{L,H\}} p_i^F [\theta_i - c'(q_i) - y_i] \tag{6}$$

Then, for a_i:

$$[a_L]: \quad \theta_L - c'(q_L) \geq \alpha_L + \beta_L(\theta_L - \underline{\theta}) + y_L \tag{7}$$

with equality if $a_L < 0$, and:

$$[a_H]: \quad \theta_H - c'(q_H) \leq \alpha_H + \beta_H(\theta_H - \overline{\theta}) \tag{8}$$

with equality if $a_H > 0$.

We prove, from now, that q_L is underprovided.

q_L **is underprovided —** $a_L = 0$**.** If $q_L = 0$, this is obvious. So we prove it for $q_L > 0$ — thus, $y_L = 0$. First, if $\theta_L < \theta_H$, equation 6 and strict convexity of c imply:

$$c'(q_L) < \sum_i p_i^F c'(q_i) = \theta_o(\psi) \leq \theta_L$$

By contrast, if $\theta_L = \mu$, by 6:

$$c'(kH'(\mu) + \psi) = \theta_o(\psi)$$

By a simple application of Topkis' lemma, $\theta_o(\psi)$ is nonincreasing with ψ. Define: $g(\psi) = c'(kH'(\mu)+\psi) - \theta_o(\psi)$, which is then strictly increasing with ψ. Additionally, $g(-kH'(\mu)) = -\mu < 0$, and g is unbounded — because $\theta_o(\psi) \leq \theta_L$. Thus, there is a unique $\psi > -kH'(\mu)$ with $g(\psi) = 0$. Because $\psi > -kH'(\mu)$, $\theta_o(\psi) < \mu$ and we have:

$$c'(q) = \theta_o(\psi) < \mu$$

as we wanted to prove. We finish by arguing that this implies $a_L = 0$. Assume $a_L < 0$. Because we just proved that q_L is underprovided, we can increase a_L, which would increase q_L, improving profits, while not violating any constraint. Then, it must be that $a_L = 0$.

First Order Conditions for θ_i**.** Now, define $L_i \equiv \theta_i q_i - c(q_i) - (kH(\theta_i) + \psi\theta + \max_{v\in\Theta}\{-kH(v) - \psi v\})$. Also, let $p^F \equiv p_H^F$. For $\theta_H > \theta_L$:

$$\frac{dp^F}{d\theta_H} = -\frac{p^F}{\theta_H - \theta_L} \qquad \text{and} \qquad \frac{dp^F}{d\theta_L} = -\frac{(1-p^F)}{\theta_H - \theta_L}$$

We start by focusing on the case in which $\theta_i \neq \mu$, $i \in \{L, H\}$, so $\theta_H > \theta_L$. Take first order conditions of $\mathcal{L}$ with respect to θ_i to obtain:

$$[\theta_H]: \qquad -\frac{p^F}{\theta_H - \theta_L}(L_H - L_L) + p^F\left(a_H + (\theta_H - c'(q_H))\,kH''(\theta_H)\right) - p^F\beta_H a_H \geq 0 \tag{9}$$

with equality if $\theta_H < \overline{\theta}$, and:

$$[\theta_L]: \qquad -\frac{(1-p^F)}{\theta_H - \theta_L}(L_H - L_L) + (1-p^F)\left(\theta_L - c'(q_L) - y_L\right)kH''(\theta_L) \leq 0 \tag{10}$$

with equality if $\theta_L > \underline{\theta}$.

We now prove the main results of the proposition.

High Quality is underprovided. From the first order conditions for θ_i:

$$(1-\beta_H)a_H + (\theta_H - c'(q_H))kH''(\theta_H) \geq (\theta_L - c'(q_L) - y_L)kH''(\theta_L) \tag{11}$$

with equality if $\underline{\theta} < \theta_L < \theta_H < \overline{\theta}$.

We proceed by analysing two cases. Start with $a_H = 0$. Then, the inequality above becomes:

$$(\theta_L - c'(q_L) - y_L)kH''(\theta_L) \leq (\theta_H - c'(q_H))kH''(\theta_H)$$

Because the left hand side of 6 is positive and H is strictly convex, we have that at least one of the sides of inequality above is positive. Thus $c'(q_H) < \theta_H$.

Now, assume $a_H > 0$. In this case, notice that it must be $\theta_H = \overline{\theta}$ and $\alpha_H = 0$. Then, by 8, $\theta_H = c'(q_H)$, and we are done.

Aggregate distortions. We can rewrite 6 as:

$$\sum_{i\in\{L,H\}} p_i^F c'(q_i) = \theta_o(\psi) - p_L^F y_L \leq \max\{\theta_L, p_H^F \theta_H\} = \sum_{i\in\{L,H\}} p_i^F c'(q_i^s)$$

where the first inequality comes from $q_L \geq 0$ and the definition of $\theta_o(\psi)$, and the last equality by the known pure screening solution. To see that the inequality in the middle holds, consider the following two cases. First, if $q_L > 0$, $y_L = 0$ and the inequality is true by $\theta_o(\psi) \leq \theta_L$. Now, if $y_L \neq 0$, we have, by 6:

$$\tau(F,q) = p_H^F c'(q_H) \leq p_H^F \theta_H$$

using the result that q_H is underprovided. That proves the aggregate distortions result. We next prove these two results when $\theta_i = \mu$ for some i.

No Acquisition: $\operatorname{supp} F = \{\mu\}$**.** We proved before that:

$$c'(q) = \theta_o(\psi) < \mu$$

This shows that both the underprovision and the aggregate distortion results hold strictly in this case.

k large enough. Recall that, q_i, $i \in \{L,H\}$ is bounded — uniformly on k by strict convexity of c. Just as in the proof of existence in Proposition 1, this implies that $\{kH'(\theta_i)\}_{i\in\{L,H\}}$ is uniformly bounded on k. Thus, for large enough k, $kH'(\theta_H) < kH'(\overline{\theta})$, as the term in the right grows unbounded. Thus, because we proved that $c'(q_H) = \theta_H$ only when $\theta_H = \overline{\theta}$, we have that, for high enough k, $c'(q_H) < \theta_H$.

Notice that, with the same argument as above, we can conclude that, for high enough k, $\theta_L, \theta_H \in (\underline{\theta}, \overline{\theta})$. Recall that 6 can be written as

$$\sum_i p_i^F c'(kH'(\theta_i) + \psi) = \theta_o(\psi) - p_L^F y_L$$

If $q_L = 0$, because $\overline{\theta} > \theta_H$, $c'(q_H) < \theta_H$ so $\sum_i p_i^F c'(kH'(\theta_i)+\psi) < p_H^F \theta_H$. If, on the other hand, $q_L > 0$, $y_L = 0$ and because $\theta_L > \underline{\theta}$, we have $\theta_L > \theta_o(\psi)$. Applying both of these arguments to the equation above, we get:

$$\sum_i p_i^F c'(q_i) < \max\{\theta_L, p_H^F \theta_H\} = \sum_i p_i^F c'(q_i^s)$$

H is UMC. If that is the case, then for $k > 0$ it is clear that $\theta_i \in (\underline{\theta}, \overline{\theta})$ for $i \in \{L, H\}$. Thus the same argument holds as for when k is large enough.

■

Proof of Proposition 4

Let $Q(k)$ be the value function of P at k. We can apply the envelope theorem to 5 to conclude that the derivative of the profit function satisfies:

$$Q'(k) = -\sum_{i \in \{L,H\}} p_i^F \left\{ H(\theta_i^k) - H(\theta_o(\psi^k)) - (\theta_i^k - c'(q_i^k) - y_i^k) H'(\theta_i^k) \right\} \tag{12}$$

where superscript k denotes that the variable solves P for k. We start the proof showing that profits are decreasing for small k and increasing for large k. We finally prove that profits are quasiconvex, which established the U-shape. Then we move on to consumer surplus.

Profits decrease for small k. When $k = 0$, the solution to P is full information and the optimal contract is the second-best contract for that information: $\operatorname{supp} F^0 = \operatorname{supp} \overline{F} = \{\underline{\theta}, \overline{\theta}\}$, $q^0 = q^s$. Notice that this implies $\theta_o(\psi^0) = \theta_L$. We know $c'(q_H^0) = \overline{\theta}$. Applying this in 6, we have

$$\theta_L - c'(q_L^0) - y_L = \frac{p_H^{\overline{F}}}{p_L^{\overline{F}}} (\overline{\theta} - \underline{\theta})$$

Plugging these into 12:

$$Q'(0) = -p_H^F \{H(\overline{\theta}) - H(\underline{\theta}) - H'(\underline{\theta})(\overline{\theta} - \underline{\theta})\} < 0$$

where the inequality is due to strong convexity of H. Then, we obtain the intended result.

Profits increase for large k. In Lemma OA 1, we prove that there is $\overline{k}$ such that, for $k \geq \overline{k}$ $\operatorname{supp} F^k = \{\mu\}$. It is direct that $\theta_o(\psi^k) < \mu$ for any finite k. Thus, using 6:

$$Q'(k) = -\{H(\mu) - H(\theta_o(\psi^k)) - H'(\mu)(\mu - \theta_o(\psi^k)\} > 0$$

for $k \geq \overline{k}$, where we used, again, strict convexity of H.

Profits are quasiconvex. For $k > 0$, define an auxiliary variable $\tilde{\psi} = \frac{\psi}{k}$. With that, we can rewrite TP in a different form:

$$U_i = k\{H(\theta_i) + \tilde{\psi} + \max_{v \in \Theta}\{-H(v) - \tilde{\psi} v\}\}, \quad \text{for all } \; i \in \{L, H\} \tag{13}$$

Additionally, we add variables a_i to write M as

$$q_i - kH'(\theta_i) - \psi - a_i \mathbb{1}_{\theta_i = \vartheta_i} = 0, \quad \text{for all } \; i \in \{L, H\} \tag{14}$$

By inspection, it should be clear that the following problem is equivalent to P:

$$\max_{F, \{U_i, q_i, a_i\}_{i \in \{L,H\}}, \tilde{\psi}} \Big\{ \sum_{i \in \{L,H\}} p_i^F \left[\theta_i q_i - c(q_i) - U_i\right] : 14, \;\; 13, \;\; \mathbb{1}_{a_L > 0} \leq 0, \mathbb{1}_{a_H < 0} \leq 0 \;\; \text{ and } \;\; BC \Big\}$$

Here is the rationale for this new version of the problem: we added two auxiliary variables that only have a role when a state is fully revealed. These variables make the former inequalities in M into equations, at a cost of an additional constraint. The constraints are such that $a_L \leq 0 \leq a_H$, but we encode them into slightly more cumbersome notation for a reason that becomes apparent in the next paragraph. Additionally, we use the fact that ψ is not profit-relevant, to do a change of variables.

For brevity, define $X \equiv \{F, \{U_i, q_i, a_i\}_{i \in \{L,H\}}, \tilde{\psi}\}$. Notice that we can write the constraint set as a vector inequality: $g(X, k) \leq 0$. g is clearly continuous in k. It is also concave in k, as all constraints are affine in k. The addition of $\{a_i\}$ and of the inequality constraints in the above form guarantee that there is always a solution X^* for this reformulated problem such that $g(X^*, p) = 0$. To see that, notice that choosing $a_i = 0$ if $\theta_i \notin \{\underline{\theta}, \overline{\theta}\}$ gives equality in all constraints in that case. In the opposite case, equalities are guaranteed by definition.

We can then apply Theorem 3.1, part (b) in Xu (2001) to obtain that the value function of the problem above is quasiconvex.

Consumer Surplus increases for small k. First, at $k = 0$ there is full information. If $q_L^s = 0$ for full information, we have that consumer surplus is zero at $k = 0$. Because consumer surplus is always positive at positive k, we have our result. By usual arguments, $q_L^s = 0$ if and only if

$$\underline{\theta} - \frac{p^{\overline{F}}}{1 - p^{\overline{F}}}(\overline{\theta} - \underline{\theta}) \geq 0$$

which can be rewritten as a function of the prior mean to obtain condition 1 in Assumption 1.

Henceforth, we prove the result holds for $q_L^s > 0$. Under Assumption 1, H is BMC. In this case, by Lemma OA 2 in the Online Appendix, there is always a $\underline{k} > 0$ such that $\operatorname{supp} \overline{F} = \operatorname{supp} F^k = \{\underline{\theta}, \overline{\theta}\}$ and $q^k = q^s$ for that distribution, for all $k \leq \underline{k}$. Let $W(k)$ and $CS(k)$ denote the welfare and consumer surplus obtained at the optimal solution, respectively. Then, notice that because the distribution and optimal quality do not change in this interval, welfare changes only to the extent that acquisition costs increase. We can then calculate the derivative of consumer surplus at zero using the fact that profits, Q are welfare minus consumer surplus:

$$\begin{aligned} CS'(0) = W'(0) - Q'(0) &= -\sum_{i \in \{L,H\}} p_i^{\overline{F}}[H(\theta_i)] + p_H^{\overline{F}}\{H(\overline{\theta}) - H(\underline{\theta}) - H'(\underline{\theta})(\overline{\theta} - \underline{\theta})\} \\ &= -H(\underline{\theta}) - p_H^{\overline{F}} H'(\underline{\theta})(\overline{\theta} - \underline{\theta}) > 0 \end{aligned}$$

where the last inequality comes from strong convexity of H and the fact that $H(\mu) = 0$. Moreover, when H is BMC, the derivative of profits and welfare is continuous for $k \leq \overline{k}$, which shows that consumers' surplus is increasing for sufficiently small k.

Consumer Surplus decreases for high k We know from Lemma OA 1 in the Online Appendix that there is $\overline{k} > 0$ such that $\operatorname{supp} F^k = \{\mu\}$ for all $k \geq \overline{k}$. Additionally, for k sufficiently high, we have $\theta_o^k \equiv \theta_o(\psi^k) > \underline{\theta}$. Then, by 6.

$$c'\left(kH'(\mu) - kH'(\theta_o^k)\right) = \theta_o^k$$

Differentiation then provides:

$$\frac{d\theta_o^k}{dk} = \frac{c''(q^k)\frac{q^k}{k}}{1 + c''(q^k)kH''(\theta_o^k)}$$

We can apply that in the derivative of $CS(k)$ to obtain:

$$CS'(k) = \frac{CS}{k}(k) - \frac{H''(\theta_o^k)c''(q^k)q^k}{1 + c''(q^k)kH''(\theta_o^k)}(\mu - \theta_o^k)$$

By the mean value theorem, there is $m_k \in [\theta_o^k, \mu]$ such that $\frac{CS}{k}(k) = H''(m_k)\frac{(\mu-\theta_o^k)^2}{2}$. We then have that the derivative above can be rewritten as:

$$CS' = \left(H''(m_k)\frac{(\mu-\theta_o^k)}{2} - \frac{H''(\theta_o^k)c''(q^k)q^k}{1+c''(q^k)kH''(\theta_o^k)}\right)(\mu-\theta_o^k)$$

The first term in parentheses converges to zero, but not the second, which means that as k is large enough the whole derivative is negative.

■

Proof of Proposition 5

If the principal wants to induce no information acquisition, she offers a single contract $\{q,t\}$ to be accepted by the ex-ante agent. By incentive compatibility, there exists a cutoff type, θ_u such that $\mu \geq \theta_u$ such that the expected payoff of any agent with interim posterior $\theta > \theta_u$ is $q(\theta-\theta_u)$. Because the agent can only choose between this contract or no contract, he will choose to sign the contract if he receives the high signal, θ_H, and to opt-out otherwise — in other words, $\theta_L \leq \theta_u$. Thus, his information acquisition problem is:

$$\max_{\theta_L \leq \theta_u \leq \mu \leq \theta_H} \frac{\mu-\theta_L}{\theta_H-\theta_L} q(\theta_H-\theta_u) - \kappa(\theta_L,\theta_H), \tag{IA'}$$

where $\frac{\mu-\theta_L}{\theta_H-\theta_L}$ is the probability of the high signal. If acquiring no information is optimal, we abuse notation to say μ solves IA'. The problem of the principal is then:

$$\max_{q,\theta_u} \{\mu q - c(q) - q(\mu-\theta_u) : \mu \text{ solves } IA'\}$$

We first show that $\theta_u < \mu$. For a contradiction, assume $\theta_u = \mu$. In that case, consider a deviation from an uninformative information structure $\{\mu, \mu+\delta\}$ to $\{\mu-\varepsilon, \mu+\delta\}$. For sufficiently small ε, the change in agent's utility is approximately the derivative of the objective function in IA' evaluated at $\{\mu,\mu+\delta\}$:

$$\left(q - \frac{\partial \kappa}{\partial \theta_L}(\mu,\mu+\delta)\right)\varepsilon$$

By making δ close to zero, the second term in the parentheses converges to zero by 2, whereas the first term is positive. Thus, the agent can benefit from acquiring a small amount of information. This concludes that for no information to be optimal for the consumer, $\theta_u < \mu$.

We conclude by showing that this implies underprovision. Indeed, assume for a contradiction that $c'(q^*) \geq \mu$. Notice that the objective function in the principal's problem, for some θ_u is:

$$\theta_u q - c(q).$$

Thus, reducing q^* clearly increases the objective function, since $\theta_u < \mu$. We just need to prove that it is feasible to do so. Feasibility is equivalent to:

$$q(\mu - \theta_u) - \frac{\mu - \theta_L}{\theta_H - \theta_L} q(\theta_H - \theta_u) - \kappa(\theta_L, \theta_H) \geq 0,$$

For all $\theta_L < \mu < \theta_H$.

For sufficiently small $\varepsilon > 0$, consider $q' = q^* - \varepsilon > 0$. Then:

$$\begin{aligned} q'(\mu - \theta_u) - \frac{\mu - \theta_L}{\theta_H - \theta_L} q'(\theta_H - \theta_u) - \kappa(\theta_L, \theta_H) \geq \\ -\varepsilon\left(\mu - \theta_u - (\mu - \theta_L)\frac{\theta_H - \theta_u}{\theta_H - \theta_L}\right) = \\ -\varepsilon\frac{(\theta_H - \mu)(\theta_L - \theta_u)}{\theta_H - \theta_L} \geq 0, \end{aligned}$$

where the first inequality follows from feasibility of q^*, the equality collects terms, and the last inequality follows from $\theta_H \geq \mu > \theta_u \geq \theta_L$.

We have then proved that q' is feasible and, for small enough ε it increases the principal's objective, contradicting optimality of q^*. Therefore, $c'(q^*) < \mu$.

■

References

Abaluck, J., and Gruber, J. (2011). Choice inconsistencies among the elderly: Evidence from plan choice in the medicare part d program. *American Economic Review*, *101*(4), 1180–1210.

Aumann, R. J., and Maschler, M. (1995). *Repeated Games with Incomplete Information* (Vol. 1). The MIT Press.

Bergemann, D., and Välimäki, J. (2002). Information acquisition and efficient mechanism design. *Econometrica*, *70*(3), 1007–1033.

Bloedel, A. W., and Segal, I. (2020). *Persuading a rationally inattentive agent* (tech. rep.). Working Paper.

Boleslavsky, R., and Kim, K. (2018). Bayesian persuasion and moral hazard. *Available at SSRN 2913669*.

Boyacı, T., and Akçay, Y. (2018). Pricing when customers have limited attention. *Management Science*, *64*(7), 2995–3014.

Brancaccio, G., Li, D., and Schürhoff, N. (2017). Learning by trading: The case of the us market for municipal bonds. *Unpublished paper. Princeton University*.

Brown, Z. Y. (2017). An empirical model of price transparency and markups in health care. *Manuscript*.

Brown, Z. Y. (2019). Equilibrium effects of health care price information. *Review of Economics and Statistics*, *101*(4), 699–712.

Brown, Z. Y., and Jeon, J. (2020). *Endogenous information and simplifying insurance choice* (tech. rep.). Mimeo, University of Michigan.

Caplin, A., Dean, M., and Leahy, J. (2022). Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *Journal of Political Economy*, *130*(6), 1676–1715.

Crémer, J., and Khalil, F. (1992). Gathering information before signing a contract. *The American Economic Review*, 566–578.

Doval, L., and Skreta, V. (2022). Mechanism design with limited commitment. *Econometrica*, *90*(4), 1463–1500.

Dworczak, P., and Martini, G. (2019). The simple economics of optimal persuasion. *Journal of Political Economy*, *127*(5), 1993–2048.

Ellison, G. (2005). A model of add-on pricing. *The Quarterly Journal of Economics*, *120*(2), 585–637.

Ellison, G., and Wolitzky, A. (2012). A search cost model of obfuscation. *The RAND Journal of Economics*, *43*(3), 417–441.

Gentzkow, M., and Kamenica, E. (2016). A rothschild-stiglitz approach to bayesian persuasion. *American Economic Review*, *106*(5), 597–601.

Georgiadis, G., and Szentes, B. (2020). Optimal monitoring design. *Econometrica*, *88*(5), 2075–2107.

Guo, Y., Li, H., and Shi, X. (2018). *Optimal discriminatory disclosure* (tech. rep.). Working paper.

Hackethal, A., Haliassos, M., and Jappelli, T. (2012). Financial advisors: A case of babysitters? *Journal of Banking & Finance*, *36*(2), 509–524.

Handel, B. R., and Kolstad, J. T. (2015). Health insurance for" humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, *105*(8), 2449–2500.

Hébert, B. M., and La'O, J. (2020). *Information acquisition, efficiency, and non-fundamental volatility* (Working Paper No. 26771). National Bureau of Economic Research.

Hébert, B. M., and Woodford, M. (2020). *Neighborhood-based information costs* (Working Paper No. 26743). National Bureau of Economic Research.

Hefti, A. (2018). Limited attention, competition and welfare. *Journal of Economic Theory*, *178*, 318–359.

Hellwig, M. F. (2010). Incentive problems with unidimensional hidden characteristics: A unified approach. *Econometrica*, *78*(4), 1201–1237.

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

Li, H., and Shi, X. (2017). Discriminatory information disclosure. *American Economic Review*, *107*(11), 3363–85.

Luenberger, D. G. (1997). *Optimization by vector space methods* (1st). John Wiley & Sons, Inc.

Mackowiak, B., Matejka, F., Wiederholt, M., et al. (2020). Rational inattention: A review. *CEPR Discussion Papers*, (15408).

Martin, D. (2017). Strategic pricing with rational inattention to quality. *Games and Economic Behavior*, *104*, 131–145.

Maskin, E., and Riley, J. (1984). Monopoly with incomplete information. *RAND Journal of Economics*, *15*(2), 171–196.

Matějka, F., and McKay, A. (2012). Simple market equilibria with rationally inattentive consumers. *American Economic Review*, *102*(3), 24–29.

Matějka, F., and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, *105*(1), 272–98.

Matějka, F., and Tabellini, G. (2016). Electoral competition with rationally inattentive voters. *Journal of the European Economic Association*.

Mensch, J. (2022). Screening inattentive buyers. *American Economic Review*, *112*(6), 1949–84.

Mensch, J., and Ravid, D. (2022). Monopoly, product quality, and flexible learning. *arXiv preprint arXiv:2202.09985*.

Mussa, M., and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, *18*(2), 301–317.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, *6*(1), 58–73.

Ostrizek, F. (2020). Vague by design: Performance evaluation and learning from wages.

Petrikaitė, V. (2018). Consumer obfuscation by a multiproduct firm. *The RAND Journal of Economics*, *49*(1), 206–223.

Pomatto, L., Strack, P., and Tamuz, O. (2023). The cost of information: The case of constant marginal costs. *American Economic Review*, *113*(5), 1360–93.

Ravid, D. (2020). Ultimatum bargaining with rational inattention. *American Economic Review*, *110*(9), 2948–63.

Ravid, D., Roesler, A.-K., and Szentes, B. (2022). Learning before trading: On the inefficiency of ignoring free information. *Journal of Political Economy*, *130*(2), 346–387.

Roesler, A.-K., and Szentes, B. (2017). Buyer-optimal learning and monopoly pricing. *American Economic Review*, *107*(7), 2072–80.

Shi, X. (2012). Optimal auctions with information acquisition. *Games and Economic Behavior*, *74*(2), 666–686.

Skreta, V. (2006). Mechanism design for arbitrary type spaces. *Economics Letters*, *91*(2), 293–299.

Szalay, D. (2009). Contracts with endogenous information. *Games and Economic Behavior*, *65*(2), 586–625.

Taubinsky, D., and Rees-Jones, A. (2018). Attention variation and welfare: Theory and evidence from a tax salience experiment. *The Review of Economic Studies*, *85*(4), 2462–2496.

Winkler, G. (1988). Extreme points of moment sets. *Mathematics of Operations Research*, *13*(4), 581–587.

Xu, H. (2001). Level functions of some optimal value functions. In A. Rubinov and B. Glover (Eds.), *Optimization and related topics* (pp. 135–152). Springer US.

Yang, M. (2019). Optimality of Debt under Flexible Information Acquisition. *The Review of Economic Studies*, *87*(1), 487–536.