School of Economic Sciences

Working Paper Series
WP 2017-15

Nonlinear Pricing with Costly Information Acquisition

Brett R. Devine and Felix Munoz-Garcia

December 2017

Nonlinear Pricing with Costly Information Acquisition*

Brett R. Devine† and Felix Munoz-Garcia‡

School of Economic Sciences
Washington State University

November 10, 2017

Abstract

This paper examines a nonlinear pricing model where the firm can choose to acquire costly information prior to offering contract menus to consumers; such as paying a consultant or investing in machine learning technologies. Information provides the firm with a signal about consumers types, whose accuracy increases as the firm acquires larger amounts of information. We show that the firm chooses to acquire information, only if it can purchase a sufficient amount that could alter its initial prior beliefs. Relative to standard settings where firms cannot acquire information, we identify how information acquisition changes optimal contract offers, equilibrium profits, information rents, and welfare. A better-informed firm increases its expected profits, but it can also increase expected utility when the cost of information is intermediate. Our results recommend balanced online privacy laws.

Keywords: Nonlinear pricing, Price discrimination, Information acquisition, Entropy, Monopolist
JEL Codes: L12, L15, D82, D83

*We thank Ana Espinola-Arredondo, Jill McCluskey, Ben Cowan, and Jonathan Yoder for their insightful discussions and suggestions. We are also grateful to seminar participants at Washington State University for their comments.

†323D Hulbert Hall, Washington State University, Pullman, WA 99163, E-mail: brett.devine@wsu.edu
‡103G Hulbert Hall, Washington State University, Pullman, WA 99164. E-mail: fmunoz@wsu.edu.

1 Introduction

In most settings, firms cannot directly observe consumer demand, leading them to practice second-degree price discrimination by offering menus of contracts. The literature on price discrimination is extensive, analyzing the separation of different types of consumers according to their willingness to pay, search costs, patience, and the quantity of information consumers have about their own preferences.[1] However, this literature assumes that information about consumer demand is given.[2] Instead, we allow for the firm to invest in information acquisition about consumer demand prior to offering menus of contracts. Our model then builds on Bergemann et al. (2015), which analyzes how price discrimination is affected when firms exogenously receive more information. Focusing on second-degree price discrimination, we allow for firms to endogenously acquire information about consumer demand, a commonly observed practice. Many large companies, for instance, currently gather consumer information, invest in new data sets, and in new tools to process this data (such as predictive analytics) to help the firm better infer consumer demand.[3] Our paper shows that allowing for information acquisition can alter firm's incentives to offer menus of contracts, but only when such information is sufficiently inexpensive to change the firm's prior beliefs about consumer types. Intuitively, the firm needs to acquire a sufficiently large amount of information to confirm that its initial beliefs are correct or, instead, that they are incorrect. Otherwise, the firm prefers to not invest in information acquisition, thus remaining as poorly informed as in standard non-linear pricing models where firms cannot acquire information.

We demonstrate that a more accurately informed firm can increase both its profits and consumers expected utility, yielding a Pareto improving outcome. However, this occurs only when the firm becomes better informed, but not perfectly informed. Our results, therefore, contribute to the debate about online purchasing privacy, suggesting that extreme policy approaches such as banning firms from accumulating any form of customer data or letting firms freely share this information with other retailers may be welfare reducing; while balanced policies can be welfare improving.

Our model considers that, in the first stage, the firm chooses how much information to acquire, such as consultants, purchase of data bases, tracking of IP addresses and, generally, any investment seeking to identify consumers willingness to pay for the good the firm sells. In the second stage, the firm practices nonlinear pricing to separate consumer types. When the firm acquires no information, its prior beliefs about consumers are unaffected, and it thus solves the standard nonlinear pricing problem. To separate types, the firm distorts the low-type output downwards relative to complete information, leaving no surplus for this buyer; but allows for a positive information rent for the high-type buyer. When the firm acquires a positive amount of information, however, it receives a signal about the consumers type, high or low, which helps the firm update its posterior beliefs. As the firm acquires more information, the signal's reliability increases, ultimately moving the firm to the complete information setting. In

1 See (Stokey, 1979; Mussa and Rosen, 1978; Riley and Zeckhauser, 1983; Spulber, 1992) for studies separating consumers according to their willingness to pay for the good; (Lewis and Sappington, 1994) for willingness to pay combined with how well consumers know their willingness to pay; (Salop, 1977) for a paper using consumer's different search costs as a separation tool; (Chiang and Spatt, 1982) for patience; (Wilson, 1988) for time of arrival; and (Stiglitz, 1977) for risk and risk aversion.

2 In (Courty and Li, 2000) and (Krähmer and Strausz, 2015), however, the firm practices sequential screening, and obtains information about consumer demand through interactions with them. Additionally, the firm itself can provide information to incompletely informed consumers as studied by the "partial disclosure" literature; see (Gentzkow and Kamenica, 2014; Hedlund, 2017; Li and Shi, 2017).

3 For instance, International Data Corporation estimated that worldwide revenues for big data and business analytics were US $130.1 billion in 2016, and could grow to US $203 billion in 2020. Banking was the industry with the largest investment in big data and business analytics solutions (nearly US $17 billion in 2016), and it is expected to experience the fastest spending growth; see Forbes (January 20th, 2017).

this case, upon receiving a signal of a consumer type, the firm responds offering the complete information contract for this type of buyer.

In the incomplete information setting, we show that the benefit of acquiring more information is weakly positive. Intuitively, as the firm acquires more information, its posterior probabilities of dealing with a low- or high-type consumer become more extreme. Upon receiving a low signal, the firm is more convinced of facing a low-type buyer, and responds by reducing the output distortion on this buyer's contract while increasing the information rents to the unlikely high-type buyer. Overall, expected profits increase since the expected increase in profits from the (more likely) low type offset the expected decrease in profits from the (less likely) high type. An analogous argument applies if, instead, the firm receives a high signal. In this setting, the firm lowers the information rent on the high-type buyer, as he became more likely after the signal; and increases the output distortion on the low-type, since the latter became less likely after the signal. Again, expected profits increase from adjusting the menu of contracts.

In addition, we show that, when the high-type buyer is likely enough, the firm would ignore the low type under the standard incomplete information model, choosing the complete-information high-type contract to extract all surplus from this buyer. In this scenario, signals that yield high enough posteriors (so the firm keeps assigning a high probability weight on this type of consumer) induce no contract changes, and no profit gain from the high-type buyer, but larger expected profits. However, if the firm acquires a large amount of information and receives a low signal, its posterior beliefs will be affected. Intuitively, the signal must be reliable given the large investment in information, implying that the high-type buyer cannot be that likely, ultimately tilting the firm to offer a menu of contracts.

We then analyze optimal information acquisition, showing that it increases as information becomes cheaper. Importantly, we find settings for which the firm may optimally choose to not acquire any information at all. In words, the firm can anticipate that, acquiring small amounts of information would lead to no subsequent changes in its contract offer, thus entailing zero expected benefits from information acquisition but certain costs. This happens when information is relatively expensive, and thus the firm cannot acquire a sufficiently large amount of information that makes signals reliable enough to alter its subsequent contract offers. We then examine settings in which the firm acquires a positive amount of information, and how the latter alters its contract offers. When priors are low, the firm offers a menu of contracts, which is adjusted as the firm receives more information about customer types. When information is sufficiently inexpensive, the firm acquires a relatively large amount of information, which produces relatively reliable signals. In this context, the firm can either receive a low signal (which confirms its initial low beliefs, and only leads to contract adjustments) or a high signal (which now induces the firm to ignore the low-type buyer, and offer units to the high-type customer alone). A similar argument applies when priors are relatively high, whereby the firm focuses on the high-type buyer alone. As information becomes inexpensive, the firm chooses to acquire a large amount, which in the event of a low signal, alters the firms pricing decision, namely, it offers a contract menu which induces self-selection.

Finally, we investigate how customer's utility, profits, and overall welfare are affected by the firms decision to acquire further bits of information. When priors are low, the firm starts offering a menu when information acquisition is banned (or prohibitively expensive), which leaves the high-type buyer with an information rent as in the standard model of nonlinear pricing. When information becomes inexpensive, the firm acquires information, which lets it adjust the menu of contracts, decreasing the output distortion on the low-type and the information rent that the high-type buyer earns. As discussed above, profits increase from such a contract adjustment. Overall, welfare initially decreases as the firm becomes better informed (when information is relatively costly) since the decrease in utility offsets the increase in profits. However, when the firm becomes more informed (i.e., when information is inexpensive) welfare can increase, given that the increase in profits offsets the decrease in utility. When priors are high, a subtler result emerges. When information is prohibitively expensive, the firm focuses on the high-type buyer alone, leaving the latter with no rents. When information becomes cheaper, the firm starts to offer menus, leaving an information rent to the high-type customer. In expectation, this increases both rents and profits, thus becoming a Pareto improvement since both agents gain from the firm being better informed about customers types. When information is more inexpensive, the firm acquires more information, signals become more reliable, helping the firm reduce the information rent it offers to the high-type buyer. As a result, his expected utility decreases while profits increase, yielding nonetheless an increase in expected welfare.

Our results can help in the debate over FCC rules on internet privacy. On October 28th 2017, President Trump signed S.J. Resolution 34, which nullifies an Obama administration rule requiring internet service providers for customer consent before sharing or selling their information to third parties, such as geolocation data, financial and health information, web browsing, and app-usage data. While other pieces of private information, such as e-mail addresses, are still protected under the new law, others are not, such as web-browsing history, allowing online retailers to display ads personalized to an individuals browsing history as he surfs the web. Many other privacy laws affect online customers, ultimately impacting a sellers ability to acquire information about their demand for different products. Our findings suggest that regulations that, essentially, decrease a sellers cost of acquiring information can yield utility and profit gains, if they do not reduce such a cost to negligible levels. If instead, privacy laws entail free customer information (e.g., setting no restrictions on how sellers can share and sell it to other parties), profits would increase at the expense of large utility reductions.

1.1 Related Literature

1.1.1 Nonlinear Pricing

Monopolists can learn their demand curve over time through experimenting with prices and observing outcomes (Clower, 1959; Grossman et al., 1977; Trefler, 1993). The effectiveness of this method is weakened, however, when demand relationships change significantly over time. In addition, the outcomes of the monopolist's "experiments" are publicly observable by potential entrants and the threat of entry can affect the monopolist's freedom to experiment optimally (Dimitrova and Schlee, 2003).[4]

In general, when lacking aggregate information about prices and quantity demanded, the monopolist can invest, at a cost, in an informative signal.[5] Kihlstrom (1976) models a firm's decision to choose an information structure with more informative information structures coming at greater costs. After observation of a signal produced by a chosen information structure, the monopolist makes profit maximizing decisions. We consider a similar situation, except in our model the monopolist acquires individualized information about consumer types and makes nonlinear, rather than linear, pricing decisions to price discriminate. The distinction is nontrivial, as satisfaction of participation and incentive compatibility constraints alter both the effect and benefit of additional units of information on the monopolist's decision problem.

Under proper conditions[6], the monopolist can make use of individualized information (rather

[4] Alternatively, the monopolist can acquire demand information through "market research", conducting costly samples of consumers to generate a sufficient statistic useful for pricing decisions (Manning, 1979).

[5] The signal could be the outcome of a price experiment, a consultant's research report, trade-journal reports, focus-group results, etc.

[6] Inability for consumers to engage in arbitrage, no anti-price discrimination laws.

than aggregated information) about consumer's maximum willingness to pay. Under first-degree price discrimination the monopolist achieves the greatest possible profit and eliminates any dead-weight losses caused by quantity restriction. Hence, overall welfare is increased at the expense of consumers receiving zero surplus.[7]

Under incomplete information, the firm seeks to maximize profits given its prior beliefs through a nonlinear contract.[8] The set of contracts offered in screening models is sensitive to how much information the firm has about consumer valuations, how well consumers know their own valuations, the information about product quality possessed by both firms and consumers, and information about the degree of bounded rationality of agents. In particular, mechanism design literature, such as (Cremer and Khalil, 1992; Cremer and McLean, 1985; Cremer et al., 1998; Khalil and Rochet, 1998; Bergemann and Välimäki, 2002; Szalay, 2009) pays a great deal of attention to the problem of optimal nonlinear pricing schemes when consumers (or contract agent) have incomplete information about their own type and can acquire information. In contrast, we analyze the situation in which consumers have complete information about their types and the monopolist endogenously acquires additional information about consumer types at cost. Overall, price discriminating behaviors are sensitive to the specific "information structure" that every agent faces in the incomplete information problem.

Recently, the literature has expanded to include dynamic price discrimination settings, in which firms acquire information about consumer types through interacting with them repeatedly (Acquisti and Varian, 2005; Bonatti, 2011; Courty and Li, 2000). In relation to our paper, this literature highlights issues involving consumer privacy in a dynamic pricing situation in which firm's dynamically screen customers. While it is possible for consumers to benefit from this behavior, consumers can act strategically to nullify the firm's efforts to establish nonlinear prices. In our model this situation cannot occur, as the "sequential" nature of the firm's decision problem involves reactions to an outside signal of consumer type and not a consumer's reaction to a proposed price schedule. As a consequence, the monopolist still finds it optimal to engage in nonlinear pricing (as opposed to a single price) after observing an informative signal.

Researchers have long been interested in the welfare and efficiency implications of price discrimination (Chiang and Spatt, 1982; Roberts, 1979; Varian, 1985) as the behavior has the capacity to decrease or increase total surplus and transfer it between consumers and producers. Recent work by Bergemann et al. (2015) analyzes the limits of price discrimination efforts by firms and the concordant welfare effects. While focusing primarily on third-degree price discrimination, they investigate a special case of nonlinear screening of two types of agents, as we do in this paper. Their results demonstrate some of the welfare affects associated with exogenous changes in the monopolist's information about consumer types. In particular, different information structures in this setting can generate outcomes in which firms benefit and consumers are hurt, both consumers and firms benefit. Our results provide additional insight that these effects can occur in equilibrium when the firm endogenously acquires its information at cost. The existence and robustness of these Pareto improving outcomes is important for

---

7 Perfect price discrimination requires the monopolist to possess complete information about consumer preferences. In practice, such a quantity of information is either impossible to obtain or prohibitively expensive. As a consequence, incomplete information seems a more prevalent scenario and the literature has focused on "second-best" strategies to reduce the impact of incomplete information.

8 For example monopolists can engage in priority pricing, limited initial quantity discounts, auctions (Dana, 2001; Harris and Raviv, 1981), commodity bundling (Adams and Yellen, 1976), two-part tariffs (Leland and Meyer, 1976), block pricing (Leland and Meyer, 1976; Cremer and McLean, 1985) and even price dispersion to discriminate on differences in search costs and time costs of consumers (Salop, 1977; Chiang and Spatt, 1982). See (Stiglitz, 1977; Mussa and Rosen, 1978; Maskin and Riley, 1984; Spulber, 1992; Armstrong, 1996; Spiegel and Wilkie, 2000; Bonatti, 2011). Cases in which consumers are imperfectly informed and/or can acquire costly information about their types are considered in (Cremer and McLean, 1985; Cremer and Khalil, 1992; Cremer et al., 1998, 2003)

informing policy debate on issues of firm's use of consumer information and their freedom to price discriminate.

1.1.2 Information and its Structure

In a Bayesian context, an information structure typically specifies how agents obtain their posterior beliefs. For example, beliefs could be revised after observing the outcome of an experiment or calculating the sufficient statistic of a given sample. The outcome and sufficient statistic represent a "signal" and the experimental form and sampling procedure represent the information structure that generates signals. Blackwell (1953) established strict conditions for comparing both the informativeness and value of an information structure.[9] Blackwell's framework has been extended by several researchers (Karlin and Rubin, 1956; Lehmann, 1988). The ordering of information structures arising from Blackwell's results, is however, incomplete as many structures are not comparable.[10]

In contrast, entropy-based measures of information provide a complete ordering as the information content of all information structures can be compared. However, the co-monotonic relationship between information and expected utility may no longer hold. Recently, Cabrales et al. (2013) establish conditions under which an investor's willingness to pay for an information structure increases in the informativeness of that structure as measured by entropy, implying that a more informative structure yields a higher expected utility.

We use entropy-based measures of information to compare the informativeness of information structures and assign their cost similar to (Cabrales et al., 2013). Unlike Cabrales et al. (2013), while more informative structures cost more, they do not necessarily increase a monopolist's expected profits, nor decrease consumer's expected utility. The remainder of the paper proceeds as follows. First, we describe the monopolist's linear prices under complete information, and its non-linear prices under incomplete information (stage-two) and analyze its properties. Second, we identify the profit gain from acquiring information about consumer types prior to solving the pricing problem. Third, the optimal costly information acquisition problem (stage-one) of the monopolist is developed and analyzed. Fourth, comparative statics of the results on equilibrium values is numerically simulated.

2 Model

Consider a monopolist that sells a product sold in bundles. The monopolist chooses both the quantity of product in a bundle and its price. For simplicity, we allow for two types $\Theta = \{\theta_H, \theta_L\}$, where $\theta_H > \theta_L$, which affect consumer preferences for the monopolist's product according to utility function,

$$U_k = \theta_k u(q) - T, \quad \text{for all } k = \{H, L\} \tag{1}$$

[9] If structure A is equivalent to structure B plus some noise (this is often referred to as A being a "garbled" version of B), then B is strictly more informative than A. Furthermore, an economic decision maker can obtain at least as great an expected utility by making decisions using B as it can using A, for any decision problem. Intuitively, a decision maker could just ignore the additional information in B and get the same outcome as A.

[10] When we restrict the firm to choosing information structures using a single value, x, every increase in x yields a uniformly more informative information structure, i.e., every pair of information structures are then comparable using Blackwell's ordering. However, if we allow the firm to choose a pair (x, y), we could find situations where larger x does not necessarily yield a more informative information structure in the Blackwell sense; although it could be more informative in the mutual information sense, as in Cabrales et al. (2013).

where q is the quantity (quality) of product, T is the price for the quantity (bundle price) [11], and $u(\cdot)$ is an increasing and concave function of the units that the individual consumes, q. Consumers of either type only buy the product if $U_k \geq 0$. For bundle (q, T) the firm earns profit $T - cq$ where $c > 0$ denotes a constant marginal cost, and T enters as the firm's revenue.

2.1 Complete Information

Under complete information, consumer type is known prior to bundle offer. This information gives the monopolist the ability to perfectly distinguish each type by simple observation and subsequently offer a type-specific profit maximizing bundle. The contract $\{q^*, T^*\}$ that the firm offers is efficient; as Lemma 1 describes. All proofs are relegated to the appendix.

Lemma 1. *Under complete information the monopolist offers contract $\{q_k^*, T_k^*\}$ where output q_k^* solves $\theta_k u'(q_k^*) = c$ and price T_k^* solves $T_k^* = \theta_k u(q_k^*)$. The monopolist's equilibrium profits are*

$$\pi_k^* = \theta_k v(q_k^*) - cq_k^*$$

Hence, for every consumer of type k, the monopolist offers output level q_k^* for which the consumer's marginal utility coincides with the monopolist's marginal cost of production. In addition, the transfer extracts all surplus from each type of consumer.

Example 1. Suppose $u(q) = 2\sqrt{q}$, then $q_k^* = (\theta_k/c)^2$ and $T_k^* = 2\theta_k^2/c$. The profits to the firm under each type are $\pi_k^* = \theta_k^2/c$. □

Prior to observing a consumer's type expected profits when $Pr(\theta_H) = \beta$ are

$$\Pi(\beta) = \beta\pi_H^* + (1-\beta)\pi_L^* \tag{2}$$

2.2 Incomplete Information

When consumer types are private information and the complete information contracts $\{q_H^*, T_H^*\}$ and $\{q_L^*, T_L^*\}$ are not incentive compatible since the high-type can reach a higher utility level by choosing the contract meant for the low-type. Assuming the monopolist has a prior beliefs β that the consumer is type θ_H, the firm solves

$$\max_{q_H, T_H, q_L, T_L} \beta(T_H - cq_H) + (1-\beta)(T_L - cq_L) \tag{3}$$

subject to standard participation and incentive compatibility constraints.[12] The firm maximizes expected profits by offering contracts $\{q_H(\beta), T_H(\beta)\}$ and $\{q_L(\beta), T_L(\beta)\}$ achieving self-selection as the next lemma describes.

Lemma 2. *Under incomplete information and beliefs $\beta \leq \theta_L/\theta_H$, the monopolist offers contracts*

[11] Traditionally this "price" of the bundle is referred to as a transfer and denoted by T or t instead of the customary p used ubiquitously for unit price in economic models.

[12] For a complete description of this problem, including the full set of constraints and the solution, see proof of Lemma 2 in the appendix.

$\{q_H(\beta), T_H(\beta)\}$ and $\{q_L(\beta), T_L(\beta)\}$ that solve

$$\theta_H u'[q_H(\beta)] = c, \qquad \theta_L u'[q_L(\beta)] = c + (\theta_H - \theta_L)u'[q_L(\beta)]\frac{\beta}{1-\beta}$$
$$T_H(\beta) = \theta_H u[q_H(\beta)] - (\theta_H - \theta_L)u[q_L(\beta)] \qquad T_L(\beta) = \theta_L u[q_L(\beta)]$$

If instead, $\beta > \theta_L/\theta_H$ the monopolist only offers the contract $\{q_H^*, T_H^*\}$ from Lemma 1, thus ignoring the low-type consumer.

The output of the high-type consumer satisfies the complete information first-order condition, i.e., $q_H(\beta) = q_H^*$; also referred to as "no distortion at the top". The output of the low-type contract is, however, weakly lower under incomplete information. Finally, the transfers in this setting entail a positive (zero) surplus for the high (low) type of consumer, which arises because of the seller's need to provide incentives, $R(\beta) = (\theta_H - \theta_L)u[q_L(\beta)]$, to the high-type consumer to truthfully reveal his type (information rents). Note that when the frequency of high-types is large enough, $\beta > \theta_H/\theta_L$, the monopolist focuses on the high-type consumers, thus offering the same contract as under complete information, which extracts all information rents from this type of consumer.[13] Expected profits under the optimal incomplete information contracts are

$$V(\beta) = \beta[T_H(\beta) - cq_H(\beta)] + (1-\beta)[T_L(\beta) - cq_L(\beta)] \tag{4}$$

Example 2. When $u(q) = 2\sqrt{q}$ and $\beta > \theta_L/\theta_H$. The firm focuses on high-type buyers alone, offering $q_H^* = \left(\frac{\theta_H}{c}\right)^2$ and $T_H^* = \frac{2\theta_H^2}{c}$ as shown in Example 1. However, when $\beta \leq \theta_L/\theta_H$ the monopolist offers the following menu of contracts:

$$\left\{q_L(\beta) = \frac{(\theta_L - \beta\theta_H)^2}{c^2(1-\beta)^2},\ T_L(\beta) = \frac{2\theta_L(\theta_L - \beta\theta_H)}{c(1-\beta)}\right\} \text{ for type } \theta_L \text{ consumers, and}$$
$$\left\{q_H(\beta) = \left(\frac{\theta_H}{c}\right)^2,\ T_H(\beta) = \frac{2(\theta_H^2 + \theta_L^2 - \theta_H\theta_L(1+\beta))}{c(1-\beta)}\right\} \text{ for type } \theta_H \text{ consumers.}$$

Hence, expected profits become

$$V(\beta) = \begin{cases} \frac{\beta\theta_H^2 - 2\beta\theta_H\theta_L + \theta_L^2}{c(1-\beta)} & \text{if } \beta \leq \theta_L/\theta_H \\ \frac{\beta\theta_H^2}{c} & \text{if } \beta > \theta_L/\theta_H \end{cases}$$

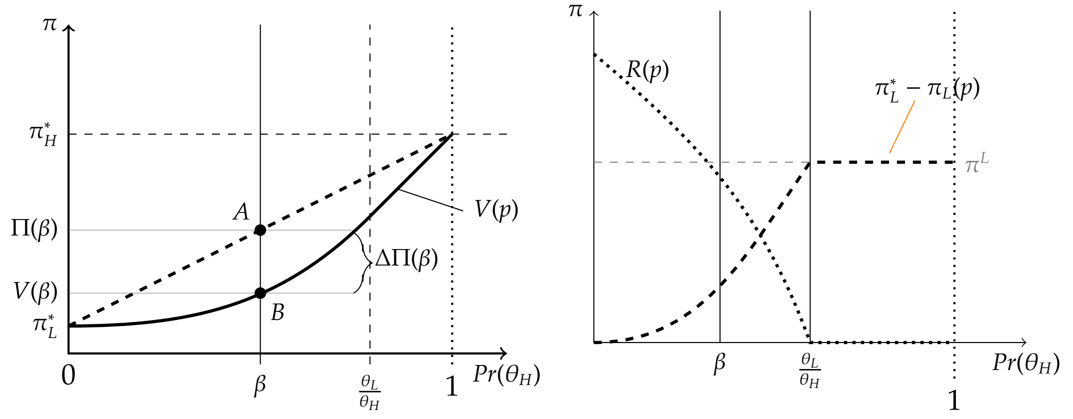
In contrast, under complete information, the monopolist can offer a contract that leads to maximal profits $\beta\theta_k^2/c$ from any type of buyer k. Therefore, the "expected maximal profits" from operating in a complete information scenario are $\Pi(\beta) = \frac{1}{c}\left(\beta\theta_H^2 + (1-\beta)\theta_L^2\right)$ which exceeds both $V(\beta)$ in both cases. □

Figure 1a shows how expected profits in the complete and incomplete information context vary as a function of the firm's prior belief. The loss of expected profit to the monopolist attributed to information rents can thus be measured by the difference $\Delta\Pi(\beta) \equiv \Pi(\beta) - V(\beta)$, as depicted in Figure 1b. The shape of $\Delta\Pi$ depends on the size of β relative to θ_L/θ_H.

13 In this case, contract (q_H^*, T_H^*) is accepted by the high-type consumer, but rejected by the low-type consumer.

Figure 1: **Profit Loss from Incomplete Information**.

Figure 1a Figure 1b



$$\Delta\Pi(\beta) = \begin{cases} \beta R(\beta) + (1-\beta)[\pi_L^* - \pi_L(\beta)] & \text{if } \beta \leq \theta_L/\theta_H \\ (1-\beta)\pi_L^* & \text{if } \beta > \theta_L/\theta_H \end{cases}$$

where $\pi_L(\beta) \equiv \beta u[q_L(\beta)] - cq_L(\beta)$ denotes the profit from the low-type buyer. Specifically, when $\beta \leq \theta_L/\theta_H$ the profit loss includes the "information rent" pad to the high-type consumer, $R(\beta) \equiv [\theta_H - \theta_L]u[q_L(\beta)]$ and the expected profit that the firm loses from distorting the low-type consumer's contract away from its complete information level, π_L^*. In words, the firm can only serve both types if it attracts the high-type buyer with information rents, and if it distorts the contract to the low-type buyer. As the probability of a high-type buyers increases, the firm reduces the information rent that it will likely pay to this consumer, as depicted in the decreasing $R(p)$ curve in Figure 1b; but increases the output distortion on the low-type contract, since this type of consumer now becomes less likely, as illustrated by the increasing curve $\pi_L^* - \pi_L(p)$. When the probability of a high-type, β, is sufficiently high, $\beta > \theta_L/\theta_H$, the firm ignores the low-type, paying no rents to the high-type. In this setting, the loss in expected profits coincides with the expected profit it would make from the low-type buyer it ignores.

Expected profits under incomplete information are increasing in β as the following lemma describes.

Lemma 3. *The value function $V(\beta)$ is increasing in β, convex for all $\beta \leq \theta_L/\theta_H$, but constant for all $\beta > \theta_L/\theta_H$, since*

$$\frac{\partial V(\beta)}{\partial \beta} = \begin{cases} \pi_H^* - R(\beta) - \pi_L(\beta) & \text{if } \beta \leq \frac{\theta_L}{\theta_H} \\ \pi_H^* & \text{otherwise} \end{cases}$$

Intuitively, maximal expected profits are increasing in β as the high-type buyer is more profitable than the low-type buyer. When β is small, the firm knows the high-type buyer is unlikely and offers a menu of contracts. Instead of collecting the full profit π_H^* on a high-type buyer, the firm loses expected profits on the high-type contract (information rent) $R(\beta)$, but gains expected profit from serving the low-type buyer; receiving $\pi_L(\beta)$ on the distorted low-type contract.

As β increases toward the cutoff θ_L/θ_H, both $R(\beta)$ and $\pi_L(\beta)$ approach zero as depicted in Figure

1b, since the monopolist focuses on the high-type alone, leaving the firm with π_H^* additional profits. The decision to ignore the low-type when $\beta \geq \theta_L/\theta_H$ eliminates the need for the firm to account for incentive compatibility in its pricing decision, changing the form of the second-stage value function $V(\beta)$. The abrupt change in the second-stage pricing policy creates a "kink" in the value function at $\beta = \theta_L/\theta_H$.

Hence if, prior to designing contracts, the monopolist reduces its uncertainty about consumer's types, it could capture a portion of the expected profit loss $\Delta\Pi(\beta)$ identified above. In particular, uncertainty can be reduced if the monopolist acquires, at a cost, information through a signal about the consumer's type; as studied in the next section.

3 Information Acquisition

3.1 Information Structures

Consider the set of consumer types $\Theta = \{\theta_H, \theta_L\}$ which are unobservable to the firm and a set of signals $S = \{s_H, s_L\}$ which are observable. A stochastic information structure is a joint density $f(\theta, s)$ over $\Theta \times S$. Furthermore, let $g_s(s)$ and $g_\theta(\theta)$ be the respective marginal densities. The quantity of information, $I(\Theta, S)$, the firm gains about Θ from observing signals in S is called the mutual information and is given, in this case, by

$$I(\Theta, S) = \sum_{\Theta} \sum_{S} f(\theta, s) \log \left(\frac{f(\theta, s)}{g_\theta(\theta) \cdot g_s(s)} \right) \tag{5}$$

The mutual information measures the degree of functional dependence of the random variables Θ and S in units of information, such as bits.[14] Mutual information increases as signals become more dependent upon types. In this context, the firm determines the expected quantity of information received about consumer types through selection of f. Given prior beliefs over consumer types, $g_\theta(\theta)$, the firm can alternatively select the information structure f by choosing a conditional density $f(s|\theta)$.[15]

$$f(\theta, s) = f(s \mid \theta) g_\theta(\theta)$$

Let $f(s_k|\theta_k) = m(x)$ for all $k = \{H, L\}$ where s_H denotes a high-type buyer and s_L a low-type buyer signal.[16] The firm's choice of x determines the information structure. The quantity of information acquired by the firm monotonically increases in the level of x. When x is chosen at its lower bound, ($x = 0$) the firm acquires no information. When x is chosen at its upper bound, ($x = \infty$) the firm approaches complete information.

Table 1 shows the joint and marginal probabilities specified by the firm's choice of x and constitutes the information structure that will generate signals communicating information about consumer's types. The cell corresponding to (θ_k, s_j) is the joint probability $f(\theta_k, s_j)$ as a function of x. We make three assumptions on the function $m(\cdot)$ mapping the firm's choice of x to conditional probabilities.

14 For instance, if types and signals are independent random variables then observing signals cannot provide any information about types and $I(\Theta, S) = 0$. Independence allows factoring of the joint density $f(\theta, s) = g_\theta(\theta) \cdot g_s(s)$. Therefore, for every prior $(\theta, s) \in \Theta \times S$ the logarithm is $\log_2(1) = 0$, which implies that $I(\Theta, S) = 0$, i.e., the information about types can be inferred from observing the signal.

15 The motivation behind the term *stochastic* information structure can be clearly seen here: for a given value of θ, a number of different signals s can be generated with positive probability.

16 (Blackwell, 1953; Marschak and Miyasawa, 1968) show that these conditional likelihoods can be both used to characterize and compare information structures.

		Signals	
		s_H	s_L
Types	θ_H	$m(x)\beta$	$[1-m(x)]\beta$
	θ_L	$[1-m(x)](1-\beta)$	$m(x)(1-\beta)$

Table 1: **Information structure (joint probabilities) from the firm's choice of x**

Assumption 1 (Feasibility and Non-redundant Range)**.** $m : [0, \infty) \to [\frac{1}{2}, 1]$.

Intuitively, as $m(x)$ approaches 1, the information structure's signals tell the truth with certainty, while as $m(x)$ approaches zero, the signals lie. Since both such information structures are equally useful for determining consumer type, we eliminate redundant information structures. In addition, $m(x)$ must be a valid probability for all x.

Assumption 2 (Increasing Accuracy)**.** $m(0) = \frac{1}{2}$ and $\lim_{x\to\infty} m(x) = 1$ and for all $x \in [0, \infty)$, $m'(x) > 0$.

In words, as x approaches its lower bound (i.e., no information acquired), the conditional probability of receiving a signal that coincides with the consumer's true type becomes 1/2, thus making the signal uninformative. However, as x increases, conditional probability $m(x)$ increases, thus making the signal informative.[17] Importantly, any increase in x, will increase the dependency of Θ and S increasing the quantity of information about consumer type defined in (5).

Assumption 3 (Diminishing Returns)**.** $\lim_{x\to 0} m'(x) = 0$ and $m''(x) < 0$ for all $x \in [0, \infty)$.

As the firm increases its investment in information, the resulting signals increase in accuracy, but at a diminishing rate.

3.2 Posterior Beliefs

From Table 1 the marginal probability that the monopolist receives signal s_H is $\rho_H(x) = m(x)\beta + [1-m(x)](1-\beta)$, as illustrated in the left column of the table, while the probability of signal s_L is $\rho_L(x) = [1-m(x)]\beta + m(x)(1-\beta)$, in the right column.

When the monopolist doesn't acquire information, $x = 0$, we obtain $\rho_H = \rho_L = 1/2$, regardless of the prior β. Conversely, when $x = 0$ (perfectly informative information structure) we have $m(x) = 1$ and, therefore, $\rho_H(x) = \beta$ and $\rho_L(x) = 1-\beta$. As a consequence, if $\beta > 1/2$ then $\rho_H(x)$ is increasing in x; and if $\beta < 1/2$, $\rho_H(x)$ is decreasing in x. In words, investing in more units of x increases the probability of receiving a signal that confirms the initial inclination, if any, of the firm.[18]

Example 3. The marginal probability distribution over signals when prior beliefs are $\beta = 0.6$

[17] At the limit when $m(x) = 1$, the predictions made by the information structure are perfectly reliable, i.e., when the true type is θ_k the information structure will produce signal s_k with probability 1.

[18] In the case that priors satisfy $\beta = 1/2$, the acquisition of more units of information does not alter the marginal probability of receiving a high-type signal $\rho_H(x)$, nor that of receiving a low signal, $\rho_L(x)$, since in this case both marginal probabilities collapse to $\rho_H(x) = \rho_L(x) = 1/2$.

can be calculated as

$$\rho_H(x) = m(x)[2(0.6) - 1] + (1 - 0.6) = 0.2m(x) + 0.4$$
$$\rho_L(x) = -m(x)[2(0.6) - 1] + 0.6 = -0.2m(x) + 0.6$$

which assigns a smaller weight on receiving a high-type signal than its initial belief, $0.2m(x) + 0.4 < 0.6 = \beta$ since $1/2 \leq m(x) \leq 1$; but a larger weight on receiving a low-type signal than its initial belief $-0.2m(x) + 0.6 > 0.4 = 1 - \beta$ for all admissible x. □

For a given signal s_k, the firm uses marginal probabilities $\rho_H(x)$ and $\rho_L(x)$ to update its beliefs. Specifically, the posterior probability of a high-type buyer given the observed signal s_k, is

$$\psi_H(x) = \frac{m(x)\beta}{\rho_H(x)} \quad \text{and} \quad \psi_L(x) = \frac{[1 - m(x)]\beta}{\rho_L(x)}$$

Under our assumptions, $\partial\psi_H(x)/\partial x > 0$ and $\partial\psi_L(x)/\partial x < 0$, indicating that, as the monopolist chooses a higher x the posterior probability that, upon receiving a high signal, the consumer is indeed of high-type increases toward 1. Conversely, the probability of a high-type consumer after receiving a low-type signal decreases toward 0. Intuitively, a more accurate information structure makes signals more informative. In addition, when $x = 0$, we obtain $\psi_H(x) = \psi_L(x) = \beta$ and for all $x > 0$, posterior beliefs satisfy $\psi_L(x) < \beta < \psi_H(x)$.[19]

3.3 Optimal Information Acquisition

Let us next examine the monopolist's expected profit maximization problem. Let $V(\psi_k)$ be the expected profits that the monopolist obtains after receiving a signal s_k, updating its posterior beliefs to ψ_k and then solving the profit maximization problem in Lemma 2. If his posterior beliefs are unaffected by the signal, $\psi_k = \beta$, then expected profits are $V(\beta)$, thus coinciding with those under the incomplete information setting where information acquisition is not possible. If, instead, posterior beliefs are higher than its prior, $\psi_k > \beta$, the firm assigns a larger probability on the buyer being a high-type, and expected profits increase relative to the setting without information acquisition, i.e., $V(\psi_k) > V(\beta)$. For presentation purposes, we separately describe the benefits and costs of acquiring information.[20]

3.3.1 Benefit of Increasing x.

The *ex-ante* expected profits from acquiring x are

$$B(x) = \rho_H(x)V[\psi_H(x)] + \rho_L(x)V[\psi_L(x)] \tag{6}$$

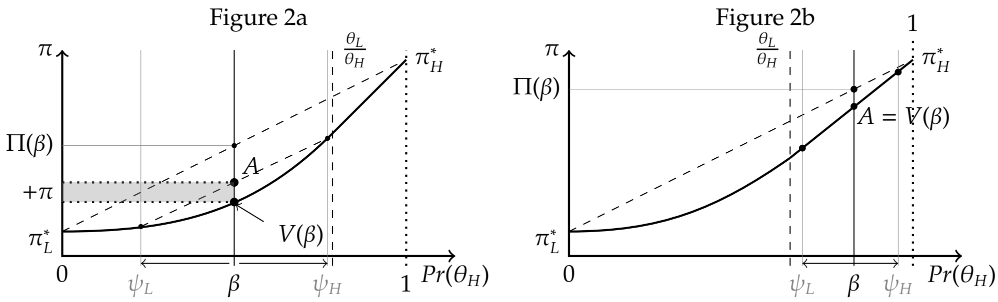
where the first term indicates the highest expected profits the monopolist can expect after receiving a s_H signal weighted by the probability, $\rho_H(x)$, of receiving such a signal s_H. The second term is analogous but in the case of receiving signal s_L.

Figure 2a depicts the case in which initial beliefs, β, lie in the convex portion of the value function $V(p)$. First, under complete information the firm makes a profit of either π_H^* with

[19] From the ex-ante viewpoint of the firm deciding on an information structure, the expected posterior probability of a high-type buyer, $\rho_H(x)\psi_H(x) + \rho_L(x)\psi_L(x)$, coincides with the firm's prior probability, β.

[20] The opposite argument applies when posterior beliefs satisfy $\psi_k < \beta$, where expected profits are lower $V(\psi_k) < V(\beta)$; which follows from Lemma 3.

Figure 2: **Expected Profit Gain from Information**.



probability β, or π_L^* with probability $1-\beta$, entailing expected profits of $\Pi(\beta)$. Under incomplete information (with no information acquisition) the firm's prior is β and, since it does not receive signals, its expected profits are $V(\beta)$; as shown in Lemma 2. However, if the firm can acquire information its profits become $V(\psi_H)$ upon receiving a high signal (with probability ρ_H) or $V(\psi_L)$ upon receiving a low signal (with probability ρ_L), yielding an expected profit at point A in the figure. Expected profits increase relative to the case in which information acquisition is not allowed, as depicted in the shaded area of Figure 2a.

Figure 2b represents the case where initial beliefs lie in the linear portion of the value function $V(p)$. When acquiring no information the firm earns profits $V(\beta)$. Unlike the previous case, when the firm acquires a small amount of x such that $\theta_L/\theta_H < \psi_L(x) < \psi_H(x)$, expected profits are the same as before, i.e., $\mathbb{E}[V(\psi_k)] = V(\beta)$.

Lemma 4. *The benefit of acquiring information, $B(x)$, satisfies $B'(x) \geq 0$ for all x, and $B''(x) \leq 0$ if and only if*

$$\begin{cases} m''(x)\mathcal{D}(x) < -[m'(x)]^2 \left[\frac{\partial D_H(x,\beta)}{\partial m} - \frac{\partial D_L(x,\beta)}{\partial m}\right] & \text{if } \psi_L(x) < \psi_H(x) < \hat{\theta} \\ m''(x)\left[\beta\pi_H^* - D_L(x,\beta)\right] < [m'(x)]^2 \left[\frac{\partial D_L(x,\beta)}{\partial m}\right] & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ 0 & \text{if } \hat{\theta} < \psi_L(x) < \psi_H(x) \end{cases} \tag{7}$$

where $D_k(x,\beta) \equiv \beta\pi_H[\psi_k(x)] - (1-\beta)\pi_L[\psi_k(x)]$ for all $k = H, L$ and $\mathcal{D}(x) = D_H(x,\beta) - D_L(x,\beta)$.

An increase in x strictly increases expected profits if and only if at least one of the posterior beliefs (ψ_H or ψ_L) lies in the strictly convex region of $V(\beta)$. Otherwise, expected profits remain unchanged, i.e., $B'(x) = 0$. Finally, the expected benefit from information is concave, $B''(x) \leq 0$, if the diminishing accuracy of additional units of information (left-hand side of (7)) offsets the associated profit from such information (right-hand side of (7)).[21] For instance, when $m(x) = x$, additional units of information provide a constant increase in accuracy, i.e., $m'(x) = 1$ and $m''(x) = 0$. In that context, expression (7) collapses to

$$\begin{cases} \frac{\partial D_H(x,\beta)}{\partial m} < \frac{\partial D_L(x,\beta)}{\partial m} & \text{if } \psi_L(x) \leq \psi_H(x) \leq \hat{\theta} \\ -\frac{\partial D_L(x,\beta)}{\partial m} < 0 & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ 0 & \text{if } \hat{\theta} < \psi_L(x) < \psi_H(x) \end{cases}$$

In words, $D_k(x,\beta)$ captures the probability-adjusted difference in profits from the high and low-types when the firm receives a s_k signal, i.e., the high-type premium. An increase in x increases $D_H(x,\beta)$, the difference in profits from the two types after receiving signal s_H; while

[21] Generally, expression (7) holds for sufficiently high values of x. In the limit as $x \to \infty$, $m'(x) \to 0$, entailing that $\psi_L(x) < \hat{\theta} < \psi_H(x)$ and expression (7) reduces to $\lim_{x\to\infty} m''(x)[\beta\pi_H^* - D_L(x,\beta)] \leq 0$, which holds since $m''(x) \leq 0$ by definition and $\beta\pi_H^* \geq D_L(x,\beta)$ for all x.

decreasing $D_L(x, \beta)$ the difference, in profits from the two types after signal s_L. Therefore, the conditions in expression (7) do not hold when $m(x) = x$, under any parameter values. In words, the benefit from acquiring information is convex for all x.

Figure 3: **Rent vs. Profit**.

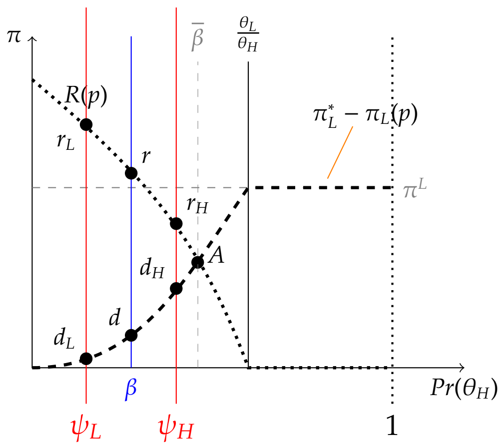


Figure 3 uses Figure 1b to depict the changes in information rents and low-type profits the firm experiences as its beliefs about the probability of the high-type buyer change after acquiring information. Under prior belief β, the firm offers a menu involving an information rent at r, and a low-type contract entailing a profit loss on this type of buyers of d. After acquiring information ($x > 0$) the firm's posterior beliefs shift outward from β. A low-signal s_L yields beliefs ψ_L indicating the high-types are less likely. As a result, the firm offers a more generous rent $r_L > r$ (which is, however, rarely paid), while decreasing the output distortion on the (now more likely) low-type buyer, $d_L < d$. Overall, the decreased probability ($\psi_L < \beta$) of paying the information rent increases expected profits. If the firm receives a high signal, s_H, updated beliefs become $\psi_H > \beta$. The firm offers a menu with a smaller information rent $r_H < r$, but a larger output distortion on the low-type contract $d_H > d$.[22]

3.3.2 Cost of Increasing x.

Greater investments in signal informativeness (higher x) represent larger outlays. We assume that the total cost of acquiring information x, is $C(x)$ where $C(\cdot)$ is a positive, increasing and convex function of x such that $C(0) = 0$. For generality, section 6 discusses how our results are affected if the cost of increasing x is a function of entropy; as in (Cabrales et al., 2013).

3.3.3 Ex-Ante Profit Maximization

The monopolist's optimal acquisition of information involves a trade-off between larger second-stage expected profits (increase in 6) as the firm becomes better informed and the cost of acquiring further information $C(x)$. In particular, the monopolist chooses x to solve

$$\max_{0 \le x \le \infty} \rho_H(x)V[\psi_H(x)] + \rho_L(x)V[\psi_L(x)] - C(x) \tag{8}$$

[22] Figure 3 considers that posteriors ψ_H lies to the left of point A. If, however, the firm acquires sufficient information to position ψ_h to the right of point A, rents become extremely low and the output distortion on the low-type buyer becomes larger. However, since the latter type of consumer is rare, expected profits increase by a large amount. This is consistent with the fact that expected profit $V(\beta)$ is convex, thus exhibiting its steepest segment immediately before the kink.

For compactness, let $\hat{\theta} \equiv \theta_L/\theta_H$ as well as $\Delta R(x) \equiv R[\psi_L(x)] - R[\psi_H(x)]$ be the difference in the rent paid to the high-type buyer when the firm receives a low or high signal and similarly let $\Delta\pi_L(x) \equiv \pi_L[\psi_L(x)] - \pi_L[\psi_H(x)]$ be the difference in profits originating from the low-type when the firm receives a low or high signal.

Proposition 1. *The firm chooses an information structure quality x to solve first order conditions (9), (10), or (11)*

$$m'(x)\{\beta\Delta R(x) + (1-\beta)\Delta\pi_L(x)\} \geq C'(x) \qquad \text{if} \quad \psi_L(x) < \psi_H(x) < \hat{\theta} \tag{9}$$

$$m'(x)\{\beta R[\psi_L(x)] + (1-\beta)\pi_L[\psi_L(x)]\} \geq C'(x) \qquad \text{if} \quad \psi_L(x) < \hat{\theta} < \psi_H(x) \tag{10}$$

$$0 \geq C'(x) \qquad \text{if} \quad \hat{\theta} < \psi_L(x) < \psi_H(x) \tag{11}$$

and its corresponding second order condition (12), (13), or (14),

$$m''(x)\Upsilon(x) + m'(x)\frac{\partial\Upsilon(x)}{\partial x} < C''(x) \qquad \text{if} \quad \psi_L(x) < \psi_H(x) < \hat{\theta} \tag{12}$$

$$m''(x)Z(x) + m'(x)\frac{\partial Z(x)}{\partial x} < C''(x) \qquad \text{if} \quad \psi_L(x) < \hat{\theta} < \psi_H(x) \tag{13}$$

$$0 \leq C''(x) \qquad \text{if} \quad \hat{\theta} < \psi_L(x) < \psi_H(x) \tag{14}$$

where $\Upsilon(x) \equiv \beta\Delta R(x) + (1-\beta)\Delta\pi_L(x)$ and $Z(x) \equiv \beta R[\psi_L(x)] + (1-\beta)\pi_L[\psi_L(x)]$. If x_i and x_j satisfy either of the first order conditions and second order conditions, where $x_i \neq x_j$, the firm chooses x_i if and only if $\mathbb{E}\Pi(x_i) > \mathbb{E}\Pi(x_j)$.

First order conditions (9) - (11) are satisfied when the firm chooses x to equate the marginal increase in expected profits from acquiring additional units of information through signals (left-hand side) with the marginal cost. The kink at the marginal benefit function produces cases (9) - (11), as we discuss next.[23]

In case (9), when priors are low enough, $\beta < \hat{\theta}$, small values of x yield posterior beliefs $\psi_L(x) < \psi_H(x) \leq \beta$. Under this condition, the firm continues to offer consumers a menu of bundles after receiving either signal, i.e., s_H and s_L. Specifically, the acquisition of information leads the firm to offer a menu, which entails information rents paid to the high-type and profits stemming from the low-type, both of which change depending on the signal that the firm receives; as captured in the left-hand side of (9).

In case (10), for any β, large enough values of x permit posterior beliefs $\psi_L(x) < \hat{\theta} < \psi_H(x)$. In this case, the firm pays an information rent and earns profits from the low-type only upon receiving a low signal (s_L) that leads to a menu offer.

Finally, in case (11) if priors satisfy $\beta > \hat{\theta}$, small values of x permit posterior beliefs $\hat{\theta} \leq \psi_L(x) < \psi_H(x)$ and the firm focuses only on the high-type buyer after receiving either signals, s_H and s_L. In this case, only $x = 0$ (no information acquisition) can satisfy the first order condition.

Non-concavities in the firm's objective function allow for local minima as well. Second order conditions (12) - (14) guarantee that any amount of information acquisition satisfying a first order condition will be a local maximum. Finally, we need a condition on profits to ensure that a local maximum is global: the firm obtains more profits acquiring information as prescribed by the local maximum than by not investing in information acquisition at all.

Corollary 1. *The firm's profit maximizing choice of x is smaller under cost $C_1(x)$ than under $C_2(x)$ if and only if $C_1(x) \geq C_2(x)$ for all x.*

[23] Graphically, the marginal gains to the firm from information depend on whether the firm's posterior beliefs fall in the strictly convex or linear portions of the firm's post-signal value function.

As expected, an increase in the cost of every unit of information induces the firm to acquire fewer units. If such cost is sufficiently high, the firm may not acquire any information at all.

3.3.4 Effect of Information Acquisition on Consumer's Utility

The firm's menu pricing behavior extracts all surplus from low-type consumers and provides an information rent to high-types. The expected utility of consumers is a function of the expected information rent collected by the high-type. Specifically,

$$\mathbb{E}U = \begin{cases} \beta\,\{m(x)R(\psi_H) + [1-m(x)]R(\psi_L)\} & \text{if } \psi_L(x) < \psi_H(x) < \hat{\theta} \\ \beta[1-m(x)]R(\psi_L) & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ 0 & \text{if } \hat{\theta} < \psi_L(x) < \psi_H(x) \end{cases}$$

Intuitively, when priors are sufficiently high (low), the firm ignores the low-type buyer extracting all rents from the high-type consumer (offers a menu thus leaving rents for the high-type buyer). When priors are intermediate, the firm offers a menu after receiving a low signal, which allows the high-type to retain a rent; but ignores the low-type after receiving a high signal, which leaves the high-type buyer with no rents. The following Corollary examines how the high-type information rents are affected by a marginal increase in the firm's information acquisition, x.

Corollary 2. *A marginal increase in x changes expected utility as follows*

$$\frac{\partial\,\mathbb{E}U}{\partial x} = \begin{cases} \beta\,\{-m'(x)\Delta R(x) + \partial\mathcal{R}(x)\} & \text{if } \psi_L(x) < \psi_H(x) < \hat{\theta} \\ -\beta m'(x)R[\psi_L(x)] + \beta[1-m(x)]\frac{\partial R(\psi_L)}{\partial q_L}\frac{\partial q_L(\psi_L)}{\partial \psi}\frac{\partial \psi_L}{\partial m}m'(x) & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ 0 & \text{if } \hat{\theta} < \psi_L(x) < \psi_H(x) \end{cases}$$

where

$$\partial\mathcal{R}(x) \equiv m'(x)\left\{m(x)\frac{\partial R(\psi_H)}{\partial q_L}\frac{\partial q_L(\psi_H)}{\partial \psi}\frac{\partial \psi_H}{\partial m} + [1-m(x)]\frac{\partial R(\psi_L)}{\partial q_L}\frac{\partial q_L(\psi_L)}{\partial \psi}\frac{\partial \psi_L}{\partial m}\right\}$$

When the firms priors are sufficiently high, the firm does not change its pricing strategy after acquiring one additional unit of information, leaving the high-type buyer with no information rents. When priors are intermediate (i.e., $\beta = \theta_L/\theta_H$), the firms strategy depends on the specific signal it receives: after a high signal, the firm keeps ignoring the low-type buyer, thus leaving the high type with no rents; but after a low signal, the firm starts offering a menu, which allows the high-type buyer to gain an information rent. Finally, when priors are sufficiently low, the firms pricing strategy depends on the signal it receives and the information it acquired up to that point. Starting from no information acquisition, the firm keeps offering a menu after a marginal increase in x. However, it adjusts the menu depending on the signal it receives. Once the firm acquires enough information, a further increase in x can have more radical effects on the firms strategy: upon a low signal, the firm keeps offering a menu; but upon a high signal, the firm ignores the low type, extracting all rents from the high type, as it now trusts the reliability of the signal.[24] As discussed in Proposition 1, the firm only acquires information when it is profitable, i.e., profits weakly increase in x. Combining this result with that in

[24] These results hold regardless of the initial amount of information the firm sustains, including $x = 0$ as a special case where the firm initially operates as in the standard model where information acquisition is not allowed.

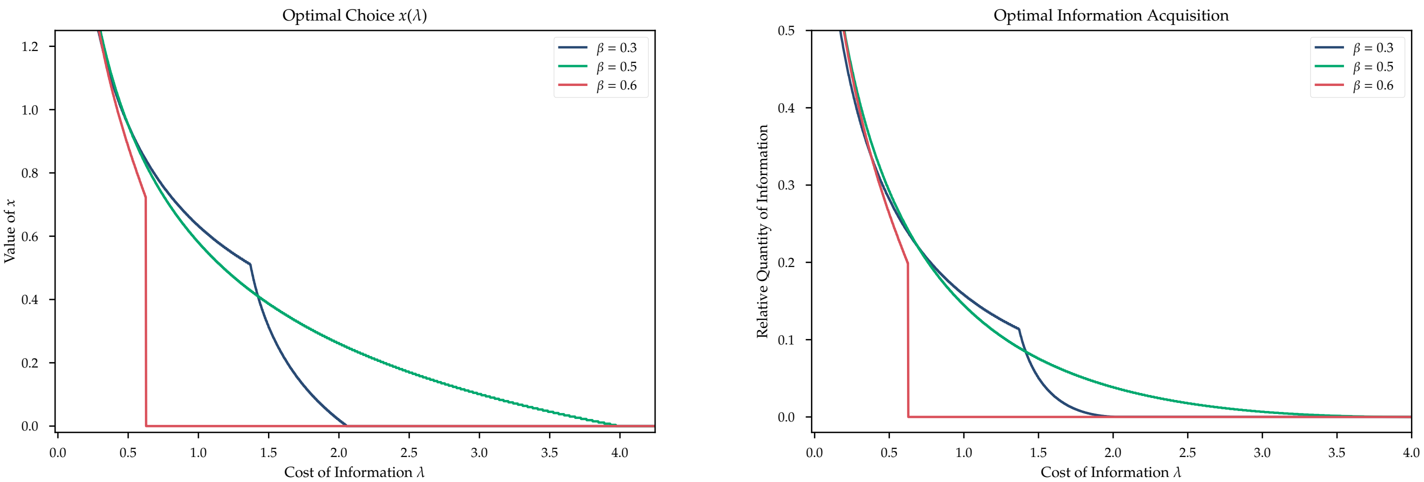
Corollary 2, overall welfare weakly increases when priors are sufficiently high. When priors are relatively low, however, expected utility weakly decreases in x. Therefore, welfare can increase or decrease as the firm becomes better informed, depending on whether the increase in profits offsets the decrease in expected utility. A similar argument applies when priors are intermediate, since expected utility is non-monotonic in x. The next section illustrates the optimal choice of x, the countervailing effect of increasing x in profits and utility, and the overall effect on welfare.

4 Numerical Simulations

In this section, we develop a parametric example to provide explicit solution for the firm's optimal information acquisition, x^*, the menu of contracts that the firm offers under different parameter conditions, information rents, profits, and welfare in equilibrium. For consistency, we consider that consumers' utility function is the same as in Examples 1-3, $u(q) = 2\sqrt{q}$, and now assume that $x \geq 0$, that $\theta_L/\theta_H = 1/2$, and that the conditional probability function $m(x)$ is given by $m(x) = 1 - \frac{1}{2}\exp^{-x}$.[25] Finally, the cost function is $C(x) = \lambda x^2$.

Figure 4: **Optimal Choice of x over λ**

(a) (b)



Optimal choice of x. Figure 4a depicts the optimal information acquisition x^* as a function of the cost of information λ. For presentation purposes, Figure 4b plots in its vertical axis the relative quantity of information $\frac{I(x^*,\Theta)}{\mathcal{H}(\Theta)}$, where $I(x^*,\Theta)$ measures the mutual information between signals and consumer types upon acquiring x^*, whereas $\mathcal{H}(\Theta) \equiv -[\beta \log \beta + (1-\beta)\log(1-\beta)]$ denotes the entropy of types given prior belief β. Intuitively, when the firm acquires as much information as possible, mutual information satisfies $I(x^*,\Theta) = \mathcal{H}(\Theta)$, implying that relative information acquisition is 1; while when the firm does not acquire information, $I(x^*,\Theta) = 0$, entailing that relative information is also zero. In other words, the vertical axis can be understood as how close the firm approaches the complete information setting, being 1 when it acquires full information and zero when it does not acquire any.

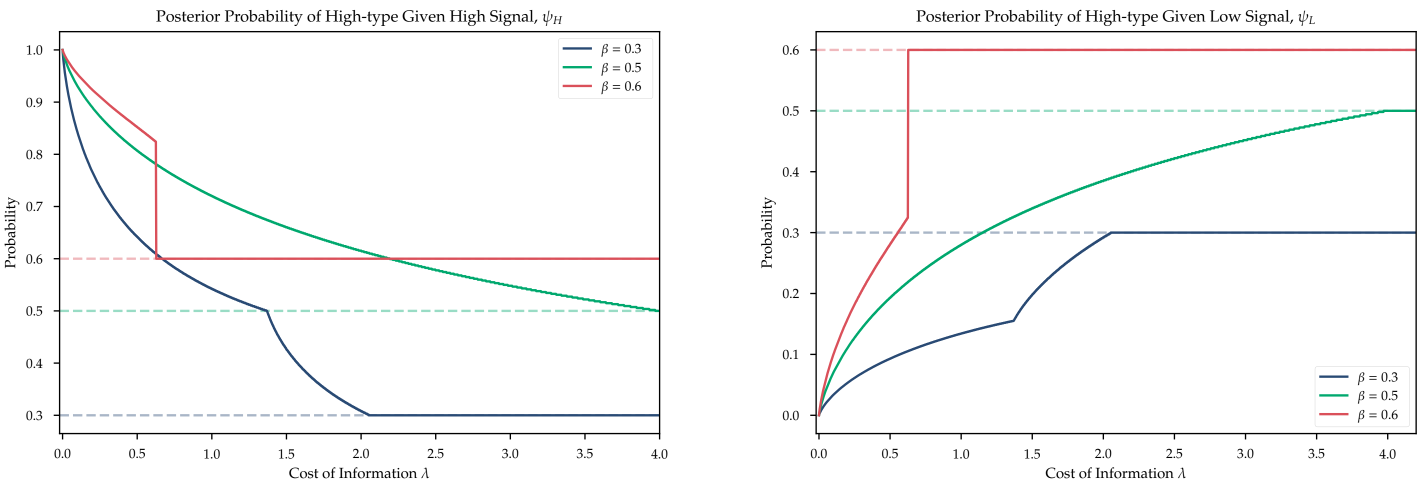
In the case that priors satisfy $\beta = 0.3$, Figure 4b illustrates that, when information is extremely

[25] As required by assumptions 1-3, this functional form satisfies $m(0) = 1/2$ when the firm acquires no information, $\lim_{x\to\infty} m(x) = 1$ when it acquires full information, and $m(x)$ is increasing and concave in x.

costly (high values of λ in the right-hand side of the figure) the firm acquires no information; and thus keeps offering a menu of contracts. When λ decreases sufficiently, the firm starts acquiring information. As discussed in previous sections, when the firm acquires small amounts of information, posterior ψ_H lies below the kink of the $V(\beta)$ function, which helps the firm reduce rents and increase its expected profits. As information becomes cheaper (i.e., $\lambda < 1.37$), posterior ψ_H lies above the kink, leading the firm to ignore the low type upon receiving a high signal. Intuitively, when $\lambda > 1.37$, cheaper information allow the firm to receive signals that help adjust both the offers to low- and high-type consumers. However, when $\lambda < 1.37$, cheaper information only leads the firm to adjust the offers to the low-type buyer upon receiving a low-type signal. In this case, the firm offers the complete information contract to the high-type buyer, ignoring the low type, thus exhausting all profitable adjustments. When β lies at the kink θ_L/θ_H, $\beta = 0.5$, the kink in the $x^*(\lambda)$ function occurs at the endpoint, i.e., when λ is extremely high. In this context, the firm offers a menu upon receiving a low signal but ignores the low type otherwise. When β increases to $\beta = 0.6$, the kink in the $x^*(\lambda)$ function happens at lower values of λ. Intuitively, information must be cheap enough for the firm to acquire a sufficiently large amount of information, so that signals become accurate predictors of consumer types. Upon receiving a low signal, the firm may then be persuaded to offer a menu, rather than ignoring the low type as it did under incomplete information.

Figure 5: **Posterior Probabilities with Optimal x**

(a) (b)

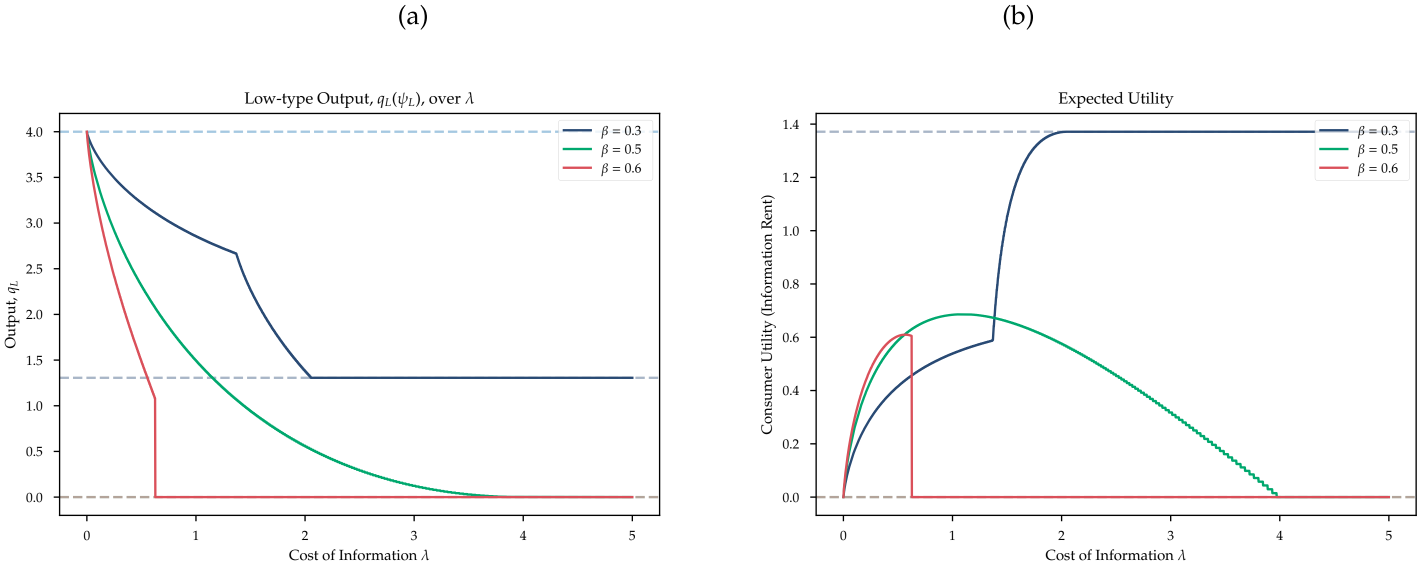


Posteriors. The equilibrium posterior probability of a high-type consumer given a high (low) signal decreases (increases) as the cost of information grows. The behavior of posteriors for priors 0.3, 0.5, and 0.6 is shown in Figure 5. When information is close to free (far left-hand side of the graph), firms acquire sufficient signal reliability to ensure that the posterior probability of a high-type is 1 (0) after observing a high (low) signal.

Equilibrium Output. The equilibrium quantity offered to the low-type buyer generally decreases in the cost of information. Figure 6a shows the behavior of the low-type's quantity under various parameter values. When information is becomes cheaper, incomplete information scenarios where the low-type buyer is ignored (zero output $q_L(\psi) = 0$) change into information acquisition situations where the low-type is offered a contract.

Information Rents The acquisition of information by the firm does not necessarily decrease the expected utility of the high-type consumer as depicted in Figure 6b.

Figure 6: **Expected Utility and Low-type Output**



When this type of buyer is unlikely, $\beta = 0.3$, the firm offers a menu of contracts if information is costly (high λ), providing a large information rent. When information becomes cheaper, however, the firm starts acquiring it, which leads to adjustments in the contract, decreasing this buyer's rents. When information becomes sufficiently cheap, the firm ignores the low-type buyer upon receiving a high signal, which produces further reductions in the high-type buyer expected rent, ultimately becoming zero when the firm is fully informed. Therefore, a better informed firm yields a decrease in expected rents in this context. This argument does not necessarily apply when priors are higher. Specifically, when $\beta = 0.5$, the firm ignores the low-type, leaving the high-type buyer with no information rents in the incomplete information context (which is equivalent to extremelly high λ). When information becomes cheaper, the firm acquires a positive amount, offering a contract menu upon receiving a low signal, which entails a positive information rent to the high-type buyer. As the firm acquires more information, signal reliability improves, reducing the probability that the firm receives a low signal, ultimately reducing rents. Overall, when high-type buyers are likely, they may have incentives for the firm to acquire some, but not full, information.[26]

Expected Profits The expected profits of the firm, are, not surprisingly, decreasing in the cost of information. When information is expensive, the firm chooses not to acquire, achieving the profits from the standard incomplete information problem (dotted horizontal lines in Figure 7). When information is inexpensive, the firm acquires information that permits more optimal design of its bundles leading to greater expected profits until the firm reaches the level of profits achieved under complete information, as a special case (dashed horizontal lines in Figure 7).

Welfare. Figure 8a summarizes our above results, depicting expected utility, profits, and welfare in the case of low priors, $\beta = 0.3$. As discussed above, expected utility unambigously decreases in this setting, since the firm moves from offering a contract menu to adjusting it

[26] A similar argument applies when priors are higher, at $\beta = 0.6$, whereby the firm ignores the low-type buyer under incomplete information. In this setting, however, the firm needs information to be extremely cheap to start acquiring a positive amount of it. At that point, the high-type buyer can capture an information rent when the firm receives a low signal. As information becomes cheaper, signal reliability in this context is high enough to entail an unambiguous reduction in the rents to the high-type buyer. Graphically, this buyer's expected utility decreases as λ decreases, as opposed to his expected utility when $\beta = 0.5$ which exhibited an increasing followed by a decreasing section.

Figure 7: **Expected Profits and Market Welfare**

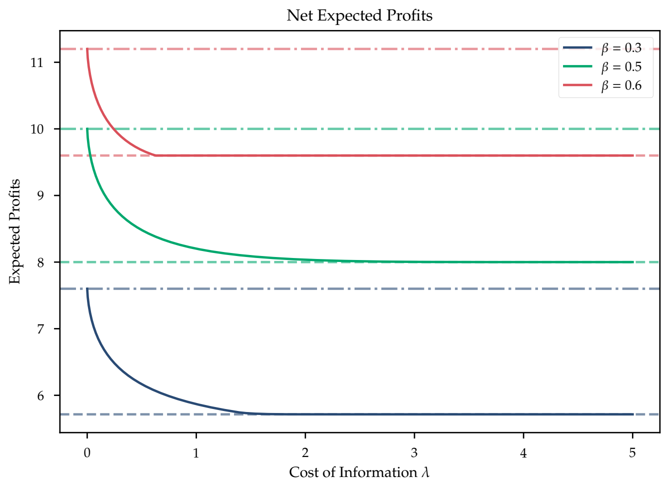
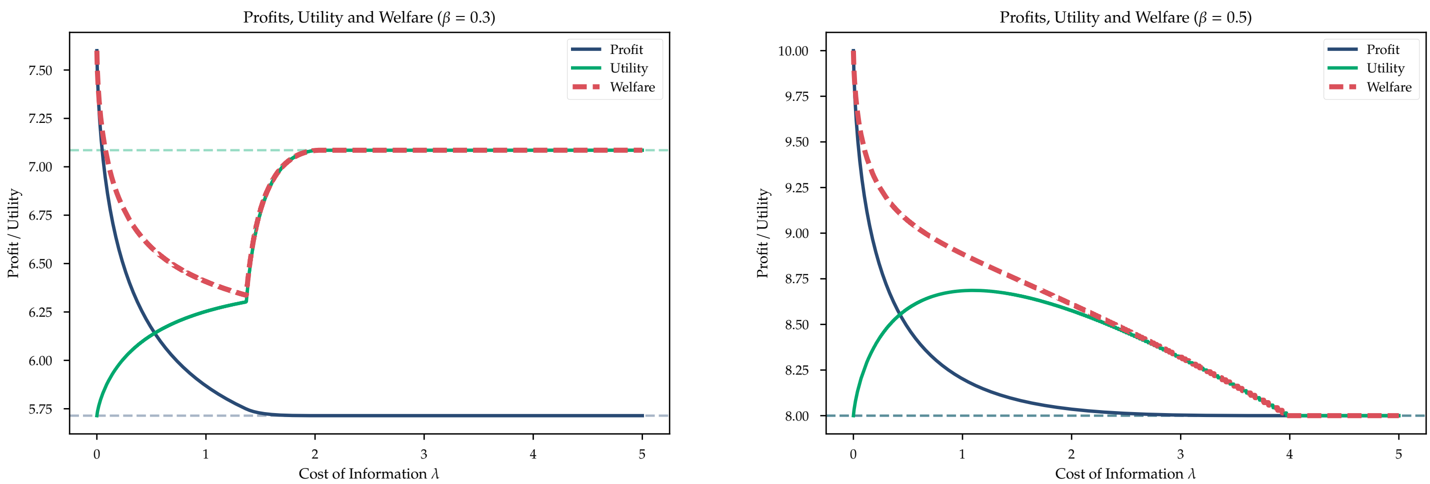


Figure 8: **Firm vs Consumer Welfare**

(a) (b)



(as signals become more reliable), and to ignoring the low-type buyer entirely. Profits, in contrast, unambigously increase as information becomes cheaper. Overall, expected welfare decreases when λ is relatively high (reflecting that expected utility decreases more than profits increase), but eventually increases when λ becomes lower since the increase in profits offset the decrease in utility. Figure 8b depicts a similar comparison, but when priors are relatively high, $\beta = 0.6$. In this context, the firm ignores the low-type buyer under incomplete information but, as information becomes cheaper, it offers a menu upon receiving a low signal, which increases both the consumer's expected utility and profits. When information is sufficiently cheap, however, signals are realiable enough to decrease expected rents; entailing that the overall increase in welfare is driven by increases in profits that offset reductions in utility.

5 Application to Entropy-based Costs

In this section, we apply our model to a cost function $C(x)$ where the cost of acquiring more informative signals is a function of entropy, and evaluate our equilibrium results in that context. The rational inattention literature often considers cost function $C_1(x) = \lambda I(\Theta; S)$, where $\lambda \geq 0$. Term $I(\Theta, S)$, as described in the previous section, denotes mutual information; see Maejka and McKay (2015) and (Cover and Thomas, 2006). In this function, the cost of acquiring information increases only as the firm increases its mutual information (i.e., receiving more reliable signals). However, it is not convex in information acquisition, x. We then normalizes $C_1(x)$ over the conditional entropy[27] $H(\Theta|S)$ as follows

$$C_2(x) = \lambda \frac{I(\Theta, S)}{H(\Theta|S)}.$$

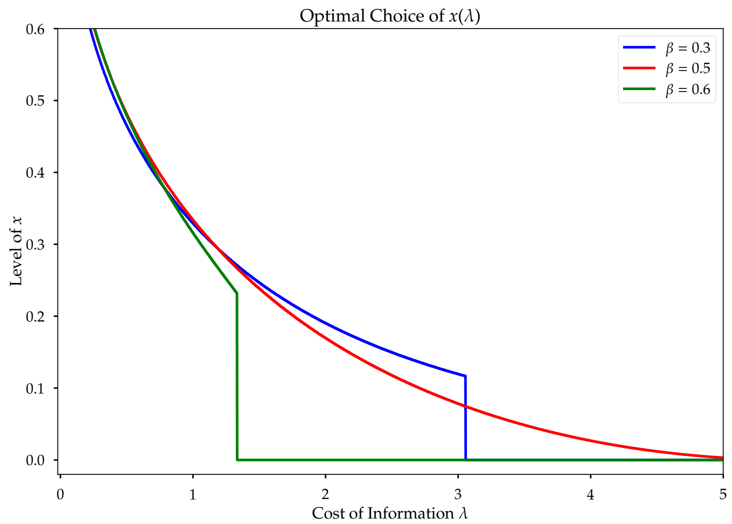
Intuitively, when the firm acquires no information, the mutual information between types and signals, $I(\Theta, S)$, approaches zero, and so does $C_2(x)$. In words, the marginal cost of acquiring the first unit of information is negligible. In contrast, when the firm acquires more information, mutual information $I(\Theta, S)$ increases, approaching $H(\Theta)$. Since $H(\Theta) = I(\Theta; S) + H(\Theta|S)$, term $H(\Theta|S)$ approaches zero in this setting. As a result, when the firm is almost perfectly informed, the cost of acquiring one more unit of information becomes infinite. Overall, cost function $C_2(x)$ is convex in information acquisition, x; as required. In this setting, the marginal cost of acquiring information (the right-hand term in first-order conditions (9)-(11) in Proposition 1) is

$$C_2'(x) = -\lambda \frac{H(\Theta)}{H(\Theta|S)^2} \cdot \frac{\partial H(\Theta|S)}{\partial m} \cdot m'(x) \tag{15}$$

Figure 9 re-evaluates the firm's optimal choice of x at cost function $C_2(x)$. Our results in Figure 4b are then robust to different cost functions, such as $C_2(x)$. For high prior beliefs ($\beta = 0.5$ and $\beta = 0.6$), the optimal quantity of information acquired by the firm remains essentially unaffected. In the case of low priors ($\beta = 0.3$), however, the firm finds it optimal to not acquire information until it becomes cheap enough to acquire a large block of it (enough to ensure that $\psi_H(x) > \hat{\theta}$).

[27] Conditional entropy is the uncertainty about consumer types that remains after the firm receives a signal. In particular, if $H(\Theta)$ is the entropy (i.e., the uncertainty about consumer types that the firm faces prior to receiving a signal), we can write $H(\Theta) = I(\Theta; S) + H(\Theta|S)$. See (Cover and Thomas, 2006) for references.

Figure 9: **Simulation of Entropy-based Cost**



6 Discussion

Stronger Demand Our results help us identify the effect of demand intensity on contract offers. Under standard nonlinear pricing models with no information acquisition, an increase in $\hat{\theta}$ makes the low-type buyer more attractive for the firm, expanding (shrinking) the region of parameter values for which the firm offers a menu of contracts (ignores the low-type buyer, respectively). Graphically, an increase in $\hat{\theta}$ shifts the kink in the $V(\beta)$ function rightward, shrinking the constant segment of this curve. When information acquisition is allowed, however, our results show that an increase in $\hat{\theta}$ induces the firm to acquire positive amounts of information under larger parameter conditions. Intuitively, the low-type buyer becomes more attractive as his demand grows, leading the firm to offer a menu under larger parameter conditions; a menu with a small output distortion for the low type, entailing small information rents for the high type. The opposite argument applies when $\hat{\theta}$ decreases. In this context, the low-type buyer assigns a low value to the good, driving the firm to ignore him under larger parameter conditions.

Privacy Laws Our findings also help understand the effect of policies that facilitate the identification of consumer types, such as mandatory disclosing laws (e.g., salary of public employees, or price of a purchased property). For a given amount of information received by x, these policies make posterior beliefs more extreme; alternatively, a given increase in x moves ψ_L and ψ_H closer to the endpoints, 0 and 1, thus indicating that the information that the firm acquired becomes more reliable. When priors are sufficiently high, $\beta > \hat{\theta}$, the region where the firm ignores the low type shrinks. Intuitively, signals become informative for lower values of x, driving the firm to offer a menu of contracts under larger parameter conditions. The opposite argument applies when priors are relatively low. In this case, the region where the firm offers a menu shrinks, as signals about high-type buyers become more reliable. Overall, the effect of these policies is not unambiguous: they can provide consumers with larger information rents when firms priors are relatively high, but are more likely to keep all information rents in the firm's hands when priors are low.

Two ways to separate consumers Welfare behavior can be better understood by considering that the firm can separate consumer types in two ways: paying consumers an information rent;

or investing resources in an information structure to receive an outside signal. When the cost of signals is high, it is cheaper to just pay an information rent. In this context, the firm doesn't acquire any information, but instead pays consumers to self-select. This situation can be welfare improving for consumers if the firm offers a menu, but comes at a cost to the firm. When information costs are intermediate, the firm may find it profitable to use a combination of investing in a signal and paying consumers to separate themselves. In this case consumers are likely to benefit because, even if the firm would ordinarily ignore the low-type, acquisition of partial information can lead the firm to offer a menu and thus an information rent. The payment of the information rent represents a transfer from the firm to consumers. In addition, note that the distortion of the low-type's contract is a deadweight loss, since the profit loss that the firm suffers from the distortion is not transferred to consumers. As the firm acquires moderate quantities of information, the expected distortion can decrease; leading to welfare gains. Finally, when signals are cheap, the firm finds investing in a signal to be more profitable than paying information rents, acquiring a great deal of information. As information approaches completeness, consumer utility can decrease in overall level as the firm saves on information rents. Hence, the increases in overall welfare associated with the cheapest information costs is driven primarily by gains in expected profits as the firm funnels its payments toward information structure investment instead of information rents.

7 Conclusions

Our results show that, as the cost of information decreases (a growing trend in recent years), firms acquire more information about consumer's preferences. One might suspect that, as the monopolist becomes better informed, consumers lose because the high-type consumer's information rent is reduced. Counter to this intuition, consumer welfare can actually increase in expectation when information cost is not prohibitively high, nor exceedingly low. High-type consumer's benefit from the "bad predictions" generated by an information structure with positive probability. The monopolist does not mind the potential bad signals as profits are increased in expectation.

Overall market welfare increases with declining costs of information and the associated increase in the quantity of information acquired. However, such increases in welfare are attributed to either large gains in utility to the high-type consumer, or a large offset of consumer utility with firm profits. When information has a medium cost, the monopolist acquires partial information about types and this benefits the high-type consumer while also providing the firm with a modest increase in profits. However, as information becomes cheaper, the firm's increased information acquisition begins to extract more surplus from consumers while reducing the probability of a bad signal. Consumer welfare drops off dramatically toward zero and firm profits increase enough to more than offset the loss to consumers.

Further research. Our model considers, for simplicity, a model with two types of customer. As we discussed, under certain conditions the firm may choose to ignore the low-type buyer to extract as much surplus as possible from the high-type customer. Our setting can be extended to more types of buyers, where the firm may also ignore the customer/s with the lowest valuation. In adition, the model could allow firms to acquire information from two sources (x and y, rather than x alone), and analyze whether the firm chooses to combine information from both sources or focus all its resources into a single source. Firms often hire reports from more than one consultant, with the goal to contrast information across sources, so this extended model could better fit observed firm behavior. Alternatively, our model could consider several firms offering menus to customers. If the goods they sell cannot be resold (so arbitrage is impossible), such a setting resembles a common agency model in which two firms (principals) offer menus to a common agent (customer) whose type they cannot observe.

Proof of Lemma 1

The monopolist solves

$$\max_{q,T} \; T - cq \qquad s.t.\; \theta_k u(q) - T \geq 0 \; (PC)$$

The participation constraint (PC) holds with equality, otherwise the firm will not maximize profit as they could strictly increase the price of the contract and still satisfy the constraint. Hence, the seller's problem collapses to,

$$\max_{q} \; \theta_k u(q) - cq$$

which yields $\theta_k u'(q) = c$. Therefore, the monopolist designs a contract (q_k^*, T_H^*) for every consumer k, where $T_k^* = \theta u(q_k^*)$ and q_k^* solves $\theta u'(q_k^*) = c$. ■

Proof of Lemma 2

The monopolist achieves self-selection and participation of consumers by satisfying conditions,

$$\theta_L u(q_L) - T_L \geq 0 \qquad \text{P.C.}_L \tag{16}$$
$$\theta_H u(q_H) - T_H \geq 0 \qquad \text{P.C.}_H \tag{17}$$
$$\theta_L u(q_L) \geq \theta_L u(q_H) - T_H \qquad \text{I.C.}_L \tag{18}$$
$$\theta_H u(q_H) \geq \theta_H u(q_L) - T_L \qquad \text{I.C.}_H \tag{19}$$

The monopolist needs to ensure that the low-type just participates, and that the high-type's incentive compatibility constraint holds exactly. As a result constraints 16 and 19 hold with equality and the monopolist chooses $T_L = \theta_L u(q_L)$ and

$$\begin{aligned} T_H &= \theta_H[u(q_H) - u(q_L)] + T_L \\ &= \theta_H[u(q_H) - u(q_L) + \theta_L u(q_L) \end{aligned} \tag{20}$$

Substitution of T_L and expression (20) into the firm's expected profits (3) redefines the firm's constrained optimization problem into an unconstrained one in terms of q_L and q_H.

$$\max_{q_H, q_L} \; \beta \left\{ \theta_H[u(q_H) - u(q_L)] + \theta_L u(q_L) - cq_H \right\} + (1-\beta)\left[\theta_L u(q_L) - cq_L\right] \tag{21}$$

The first order derivatives of the objective functions are,

$$\beta\theta_H u'(q_H) - \beta c = 0$$
$$-\beta\theta_H u'(q_L) + \beta\theta_L u'(q_L) + (1-\beta)\theta_L u'(q_L) - (1-\beta)c = 0$$

Rearranging the derivatives above we obtain the first order conditions for the optimal contracts for $\beta < \theta_L/\theta_H$.

$$\theta_H u'(q_H) = c$$
$$\theta_L u'(q_L) = c + [\theta_H - \theta_L]\, u'(q_L)\frac{\beta}{1-\beta}$$

Note that the second condition can be rearranged as,

$$u'(q_L) = \frac{(1-\beta)c}{\theta_L - \beta\theta_H}$$

From the above expression, it is clear that when $\beta = \theta_L/\theta_H$ the denominator is zero and the condition is undefined at infinity. Because the marginal utility $u'(q_L)$ goes to infinity as $q_L \to 0$ the firm decreases q_L as β approaches θ_L/θ_H from below and then remains at zero for all $\beta > \theta_L/\theta_H$ as the firm cannot offer a negative quantity. ■

Proof of Lemma 3

The objective function evaluated at the optimum as a function of β is

$$\begin{aligned} V(\beta) &= \beta\pi_H[q_H^*, q_L(\beta); \beta] + (1-\beta)\pi_L[q_L(\beta); \beta] \\ &= \beta\left(\theta_H u[q_H^*] - cq_H(\beta) - (\theta_H - \theta_L)u[q_L(\beta)]\right) + (1-\beta)\left(\theta_L u[q_L(\beta)] - cq_L(\beta)\right) \end{aligned}$$

Recall that the shape depends upon the relative value of β to θ_L/θ_H. Consequently, the piece-wise value function is,

$$V(\beta) = \begin{cases} \beta\left(\theta_H u[q_H^*] - cq_H^* - (\theta_H - \theta_L)u[q_L(\beta)]\right) + (1-\beta)\left(\theta_L u[q_L(\beta)] - cq_L(\beta)\right) & \text{if } \beta \leq \theta_L/\theta_H \\ \beta(\theta_H u[q_H^*] - cq_H^*) & \text{if } \beta > \theta_L/\theta_H \end{cases} \tag{22}$$

By the chain rule and the satisfaction of the first order condition for a maximum, we can rely on the envelope theorem to simplify the slope of the value function. The derivative in the case where $\beta \leq \theta_L/\theta_H$ can be simplified by application of the envelope theorem as it is a value function; allowing us to ignore the dependence of q_L on β. This derivative is $V(\beta) = \theta_H u[q_H(\beta)] - (\theta_H - \theta_L)u[q_L(\beta)] - u[q_L(\beta)] - cq_L(\beta)$. When $\beta > \theta_L/\theta_H$ the firm ignores the low-type and offers only the complete information version of the high-type's contract yielding expected profit $\beta\pi_H^*$. For compactness, note that the information rent is $R(\beta) \equiv (\theta_H - \theta_L)u[q_L(\beta)]$ and $\pi_L(\beta) \equiv \theta_L u[q_L(\beta)] - cq_L(\beta)$ and $\pi_H^* = \theta_H u[q_H^*] - cq_H^*$. The piece-wise slopes of $V(\beta)$ are then,

$$\frac{\partial V(\beta)}{\partial \beta} = \begin{cases} \pi_H^* - R(\beta) - \pi_L(\beta) & \text{if } \beta \leq \theta_L/\theta_H \\ \pi_H^* & \text{if } \beta > \theta_L/\theta_H \end{cases} \tag{23}$$

The sign of $V'(\beta)$ is clearly positive when $\beta > \theta_L/\theta_H$, but in the case $\beta \leq \theta_L/\theta_H$ the sign is less clear. We require that $\pi^H \geq R(\beta) + \pi_L(\beta)$.

$$
\begin{aligned}
R(\beta) + \pi_L(\beta) &\leq \pi^*_H \\
(\theta_H - \theta_L)u[q_L(\beta)] + \theta_L u[q_L(\beta)] - cq_L(\beta) &\leq \pi^*_H \\
\theta_H u[q_L(\beta)] - cq_L(\beta) &\leq \pi^*_H
\end{aligned}
\tag{24}
$$

Consider that $\pi^*_H = \theta_H u[q^*_H] - cq^*_H$ and that q^*_H is the quantity that maximizes profit on the high-type's complete information contract. In other words, q^*_H solves $\theta_H u'[q_H] = c$. We now that for any β, we have $q_L(\beta) < q^*_H$ which implies that

$$\theta_H u[q_L(\beta)] - cq_L(\beta) < \theta_H u[q^*_H] - cq^*_H$$

and therefor the condition $R(\beta) + \pi_L(\beta) < \pi^*_H$ holds for all $\beta \in [0, \theta_L/\theta_H]$ which demonstrates that the first derivative of the value function with respect to β is strictly positive, proving statement 1.

We now look at the second derivative, noting that q^*_H is not a function of β.

$$
\begin{aligned}
\frac{\partial^2 V(\beta)}{\partial \beta^2} &= -\theta_H u'[q_L(\beta)]\frac{\partial q_L(\beta)}{\partial \beta} + c\frac{\partial q_L(\beta)}{\partial \beta} \\
&= -\underbrace{\left(\frac{\partial q_L(\beta)}{\partial \beta}\right)}_{-} \cdot \underbrace{(\theta_H u'[q_L(\beta)] - c)}_{+}
\end{aligned}
$$

The complete information case tells us that $\theta_H u'(q^*_H) - c = 0$ and since $q_L(\beta) < q^*_H$ and $u(\cdot)$ is concave, we have $u'[q_L(\beta)] > u'[q^*_H]$ which implies that $\theta_H u'[q_L(\beta)] - c > 0$. In addition, we have $q_L(\beta)$ decreasing toward zero as β increases toward θ_L/θ_H, leading to an overall positive derivative. When considered as a univariate function of β, since the value function $V(\beta)$ has positive first and second derivatives in the interval $[0, \theta_L/\theta_H]$ it is increasing and strictly convex over this range – proving statement 2.

When $\beta > \theta_L/\theta_H$, the firm will ignore the low-type, setting $q_L(\beta) = 0$. As previously noted, $V'(\beta) = \pi^H$ in this case implying that the second derivative in the region $\beta \in (\theta_L/\theta_H, 1)$ is zero. Therefore, the value function is strictly positive over the interval $(\theta_L/\theta_H, 1]$ and has constant slope equal to $\theta_H u[q^*_H] - cq^*_H$ – proving the final result. ■

7.1 Proof of Lemma 4

Let $V[\psi_k(x)]$ be the value to the firm of having posterior belief $\psi_k(x)$ when the firm has chosen x, then

$$
V[\psi_k(x)] = \begin{cases} \psi_k(x)\pi_H[\psi_k(x)] + [1 - \psi_k(x)]\pi_L[\psi_k(x)] & \text{if } \psi_k(x) \leq \hat{\theta} \\ \psi_k(x)\pi^*_H & \text{if } \psi_k(x) > \hat{\theta} \end{cases}
$$

The probability of the firm receiving $V[\psi_k(x)]$ is the marginal probability, $\rho_k(x)$, of receiving signal s_k. The *ex-ante* expected profits to the firm from choosing the information structure associated with x, is

$$B(x) = \begin{cases} \rho_H(x)V[\psi_H(x)] + \rho_L(x)V[\psi_L(x)] & \text{if } \psi_L(x), \psi_H(x) \leq \hat{\theta} \\ \rho_H(x)\psi_H(x)\pi_H^* + \rho_L(x)V[\psi_L(x)] & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ \beta\pi_H^* & \text{if } \hat{\theta} < \psi_L(x), \psi_H(x) \end{cases}$$

In the last case, the firm ignores the low-type after either signal (s_H and s_L) and the derivatives over this section of the piece-wise function are trivial. Let $B_n(x)$ correspond to the the benefit function in condition n for the three conditions given above. Then $B_3'(x) = B_3''(x) = 0$. In cases 1 and 2, however, the derivative will be a sum of the derivatives of signal dependent outcomes. We begin by differentiating each term, simplifying it, and combining terms back into the piece-wise derivative function, $B'(x)$.

Derivatives of $\rho_k V(\psi_k)$ We first rephrase the terms $\rho_k V(\psi_k)$ in terms of the sum of joint probability weighted contract profits. As noted in Table (1 the joint probabilities are specified by a value of x so that we can rewrite the last expression in terms of β and $m(x)$. For example, we have

$$\rho_H(x)\psi_H(x) = m(x)\beta \qquad \rho_H(x)[1 - \psi_H(x)] = [1 - m(x)](1 - \beta)$$
$$\rho_L(x)\psi_L(x) = [1 - m(x)]\beta \qquad \rho_L(x)[1 - \psi_L(x)] = m(x)(1 - \beta)$$

Using the above forms of the joint probabilities, we rewrite the terms $\rho_k V(\psi_k)$ as

$$\rho_H(x)V[\psi_H(x)] = m(x)\beta\pi_H[\psi_H(x)] + [1 - m(x)](1 - \beta)\pi_L[\psi_L(x)]$$
$$\rho_L(x)V[\psi_L(x)] = [1 - m(x)]\beta\pi_H[\psi_L(x)] + m(x)(1 - \beta)\pi_L[\psi_L(x)]$$

We now differentiate both equations above to demonstrate the simplification through the firm's second-stage first order conditions.

$$\begin{aligned} \frac{\partial}{\partial x}\left(\rho_H(x)V[\psi_H(x)]\right) &= m'(x)\beta\pi_H[\psi_H(x)] + m(x)\beta\frac{\partial\pi_H[\psi_H(x)]}{\partial q_L}\frac{\partial q_L[\psi_H(x)]}{\partial m}m'(x) \\ &\quad - m'(x)(1 - \beta)\pi_L[\psi_H(x)] + [1 - m(x)](1 - \beta)\frac{\partial\pi_L[\psi_H(x)]}{\partial q_L}\frac{\partial q_L[\psi_H(x)]}{\partial m}m'(x) \end{aligned}$$

collecting the terms involving $\partial\pi_k/\partial q_L$ we have,

$$\begin{aligned} &\rho_H(x)\left(\psi_H(x)\frac{\partial\pi_H[\psi_H(x)]}{\partial q_L} + [1 - \psi_H(x)]\frac{\partial\pi_L[\psi_H](x)}{\partial q_L}\right)\frac{\partial q_L[\psi_H(x)]}{\partial m}m'(x) \\ &= \rho_H\,(0)\,\frac{\partial q_L[\psi_H(x)]}{\partial m}m'(x) \text{ by FOC} \\ &= 0 \end{aligned}$$

From this application of the envelope theorem, the partial derivatives simplifies to

$$\frac{\partial}{\partial x}\left(\rho_H(x)V[\psi_H(x)]\right) = m'(x)\left(\beta\pi_H[\psi_H(x)] - (1 - \beta)\pi_L[\psi_H(x)]\right)$$

A similar approach for $\rho_L V(\psi_L)$ yields,

$$\frac{\partial}{\partial x}\left(\rho_L(x)V[\psi_L(x)]\right) = m'(x)\left(-\beta\pi_H[\psi_L(x)] + (1-\beta)\pi_L[\psi_L(x)]\right)$$

Note that in case two, the derivative $B_2'(x)$ involves the partial derivative of $m(x)\beta\pi_H^*$ plus $\partial/\partial x\left(\rho_L(x)V[\psi_L(x)]\right)$. We can now define the piece-wise derivative function consisting of $B_1'(x), B_2'(x), B_3'(x)$. For compactness, we define $D_k(x,\beta) \equiv \beta\pi_H[\psi_k(x)] - (1-\beta)\pi_L[\psi_k(x)]$.

$$B'(x) = \begin{cases} m'(x)\left[D_H(x,\beta) - D_L(x,\beta)\right] & \text{if } \psi_L(x) < \psi_H(x) < \hat{\theta} \\ m'(x)\left[\beta\pi_H^* - D_L(x,\beta)\right] & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ 0 & \text{if } \hat{\theta} < \psi_L(x) < \psi_H(x) \end{cases}$$

We now sign these derivatives for cases 1 and 2. When $\psi_L(x) < \psi_H(x) < \hat{\theta}$, $D_H(x,\beta) \geq D_L(x,\beta)$ entails

$$\beta[\pi_H(\psi_H) - \pi_H(\psi_L)] \geq (1-\beta)[\pi_L(\psi_H) - \pi_L(\psi_L)] \tag{25}$$

The profit on the high-type contract weakly increases with the probability of a high-type (ψ_H). The profit on the low-type contract weakly increases with the probability of a low-type. Because $\psi_H(x) \geq \psi_L(x)$ for all x, the left-hand side of inequality (25) is positive for values of x such satisfying $\psi_L(x), \psi_H(x) < \hat{\theta}$ while the right-hand side is negative for the same values of x. Therefore the condition always holds and the marginal benefit satsifies $B_1'(x) \geq 0$.

When x satisfies $\psi_L(x) < \hat{\theta} < \psi_H(x)$, the sign of derivative $B_2'(x)$ is determined by $\beta\pi_H^* \geq D_L(x,\beta)$, or

$$\beta[\pi_H^* - \pi_H[\psi_L(x)]] \geq -(1-\beta)\pi_L[\psi_L(x)] \tag{26}$$

Complete information profits are weakly greater than incomplete information profits on the high-type contract. Hence, the left-hand side of (26) is nonnegative for any choice of x. Since the right-hand side of (26) is negative the inequality always holds, and therefore $B_2'(x) \geq 0$. This establishes the fact that $B(x)$ weakly increases in the firm's choice of x.

Second Derivative of Benefit $B(x)$ As we did with the first derivative, we shall proceed case-by-case and then sign each case. For convenience, we define $\mathcal{D}(x) \equiv D_H(x,\beta) - D_L(x,\beta)$. When x satisfies $\psi_L(x) < \psi_H(x) < \hat{\theta}$,

$$B_1''(x) = m''(x)\mathcal{D}(x) + [m'(x)]^2\frac{\partial\mathcal{D}(x)}{\partial m}$$

When x satisfies $\psi_L(x) < \hat{\theta} < \psi_H(x)$, the second derivative is

$$B_2''(x) = m''(x)[\beta\pi_H^* - D_L(x,\beta)] + [m'(x)]^2\left[\frac{\partial D_L(x,\beta)}{\partial m}\right]$$

Therefore, the piece-wise second derivative $B''(x)$ is

$$B''(x) = \begin{cases} m''(x)\mathcal{D}(x) + [m'(x)]^2\left[\frac{\partial\mathcal{D}(x)}{\partial m}\right] & \text{if } \psi_L(x) < \psi_H(x) < \hat{\theta} \\ m''(x)\left[\beta\pi_H^* - D_L(x,\beta)\right] - [m'(x)]^2\left[\frac{\partial D_L(x,\beta)}{\partial m}\right] & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ 0 & \text{if } \hat{\theta} < \psi_L(x) < \psi_H(x) \end{cases}$$

The sign of the partial derivative $\partial\mathcal{D}(x)/\partial m$ will depend upon the signs and magnitudes of the partial derivatives of $D_L(x,\beta)$ and $D_H(x,\beta)$.

When x satisfies $\psi_L(x) \leq \psi_H(x) \leq \hat{\theta}$, the second derivative $B''(x)$ is negative if and only if,

$$m''(x)\mathcal{D}(x) < -[m'(x)]^2 \left[\frac{\partial\mathcal{D}(x)}{\partial m}\right]$$

When x instead satisfies $\psi_L(x) \leq \hat{\theta} \leq \psi_H(x)$, the second derivative $B''(x)$ is negative if and only if,

$$m''(x)[\beta\pi_H^* - D_L(x,\beta)] < [m'(x)]^2 \frac{\partial D_L(x,\beta)}{\partial m}$$

■

7.2 Proof of Proposition 1

Let $B(x) = \rho_H(x)V(\psi_H(x)) + \rho_L(x)V(\psi_L(x))$ be the expected second-stage profits of the firm and $C(x)$ the cost of acquiring the information structure with the informativeness implied by x. Then the firm's objective function is

$$J(x) = B(x) - C(x)$$

Maximization of this function over the domain $[\underline{x},\overline{x}]$ will involve finding a value of x that satisfies a first order condition (critical point), a second order condition (ensure local maximum) and a third order condition (guarantee global maximum).

The general form of the first order condition is

$$B'(x) \geq C'(x)$$

Given the piece-wise nature of the benefit function and the results previously derived in the proof for Lemma 4, we obtain

$$J'(x) = \begin{cases} m'(x)\left[\mathcal{L}_L(x,\beta) - \mathcal{L}_H(x,\beta)\right] \geq C'(x) & \text{if } \psi_L(x), \psi_H(x) \leq \hat{\theta} \\ m'(x)\mathcal{L}_L(x,\beta) \geq C'(x) & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ 0 \geq C''(x) & \text{if } \hat{\theta} < \psi_L(x), \psi_H(x) \end{cases} \tag{27}$$

Recalling the definition $\mathcal{L}_k(x,\beta) = \beta R(\psi_k) + (1-\beta)\pi_L(\psi_k)$ which permits rearranging of the differences $\mathcal{L}_L - \mathcal{L}_H$ and make use of the definitions $\Delta R(x) = R[\psi_L(x)] - R[\psi_L(x)]$ and $\Delta\pi_L(x) = \pi_L[\psi_L(x)] - \pi_L[\psi_H(x)]$. The first order condition can now be expressed,

$$J'(x) = \begin{cases} m'(x)\left[\beta\Delta R(x) + (1-\beta)\Delta\pi_L(x)\right] \geq C'(x) & \text{if } \psi_L(x), \psi_H(x) \leq \hat{\theta} \\ m'(x)\left[\beta R[\psi_L(x)] + (1-\beta)\pi_L[\psi_L(x)]\right] \geq C'(x) & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ 0 \geq C''(x) & \text{if } \hat{\theta} < \psi_L(x), \psi_H(x) \end{cases}$$

Again, recall from Lemma 4 that $B'(x) = m'(x)\Gamma(x)$ where we let $\Gamma(x)$ be defined as,

$$\Gamma(x)=\begin{cases}\beta\Delta R(x)+(1-\beta)\Delta\pi_L(x) & \text{if } \psi_L(x),\psi_H(x)\leq\hat{\theta}\\ \beta R[\psi_L(x)]+(1-\beta)\pi_L[\psi_L(x)] & \text{if } \psi_L(x)<\hat{\theta}<\psi_H(x)\\ 0 & \text{if } \hat{\theta}<\psi_L(x),\psi_H(x)\end{cases}$$

Using this notation we define the second derivative of the objective function as,

$$J''(x)=m''(x)\Gamma(x)+[m'(x)]^2\frac{\partial\Gamma(x)}{\partial m(x)}-C''(x)$$

where the term $\Gamma'(x)$, like $\Gamma(x)$ depends upon the values ψ_H and ψ_L relative to $\hat{\theta}$.

$$\Gamma'(x)=\begin{cases}\beta\Delta R'(x)+(1-\beta)\Delta\pi_L'(x) & \text{if } \theta_L(x),\theta_H(x)<\hat{\theta}\\ \beta R'[\psi_L(x)]+(1-\beta)\pi_L'[\psi_L(x)] & \text{if } \theta_L(x)<\hat{\theta}<\psi_H(x)\\ 0 & \text{if } \hat{\theta}<\psi_L(x),\psi_H(x)\end{cases}$$

where the following derivatives can be expanded,

$$\Delta R'(x)=m'(x)(\theta_H-\theta_L)\left\{\frac{\partial u[q_L(\psi_L[x])]}{\partial q_L}\cdot\frac{q_L(\psi_L[x])}{\partial\psi_L}\cdot\frac{\partial\psi_L(x)}{\partial m}-\frac{\partial u[q_L(\psi_H[x])]}{\partial q_L}\cdot\frac{\partial q_L(\psi_H[x])}{\partial\psi_H}\cdot\frac{\partial\psi_H(x)}{\partial m}\right\}$$
$$\Delta\pi_L'(x)=m'(x)\left\{\left[\theta_L\frac{\partial u[q_L(\psi_L[x])]}{\partial q_L}-c\right]\frac{\partial q_L[\psi_L(x)]}{\partial\psi_L}\cdot\frac{\partial\psi_L(x)}{\partial m}-\left[\theta_L\frac{\partial u[q_L(\psi_H[x])]}{\partial q_L}-c\right]\frac{\partial q_L[\psi_H(x)]}{\partial\psi_H}\cdot\frac{\partial\psi_H(x)}{\partial m}\right\}$$

For any x satisfying $\psi_L(x)\leq\hat{\theta}\leq\psi_H(x)$ to be a local maximum of the objective function it must satisfy,

$$m'(x)[\beta\Delta R(x)+(1-\beta)\Delta\pi_L(x)]\geq C'(x)$$
$$m''(x)[\beta\Delta R(x)+(1-\beta)\Delta\pi_L(x)]+[m'(x)]^2[\beta\Delta R'(x)+(1-\beta)\Delta\pi_L'(x)]\leq C''(x)$$

For any x satisfying $\psi_L(x)\leq\hat{\theta}\leq\psi_H(x)$ to be a local maximum of the objective function it must satisfy,

$$m'(x)[\beta R[\psi_L(x)]+(1-\beta)\pi_L[\psi_L(x)]]\geq C'(x)$$
$$m''(x)[\beta R[\psi_L(x)]+(1-\beta)\pi_L[\psi_L(x)]]+[m'(x)]^2[\beta R'[\psi_L(x)]+(1-\beta)\pi_L'[\psi_L(x)]]\leq C''(x)$$

The only x which can satisfy the first order conditions when x satisfies $\hat{\theta}\leq\psi_L(x)\leq\psi_H(x)$ is $x=\underline{x}$, implying no information acquisition.

Finally, because the firm always has the option to acquire no information at no cost (i.e, $x=\underline{x}$ and $C(\underline{x})=0$), any x which is a local interior maximum of the objective function will need to be compared to the profits the firm receives when acquiring no information to ensure global optimality. Hence, if x^* satisfies the first and second order conditions for the particular regions its posteriors $\psi_L(x^*)$ and $\psi_H(x^*)$, it must also satisfy the global condition, $J(x^*)\geq J(\underline{x})$. ■

7.3 Proof of Corollary 1

Let $C_1(x)$ and $C_2(x)$ be two cost functions satisying the conditions assumed in section 3.3.2 and in addition we have

$$C_1(x) > C_2(x) \text{ for all } x.$$

Letting $X \subset \mathbb{R}^+$ denote the domain of $C_j(x)$ consider the number $\lambda > 0$ such that $\lambda \equiv \min_{\mathbb{R}^+} \{C_1(x) - C_2(x) : x \in X\}$, so that $\lambda > 0$ is the minimum difference in the cost functions. Then we can construct the cost function $\hat{C}(x) \equiv \lambda C_2(x)$ where it is also true that $C_1(x) \geq \hat{C}(x) > C_2(x)$ for all $x \in X$.

The firm's objective function in terms of $\hat{C}(x)$ is then, $J(x) = B(x) - \lambda C_2(x)$. Note the sign of the following cross-partial derivative of $J(x)$,

$$\frac{\partial^2 J(x)}{\partial x \partial \lambda} = -C'(x) < 0 \quad \text{for all } x$$

This implies that the objective function is *submodular* in λ and, because the domain $X \subset \mathbb{R}^+$ is a lattice, Topkis' theorem ensures that the optimal choice $x(\lambda)$ is nonincreasing in λ.

To complete the result, note that because $\hat{C}(x) > C_2(x)$ we know that if the above result is true for $\hat{C}(x)$ is it also true for $C_1(x) \geq \hat{C}(x)$. Because C_1 and C_2 were arbitrary, this applies to any two such functions. ■

7.4 Proof of Corollary 2

Recall that for any belief β, the contract the firm offers to the low-type extracts all surplus, ensuring the low-type receives zero utility. When ignoring the low-type, the high-type will be offered the complete information contract (q^*_H, T^*_H), leaving with zero utility. However, when the firm does offer a menu, the high-type's contract will involve a positive utility surplus (the information rent) of $R(\beta) = (\theta_H - \theta_L)u[q_L(\beta)]$.

The expected utility over signals and types is equivalent to the expected information rent,

$$\begin{aligned}
\mathbb{E}U &= \sum_j \sum_k p(\theta_k, s_j)U(\theta_k, s_j) = \sum_j p(\theta_H, s_j)U(\theta_H, s_j) \\
&= \sum_j p(\theta_H, s_j)R(\psi_j) = p(\theta_H, s_H)R(\psi_H) + p(\theta_H, s_L)R(\psi_L) \\
&= m(x)\beta R(\psi_H) + [1 - m(x)]\beta R(\psi_L) \\
&= \beta \left(m(x)R[\psi_H(x)] + [1 - m(x)]R[\psi_L(x)] \right)
\end{aligned}$$

Whenever $\psi_H(x) > \hat{\theta}$ the firm ignores the low-type and $R[\psi_H(x)] = 0$ reducing the expected entropy to $\beta[1 - m(x)]R[\psi_L(x)] > 0$. This allows us to describe the consumer's piece-wise expected utility as a function of the firm's choice of x.

$$\mathbb{E}U = \begin{cases} \beta \{m(x)R[\psi_H(x)] + [1 - m(x)]R[\psi_L(x)]\} & \text{if } \psi_L(x) < \psi_H(x) < \hat{\theta} \\ \beta[1 - m(x)]R[\psi_L(x)] & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ 0 & \text{if } \hat{\theta} < \psi_L(x) < \psi_H(x) \end{cases} \tag{28}$$

We now compute the piece-wise change in expected consumer utility as the firm changes its choice x. In the first case, where $\psi_L(x) < \psi_H(x) < \hat{\theta}$ we have,

$$\beta m'(x)\left\{R[\psi_H(x)] + m(x)\frac{\partial R[\psi_H(x)]}{\partial \psi}\cdot\frac{\partial \psi_H(x)}{\partial m} - R[\psi_L(x)] + [1-m(x)]\frac{\partial R[\psi_L(x)]}{\partial \psi}\cdot\frac{\partial \psi_L(x)}{\partial m}\right\}$$

Collecting the terms involving undifferentiated information rents yields,

$$\begin{aligned} R[\psi_H(x)] - m'(x)R[\psi_L(x)] &= (R[\psi_H(x)] - R[\psi_L(x)]) \\ &= -(R[\psi_L(x)] - R[\psi_H(x)]) \\ &= \Delta R(x) \end{aligned}$$

We collect the remaining terms involving the changes in the rents themselves as,

$$\partial \mathcal{R}(x) = m(x)\frac{\partial R[\psi_H(x)]}{\partial \psi}\frac{\partial \psi_H(x)}{\partial m} + [1-m(x)]\frac{\partial R[\psi_L(x)]}{\partial \psi}\frac{\partial \psi_L(x)}{\partial m}$$

In the second case we have

$$\begin{aligned} &= -\beta m'(x)R[\psi_L(x)] + \beta[1-m(x)]\frac{\partial R[\psi_L(x)]}{\partial \psi}\frac{\partial \psi_L(x)}{\partial m}m'(x) \\ &= \beta m'(x)\left[-R[\psi_L(x)] + [1-m(x)]\frac{\partial R[\psi_L(x)]}{\partial \psi}\frac{\partial \psi_L(x)}{\partial m}m'(x)\right] \end{aligned}$$

Combining all the results we have the piece-wise representation of the change in consumer expected utility with a change in the firm's choice of x.

$$\frac{\partial \mathbb{E}U}{\partial x} = \begin{cases} \beta\left[-m'(x)\Delta R(x) + \mathbb{E}[R'(x)|\theta = \theta_H]\right] & \text{if } \psi_L(x) < \psi_H(x) < \hat{\theta} \\ \beta\left[-m'(x)R[\psi_L(x)] + [1-m(x)]\frac{\partial R[\psi_L(x)]}{\partial \psi}\right] & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ 0 & \text{if } \hat{\theta} < \psi_L(x) < \psi_H(x) \end{cases}$$

■

References

Acquisti, Alessandro and Hal R. Varian (2005), "Conditioning prices on purchase history." *Marketing Science*, 24, 367–381.

Adams, William James and Janet L Yellen (1976), "Commodity bundling and the burden of monopoly." *The Quarterly Journal of Economics*, 90, 475–498.

Armstrong, Mark (1996), "Multiproduct nonlinear pricing." *Econometrica*, 64, 51–75.

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris (2015), "The Limits of Price Discrimination." *American Economic Review*, 105, 921–957.

Bergemann, Dirk and Juuso Välimäki (2002), "Information acquisition and efficient mechanism design." *Econometrica*, 70, 1007–1033.

Blackwell, David (1953), "Equivalent comparisons of experiments." *The Annals of Mathematical Statistics*, 24, 265–272.

Bonatti, Alessandro (2011), "Menu pricing and learning." *American Economic Journal: Microeconomics*, 3, 124–163.

Cabrales, Antonio, Olivier Gossner, and Roberto Serrano (2013), "Entropy and the value of information for investors." *American Economic Review*, 103, 360–377.

Chiang, Raymond and Chester S. Spatt (1982), "Imperfect price discrimination and welfare." *The Review of Economic Studies*, 49, 155–181.

Clower, R. W. (1959), "Some theory of an ignorant monopolist." *The Economic Journal*, 69, 705–716.

Courty, Pascal and Hao Li (2000), "Sequential Screening." *Review of Economic Studies*, 67, 697–717.

Cover, Thomas M and Joy A Thomas (2006), *Elements of Information Theory*, 2 edition. Wiley-Interscience, John Wiley and Sons Inc.

Cremer, Jacques and Fahad Khalil (1992), "Gathering information before signing a contract." *The American Economic Review*, 82, 566–578.

Cremer, Jacques, Fahad Khalil, and Jean-Charles Rochet (1998), "Strategic information gathering before a contract is offered." *Journal of Economic Theory*, 81, 163–200.

Cremer, Jacques and Richard P. McLean (1985), "Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent." *Econometrica*, 53, 345–361.

Cremer, Jacques, Yossi Spiegel, and Charles Z. Zheng (2003), "Optimal selling mechanisms with costly information acquisition." *Discussion Paper, Northwestern University*.

Dana, James D., Jr. (2001), "Monopoly price dispersion under demand uncertainty." *International Economic Review*, 42, 649–670.

Dimitrova, Magdalena and Edward E. Schlee (2003), "Monopoly, competition and information acquisition." *International Journal of Industrial Organization*, 21, 1623–1642.

Gentzkow, Matthew and Emir Kamenica (2014), "Costly Persuasion." *The American Economic Review*, 104, 457–462.

Grossman, Sanford J., Richard E. Kihlstrom, and Leonard J. Mirman (1977), "A bayesian approach to the production of information and learning by doing." *The Review of Economic Studies*, 44, 533–547.

Harris, Milton and Artur Raviv (1981), "A theory of monopoly pricing schemes with demand uncertainty." *The American Economic Review*, 71, 347–365.

Hedlund, Jonas (2017), "Bayesian persuasion by a privately informed sender." *Journal of Economic Theory*, 167, 229–268.

Karlin, Samuel and Herman Rubin (1956), "The theory of decision procedures for distributions with monotone likelihood ratio." *The Annals of Mathematical Statistics*, 27, 272–299.

Khalil, Fahad and Jean-Charles Rochet (1998), "Contracts and productive information gathering." *Games and Economic Behavior*, 25, 174–193.

Kihlstrom, Richard E. (1976), "Firm demand for information about price and technology." *Journal of Political Economy*, 84, 1335–1341.

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

Lehmann, E. L. (1988), "Comparing location experiments." *The Annals of Statistics*, 16, 521–533.

Leland, Hayne E. and Robert A. Meyer (1976), "Monopoly pricing structures with imperfect discrimination." *The Bell Journal of Economics*, 7, 449–462.

Lewis, Tracy R and David EM Sappington (1994), "Supplying information to facilitate price discrimination." *International Economic Review*, 35, 309–327.

Li, Hao and Xianwen Shi (2017), "Discriminatory Information Disclosure." *The American Economic Review*, 107, 3363–3385.

Manning, R. (1979), "Market research by a monopolist: A bayesian analysis." *Economica*, 46, 301–306.

Marschak, Jacob and Koichi Miyasawa (1968), "Economic comparability of information systems." *International Economic Review*, 9, 137–174.

Maskin, Eric and John Riley (1984), "Monopoly with incomplete information." *The RAND Journal of Economics*, 15, 171–196.

Mussa, Michael and Sherwin Rosen (1978), "Monopoly and product quality." *Journal of Economic Theory*, 18, 301–317.

Riley, John and Richard Zeckhauser (1983), "Optimal Selling Strategies: When to Haggle, When to Hold Firm." *The Quarterly Journal of Economics*, 98, 267–289.

Roberts, Kevin W. S. (1979), "Welfare considerations of nonlinear pricing." *The Economic Journal*, 89, 66–83.

Salop, Steven (1977), "The noisy monopolist: Imperfect information, price dispersion and price discrimination." *The Review of Economic Studies*, 44, 393–406.

Spiegel, Yossi and Simon Wilkie (2000), "Optimal multiproduct nonlinear pricing with correlated consumer types." *SSRN Working Paper*.

Spulber, Daniel F. (1992), "Optimal nonlinear pricing and contingent contracts." *International Economic Review*, 33, 747–772.

Stiglitz, Joseph E. (1977), "Monopoly, non-linear pricing and imperfect information: The insurance market." *The Review of Economic Studies*, 44, 407–430.

Stokey, Nancy L (1979), "Intertemporal Price Discrimination." *The Quarterly Journal of Economics*, 93, 355–371.

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

Trefler, Daniel (1993), "The ignorant monopolist: Optimal learning with endogeneous information." *International Economic Review*, 34, 565–581.

Varian, Hal R (1985), "Price Discrimination and Social Welfare." *The American Economic Review*, 75, 870–875.

Wilson, Charles A. (1988), "On the optimal pricing policy of a monopolist." *Journal of Political Economy*, 96, 164–176.