Nonlinear Pricing with Costly Information Acquisition*

Brett R. Devine† and Felix Munoz-Garcia‡

School of Economic Sciences
Washington State University

November 16, 2018

Abstract

This paper examines a nonlinear pricing model where the firm can choose to acquire costly information prior to offering contract menus to consumers. Information provides the firm with a signal about consumer types, whose accuracy increases as the firm acquires larger amounts of information. We show that the firm acquires information only if it alters its initial beliefs. We identify how information acquisition changes optimal contracts, profits, information rents, and welfare; showing that information weakly increases profits and can also increase expected utility when information costs are 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 Lei He and other seminar participants at the IIOC 2018, and 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

The literature on price discrimination analyzes 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] While this literature is extensive, it generally 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 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 firms' 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 at all, thus remaining as poorly informed as in standard nonlinear 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 improvement. 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 purchase of data bases, tracking of IP addresses and, generally, any investment seeking to identify consumers willingness to pay for its product. 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 consumer's type, 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.

[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) who uses 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. This differs from our paper in that consumers may act strategically to avoid informative actions. 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).

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 type of buyer. In this scenario, signals that yield high enough posteriors induce no contract changes, and no additional profit gain from the high-type buyer. However, if the firm acquires a large amount of information and receives a low signal, its posterior beliefs will be significantly 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. This happens when information is relatively expensive, and thus the firm cannot acquire a sufficiently large amount to make signals reliable enough to alter its subsequent contract offers.

Finally, we investigate how consumer utility, profits, and overall welfare are affected by the firm's decision to acquire 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 nonlinear pricing model. 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 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 less expensive, signals become more reliable, helping the firm reduce the expected 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 and the European General Data Protection Regulation law which came into effect on May 25th, 2018. 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 individual's 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 seller's 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 the demand for their product 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. After observing a signal from an 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 the benefit of acquiring additional units of information to the monopolist.

The monopolist can also make use of individualized information (rather than aggregated information) about the consumer's maximum willingness to pay. Under first-degree price discrimination the monopolist achieves the highest profit and eliminates any deadweight losses. Hence, overall welfare is increased at the expense of consumers receiving zero surplus. Our results are consistent with this finding when the monopolist becomes perfectly informed.[6]

Under incomplete information, the firm seeks to maximize profits given its prior beliefs through a nonlinear contract.[7] Mechanism design literature, such as (Cremer and Khalil, 1992; Cremer

[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] 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.

[7] 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

and McLean, 1985; Cremer et al., 1998; Khalil and Rochet, 1998; Bergemann and Välimäki, 2002; Szalay, 2009) focuses on nonlinear pricing schemes when agents have incomplete information about their own type and can acquire information. In contrast, we assume consumers have complete information about their types and the monopolist endogenously acquires additional information about consumer types at a cost.

Recently, the literature has expanded to include dynamic price discrimination settings, in which firms acquire information about consumer types through interacting with them repeatedly (Courty and Li, 2000; Acquisti and Varian, 2005; Bonatti, 2011). In relation to our paper, this literature highlights issues involving consumer privacy when the firm dynamically screens customers. While it is possible for consumers to benefit from this behavior, they 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 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 (Roberts, 1979; Chiang and Spatt, 1982; Varian, 1985) as its practice can 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 resulting 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 effects 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, or 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 a cost. The existence and robustness of these Pareto improving outcomes is important for informing the policy debate on firm's use of consumer information and their ability 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. While the outcome and sufficient statistic represent a "signal", the experimental form and sampling procedure represent the "information structure" that generates signals. Blackwell (1953) establishes strict conditions for comparing both the informativeness and value of an information structure.[8] The ordering of information structures arising from Blackwell's results, is, however, incomplete.

While, entropy-based measures of information provide a complete ordering (the informative content of all information structures can be compared), 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

et al., 1998, 2003)

8 In words, Blackwell's conditions for comparing information structures say that 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 with A. Blackwell's framework has been extended by several researchers (Karlin and Rubin, 1956; Lehmann, 1988).

in the informativeness of that structure, as measured by entropy, implying that a more informative structure yields a higher expected utility. In the last section of the paper, we also use entropy-based measures of information to compare the informativeness of information structures. Unlike Cabrales et al. (2013), we show that more informative structures do not necessarily increase a monopolist's expected profits, nor decrease consumer's expected utility.

The remainder of the paper proceeds as follows. We first describe the monopolist's linear prices under complete information, and its nonlinear prices under incomplete information (stage-two) and analyze its properties. We also identify the profit gain from acquiring information about consumer types prior to solving the pricing problem. Section 3 develops the optimal information acquisition problem (stage-one) of the monopolist and also investigates its welfare implications. Sections 4-6 explore comparative statics of our results, applies our analysis to entropy-based costs of information, and discuss our results.

2 Model

Consider a monopolist that sells a product, choosing both the quantity of product in a bundle and its price. Every consumer $k = \{H, L\}$ has utility function

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

where $\theta_H > \theta_L$, q is the quantity of product, T is the price for the quantity (bundle price) and $u(\cdot)$ is an increasing and concave function of the units that the individual consumes, q. Every type k consumer only buys the product if $U_k \geq 0$, yielding profit $T - cq$ for the firm, where $c > 0$ denotes a constant marginal cost, and T enters as the firm's revenue.

2.1 Complete Information

Under complete information, the firm observes consumer type before making a bundle offer, leading to an efficient outcome 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*

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

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. Prior to observing a consumer's types, and the probability of a high-type buyer is $Pr(\theta_H) = \beta$, expected profits are

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

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 $\pi_k^* = \theta_k^2/c$. □

2.2 Incomplete Information

When consumer types are private information, 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. In this setting, the monopolist maximizes its expected profit.

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

subject to standard participation and incentive compatibility constraints, leading to the next result.[9]

Lemma 2. *Under incomplete information, if prior belief β satisfies $\beta \leq \theta_L/\theta_H$, the monopolist offers contracts $\{q_H(\beta), T_H(\beta)\}$ 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. The rent to the high-type consumer is $R(\beta) = (\theta_H - \theta_L)u[q_L(\beta)]$, inducing him to truthfully reveal his type. 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. 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:

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

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

Hence, expected profits from expression (4) becomes

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

which are lower than the expected profit that the firm obtains under complete information (before observing consumer types), $\Pi(\beta) = \frac{1}{c}(\beta\theta_H^2 + (1-\beta)\theta_L^2)$, as found in Example 1. □

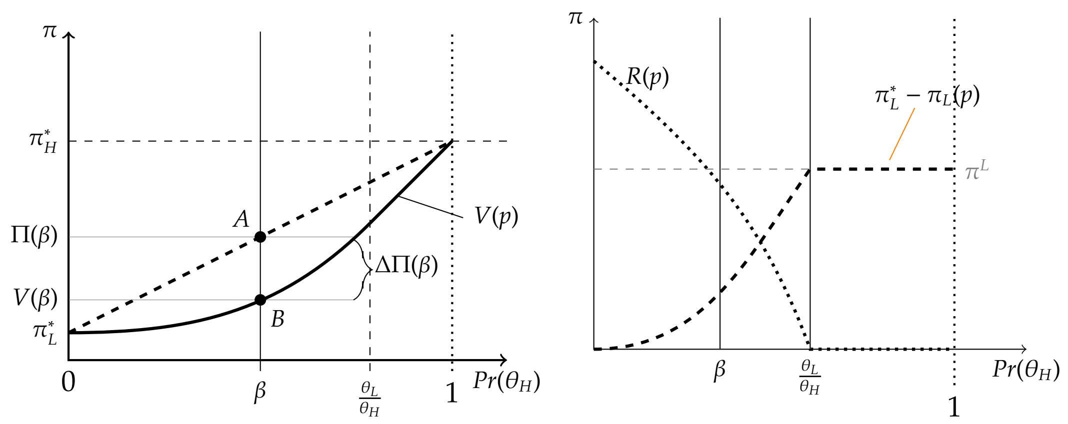
Figure 1a depicts complete information profits π_H^* and π_L^*, and the dashed line connecting them, with expected profit $\Pi(\beta)$ at point A. The figure also illustrates expected profits under incomplete information with function $V(\beta)$ at point B.

The loss of expected profit attributed to incomplete information can thus be measured by the difference $\Delta\Pi(\beta) \equiv \Pi(\beta) - V(\beta)$, as depicted in Figure 1a, and is formally defined as follows.

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

Figure 1a

Figure 1b



$$\Delta\Pi(\beta) = \begin{cases} \beta R(\beta) + (1-\beta)[\pi_L^* - \pi_L(\beta)] & \text{if } \beta \leq \theta_L/\theta_H \\ (1-\beta)\pi_L^* & \text{if } \beta > \theta_L/\theta_H \end{cases} \tag{5}$$

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" paid 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, $\pi_L^* - \pi_L(\beta)$. 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, 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 (that is, 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 as in (Cover and Thomas, 2006, p. 19,20), by

$$I(\Theta, S) = \sum_{\Theta} \sum_{S} f(\theta, s) \log\left(\frac{f(\theta, s)}{g_\theta(\theta) \cdot g_s(s)}\right) \tag{6}$$

The mutual information measures the degree of functional dependence of the random variables Θ and S in units of information, such as bits.[10] 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

[10] 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.

a conditional density $f(s|\theta)$.[11]

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

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.[12] 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.

		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

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 x to conditional probabilities.

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

Intuitively, as $m(x)$ approaches 1, signals tell the truth about consumer types with certainty, while as $m(x)$ approaches zero, always 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.[13]

Assumption 3 (Diminishing Returns)**.** $\lim_{x\to\infty} m'(x) = 0$ 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 as depicted in Figure 2.

3.2 Posterior Beliefs

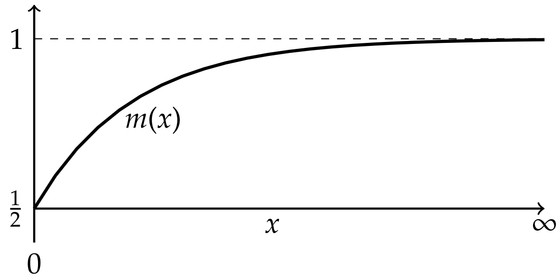
From the left column in Table 1, the marginal probability that the monopolist receives signal s_H is $\rho_H(x) = m(x)\beta + [1 - m(x)](1 - \beta)$, while the probability of signal s_L is $\rho_L(x) = [1 - m(x)]\beta + m(x)(1 - \beta)$,

[11] 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.

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

[13] 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.

Figure 2: Signal Accuracy



as illustrated in the right column of the table. When the monopolist doesn't acquire information, $x = 0$, we obtain $m(0) = 1/2$, yielding probabilities $\rho_H = \rho_L = 1/2$, regardless of the prior β. Conversely, when $x \to \infty$ (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.[14]

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

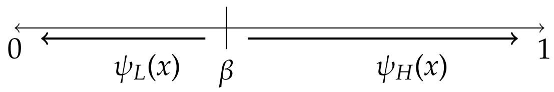
$$\psi_H(x) = \frac{m(x)\beta}{\rho_H(x)} \quad \text{and} \quad \psi_L(x) = \frac{[1 - m(x)]\beta}{\rho_L(x)}$$

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. Conversely, the probability of a high-type consumer after receiving a low-type signal decreases. Intuitively, a more accurate information structure makes signals more informative. In addition, when $x = 0$, signals are uninformative and posterior probabilities are equivalent to prior probabilities.[15] As x increases, the posterior probabilities "separate" away from the prior probability, β; as illustrated in Figure 3.

[14] 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 $\rho_H(x) = \rho_L(x) = 1/2$.

[15] 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.

Figure 3: Separation of Posterior Probabilities



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 (as if information acquisition were 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.[16]

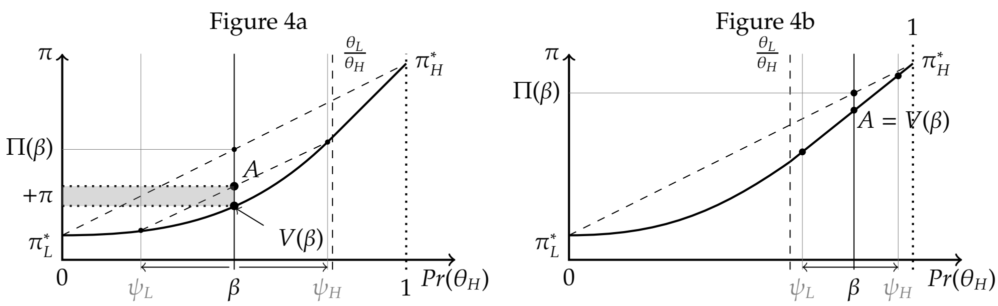
3.3.1 Benefit of Increasing x.

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

$$B(x) = \rho_H(x)V[\psi_H(x)] + \rho_L(x)V[\psi_L(x)] \tag{7}$$

where the first term indicates the highest expected profits the monopolist can anticipate 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 4: **Expected Profit Gain from Information**.



Low Priors. Figure 4a 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 probability β, or π_L^* with probability $1-\beta$, entailing expected profits of $\Pi(\beta)$. Under incomplete information (with no information acquisition) 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 4a.

[16] 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.

High Priors. Figure 4b 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 unaffected, i.e., $\mathbb{E}[V(\psi_k)] = V(\beta)$. Graphically, this occurs when posteriors ψ_L and ψ_H remain above the cutoff, θ_L/θ_H, yielding profits in the linear section of the $V(p)$ function. This indicates that information acquisition produces an increase in 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., $B'(x) = 0$. The benefits of information acquisition originate from two primary sources: (1) an increase in the likelihood of correct signals (leading to better identification of type); and (2) gains from adjusting the firm's menus to the new posterior beliefs.

Profits from Correct and Incorrect Signals. We say a signal s_k is *correct* if the true consumer type is θ_k and is *incorrect* otherwise. If signal s_k is correct, then the firm's posterior beliefs ψ_k will induce menu adjustments yielding profits $\pi_k[\psi_k(x)]$, that is, the profit from the θ_k type consumer when offered the menu associated with posterior ψ_k. Otherwise when s_k is incorrect, the true type is θ_j and profits will be $\pi_j[\psi_k(x)]$ for some $j \neq k$. Given a choice of x, the firm offers contract menus yielding an expected profit $\mathcal{G}(x)$ from correct signals and an expected profit $\mathcal{L}(x)$ from incorrect signals, where[17]

$$\mathcal{G}(x) = \begin{cases} \beta\pi_H[\psi_H(x)] + (1-\beta)\pi_L[\psi_L(x)] & \text{if } \psi_L(x) < \psi_H(x) \leq \theta_L/\theta_H \\ \beta\pi_H^* + (1-\beta)\pi_L[\psi_L(x)] & \text{if } \psi_L(x) < \theta_L/\theta_H < \psi_H(x) \\ \beta\pi_H^* & \text{if } \theta_L/\theta_H < \psi_L(x) < \psi_H(x) \end{cases} \tag{8}$$

$$\mathcal{L}(x) = \begin{cases} \beta\pi_H[\psi_L(x)] + (1-\beta)\pi_L[\psi_H(x)] & \text{if } \psi_L(x) < \psi_H(x) \leq \theta_L/\theta_H \\ \beta\pi_H[\psi_L(x)] & \text{if } \psi_L(x) < \theta_L/\theta_H < \psi_H(x) \\ \beta\pi_H^* & \text{if } \theta_L/\theta_H \leq \psi_L(x) < \psi_H(x) \end{cases} \tag{9}$$

When both posterior beliefs lie in the convex portion of the value function ($\psi_L, \psi_H < \theta_L/\theta_H$), the firm offers a full menu after either signal. This strategy results in the firm receiving positive profit from both correct and incorrect signals after either type of signal. When posteriors satisfy $\psi_L(x) < \theta_L/\theta_H < \psi_H(x)$, the high-type signals are sufficiently convincing that the firm ignores the low-type and offers only the high-type's complete information contract. After signal s_L the firm believes the high-type to be sufficient unlikely that it offers a menu of contracts to screen the two types. In this case both types of correct signals result in profits to the firm, but not with incorrect signals. Unlike the first case, an incorrect s_H causes the firm to ignore the low-type and receive zero profit. Finally, when both posteriors lie in the linear portion of the value function ($\theta_L/\theta_H < \psi_L(x) < \psi_H(x)$), the firm will ignore the low-type after both s_H and s_L and, regardless of whether the signal is correct or incorrect, the firm receives expected profit $\beta_H\pi_H^*$.

The ex-ante expected profits, $B(x)$, can then be rewritten as the expected value of correct signal profits and incorrect signal profits.

$$B(x) = m(x)\mathcal{G}(x) + [1 - m(x)]\mathcal{L}(x)$$

[17] The piece-wise nature of the functions above is a result of the firm's beliefs over consumer types and the resulting optimal nonlinear pricing scheme as specified in Lemma 2.

3.3.2 Marginal Benefit of Increasing x.

The next lemma identifies the marginal benefit of increasing x, $B'(x)$, and under which conditions such benefit is positive and decreasing.

Lemma 4. *The marginal benefit from increasing x, $B'(x)$, is*

$$B'(x) = m'(x)\,[\mathcal{G}(x) - \mathcal{L}(x)] \tag{10}$$

where $B'(x) \geq 0$ for all x, and $B''(x) \leq 0$ if and only if

$$\frac{m'(x)}{1 - m(x)}\mathcal{G}'(x) \leq -m''(x)\,[\mathcal{G}(x) - \mathcal{L}(x)] \tag{11}$$

Intuitively, the marginal benefit of increasing x comprises two positive effects: the marginal value of accuracy for correct signals, $m'(x)\mathcal{G}(x)$, and the marginal value of accuracy for incorrect signals, $m'(x)\mathcal{L}(x)$, which are both positive and generally decreasing in x. To understand the above result, first note that a small acquisition in information (x near zero) diminishes marginal accuracy (m' decreases), but the expected profit of that accuracy depends on the profit difference between correct and incorrect signals, $\mathcal{G}(x) - \mathcal{L}(x)$. If this profit difference increases faster than the decrease in marginal accuracy, then the overall marginal benefit of information increases, i.e., $B''(x) > 0$. If the firm acquires enough information, the rate at which marginal accuracy diminishes overcomes the rate at which profit differences increase, leading to decreasing marginal benefits, i.e., $B''(x) < 0$.

3.3.3 Cost of Increasing x.

Investments in signal informativeness (higher x) represent monetary 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 (Matějka and McKay, 2015).

3.3.4 Ex-Ante Profit Maximization

The monopolist's optimal information acquisition involves a trade-off between larger second-stage expected profits (increase in expression 7) as the firm becomes better informed and the cost of acquiring further information $C(x)$. In particular, the monopolist chooses x to solve

$$\max_{0 \leq x \leq \infty} \; m(x)\mathcal{G}(x) + [1 - m(x)]\mathcal{L}(x) - C(x) \tag{12}$$

Because the marginal benefit is equal to zero when $x = 0$ and because it is increasing for some values of x, solutions for the above maximization problem require not only first and second order conditions for a local maximum, but a third *individual rationality* condition. Simply put, the monopolist will only acquire information if doing so weakly increases its net expected profit. Formally, there must exist an $x > 0$ such that $B(x) - C(x) \geq V(\beta)$.

We analyze the monopolist's optimal information acquisition, starting by characterizing the individual rationality condition as a function of the cutoff θ_L/θ_H. For compactness, let $\hat{\theta} \equiv \theta_L/\theta_H$.

Proposition 1. *If $\beta \leq \hat{\theta}$, then the monopolist acquires a positive quantity of information, $x^* > 0$, if and only if there exists an x^* satisfies*

$$m(x^*)\,[\mathcal{G}(x^*) - V(\beta)] \geq [1 - m(x^*)]\,[V(\beta) - \mathcal{L}(x^*)] + C(x^*) \tag{13}$$

If $\beta > \hat{\theta}$ the monopolist acquires a positive quantity of information, $x^* > 0$, if and only x^* satisfies (13) and $x^* > \hat{x} > 0$ where $\hat{x}$ solves $\psi_L(\hat{x}) = \hat{\theta}$.

Expression (13) can be intuitively understood as that, when prior beliefs are relatively low, $\beta \leq \hat{\theta}$, the firm offers a menu under no information acquisition. Therefore, the firm only acquires information if the expected profit gains from correct signals, relative to no information acquisition, are at least as great as the expected losses from incorrect signals plus the total cost of acquiring information. If no such x^* exists, the firm chooses to remain uninformed.

When prior beliefs are high, $\beta > \hat{\theta}$, the firm offers only the high-type's complete information contract. Acquiring small quantities of information (x near zero) yields signals too uninformative to warrant offering a menu after either signal; providing zero benefit to the firm. Since, in addition, information is costly, the firm prefers zero information acquisition ($x = 0$).[18]

While proposition 1 specifies the condition for x^* to be individually rational, relative to no information acquisition, Proposition 2 below identifies the necessary and sufficient conditions for x^* to locally maximize expected profits.

Proposition 2. *The firm chooses information x^* to satisfy Proposition 1 and the following first and second order conditions*

$$\mathcal{G}(x^*) - \mathcal{L}(x^*) \geq \kappa(x^*) \quad FOC \tag{14}$$

$$[1 - m(x)]\frac{\partial \kappa(x)}{\partial x} \geq \mathcal{G}'(x^*) \quad SOC \tag{15}$$

where $\kappa(x) \equiv C'(x)/m'(x)$.

First order condition (14) holds when the firm chooses x to equate the difference in profits from correct and incorrect signals to the marginal cost of accuracy; as captured by ratio $\kappa(x) \equiv C'(x)/m'(x)$.[19] Since $C'(x)$ is the marginal cost of an additional unit of x and $m'(x)$ is its marginal increase in accuracy, the ratio $C'(x)/m'(x)$ describes the marginal cost of an additional unit of accuracy. Second-order condition (15) says that, at x^*, the firm requires that the rate of increase in the marginal cost of accuracy exceed the rate at which correct signals increase the marginal value of accuracy. The next corollary examines how the firm's optimal choice of x depends on cost function $C(x)$.

Corollary 1. *Let $C_1(x)$ and $C_2(x)$ be two cost functions such that $C_1(x) \geq C_2(x)$ for all x. Then the optimal choice x_1^* under C_1 is weakly smaller than the optimal choice x_2^* under C_2.*

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.

18 The firm must acquire a critical quantity of information ($x > \hat{x}$) before it begins to benefit. In this case, individual rationality of the firm prevents information acquisition unless there exists an information choice greater than the critical value $\hat{x}$.

19 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.

3.4 Effect of Information Acquisition on Welfare

Nonlinear pricing extracts all surplus from low-type consumers. The high-type buyer can capture, however, a positive expected utility via information rents, $R(\psi_k)$. The next corollary finds his expected utility.

Corollary 2. *The high-type buyer's expected utility is*

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

This corollary says that, if firm beliefs, ψ_k, lie below the cutoff after both signals (the first case in Corollary 2), information rents are paid and high-types are guaranteed positive utility. If the firm believes the high-type is unlikely only after signal s_L (as in the second case), then high-type buyers gain utility only after incorrect signals. Finally, if the firm places a high weight of the probability on the high-type after both signals, then no information rents are offered and no utility obtained.

3.4.1 Pareto Improvement in Welfare

Incomplete information causes a loss in expected profits, which can be expressed by rewriting expression (5) as follows

$$\Delta\Pi(\beta) = \begin{cases} \mathbb{E}U(\beta) + DWL(\beta) & \text{if } \beta \leq \hat{\theta} \\ DWL(\beta) & \beta > \hat{\theta} \end{cases}$$

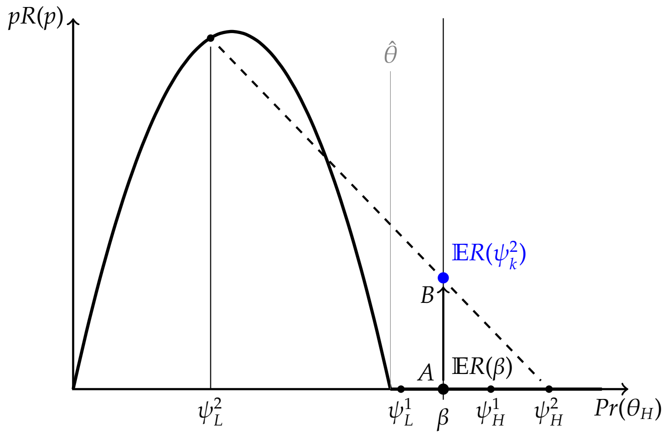
When $\beta \leq \hat{\theta}$, the firm offers a contract menu which entails an information rent to the high-type buyer, captured by $\mathbb{E}U(\beta)$; and an output distortion on the low-type buyer's output, as captured by deadweight loss $DWL(\beta)$. In contrast, when $\beta > \hat{\theta}$, the firm offers the complete information contract to high-type buyers, leading to no information rents $\mathbb{E}U(\beta) = 0$; and ignores low-type buyers, still giving rise to $DWL(\beta)$. Hence, significant reductions in deadweight loss can lead to Pareto improvements since both expected information rents to consumers and expected profits can increase.

Corollary 3. *If priors beliefs are high, $\beta > \hat{\theta}$, and the cost of information is positive, optimal information acquisition by the monopolist strictly increases expected utility of high-type buyers.*

Figure 5 depicts the result in Corollary 3. When the high-type buyer is likely, the firm offers only an efficient high-type bundle. Both types of buyers have zero surplus as shown by point A in the figure. The firm acquires information only if costs allow choices of x which satisfy the individual rationality condition (Proposition 1). The firm's optimal behavior ensures that posterior beliefs, ψ_L, are below $\hat{\theta}$, as ψ_L^2 in the figure. Information rents increase from zero (point A) to point B in the figure. Information acquisition is then Pareto improving when the firm ignores the low-type buyer in the absence of the information. However, when prior beliefs lead the firm to offer a contract menu, the results may be different, as we show below in Proposition 3.

In Figure 6a, prior beliefs satisfy $\beta < \hat{\theta}$ and the firm initially offers a rather large information rent at point (A). Information acquisition separates the posterior beliefs to (ψ_L^1, ψ_H^1) and decreases

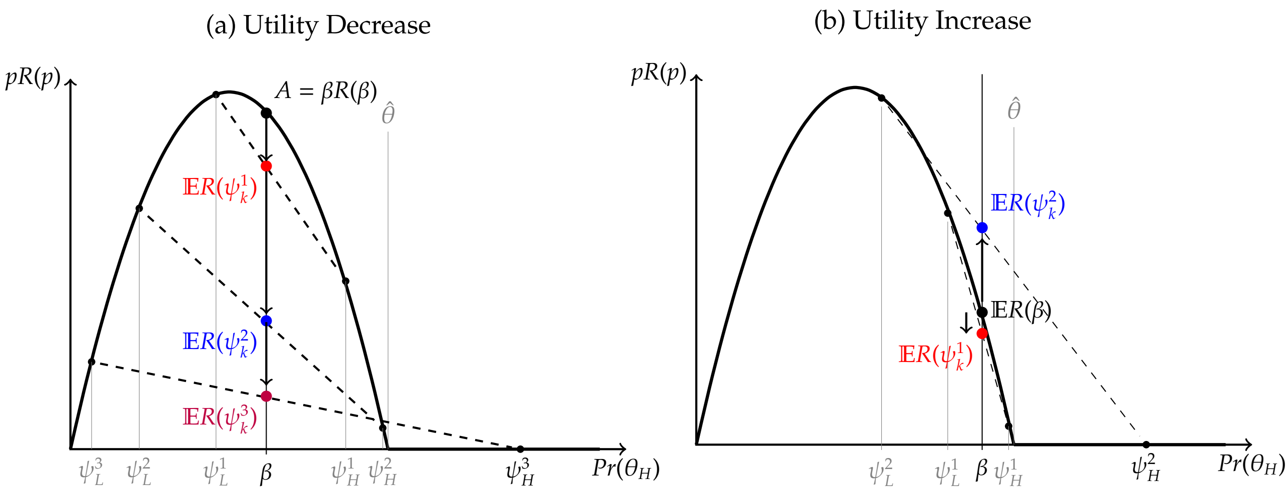
Figure 5: **When $\beta > \hat{\theta}$ information acquisition always increases expected utility.**



the expected information rent to $\mathbb{E}R(\psi_k^1)$. Greater choices of x further separate posteriors and further decrease expected rents. Figure 6b, in contrast, depicts the case where the initial information rent is much lower, which occurs when prior beliefs are close to cutoff $\hat{\theta}$. When the firm acquires a small quantity of information, posterior beliefs again separate to values (ψ_L^1, ψ_H^1) and expected information rents decrease to $\mathbb{E}R(\psi_k^1)$. However, as x increases and $\psi_H^2 > \hat{\theta}$, the rent after a high signal stops decreasing; yet rents after a low signal continue to increase. As $\psi_L R(\psi_L)$ enlarges, the expected information rent, $\mathbb{E}R(\psi_k^2)$, can exceed $\beta R(\beta)$.

Before information acquisition, if prior beliefs are low, $\beta \leq \hat{\theta}$, the firm offers a menu of contracts producing positive utility for consumers. Proposition 3 details conditions under which information acquisition can decrease or increase expected utility relative to incomplete information. Let us define some terms we use in the next proposition. If $C_0(x)$ is an initial cost of information satisfying the conditions in section 3.3.3, $C_\lambda(x) = \lambda C_0(x)$ denotes a new cost function, where $\lambda > 0$, thus indicating a higher cost than $C_0(x)$. By Corollary 1, the monopolist's choice of x is non-increasing in λ. Thus, for every $x > 0$, there is a maximum cost that supports the choice of each value of x in equilibrium.

Figure 6: **When $\beta < \hat{\theta}$, expected utility can either decrease or increase relative to incomplete information.**



Proposition 3. *Let $\beta < \hat{\theta}$, $\hat{x}$ solve $\psi_H(\hat{x}) = \hat{\theta}$ and define β_m to satisfy,*

$$\beta_m R(\beta_m) = \max_x \{\beta_m[1 - m(x)]R[\psi_L(x)]\} \;\; \textit{for } x \in [\hat{x}, \infty)$$

Define $x_m = \arg\max\{\beta[1 - m(x)]R[\psi_L(x)]\}$ for $x \in [\hat{x}, \infty)$. If prior beliefs, β, satisfy:

(a) $\beta < \beta_m < \hat{\theta}$, then information acquisition decreases consumer utility;

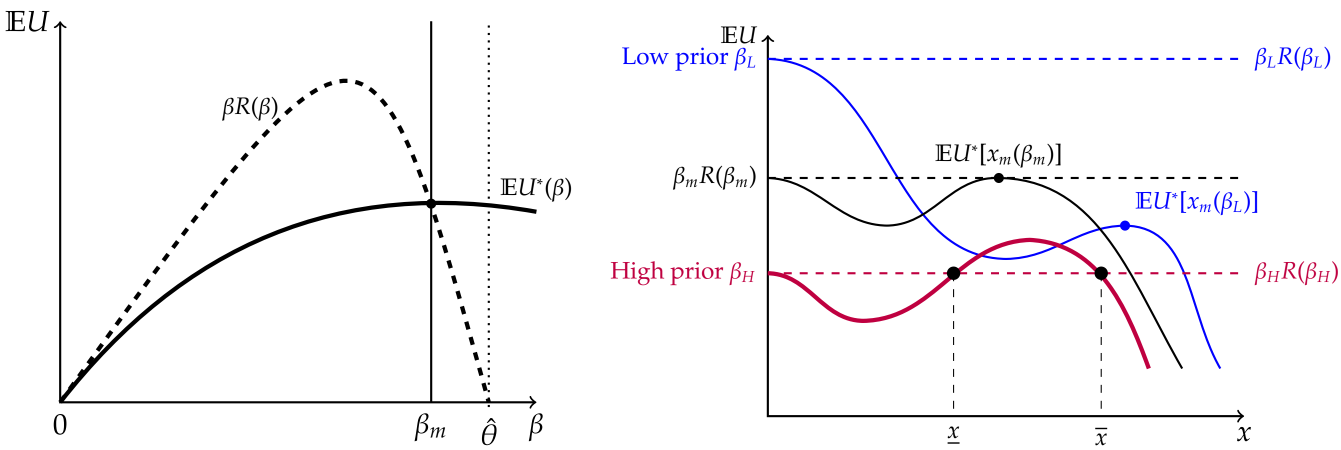
(b) $\beta_m \leq \beta < \hat{\theta}$, then for costs $[\underline{\lambda}, \overline{\lambda}]$, the monopolist chooses x in the interval $[\underline{x}, \overline{x}]$ and information acquisition increases consumer utility, where $\overline{\lambda}$ solves $\underline{x} = x(\overline{\lambda}, \beta)$, while $\underline{\lambda}$ solves $\overline{x} = x(\underline{\lambda}, \beta)$.

Figure 7a depicts the maximum expected utility curve, $\mathbb{E}U^*(\beta) = \max_x \beta[1 - m(x)]R[\psi_L(x)]$, and the initial expected utility curve, $\beta R(\beta)$, intersecting at the value β_m. For priors below β_m, information acquisition results in less utility than incomplete information, while for priors above β_m information acquisition can increase expected utility. However, this utility gain only occurs if the firm chooses to acquire an intermediate quantity of information as we describe next.

Figure 7: **Conditions for Welfare Gains when** $\beta < \hat{\theta}$

(a) Potential utility gain over β

(b) Expected Utility over x with Different Priors



As illustrated in Figure 7b (bottom curve)[20], when the cost of information is intermediate, the firm acquires an amount of x that yields an expected utility gain, i.e., λ induces the firm to choose x in the interval $[\underline{x}, \overline{x}]$.[21] In summary, optimal information can result in a Pareto improvements when prior beliefs are not too low and information is not prohibitively expensive.

Example 4. Suppose that consumer utility is defined as $u(q) = 2\sqrt{q}$ and that the accuracy of signals is determined as $m(x) = 1 - \frac{1}{2}e^{-x}$. Furthermore, assume $\theta_H = 2$, $\theta_L = 1$, and $c = 1$. For all x where $\psi_H(x) > \hat{\theta}$ the maximum expected utility curve is

$$\mathbb{E}U^*(\beta) \equiv \max_x \; \beta[1 - m(x)]R[\psi_L(x)] = \frac{\beta^{\frac{3}{2}}\left(2\beta - \beta^2 - 1\right)}{\frac{1}{2}\beta^{\frac{5}{2}} - \frac{1}{2}\sqrt{\beta} + \beta^2 - \beta} \tag{16}$$

20 In the top curve associated with a low prior $\beta_L < \beta_m$, expected utility does not exceed, for any $x > 0$, the utility levels when the firm is uninformed, at $x = 0$. The middle curve associated with prior β_m, has a maximum expected utility $\mathbb{E}U^*(x_m(\beta_m))$ for $x > 0$ which is exactly equal to initial utility $\beta_m R(\beta_m)$.

21 If more than one $\overline{\lambda}$ induces $\underline{x}$, choose the highest. Similarly, if more than one $\underline{\lambda}$ induces $\overline{x}$, then choose the lowest.

Expected utility without information acquisition, $\beta R(\beta)$, in this example is

$$\beta R(\beta) = \frac{2\beta\,(2\beta - 1)}{\beta - 1} \tag{17}$$

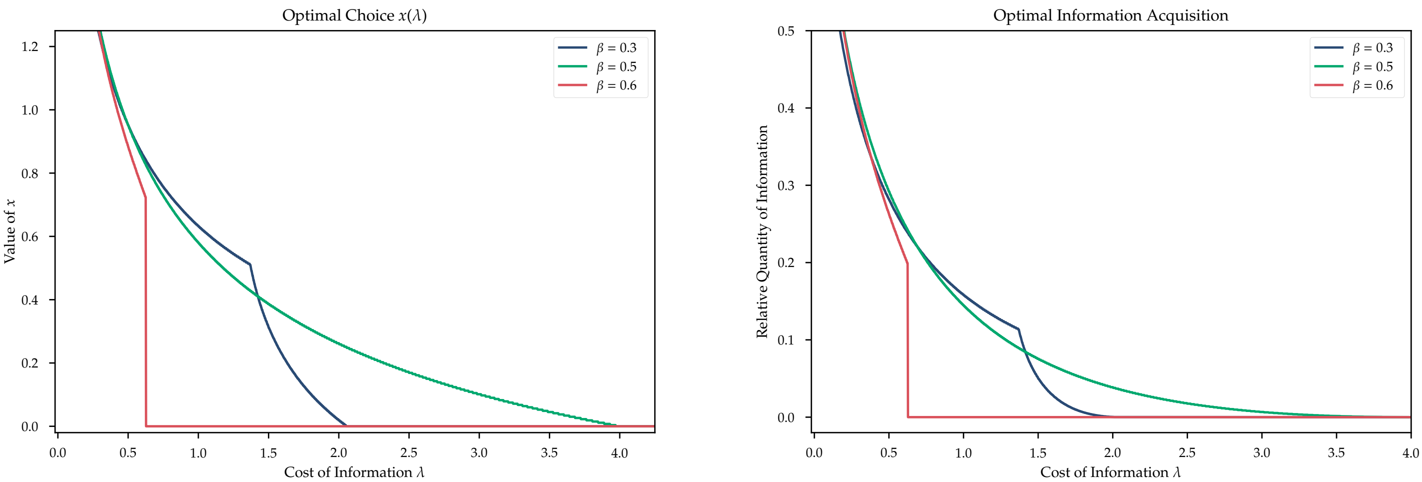
which is positive since $\beta < \hat{\theta} = \theta_L/\theta_H = 1/2$. Therefore, the cutoff prior for which expression (16) coincides with expression (17) is $\beta_m = 0.44$. For all $\beta < 0.44$, information acquisition decreases expected utility relative to $\beta R(\beta)$. □.

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 $m(x) = 1 - \frac{1}{2}\exp^{-x}$.[22] Finally, the cost function is $C(x) = \lambda x^2$, where $\lambda > 0$ denotes the cost of information.

Figure 8: **Optimal Choice of x over λ**

(a) (b)



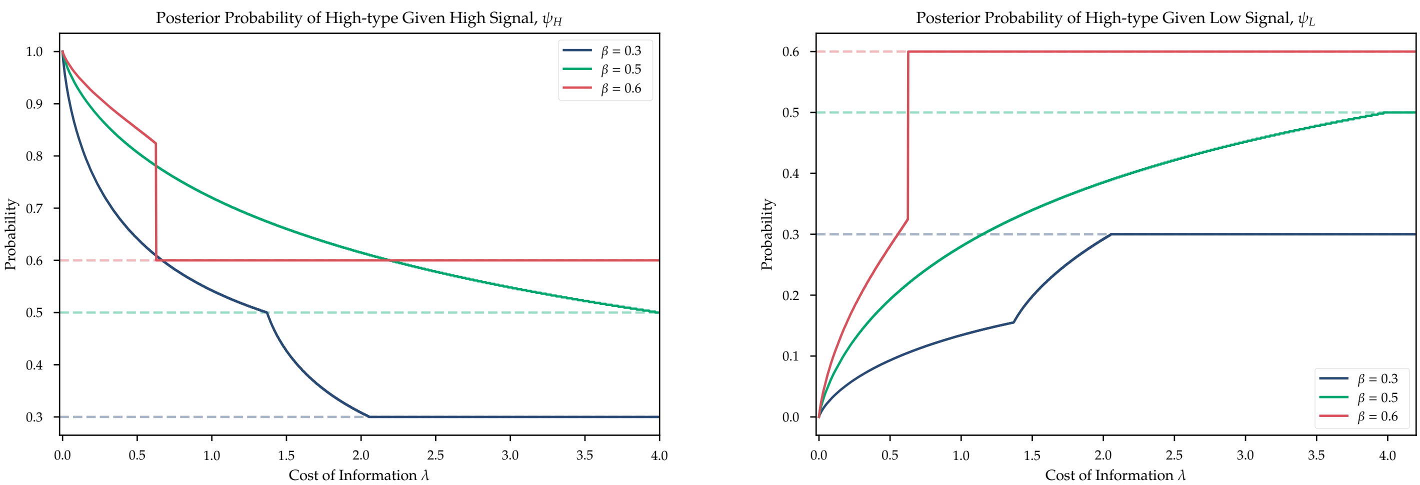
Optimal choice of x. Figure 8a depicts the optimal information acquisition x^* as a function of the cost of information λ. For presentation purposes, Figure 8b plots in its vertical axis the relative quantity of information $\frac{I(x^*,\Theta)}{\mathcal{H}(\Theta)}$, where $I(x^*,\Theta)$ measures the 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.

[22] 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.

In the case that priors satisfy $\beta = 0.3$, Figure 8b illustrates that, when information is extremely 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 9: **Posterior Probabilities with Optimal x**

(a) (b)



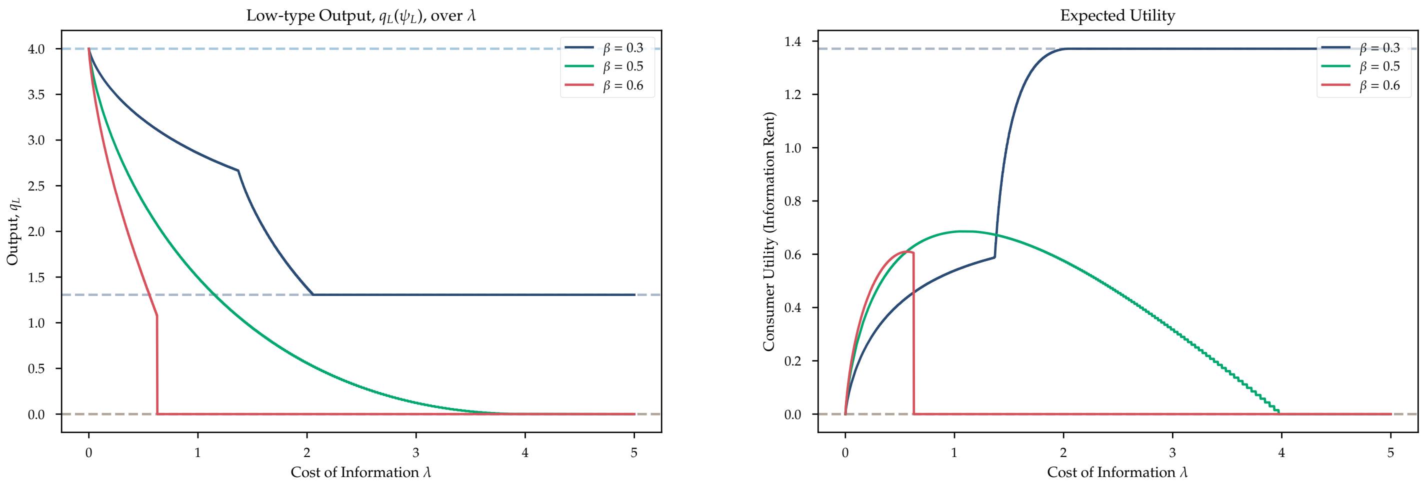
Posteriors. The equilibrium posterior probability of a high-type consumer given a high (low) signal decreases (increases) as the cost of information grows. The behavior of posteriors for priors 0.3, 0.5, and 0.6 is shown in Figure 9. When information is close to free (far left-hand side of the graph), 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 10a 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 10b.

Figure 10: **Expected Utility and Low-type Output**

(a) (b)



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 extremely 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.[23]

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 11). 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 11).

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 unambiguously decreases in this setting, since the firm moves from offering a contract menu to adjusting it

[23] 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 11: **Expected Profits and Market Welfare**

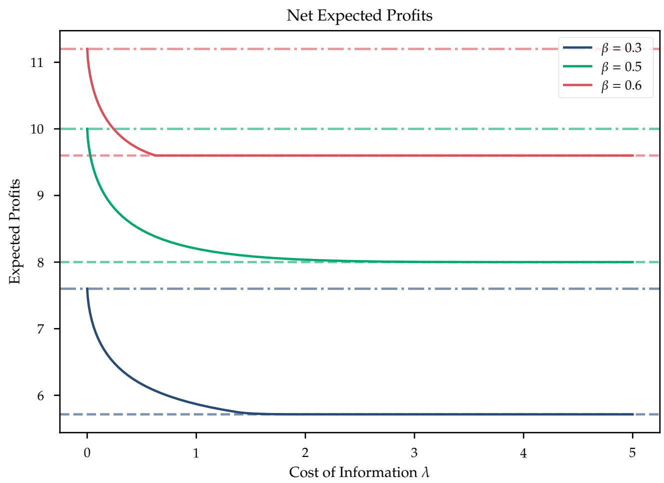
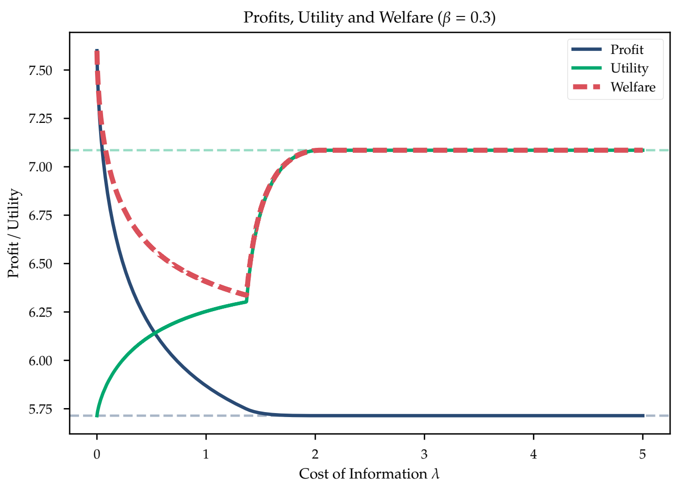
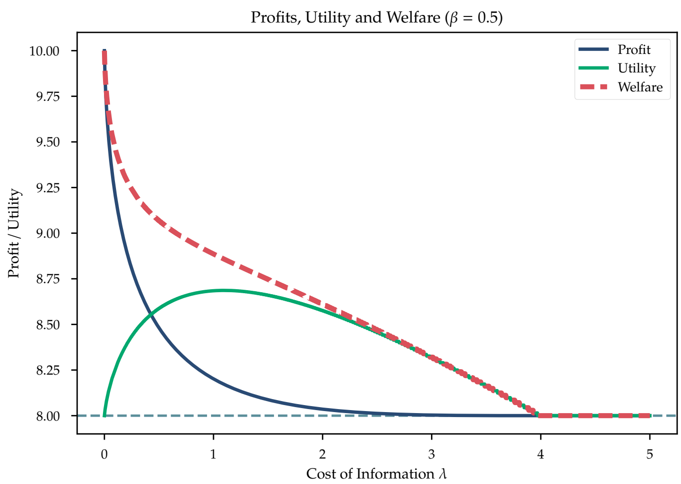


Figure 12: **Firm vs Consumer Welfare**

(a)



(b)



(as signals become more reliable), and to ignoring the low-type buyer entirely. Profits, in contrast, unambiguously 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 reliable 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 then 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[24] $H(\Theta|S)$ as follows

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

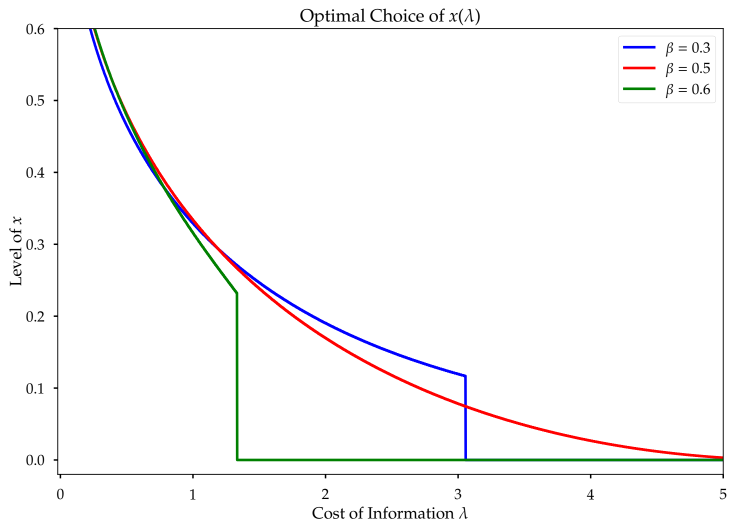
Intuitively, when the firm acquires no information, the mutual information between types and signals, $I(\Theta, S)$, approaches zero, and so does $C_2(x)$. In words, the marginal cost of acquiring the first unit of information is negligible. In contrast, when the firm acquires more information, mutual 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

$$C_2'(x) = -\lambda \frac{H(\Theta)}{H(\Theta|S)^2} \cdot \frac{\partial H(\Theta|S)}{\partial m} \cdot m'(x) \tag{18}$$

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 $\psi_H(x) > \hat{\theta}$).

[24] 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 13: **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 linear 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 the 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 consumers' 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 for incorrect signals as his profits increase in expectation.

Overall welfare increases with declining costs of information and the associated increase in information acquisition. 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 an incorrect 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 addition, 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,

$$\theta_L u(q_L) - T_L \geq 0 \qquad \text{P.C.}_L \tag{19}$$

$$\theta_H u(q_H) - T_H \geq 0 \qquad \text{P.C.}_H \tag{20}$$

$$\theta_L u(q_L) \geq \theta_L u(q_H) - T_H \qquad \text{I.C.}_L \tag{21}$$

$$\theta_H u(q_H) \geq \theta_H u(q_L) - T_L \qquad \text{I.C.}_H \tag{22}$$

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 19 and 22 hold with equality and the monopolist chooses $T_L = \theta_L u(q_L)$ and

$$\begin{aligned} T_H &= \theta_H[u(q_H) - u(q_L)] + T_L \\ &= \theta_H[u(q_H) - u(q_L) + \theta_L u(q_L) \end{aligned} \tag{23}$$

Substitution of T_L and expression (23) 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.

$$\max_{q_H, q_L} \; \beta \left\{\theta_H[u(q_H) - u(q_L)] + \theta_L u(q_L) - cq_H\right\} + (1-\beta)\left[\theta_L u(q_L) - cq_L\right] \tag{24}$$

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,

$$V(\beta) = \begin{cases} \beta\left(\theta_H u[q_H^*] - cq_H^* - (\theta_H - \theta_L)u[q_L(\beta)]\right) + (1-\beta)\left(\theta_L u[q_L(\beta)] - cq_L(\beta)\right) & \text{if } \beta \leq \theta_L/\theta_H \\ \beta(\theta_H u[q_H^*] - cq_H^*) & \text{if } \beta > \theta_L/\theta_H \end{cases} \tag{25}$$

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,

$$\frac{\partial V(\beta)}{\partial\beta} = \begin{cases} \pi_H^* - R(\beta) - \pi_L(\beta) & \text{if } \beta \leq \theta_L/\theta_H \\ \pi_H^* & \text{if } \beta > \theta_L/\theta_H \end{cases} \tag{26}$$

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 $\pi^H \geq R(\beta) + \pi_L(\beta)$.

$$
\begin{aligned}
R(\beta) + \pi_L(\beta) &\le \pi_H^* \\
(\theta_H - \theta_L)u[q_L(\beta)] + \theta_L u[q_L(\beta)] - cq_L(\beta) &\le \pi_H^* \\
\theta_H u[q_L(\beta)] - cq_L(\beta) &\le \pi_H^*
\end{aligned} \tag{27}
$$

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. ■

Proof of Lemma 4

The proof will proceed as follows. First, we will establish an envelope result from the firm's second-stage profit maximization problem. Second, we will derive the generalized marginal benefit function for the firm's information acquisition problem apply the result from the first step. Third, we derive the condition on the second derivative of the benefit function and again apply the envelope condition from the first step to arrive at our final result.

Step One. By lemma 2, for any given probability of high type ψ_k, the firm's profit maximization problem involves a first order condition for the low-type quantity,

$$(1 - \psi_k)\theta_L u'[q_L] - c = \psi_k[\theta_H - \theta_L]u'(q_L).$$

The above condition equates the expected partial derivative of the profit on the low-type with the expected partial derivative of the information rent.

$$(1-\psi_k)\frac{\partial \pi_L(\psi_k)}{\partial q_L} = \psi_k \frac{\partial R(\psi_k)}{\partial q_L}$$

which is equivalent to,

$$Pr(\theta_k, s_k')\frac{\partial \pi_L(\psi_k)}{\partial q_L} = Pr(\theta_k, s_k)\frac{\partial R(\psi_k)}{\partial q_L}$$

Recall the profits from correct and incorrect signals are defined respectively as

$$\begin{aligned}
\mathcal{G}(x) &= \beta_H \pi_H[\psi_H(x)] + \beta_L \pi_L[\psi_L(x)] \\
\mathcal{L}(x) &= \beta_H \pi_H[\psi_L(x)] + \beta_L \pi_L[\psi_H(x)].
\end{aligned}$$

The partial derivatives of these piece-wise profit functions are

$$\begin{aligned}
\mathcal{G}'(x) &= -\beta_H \frac{\partial R[\psi_H(x)]}{\partial q_L}\frac{\partial q_L[\psi_H(x)]}{\partial x} + \beta_L \frac{\partial \pi_L[\psi_L(x)]}{\partial q_L}\frac{\partial q_L[\psi_L(x)]}{\partial x} \\
\mathcal{L}'(x) &= -\beta_H \frac{\partial R[\psi_L(x)]}{\partial q_L}\frac{\partial q_L[\psi_L(x)]}{\partial x} + \beta_L \frac{\partial \pi_L[\psi_H(x)]}{\partial q_L}\frac{\partial q_L[\psi_H(x)]}{\partial x}
\end{aligned}$$

By combining terms of like posteriors, ψ_k, in the sum $m(x)\mathcal{G}'(x) + [1-m(x)]\mathcal{L}'(x)$ we obtain

$$\begin{aligned}
&\mathcal{G}'(x) + \mathcal{L}'(x) = \\
&\frac{\partial q_L[\psi_H(x)]}{\partial x}\left[-m(x)\beta_H \frac{\partial R[\psi_H(x)]}{\partial q_L} + [1-m(x)]\beta_L \frac{\partial \pi_L[\psi_H(x)]}{\partial q_L}\right] + \\
&\frac{\partial q_L[\psi_L(x)]}{\partial x}\left[-[1-m(x)]\beta_H \frac{\partial R[\psi_L(x)]}{\partial q_L} + m(x)\beta_L \frac{\partial \pi_L[\psi_L(x)]}{\partial q_L}\right]
\end{aligned}$$

The bracketed quantities are zero by the first order conditions in lemma 2. Hence, $m(x)\mathcal{G}'(x) + [1-m(x)]\mathcal{L}'(x) = 0$ which implies that $m(x)\mathcal{G}'(x) = -[1-m(x)]\mathcal{L}'(x)$ and completes the first step.

Step Two. The benefit from x is

$$B(x) = m(x)\mathcal{G}(x) + [1-m(x)]\mathcal{L}(x)$$

Differentiating the piece-wise functions with respect to x yields

$$\begin{aligned}
B'(x) &= m'(x)\mathcal{G}(x) - m'(x)\mathcal{L}(x) + m(x)\mathcal{G}'(x) + [1-m(x)]\mathcal{L}'(x) \\
&= m'(x)\left[\mathcal{G}(x) - \mathcal{L}(x)\right] + \{m(x)\mathcal{G}'(x) + [1-m(x)]\mathcal{L}'(x)\}
\end{aligned}$$

Substituting the result from step 1 into the last expression above for $B'(x)$ causes the term in curly brackets to equal zero. As a consequence, the marginal benefit, $B'(x)$, reduces to the expression below.

$$B'(x) = m'(x)\mathcal{G}(x) - m(x)\mathcal{L}(x)$$

By assumption $m'(x) \geq 0$ for all x and both $\mathcal{G}(x) \geq 0$ and $\mathcal{L}(x) \geq 0$ as each term corresponds to a nonnegative profit. Therefore $B'(x) \geq 0$ for all x, completing step two.

Step Three. Differentiating the piece-wise marginal benefit function with respect to x through the product rule yields,

$$B''(x) = m''(x)\mathcal{G}(x) + m'(x)\mathcal{G}'(x) - m''(x)\mathcal{L}(x) - m'(x)\mathcal{L}'(x)$$

Applying the envelope condition from step one again, we can substitute the equation,

$$-\mathcal{L}'(x) = \frac{m(x)}{1 - m(x)}\mathcal{G}'(x).$$

into the second derivative, $B''(x)$ to obtain,

$$\begin{aligned} B''(x) &= m''(x)[\mathcal{G}(x) - \mathcal{L}(x)] + m'(x)\mathcal{G}'(x) + m'(x)\frac{m(x)}{1 - m(x)}\mathcal{G}'(x) \\ &= m''(x)[\mathcal{G}(x) - \mathcal{L}(x)] + \frac{m'(x)}{1 - m(x)}\mathcal{G}'(x) \end{aligned}$$

The marginal benefit is nonincreasing if and only if $B''(x) \leq 0$ which holds only when x satisfies

$$\frac{m'(x)}{1 - m(x)}\mathcal{G}'(x) \leq -m''(x)[\mathcal{G}(x) - \mathcal{L}(x)]$$

■

Proof of Proposition 1

If the firm has prior beliefs β and acquires no information ($x = 0$) then its expected profits will be $V(\beta)$ as shown in equation (4) of section 2.2. The cost of the information structure specified by a choice $x > 0$ is $C(x) > 0$. Acquisition of any such x must increase net expected profits, otherwise the firm would find no information acquisition optimal. The firm's individual rationality requires,

$$m(x)\mathcal{G}(x) + [1 - m(x)]\mathcal{L}(x) - C(x) \geq V(\beta)$$

We rearrange the above condition into terms involving relative loss and gains by substituting $V(\beta) = m(x)V(\beta) + [1 - m(x)]V(\beta)$.

$$m(x)\left[\mathcal{G}(x) - V(\beta)\right] \geq [1 - m(x)]\left[V(\beta) - \mathcal{L}(x)\right] + C(x)$$

For the last part of the proof, suppose that $\beta > \theta_L/\theta_H$. Then $V(\beta) = \beta_H\pi_H^*$ and let $\hat{x}$ solve $\psi_L(\hat{x}) = \theta_L/\theta_H$. Then for all $0 \leq x < \hat{x}$ we have the correct and incorrect signal profits as

$$\mathcal{G}(x) = \mathcal{L}(x) = \beta_H\pi_H^*$$

which implies that

$$\mathcal{G}(x) - V(\beta) = V(\beta) - \mathcal{L}(x) = 0$$

This implies that for any choice of information structure $0 \leq x < \hat{x}$ the firm receives no benefit, but will incur positive cost $C(x) > 0$. Consequently, for any x which satisfies $0 \leq x < \hat{x}$, the individual rationality condition requires $0 \geq C(x)$ which only holds for $x = 0$ since by assumption $C(0) = 0$. Therefore, when $\beta > \theta_L/\theta_H$, if any information structure x^* satisfies the individual rationality condition it must be that $x^* \geq \hat{x}$. ■

Proof of Proposition 2

The firm's objective function $J(x)$ has the form,

$$J(x) \equiv m(x)\mathcal{G}(x) + [1 - m(x)]\mathcal{L}(x) - C(x)$$

Using the result for $B'(x)$ from lemma 4, the first derivative of the objective function is

$$J'(x) = m'(x)\,[\mathcal{G}(x) - \mathcal{L}(x)] - C'(x)$$

The first order condition for a local maximum requires that $J'(x) \geq 0$ leaving

$$\mathcal{G}(x) - \mathcal{L}(x) \geq \frac{C'(x)}{m'(x)}$$

Defining $\kappa(x) = C'(x)/m'(x)$ completes the first order condition.

Again recalling the second derivative $B''(x)$ from lemma 4, the second derivative of the objective function, $J''(x)$, is expressed as

$$J''(x) = m''(x)[\mathcal{G}(x) - \mathcal{L}(x)] + \frac{m'(x)}{1 - m(x)}\mathcal{G}'(x) - C''(x)$$

At any critical point, x, satisfying the first order condition, we require $J''(x^*) \leq 0$ to ensure a local maximum.

$$m''(x)[\mathcal{G}(x) - \mathcal{L}(x)] \leq -\frac{m'(x)}{1 - m(x)}\mathcal{G}'(x) + C''(x)$$

By the first order condition, $\kappa(x) \leq \mathcal{G}(x) - \mathcal{L}(x)$. At any x satisfying the first order condition we will then have, $m''(x)\kappa(x) \leq m''(x)[\mathcal{G}(x) - \mathcal{L}(x)]$ yielding condition

$$m''(x)\kappa(x) - C''(x) \leq -\frac{m'(x)}{1 - m(x)}\mathcal{G}'(x)$$

Since $\kappa(x) = C'(x)/m'(x)$ and dividing both sides by $m'(x) > 0$ we have,

$$\frac{m''(x)C'(x) - C''(x)m'(x)}{m'(x)^2} \leq -\frac{1}{1 - m(x)}\mathcal{G}'(x)$$

Multiplying both sides by negative one yields,

$$\frac{1}{1 - m(x)}\mathcal{G}'(x) \leq \frac{C''(x)m'(x) - m''(x)C'(x)}{m'(x)^2}$$

In the above expression, the right hand side is exactly $\kappa'(x)$, the increase in the marginal cost of accuracy. Therefore the second order condition for any x^* satisfying the first order condition is

$$\mathcal{G}'(x) \leq [1 - m(x)] \cdot \frac{\partial \kappa(x)}{\partial x} \tag{28}$$

■

7.1 Proof of Corollary 1

Let $C_1(x)$ and $C_2(x)$ be two cost functions satisying the conditions assumed in section 3.3.3 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, $J(x) \equiv B(x) - \lambda C_2(x)$. 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. ■

7.2 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,

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

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.

$$\mathbb{E}U = \begin{cases} \beta \left\{ m(x) R[\psi_H(x)] + [1 - m(x)] R[\psi_L(x)] \right\} & \text{if } \psi_L(x) < \psi_H(x) < \hat{\theta} \\ \beta[1 - m(x)] R[\psi_L(x)] & \text{if } \psi_L(x) < \hat{\theta} < \psi_H(x) \\ 0 & \text{if } \hat{\theta} < \psi_L(x) < \psi_H(x) \end{cases} \tag{29}$$

■

7.3 Proof of Corollary 3

Define $\hat{\theta} \equiv \theta_L/\theta_H$ and let $\beta > \hat{\theta}$. Then from Lemma 2, we know that, without information acquisition, the firm ignores the low-type consumer and offers an efficient contract to the high-type resulting in information rent $R(\beta) = 0$. The expected utility, $\beta R(\beta)$ is also zero. Posterior beliefs, ψ_k separate from the prior implying that $\hat{\theta} < \psi_L < \beta$ for all x satisfying $0 < x < \hat{x}$ where $\hat{x}$ solves $\psi_L(\hat{x}) = \hat{\theta}$. Let x^* denote the firm's expected profit maximizing choice of x. Then from Proposition 1 we know that since $\beta < \hat{\theta}$, either $x^* = 0$ or $x^* > \hat{x}$. If $x^* = 0$ the firm acquires no information and the premise of the corollary isn't satisfied, so we focus on $x^* > \hat{x}$.

For all $x > \hat{x}$, we have $\psi_L(x) < \hat{\theta} < \psi_H(\hat{x})$. After a high-signal, because $\psi_H(\hat{x}) > \hat{\theta}$, the firm offers information rent $R(\psi_H) = 0$ yielding an expected utility of zero. After a low-signal, because $\psi_L(x) < \hat{\theta}$, the firm offers a menu of contracts involving an information rent of $R(\psi_L)$. Since the low-type contract is nontrivial in this case, $q_L(\psi_L) > 0$ which guarantees that $u[q_L(\psi_L(x))] > 0$ and therefore the information rent $R[\psi_L(x)] = [\theta_H - \theta_L]u[q_L(\psi_L(x))] > 0$.

From Corollary 2, we know the expected utility of consumers for x such that $\psi_L(x) < \hat{\theta} < \psi_H(x)$ is defined as

$$\mathbb{E}U = \beta[1 - m(x)]R[\psi_L(x)]$$

We previously established that because $q_L[\psi_L(x)] > 0$ for all $x > \hat{x}$, the information rent will also be positive. However, the expected information rent is scaled by the probability of incorrect signals $1 - m(x)$. As x increases without bound we have $\lim_{x\to\infty}[1 - m(x)] = 0$ which would place the firm in complete information and result in zero utility for consumers. However, this limit is never achieved when information is costly as we now establish.

Note that the loss in expected profit from incomplete information, $\Delta\Pi(\beta)$, satisfies $\Delta\Pi(\beta) < \infty$, and represents an upper bound on the benefit from information acquisition. The cost function $C(x)$ is unbounded. The individual rationality condition requires that $B(x) \geq C(x)$. Because $B(x) \leq \Delta\Pi(\beta)$, there exists $\bar{x}$ solving $B(\bar{x}) = C(\bar{x})$. If $C(x) > 0$ for all $x > 0$ then $\bar{x} < \infty$. Hence, any x^* is bounded between finite, positive numbers, $0 < \hat{x} < x^* < \bar{x}$. This implies that $1 - m(x^*) > 0$ and therefore $\beta[1 - m(x^*)]R[\psi_L(x^*)] > 0$, ultimately entailing that the high-type consumer has positive expected utility. ■

7.4 Proof of Proposition 3

Let $\hat{\theta} = \theta_L/\theta_H$ and let the firm have prior beliefs $\beta < \hat{\theta}$. By Lemma 2, without information acquisition, the firm will offer a menu to consumers involving an information rent $R(\beta) = [\theta_H - \theta_L]u[q_L(\beta)]$. The probability of a high-type is β and the expected utility of consumers is $\mathbb{E}U(\beta) = \beta R(\beta)$. Any increase or decrease in expected utility caused by information acquisition will be compared to this baseline incomplete information value.

The proof will proceed as follows. First, we establish that when prior beliefs satisfy $\beta < \hat{\theta}$ and posterior beliefs satisfy $\psi_L(x) < \psi_H(x) < \hat{\theta}$ (call it region I), information acquisition decreases expected utility. We accomplish this by demonstrating, under the assumptions on homogeneity of the utility function, that the expected information rent function is concave for probabilities less than $\hat{\theta}$. Second, when posterior probabilities of the high-type satisfy $\psi_L(x) < \hat{\theta} < \psi_H(x)$ (call it region II), non-concavities in the expected rent function make increases in expected utility possible. If the maximum of the expected utility function in region II is less than initial expected utility $\beta R(\beta)$, information acquisition only yields a utility decrease. If, on the other hand, the maximum of expected utility function in region II exceeds the initial expected utility, $\beta R(\beta)$, then positive levels of information acquisition are required to achieve it, and there exist costs of information which will induce the firm to optimally acquire those levels, leading to welfare (and Pareto) improvements.

Part I: Finding expected utility without information acquisition Let $x^*(\lambda)$ represent the expected profit maximizing choice of x (satisfying propositions 1 and 2) given cost function C_λ.

Denote the cost of information, λ, that makes the firm just indifferent between acquisition and no acquisition as $\lambda^0 > 0$ and note that $x^0 = x^*(\lambda^0) = 0$. Define $\hat{x}$ as that x which solves $\psi_H(\hat{x}) = \hat{\theta}$ and the cost of information which induces this choice by the firm as $\hat{\lambda}$. Then for costs of information in the interval $[\hat{\lambda}, \lambda^0]$, the firm will select information x, from the interval $[0, \hat{x})$. For all $x \in [0, \hat{x})$ the posterior beliefs will satisfy,

$$\psi_L(x) \leq \beta \leq \psi_H(x) \leq \hat{\theta}$$

Consequently, from Corollary 2, the expected utility of the high-type buyer over the interval $[0, \hat{x})$ involves a positive information rent after both high and low signals,

$$\begin{aligned}
\mathbb{E}U &= \beta m(x) R[\psi_H(x)] + \beta[1 - m(x)] R[\psi_L(x)] \\
&= \rho_H(x) \left[\frac{\beta m(x)}{\rho_H(x)} R[\psi_H(x)] \right] + (1 - \rho_H(x)) \left[\frac{\beta[1 - m(x)]}{1 - \rho_H(x)} R[\psi_L(x)] \right] \\
&= \rho_H(x) \left[\psi_H(x) R[\psi_H(x)] \right] + (1 - \rho_H(x)) \left[\psi_L(x) R[\psi_L(x)] \right] && \text{group by posterior expected rent} \\
&= \rho_H(x) f(\psi_H) + (1 - \rho_H(x)) f(\psi_L)
\end{aligned}$$

where $f(\psi_k) \equiv \psi_k R[\psi_k]$. Note that the expected posterior probability of the high-type is equal to the prior probability. In other words we have,

$$\begin{aligned}
\mathbb{E}[\psi] &= \rho_H \psi_H + (1 - \rho_H) \psi_L \\
&= \rho_H \frac{m\beta}{\rho_H} + (1 - \rho_H) \frac{(1 - m)\beta}{(1 - \rho_H)} \\
&= m\beta + (1 - m)\beta = \beta
\end{aligned}$$

If the expected rent function $f(\psi_k) = \psi_k R[\psi_k]$ is concave in the probability of the high-type, then by Jensen's inequality we have,

$$f(\beta) \geq \rho_H f(\psi_H) + (1 - \rho_H) f(\psi_L)$$

which implies that, over this region, expected utility under information acquisition is lower than in the case of no information acquisition.

Part II: Showing the concavity of $f(p) = pR(p)$ and decreasing expected utility in region I We now establish the concavity of the function f under the assumptions of the proposition. To do so, we establish the first and second derivatives of the low-type bundle quantity and then show concavity. We proceed by letting the probability of the high-type (whether prior or posterior) be denoted by p.

Finding $dq_L(p)/dp$. The first order condition for the monopolist's optimal choice of low-type contract quantity, q_L^{**}, under incomplete information and probability of high-type p, satisfies

$$\begin{cases} u'(q_L)[\theta_L - p\theta_H] - (1-p)c \equiv 0 & \text{if } p < \theta_L/\theta_H \\ 0 & \text{if } p \geq \theta_L/\theta_H \end{cases}$$

Implicit differentiation of the top condition with respect to p yields

$$u''(q_L)\frac{dq_L}{dp}[\theta_L - p\theta_H] - u'(q_L)\theta_H + c = 0$$

or, rearranging,

$$\frac{dq_L}{dp} = \frac{\theta_H u'(q_L) - c}{u''(q_L)[\theta_L - p\theta_H]} \tag{30}$$

By assumption, the second derivative $u''(\cdot) < 0$ over its domain. Additionally, for all $p \in [0, \theta_L/\theta_H)$, we have $\theta_L > p\theta_H$. Hence, the denominator of dq_L/dp is negative. The first order conditions further imply that,

$$\theta_L u'(q_L) - c = (\theta_H - \theta_L)u'(q_L)\frac{p}{1-p}. \tag{31}$$

By assumption $\theta_H > \theta_L$, $u'(q) > 0$ for all q and $p \in [0, 1]$. Hence the term $\theta_L u'(q_L)$ in (31) is positive and again since $\theta_H > \theta_L$, it must also be true that $\theta_H u'(q_L) - c > 0$ from both (30) and (31) is positive. Therefore, the numerator from (30) is a strictly positive term and the entire derivative (30), is strictly negative for $p \in [0, \theta_L/\theta_H)$. For all $p \geq \theta_L/\theta_H$, the low-type quantity is identically zero. Therefore $q_L(p)$ is non-increasing and strictly decreasing for $p \in [0, \theta_L/\theta_H)$.

We now derive the second derivative of the function $q_L(p)$. We now know that the first derivative is

$$\frac{dq_L}{dp} = \frac{\theta_H u'(q_L) - c}{u''(q_L)(\theta_L - p\theta_H)}$$

Using quotient rule to find $d^2 q_L(p)/dp^2$. Implicitly differentiating this expression again we can identify the necessary components of the second derivative. First, the implicit derivative of the numerator is

$$\frac{d}{dp}\left(\theta_H u'(q_L) - c\right) = \theta_H u''(q_L)\frac{dq_L}{dp}.$$

Second, the implicit derivative of the denominator is

$$\frac{d}{dp}\left[u''(q_L)(\theta_L - p\theta_H)\right] = u'''(q_L)\frac{dq_L}{dp}(\theta_L - p\theta_H) - \theta_H u''(q_L)$$

Third, we derive the product of the derivative of the numerator and the undifferentiated denominator.

$$\begin{aligned} &\theta_H u''(q_L)\frac{dq_L}{dp}u''(q_L)(\theta_L - p\theta_H) \\ &= \theta_H [u''(q_L)]^2 \frac{\theta_H u'(q_L) - c}{u''(q_L)(\theta_L - p\theta_H)}(\theta_L - p\theta_H) && \text{substitute form of } \frac{dq_L}{dp} \\ &= \theta_H u''(q_L)(\theta_H u'(q_L) - c). \end{aligned} \tag{32}$$

Fourth, the product of the undifferentiated numerator and the derivative of the denominator is

$$
\begin{aligned}
&(\theta_H u'(q_L) - c)\left[u'''(q_L)\frac{dq_L}{dp}(\theta_L - p\theta_H) - \theta_H u''(q_L)\right] \\
&= (\theta_H u'(q_L) - c)\left[u'''(q_L)\cdot\frac{(\theta_H u'(q_L) - c)}{u''(q_L)(\theta_L - p\theta_H)}\cdot(\theta_L - p\theta_H) - \theta_H u''(q_L)\right] \quad \text{substitute the form of } \frac{dq_L}{dp} \\
&= \frac{u'''(q_L)}{u'(q_L)}(\theta_H u'(q_L) - c)^2 - \theta_H u''(q_L)(\theta_H u'(q_L) - c) \qquad (33)
\end{aligned}
$$

By the quotient rule, we take the difference of (32) and (33).

$$
\begin{aligned}
&\theta_H u''(q_L)(\theta_H u'(q_L) - c) - \left\{\frac{u'''(q_L)}{u''(q_L)}(\theta_H u'(q_L) - c)^2 - \theta_H u''(q_L)(\theta_H u'(q_L) - c)\right\} \\
&= (\theta_H u'(q_L) - c)\left\{2\theta_H u''(q_L) - \frac{u'''(q_L)}{u''(q_L)}(\theta_H u'(q_L) - c)\right\}.
\end{aligned}
$$

The denominator of the second derivative is simply $[u''(q_L)(\theta_L - p\theta_H)]^2$. The entire second derivative can now be constructed as,

$$
\begin{aligned}
\frac{d^2 q_L}{dp^2} &= \frac{(\theta_H u'(q_L) - c)\left\{2\theta_H u''(q_L) - \frac{u'''(q_L)}{u''(q_L)}(\theta_H u'(q_L) - c)\right\}}{[u''(q_L)(\theta_L - p\theta_H)]^2} \\
&= \frac{(\theta_H u'(q_L) - c)}{u''(q_L)(\theta_L - p\theta_H)}\cdot\left\{\frac{2\theta_H u''(q_L)}{u''(q_L)(\theta_L - p\theta_H)} - \frac{u'''(q_L)}{u''(q_L)}\cdot\frac{(\theta_H u'(q_L) - c)}{u''(q_L)(\theta_L - p\theta_H)}\right\} \\
&= \frac{dq_L}{dp}\cdot\left\{\frac{2\theta_H u''(q_L)}{u''(\theta_L - \theta_H\beta)} - \frac{u'''(q_L)}{u''(q_L)}\cdot\frac{dq_L}{dp}\right\} \quad \text{substitute in form of } \frac{dq_L}{dp} \\
&= \frac{dq_L}{dp}\cdot\left\{\frac{2\theta_H}{(\theta_L - p\theta_H)} - \frac{u'''(q_L)}{u''(q_L)}\cdot\frac{dq_L}{dp}\right\}
\end{aligned}
$$

which completes the derivative.

The information rent function $R(p)$ is composed of the low-type contract's quantity and the consumer utility, i.e., $R(p) = [\theta_H - \theta_L]u[q_L(p)]$. The expected rent curve is defined as $f(p) = pR(p)$. The first and second derivatives of this expected rent curve are

$$
\begin{aligned}
\frac{df}{dp} &= R(p) + pR'(p) \\
\frac{d^2 f}{dp^2} &= 2R'(p) + pR''(p) \qquad (34)
\end{aligned}
$$

The function f is concave over $[0, \theta_L/\theta_H)$ if $f''(p) \leq 0$ for all p in the interval. This condition holds if, and only if

$$
pR''(p) \leq -2R'(p) \qquad (35)
$$

By the chain rule, the derivatives $R'(p)$ and $R''(p)$ are defined as

$$
\begin{aligned}
R'(p) &= [\theta_H - \theta_L]u'[q_L(p)]\frac{dq_L}{dp} \\
R''(p) &= [\theta_H - \theta_L]\left\{u''[q_L(p)]\left(\frac{dq_L}{dp}\right)^2 + u'[q_L(p)]\frac{d^2 q_L}{dp^2}\right\}
\end{aligned}
$$

For compactness, let u, u', u'', u''' represent the values of the derivatives of the utility function evaluated at $q_L(p)$. The values $q_L(p)$, $q_L'(p)$ and $q_L''(p)$ will likewise be represented by the shortened expressions q_L, q_L', q_L''. Then (35) can be expressed as

$$p[\theta_H - \theta_L]\left\{u'' \cdot \left(q_L'\right)^2 + u' \cdot q_L''\right\} \leq -2[\theta_H - \theta_L]u' \cdot q_L'$$
$$p\left\{u'' \cdot \left(q_L'\right)^2 + u' \cdot q_L''\right\} \leq -2u' \cdot q_L'$$

We now substitute in the expressions derived for $\frac{d^2 q_L}{dp^2}$ for q_L'' and expand the results.

$$p\left\{u'' \cdot (q_L')^2 + u' \frac{2\theta_H}{(\theta_L - p\theta_H)} q_L' - u' \frac{u'''}{u''} \cdot (q_L')^2\right\} \leq -2u' \cdot q_L'$$

dividing both sides by $(q_L')^2$ and simplifying , yields

$$p\left\{\frac{u''}{u'} - \frac{u'''}{u''}\right\} \leq -\frac{2\theta_L}{(\theta_L - p\theta_H)} \frac{u'' \cdot (\theta_L - p\theta_H)}{\theta_H u' - c}. \tag{36}$$

Now continuing on with the inequality (36) we have,

$$p\left\{\frac{u''}{u'} - \frac{u'''}{u''}\right\} \leq -\frac{2\theta_L}{1} \frac{u''}{\theta_H u' - c}$$
$$p\left\{\frac{u'}{u'} \cdot (k-1) - \frac{u''}{u''} \cdot (k-2)\right\} \leq -\frac{2\theta_L}{\theta_H u' - c} u' \cdot (k-1) \qquad \text{by homogeneity of degree } k$$

by the second stage first order condition and simplifying, we obtain

$$(2k-1) \leq \frac{\hat{\theta} - p}{\hat{\theta}(1-p)} \tag{37}$$

Since the consumer utility function $u(\cdot)$ is homogenous of degree $k \leq 1/2$, the left-hand side of the inequality is less than or equal to zero, while the right-hand side is strictly positive for $p < \hat{\theta}$. Hence, the condition holds and the function $f(p) = pR(p)$ is concave in p for $p \in [0, \hat{\theta})$.

Since we know that $f(\beta) = f(\rho_H(x)\psi_H(x) + (1 - \rho_H(x))\psi_L(x)) = \beta R(\beta)$ is the expected utility of consumers, and that it is concave, when $\beta < \hat{\theta}$ and x is in the interval $[0, \hat{x})$, then by Jensen's inequality we have

$$\beta R(\beta) \geq \rho_H(x)\psi_H(x)R[\psi_H(x)] + (1 - \rho_H(x))\psi_L(x)R[\psi_L(x)]$$
$$\mathbb{E}U(x = 0) \geq \mathbb{E}U(0 \leq x < \hat{x})$$

Maximum expected utility in region II Since expected utility over $[0, \hat{x})$ is strictly below $\beta R(\beta)$, increases in expected utility, if any, must occur for values of x in the interval $[\hat{x}, \infty)$. We now search this interval for expected utility levels above $\beta R(\beta)$.

Establish a Compact Interval. We now establish, that for any positive cost of information, the firm's optimal choice of x is bounded above. The benefits of information acquisition have an upper bound equal to $\Delta\Pi(\beta)$. However, costs are increasing, convex and unbounded as x approaches infinity. Hence, for any positive cost of information, $\lambda > 0$, there exists an $x_b < \infty$ that solves $B(x_b) - C(x_b) = 0$. For all $x > x_b$ the firm is strictly worse off than choosing $x = 0$. We therefore define expected utility over the compact interval $[\hat{x}, x_b]$. The expected utility function over this interval is

$$\mathbb{E}U(x) = \beta[1 - m(x)]R[\psi_L(x)].$$

Since we have a continuous function defined over a compact set, the Extreme Value Theorem guarantees the existence of a maximum and minimum.

Maximum Post-acquisition Utility vs Pre-acquisition Utility. Define x_m as the value of x that maximizes the expected utility over the interval $[\hat{x}, x_b]$.

$$x_m = \arg\max_{[\hat{x}, x_b]} \beta[1 - m(x)]R[\psi_L(x)]$$

and the expected utility at this value of x as $\mathbb{E}U(x_m)$. We now compare this to the pre-information acquisition utility $\beta R(\beta)$.

First, if $\mathbb{E}U(x_m) \leq \beta R(\beta)$, there does not exist an $x \in [\hat{x}, x_b]$ that produces an expected utility greater than $\beta R(\beta)$. This means that any information acquisition by the firm in equilibrium reduces expected utility.

To the contrary, suppose $\mathbb{E}U(x_m) > \beta R(\beta)$. Since expected utility decreases over the interval $[0, \hat{x})$, we know that $\mathbb{E}U(\hat{x}) \leq \mathbb{E}U(x = 0) = \beta R(\beta)$. Because the expected utility function in the restricted interval is continuous, the intermediate value theorem ensures that there exists an $\underline{x}$ satisfying $\hat{x} \leq \underline{x} \leq x_m$ that solves $\mathbb{E}U(\underline{x}) = \beta R(\beta)$. Because $\mathbb{E}U(x_m) > \mathbb{E}U(x = 0)$ and $\mathbb{E}U(x) \to 0$ as $x \to \infty$, there must be an $\overline{x} > x_m$ such that $\mathbb{E}U(\overline{x}) = \mathbb{E}U(x = 0) = \beta R(\beta)$. Then for all choices of x in the interval $(\underline{x}, \overline{x})$ the high-type buyer has a strictly higher expected utility than they do before information acquisition.

The firm will choose values x in $(\underline{x}, \overline{x})$ so long as the cost of information is sufficient to induce them to do so. Let $\overline{\lambda}$ solve $\underline{x} = x^*(\overline{\lambda})$ where $x^*(\lambda)$ is the argmax of the monopolist's profit maximization problem in expression (12). Similarly, let $\underline{\lambda}$ solve $\overline{x} = x^*(\underline{\lambda})$.

Hence, $\underline{\lambda} < \overline{\lambda}$ and we define an interval of costs $[\underline{\lambda}, \overline{\lambda}]$ that lead the firm to choose $x \in [\underline{x}, \overline{x}]$ in equilibrium, resulting in consumers achieving expected utility $\mathbb{E}U(x) \geq \beta R(\beta)$.

Finally, consider the value of prior beliefs β which equates the maximum achievable expected utility, $\mathbb{E}U(x_m; \beta)$ with the pre-information acquisition level of $\beta R(\beta)$. Formally, let β_m solve,

$$\beta_m R(\beta_m) = \mathbb{E}U(x_m(\beta_m), \beta_m)$$

Then for all $\beta \leq \beta_m$, there does not exist an x (or cost λ to induce such an x) that will result in consumers having higher expected utility. Hence, information acquisition will only make consumers worse off than without information acquisition. However, for all β satisfying $\beta_m \leq \beta \leq \hat{\theta}$ there exist cost values $[\underline{\lambda}, \overline{\lambda}]$ which cause the firm to choose x in $(\underline{x}, \overline{x}]$ resulting in expected utility $\mathbb{E}U(x; \beta) > \beta R(\beta)$. Therefore, for prior beliefs satisfying $\beta_m < \beta \leq \hat{\beta}$, information acquisition can lead to higher expected utility for consumers if costs lie in $[\underline{\lambda}, \overline{\lambda}]$. ■

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.

Matějka, Filip and Alisdair McKay (2015), "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review*, 105, 272–298.

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.