Optimal Privacy with Coarse Labels in Screening Markets

Niklas Haeusle* Daniel Quigley† Casey Rothschild‡ Yanwei Sun§ Andre Veiga¶

February 05, 2026

Abstract

We study the implications of fundamental information frictions for optimal information design in settings with screening. We introduce a regulator who observes imperfect information about a buyer's private valuation for quality based on a class-level characteristic, and can commit to provide a (possibly garbled) signal of these characteristics to a monopoly seller. In a setting with binary types and classes, we fully characterize optimal signals and derive their welfare properties. In contrast to much of the recent literature studying information design in markets, we find that aggregate surplus is often not maximized at full information; less than full disclosure is optimal when class-membership statistics are sufficiently poor indicators of type. In this way, our results provide a robust rationale for privacy protection that applies even when consumers and producers are treated symmetrically. Our main analysis focuses on information design in the classic setting of Mussa and Rosen (1978), but we show that the same economics apply in a broad array of classic information economics settings and to settings with many types. Our results can be interpreted through the lens of restrictions imposed on the observables on which monopolists may discriminate to aid their screening.

*Leipzig University. Email: niklas.haeusle@uni-leipzig.de
†University of Oxford. Email: daniel.quigley@economics.ox.ac.uk
‡Wellesley College. Email: crothsch@wellesley.edu
§Imperial College London. Email: yanwei@imperial.ac.uk
¶Imperial College London. Email: a.veiga@imperial.ac.uk

1 Introduction

In many economic contexts, policymakers and institutions are faced with the questions of how much information to disclose to interested parties and whether to impose limits on the use of readily available information. The information available for these purposes is itself often incomplete, based only on coarse signals such as class-level characteristics of the economically relevant primitives such as individual preferences and traits. For example, platforms such as Amazon and Taobao may decide how much of their buyer information to share with sellers. Likewise, competition authorities and legislators may impose limits on the degree to which such information can be shared or may place other limits on the use of demographic information for price discrimination in markets such as insurance,[1] and authorities may commit to design jurisdictions within which tax schedules must be neutral with respect to identity. In these settings, uninformed market participants and institutions may attempt to resolve the asymmetry themselves via screening. When should a (Pareto) efficiency-minded policy-maker allow such uninformed participants to use class-level characteristics to augment their attempts at screening? By contrast, when and how should the use of class-membership information be limited?

This paper addresses these questions by explicitly studying the role of class-level information in models of screening. To fix ideas, our baseline analysis introduces a social planner who has the power to provide information to a monopolist in a classic trade problem.[2] As in the Mussa and Rosen (1978) model, we consider an environment in which a monopoly firm chooses the quality of goods to serve to a continuum of customers who have private information about their own willingness to pay (type). To ease the exposition, we present our main analysis for the binary-type case.[3] As in Mussa and Rosen (1978), the monopolist may address this informational asymmetry using screening contracts. Also like Mussa and Rosen (1978), but in contrast to the recent literature on information design in monopoly trading problems,[4] we assume the customers' types are truly private — in particular, they are not even observable to the planner.

Instead, we allow the planner access to only a coarse signal of the consumer's type. Our motivation is to study information design in the presence of this natural informational friction, and characterize comparative statics of design with respect to these frictions. Formally, we endow each agent with a category-level characteristic (class) which is observable to the social planner and determines the distribution from which the customers' type is drawn. In our baseline analysis, we take class membership as binary — only a low proportion of class membership A customers have the high value type, while more of class B customers have the high type; in section 6, we explain how the results extend to richer distributions of class memberships.

We assume the planner wishes to maximize a weighted sum of producer and consumer surpluses. In the spirit of the information design literature, we allow the planner to commit

[1] For instance, in 2012 the European Court of Justice outlawed the use of gender-based discrimination in insurance contexts. By contrast, health insurers in the United States are free to — and do — discriminate based on occupation.

[2] We consider applications to insurance and tax in section 5.

[3] We discuss how our results extend to many-type settings in section 6.

[4] See for instance Bergemann, Brooks, and Morris (2015), Bergemann, Heumann, and Wang (2024).

to providing the monopolist with any signal of the class. After observing the planner's signal, the monopolist offers each consumer a menu of quality-price pairs from which they subsequently choose.

Our main finding is that restricting the planner to revealing only class information can have material consequences for the incentive to regulate information. The recent literature on information design in monopoly trading problems studies the feasible distributions of producer and consumer surplus attainable when the planner can commit to provide any information structure about consumers' types. For instance, Bergemann, Brooks, and Morris (2015) (henceforth BBM) study information design in a monopoly pricing problem without screening, and find that the regulator can implement every efficient split of surpluses in which the monopolist makes at least its 'no information' profit (see Figure 1a). In a setting with screening, Bergemann, Heumann, and Wang (2024) (henceforth BHW) characterize the efficient frontier (Figure 1b). In both these settings, aggregate surplus is maximized at full information. After all, full information about types allows the monopolist to extract full surplus, and hence gives it an incentives to sell efficiently. By implication, any case for privacy protection in these settings requires that the policymaker have a differential interest in providing consumers with surplus.

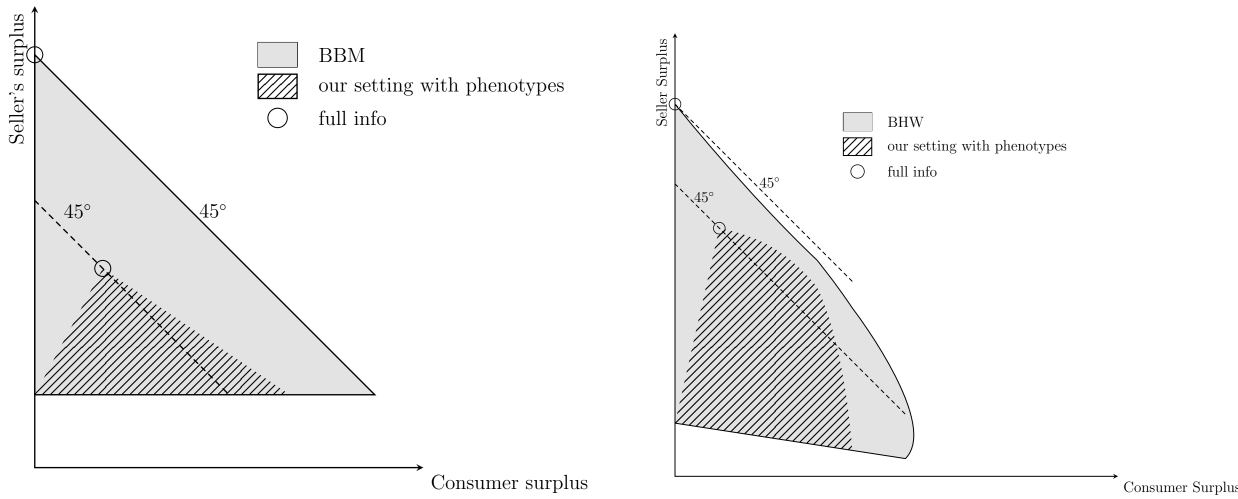
A naïve interpretation of these benchmark intuitions might suggest that, insofar as the planner's objective is total surplus, the planner should allow the monopolist to condition trade on all observable information. We show that this is not the case — when the observable information contained in the class memberships is not fully informative of the customer's type, total surplus can be enhanced by actively restricting the information available to the monopolist. As a result, informational frictions can themselves justify privacy protection policies in particular, it is not necessary for the planner to care intrinsically more about the consumer.[5]

Moreover, this is the case whether or not the monopolist screens types. In the special case without screening, the effect of imperfectly informative class membership contracts the Pareto frontier but also flattens it. In this case, there is a stark implication: the policies that maximize consumer surplus and aggregate surplus coincide (Figure 1a).[6] With screening, imperfect information can again cause the efficient outcome to deviate from the profit-maximizing, full information one – though in these cases optimal design may not go all the way to maximizing consumer surplus (note in particular the slope of the hatched frontier at the point of profit maximization in Figure 1b).

Our results clarify the structure of the planner's optimal information design and their comparative statics with respect to the distribution of classes. Under mild regularity conditions, Proposition 3 shows that the optimal information structure lies in a family of policies which correspond to revealing the identity of a randomly chosen subset of consumers from class B (i.e. the class with many high types), and otherwise pooling the remainder with class

[5] Although we believe this is an important and novel observation in the information-design context, and perhaps at first surprising, it can also be understood as a particular instance of the theory of the second best (Lipsey and Lancaster 1956): with pre-existing distortions, moving "towards" the first best does not ensure a welfare improvement. Another example in a private information context with a "total surplus" objective function is Einav and Finkelstein (2011) who show, in an insurance application, that classification based on observable characteristics like gender has ambiguous effects on deadweight loss.

[6] We discuss the case without screening in Section 5.



(a) Third-Degree Price Discrimination

(b) Second-Degree Price Discrimination

Figure 1: Pareto frontier

A consumers. Though full revelation is a member of this family, the solution often involves a strictly positive degree of pooling. In these cases, the optimal information structure affords class membership A consumers and a subset of class B consumers some privacy protection.

We further characterize the primitive conditions under which privacy protection is optimal. Lemma 5 shows how the optimality of privacy protection depends on the underlying distribution of class memberships. In particular, we show that as the proportion of low types in either class increases (holding fixed the underlying distribution of types), there comes a threshold at which full revelation is not optimal. When the planner maximizes aggregate surplus, this threshold will be interior so long as the other class is not fully informative of type. In particular, the planner is more interested in providing privacy protection when membership in class B is a poor signal of high value customers, and yet membership in class A is a sufficiently strong signal of a low value customer.

Additionally, we characterize the comparative statics of optimal information in the relative weight the planner places on consumer surplus. We find that the degree of pooling increases with this weight, so that the planner provides less information about consumers when it cares more about them. This is intuitive, and follows because the planner's solution lies along the constrained efficient frontier and the monopolist is information-loving.

To establish these results, we show that the planner's information design problem on the space of beliefs over classes can be recast as a constrained information design problem defined on the space of beliefs over types. In our baseline analysis, the planner's feasible choices are the Bayes-plausible distributions whose support lies between the posterior beliefs associated with class B and A. As a result, the solution to the planner's problem can be identified by the concavification of its objective function taken over that restricted domain.

Our result is driven in large part by the geometry of the planner's objective function, to which we refer as 'welfare'. Under natural conditions, we show that this welfare function is first convex, and then concave in the frequency of low value types. The convexity 'at the bottom' introduces an incentive for the planner to reveal some information about classes with few low types, while the concavity 'at the top' incentivizes the regulator to pool classes with many low types, thereby revealing little information about them. This trade-off is best resolved with the family of policies described above.

To understand why welfare is first convex, recall that when the seller believes the proportion of low types is sufficiently small, it induces maximal inefficiency by excluding those types from trade. Doing so extracts all the surplus from trade with high types. As the proportion of low types increases, exclusion becomes too costly and instead the seller becomes willing to trade with these low types, and this trade allows high types to earn information rent. This switch away from exclusion benefits both the seller (by revealed preference) and the consumer (who now enjoys some consumer surplus). In other words, total welfare is locally convex around the exclusion threshold.

To understand why welfare turns concave as the proportion of low types increases past the exclusion threshold, it is helpful to consider the way in which consumer surplus varies. On the one hand, with a higher proportion of low types the consumer can enjoy an information rent less often; on the other, those high types who do exist each enjoy a greater rent, as the seller offers less distorted trade to low types. Under a natural assumption on the seller's marginal costs, the latter effect is diminishing in the proportion of low types, which implies consumer surplus is concave. When the weight on consumers is not too small, this concavity carries over to welfare.

This geometry naturally generates information policies in the family of pooling schemes described above. For intuition as to why full information is not optimal in the presence of class memberships, suppose the planner wishes to maximize aggregate surplus and suppose at first that the classes are fully informative of type. In that case, full information is optimal and achieves the first best allocation in which trade with each type is efficient. Now introduce a small misclassification error into the class, so that some small proportion of low types are mislabelled as members of class B. Now, a fully informative policy that reveals class identity introduces an inefficiency: the monopolist excludes consumers who are mislabelled as low types from the market in order to limit the rents it must concede to all high types with that class membership. The planner can mitigate this cost by pooling class B with class A. Though this pooling will now incentivize the monopolist to introduce some downward distortions to the newly created pool (as there are now high types to exploit), this effect has zero cost to the first order, since the initial allocation was efficient. Hence, some degree of pooling becomes optimal as class B membership becomes a weaker signal of a low type.

In section 5, we show that our findings arise in a variety of other classic asymmetric information problems. In particular, we study a simple version of the BBM model of monopoly pricing, models of monopolistic and competitive insurance, and optimal taxation, and find similar implications for the optimal use of class-level information. In this way, we illustrate that these incentives for using information appear to arise as a common feature of asymmetric information problems, driven in large part by the strong local convexity that is induced around exclusion thresholds.

Moreover, we show that our findings are not limited to the binary structure of the type and class-membership spaces in our analysis. We show how to extend our main analysis to allow for a continuous types and a continuum of class memberships (the latter ordered along a one-dimensional space). In these cases, the tools of Dworczak and Martini 2019 can be used to characterize optimal information design as a function of welfare, and we discuss conditions under which the optimal policy forms a natural extension of the pooling policies described above.

The rest of the paper proceeds as follows. Section 2 discusses the relevant literature. Section 3 sets up the baseline model, and Section 4 derives the main theoretical results. Section 5 applies the framework to canonical information asymmetry environments, including third-degree price discrimination, monopolistic and competitive insurance markets, and tax problems. Section 6 extends the model to a continuum of class memberships and types, while Section 7 outlines its boundaries. Section 8 concludes. Most proofs are relegated to the Appendix.

2 Literature Review

This paper brings together themes from two strands of the information design literature. The first analyzes how information structures interact with price discrimination. The seminal framework of Bergemann, Brooks, and Morris (2015) studies information design under third degree discrimination and has since been extended to environments with second degree discrimination. Subsequent work, for example, uses similar geometric tools to examine how additional information affects consumer surplus (Haghpanah and Siegel 2022) and when Pareto improving segmentations can be constructed in such settings (Haghpanah and Siegel 2023). A unified framework is developed by Bergemann, Heumann, and Wang (2024), who characterize the nongeneric environments in which such segmentation fails. Related work further extends these insights to, for example, arbitrary welfare objectives and endogenously determined types (Dosis and C. Rothschild 2024). Collectively, this line of research characterizes how information shapes feasible welfare distributions when the designer in principle has access to all relevant information and can freely combine second and third degree discrimination. In practice, however, real world information is often constrained by institutional or technological limits. These constraints matter for both policy design and firm behavior, and our paper examines settings with exactly such constraints.

Our paper therefore also relates to the literature that studies information design with constrained feasible posteriors that must satisfy restrictions beyond Bayes plausibility (e.g. Bergemann and Morris 2016 and Dworczak and Martini 2019). This strand also includes work on communication constraints, for example Le Treust and Tomala (2019) and Doval and Skreta (2024), where limited or noisy channels restrict the set of feasible posteriors. Fallah et al. (2024) impose privacy restrictions in the model of Bergemann, Brooks, and Morris (2015). Intermediary-based or data-based constraints have also been explored, for example when intermediaries observe imperfect valuation signals (Galperti, Levkun, and Perego 2024) or when platforms selectively disclose consumer data (Fainmesser, Galeotti, and Momot 2023). Finally, Farboodi, Haghpanah, and Shourideh (2025) study coarse signals in the setting of Bergemann, Brooks, and Morris (2015).

While these papers differ in their specific objectives or forms of restriction, they all study constrained information design in environments without screening. Our interest is in understanding how exogenous information restrictions interact with second and third degree price discrimination. We provide a tight characterization in a tractable class of models and then show that the intuition from this characterization extends to a broad set of environments beyond that benchmark.

3 Model Setup

We begin by considering a simplified environment with a simplified utility function, binary types and binary classes. That is, individuals are split into two classes (e.g., young and old, women and men) and, within each class, individuals have either high or low demand. We will further assume that a regulator places some relative weight on consumer and producer surplus, but the weight placed on consumer's surplus is sufficiently high, in a sense we make precise below. Section 6 extends the results to continuous types and signals.

Sellers A profit-maximizing seller offers a menu of products, each characterized by a quality $q \geq 0$ and a price $t \geq 0$. A product is a pair (q, t). If the seller produces a good of quality q, they incur a cost $c(q) \geq 0$, with $c'(q) > 0$. If the seller sells a good of quality q at a price of t, their profit is

$$\pi(q, t) = t - c(q).$$

We impose the following restrictions on the cost function:

Assumption 1. *$c''(q) > 0$ and $c'''(q) \geq 0$.*

The first part of Assumption 1 posits that the cost function is convex, a standard condition ensuring the firm's optimization problem is well behaved. This is all we need for this section. When we turn to information design, we will also use the second part of Assumption 1, namely convexity of marginal costs. We discuss in that section why this technical condition is useful.

Buyers There is a unit mass of buyers. Buyers have types $\theta \in \{\theta_h, \theta_l\}$, with $\theta_h > \theta_l > 0$, which are each buyer's private information. If a buyer of type θ obtains a good of quality q at a price of t, their utility is

$$U(q, t, \theta) = \theta q - t.$$

If a buyer does not purchase the good, their utility is zero.

Classes Let $\omega \in \Omega := \{\omega_A, \omega_B\}$ denote an observable characteristic (e.g., age or gender) of the buyer. We refer to ω as the buyer's *class*. Class membership provides an imperfectly informative signal about the buyer's type θ, which is the only payoff-relevant attribute for the buyer's choices. In class ω_A, the share of individuals with type θ_l is λ_A, and analogously in class ω_B it is λ_B. A buyer in class ω_A is more likely to be of type θ_l than a buyer in class ω_B, so without loss of generality we assume $\lambda_B < \lambda_A$. Thus, in classes A and B, the shares of individuals of type θ_h are $1 - \lambda_A$ and $1 - \lambda_B$, respectively. If $\lambda_A = \lambda_B$, the two classes have the same composition, so there is no scope for information design.

Information The values λ_A and λ_B are common knowledge. Neither the seller or the regulator know the type θ of any buyer. The regulator commits to an information structure $(\pi, \mathbb{S})$, which consists of a signal space $\mathbb{S}$ and a mapping $\pi : \Omega \to \Delta(\mathbb{S})$.[7]

Belief Updating Given a signal, the seller's beliefs about class memberships become the posterior $\mu = P(\omega = \omega_A)$. A given belief μ about classes ω implies that the seller believes that the share of types θ_l in the overall population of buyers is

$$\Lambda(\mu) := \mu\lambda_A + (1 - \mu)\lambda_B, \tag{1}$$

The belief that a buyer is of type θ_h is $1 - \Lambda$. In particular, if μ_0 is the prior belief about classes then, without any additional information, the seller will believe that the share of types θ_l is $\Lambda_0 := \mu_0\lambda_A + (1 - \mu_0)\lambda_B$.

Seller's Actions A seller's beliefs about class memberships μ imply a unique distribution of types $\Lambda(\mu)$ following (1). For this distribution, the seller will maximize their profit by choosing a feasible menu of two products: one product $(q_l^*(\Lambda(\mu)), t_l^*(\Lambda(\mu)))$ aimed at types θ_l, and a product $(q_h^*, t_h^*(\Lambda(\mu)))$ aimed at types θ_h. As we discuss further below, the typical result of "no distortion at the top" implies that q_h^* does not depend on the seller's beliefs. We characterize the seller's actions and discuss their feasibility in detail in Section 3.1.

Regulator's Objective Suppose the regulator places a weight α on the consumer's surplus and a complementary weight $1-\alpha$ on the seller's profit. Then, if the seller allocates a product (q, t) to a buyer of type θ, the regulator's payoff is

$$v(q, t, \theta) = \alpha U(q, t, \theta) + (1 - \alpha)\pi(q, t).$$

Then, for a given posterior belief μ, and assuming that the seller chooses their optimal feasible actions, the regulator's pay-off is

$$H(\mu) = \Lambda(\mu)v(q_l^*(\Lambda(\mu)), t_l^*(\Lambda(\mu)), \theta_l) + (1 - \Lambda(\mu))v(q_h^*, t_h^*(\Lambda(\mu)), \theta_h) \tag{2}$$

Timing The timing of the game is as follows. First, the regulator commits to an information structure. Then, the buyer's class ω is realized. A random signal $s \sim \pi(\omega)$ is released to the seller, who (Bayesian) updates their beliefs about ω, forming a posterior belief μ. Each posterior belief μ about the classes ω implies also a belief about Λ, the proportion of types θ_l in the population of consumers. Finally, the seller offers a menu of qualities q and prices t conditional on each realization of the signal s, and the each buyer maximizes utility by choosing from the menu they are offered.

3.1 The Seller's Problem

We begin by briefly revisiting the textbook results describing the seller's behavior for a given posterior belief $\Lambda = P(\theta = \theta_l)$ about customers' types.[8] We will then use these results to

[7] There are two possible interpretations. First, the seller does not know ω (for instance ω refers to the person's income, which is unobserved). Second, the seller observes ω (e.g., gender) but the regulator constraints the seller in the extent to which ω can be used in pricing.

[8] See for instance Salanié 2005.

characterize the regulator's optimal choice of information structure.

In a first-best setting with symmetric information (i.e., if θ is known for each buyer), the seller would offer quality $q^{FB}(\theta)$ such that the marginal cost of additional quality matches the consumer's marginal utility for quality: $c'(q^{FB}(\theta)) = \theta$. Moreover, the seller would charge a price that extracts the full surplus from each type: $t^{FB}(\theta) = q^{FB}(\theta)\theta$. In such a first-best setting, there is no distortion and the seller captures the entire surplus.

Let t_h, t_l be the prices charged for qualities q_h, q_l. If the seller's belief regarding the share of types θ_l is Λ, the seller solves

$$\Pi^{seg}(\Lambda) \equiv \max_{t_h,t_l,q_h,q_l} \Lambda \cdot (t_l - c(q_l)) + (1-\Lambda)\cdot(t_h - c(q_h)). \tag{3}$$

By a standard appeal to the Revelation principle, the seller solves (3) subject to (IR) and (IC). Standard arguments imply only the H type IC and L type IR bind, and that the solution involves separation of h and l types into different contracts. This implies that prices are

$$t_h = \theta_h q_h - q_l(\theta_h - \theta_l), \qquad\qquad t_l = \theta_l q_l.$$

Substituting these into the seller's objective function and maximizing over q_h, q_l yields $q_h^* = q^{FB}(\theta_h)$ and

$$q_l^*(\Lambda) = \max\left\{0, q_l : c'(q_l) = \theta_l - \frac{1-\Lambda}{\Lambda}(\theta_h - \theta_l)\right\}. \tag{4}$$

The optimal menu induces type θ_h to consume the first-best quantity q_h^*, which is independent of the belief Λ. The first-best quality for θ_l is $q^{FB}(\theta_l)$, but at the profit-maximizing solution, type θ_l is induced to choose strictly less than the surplus-maximizing choice, unless $\Lambda = 1$ in which case all buyers are of type θ_l, so there is no private information and no distortion. The consumer surplus given belief Λ is equal to the total information rent accruing to the θ_l types, and is readily shown to be:

$$CS^{seg}(\Lambda) = (1-\Lambda)(\theta_H - \theta_L)q_l^*(\Lambda). \tag{5}$$

Because type θ_h's information rent increases in q_l, the seller distorts q_l downward.

The quality that low-demand types are induced to buy $q_l^*(\Lambda)$ is decreasing in the hazard rate $\frac{1-\Lambda}{\Lambda}$, which itself is decreasing in Λ. Therefore, the distortion is decreasing in the belief Λ. Intuitively, when Λ is large, low types are more prevalent, making it more costly to distort their allocation. Hence, Λ can be viewed as a measure of the inefficiency arising in a market.

Typically, the literature on second-price discrimination assumes all types are served, i..e, (4) is strictly positive. However, it is important in our setting to allow for the possibility that some types may not be served. From the expression for q_l^* above, the optimal contract involves not serving low types ($q_l^* = 0$) whenever

$$\Lambda \le 1 - \frac{\theta_l}{\theta_h} := \bar{\Lambda}. \tag{6}$$

The low type is excluded whenever Λ is sufficiently low (i.e. when low types are rare), or θ_l is sufficiently small relative to θ_h. When (6) holds, the profit gain from reducing the information rent for θ_h outweighs the cost of excluding θ_l. As we show below, this incentive to exclude low-demand types will have important implications for information design in this setting.

3.2 The Regulator's Problem

The regulator's problem is to choose how much of the buyers' class-specific information should be revealed to the seller in order to maximize social welfare. By committing to a particular information disclosure policy, the regulator shapes the seller's posterior belief Λ that a buyer is of the low-demand type. This belief, in turn, determines how the seller tailors the menu of contracts and the extent to which the seller distorts the allocation for low-demand buyers. The regulator's objective is to select a distribution over these posterior beliefs that maximizes their objective. At the same time, the policy must satisfy Bayes' plausibility; that is, when the seller averages over all possible posteriors, the overall belief about the buyer types must equal the prior Λ_0.

As stated above, the regulator's objective function, as a function of posterior beliefs is given by $H(\mu)$ defined by (2) for $\mu \in [0,1]$. By Kamenica and Gentzkow (2011), the welfare-maximizing information structure is to choose a distribution $\tau \in \Delta([0,1])$ over the posterior belief $\mu \in [0,1]$ to maximize this objective. Specifically,

$$\max_{\tau \in \Delta([0,1])} \mathbb{E}_{\mu\sim\tau}[H(\mu)] \qquad \text{s.t.} \;\; \mathbb{E}_{\mu\sim\tau}[\mu] = \mu_0 \tag{7}$$

We now show that this problem is equivalent to one where the objective function is $W(\Lambda)$ defined as follows. Since a given belief μ implies a unique distribution $\Lambda(\mu)$, it will be convenient in what follows to express the regulator's pay-off as a function of this induced distribution:

$$\begin{aligned} W(\Lambda) &= \Lambda v(q_l^*(\Lambda), t_l^*(\Lambda), \theta_l) + (1-\Lambda) v(q_h^*, t_h^*(\Lambda), \theta_h) \\ &= \alpha CS^{seg}(\Lambda) + (1-\alpha)\Pi^{seg}(\Lambda) \end{aligned} \tag{8}$$

for $\Lambda \in [\lambda_B, \lambda_A]$.

Lemma 1. *The welfare-maximizing information structure solves the following problem:*

$$\begin{aligned} \max_{\sigma \in \Delta([\lambda_B, \lambda_A])} \; & \mathbb{E}_{\Lambda\sim\sigma}[W(\Lambda)] \\ s.t. \;\; & \mathbb{E}_{\Lambda\sim\sigma}[\Lambda] = \Lambda_0. \end{aligned} \tag{Bayes Plausiblity}$$

Proof. Changing the variables from μ to Λ using $\Lambda = \mu\lambda_A + (1-\mu)\lambda_B$ yields Lemma 1. □

Lemma 1 describes a problem in which the regulator directly chooses a distribution over interim beliefs Λ that the seller holds over the consumer's type; we may view this induced Λ as an observable segment in which the consumer lies. The regulator's choice of segment assignment is subject to two constraints. First, the regulator cannot systematically alter the seller's mean belief in the consumer's type (each consumer having been assigned to *some* segment). This is simply a Bayes plausibility constraint. The second constraint is an informational one: the regulator cannot provide more information than it can observe itself (that is, class membership). As a result, any observable segment must have an associated Λ which corresponds to some mixture of λ_B and λ_A. Equivalently, we must have $\lambda_B \leq \Lambda \leq \lambda_A$ for all $\Lambda \in \text{supp}\,\sigma$.

We now show that the optimal information structure obtained from the concavification of $H(\mu)$ on the domain $\mu \in [0,1]$, can be identified with the concavification of $W(\Lambda)$ on the constraint set $\Lambda \in [\lambda_B, \lambda_A]$.

Lemma 2. *Let $f : [0,1] \to \mathbb{R}$, and write $co(f)$ for the smallest concave majorant of f. Let $y = ax + b$, for scalars a, b with $a > 0$. Then the function $g : [b, a+b] \to \mathbb{R}$ defined by*

$$g(y) = f \circ x^{-1}(y) = f\left(\frac{y-b}{a}\right) \tag{9}$$

has

$$co(g) = co(f) \circ x^{-1}. \tag{10}$$

Proof. We need only verify that $co(f) \circ x^{-1}$ is the smallest concave majorant of g. That $co(f) \circ x^{-1}$ is concave and majorizes g are obvious.

We show that $co(f) \circ x^{-1}$ is the smallest such majorant. Suppose for the sake of a contradiction that there exists some other concave function $h : [b, a+b] \to \mathbb{R}$ with

$$g \leq h \leq (cof) \circ x^{-1}$$

with $h(y') < (cof) \circ x^{-1}(y')$ for at least one $y' \in [b, a+b]$. Then the function $h \circ y : [0,1] \to \mathbb{R}$ is a concave majorant of f with $h \leq co(f)$, and

$$h(\frac{y'-b}{a}) < co(f)(\frac{y'-b}{a}),$$

so $co(f)$ is not the smallest concave majorant of f, a contradiction. □

Lemma 2 is convenient. While the function $H(\mu)$ varies with the parameters λ_A, λ_B, W does not. As we will see in the next section, this change of variables admits a simple geometric approach to solving for the regulator's optimal information policy as a function of the classes. Moreover, as Λ is the sufficient statistic for determining both the seller's behaviour and interim surpluses, working with it has the advantage of more directly capturing the economic trade-offs facing the regulator. The intuition is straightforward: μ and Λ are linearly related, and so the change of variables preserves the curvature of the welfare function. As a result, the two functions share essentially the same concavification—up to a linear rescaling of the domain.

4 Results

In this section, we maintain the following assumptions, which are useful for characterizing the geometry of the regulator's objective function:

Assumption 2. $\alpha > 1/3$

Assumption 2 implies that the regulator places a sufficiently large weight on the surplus of consumers. This kind of special regulatory concern for consumer protection is not unreasonable. Indeed, competition authorities such as the Federal Trade Commission (FTC) and

Competition and Markets Authority (CMA) appear to give significant priority to consumer welfare.[9] Still, we do not require that the welfare objective puts more weight on consumers than producers ($\alpha > 1/2$). Indeed, Assumption 2 admits the case of a regulator whose objective is to maximize aggregate surplus ($\alpha = \frac{1}{2}$). Beyond that, our focus is on understanding the behaviour of a regulator who does not put *too much* weight on the producer.

The second part of Assumption 1, i.e. convex marginal cost, implies that the seller's optimal supply to low types q_l^* becomes successively less sensitive to increases in virtual values (see the right side of (4)), on the range $\Lambda \geq \bar{\Lambda}$ on which low types are always served. As virtual surplus is increasing in Λ and subject to diminishing returns, these features together imply that the the seller's offer q_l^* to low types is concave. One interpretation of this is therefore that Λ has a diminishing marginal impact in correcting the monopoly distortion.

With these assumptions in hand, we are now ready to characterize the geometry of the regulator's objective in Λ.

4.1 Characterizing the Regulator's Objective

This subsection characterizes the geometry of the regulator's interim objective $W(\Lambda)$ described in (8). This geometry is useful for understanding the regulator's incentives to influence the firm's beliefs, and is thus a key input to solving its information design problem. Under Assumptions 2 and 1, we show that W is is 'S-shaped': in particular, it is first affine on $\Lambda \in [0, \bar{\Lambda}]$, reaches a convex kink at $\Lambda = \bar{\Lambda}$ and is thereafter concave on $\Lambda \in (\bar{\Lambda}, 1]$.

Proposition 1. *For $\Lambda \leq \bar{\Lambda} := 1 - \theta_l/\theta_h$, the low type is excluded, $q_l(\Lambda) = 0$, and*

$$W(\Lambda) = (1-\alpha)(1-\Lambda)\big(\theta_h q_h - c(q_h)\big),$$

which is affine and strictly decreasing in Λ.

Proof. In this region, $W(\Lambda)$ does not depend on q_l, and moreover q_h does not vary with Λ. Since $\theta_h q_h - c(q_h) > 0$, the result follows. □

In the region $\Lambda \leq \bar{\Lambda}$, the seller does not change the menu offered in response to changes in beliefs. Therefore interim welfare changes only through the direct effect of the change in beliefs about the proportion of types θ_l, and is therefore linear. Since low types are excluded from trade in this range, W decreases in their share Λ on this range.

By contrast, for $\Lambda \geq \bar{\Lambda}$, the proportion of low types is sufficiently high that the seller no longer excludes them, and moreover adapts its menu so that distortions are decreasing in Λ. The next proposition characterizes the consequences for welfare:

[9]For example, the FTC describes advancing consumers' interests as a major goal, as "...the only federal agency that deals with consumer protection and competition issues in broad sectors of the economy." (Federal Trade Commission 2025). Likewise, in describing the CMA's purpose, Chief Executive Sarah Cardell said "consumers... must have the confidence that the CMA is actively standing up for their interests" (Competition and Markets Authority 2025).

Proposition 2. *Suppose Assumptions 2 and 1 hold. For every $\Lambda > \bar{\Lambda}$, the low type is served* and

$$W''(\Lambda) \;\leq\; \Delta\theta\, q_l'(\Lambda)\,\frac{1-3\alpha}{\Lambda}.$$

W is thus strictly concave on $(\bar{\Lambda}, 1]$.

Notice that the conditions in Proposition 2 (Assumptions 2 and 1) are sufficient but not necessary for the concavity of W. As discussed in section 4, Assumption 1 is sufficient to ensure that q_l^* is concave in Λ for $\Lambda \geq \bar{\Lambda}$. It is not hard to verify that this in turn is sufficient to ensure that CS^{seg} is concave on the range $\Lambda \geq \bar{\Lambda}$. While PS^{seg} is convex on this range, we show in the Appendix that Assumption 2 implies that enough weight in W is placed on CS^{seg}, so that its concavity is inherited everywhere on $\Lambda \geq \bar{\Lambda}$.

While Propositions 1 and 2 characterize the shape of W on the domains $\Lambda \leq \bar{\Lambda}$ and $\Lambda \geq \bar{\Lambda}$ respectively, they do not fully describe the geometry of W because they are silent about its properties as Λ crosses these domains. The following Lemma completes our characterization of the geometry of W by describing this transition.

Lemma 3. *Suppose Assumptions 2 and 1 hold. The right-derivative of W at $\Lambda = \bar{\Lambda}$ strictly exceeds its left derivative:*

$$W'\left(\bar{\Lambda}_-\right) < W'\left(\bar{\Lambda}_+\right).$$

Lemma 3 establishes that W has a convex kink at $\Lambda = \bar{\Lambda}$. Importantly, this implies that W is not globally concave. If it were, then the regulator would always prefer to reveal no information about class membership at all, regardless of the values of λ_B or λ_A. By contrast, the existence of this convex kink introduces at least some incentive for the regulator to provide information. The underlying reason for the kink is simple: to the left of $\bar{\Lambda}$, low type consumers are shut out of the market and the seller fully extracts surplus from high types. However, as Λ increases above $\bar{\Lambda}$ the monopolist becomes willing to serve low types, which leads to information rents and positive consumer surplus. This induces consumers to be locally information loving (i.e., convex) at $\bar{\Lambda}$, since small increases in Λ yield them positive rents, while small reductions do not cause them to sacrifice any rent. Since the seller is always naturally information loving, W also inherits this property.

A common assumption in the literature is the case of quadratic costs, i.e. $c(q) = \frac{1}{2}q^2$. Quadratic costs imply $c''(q) = 1$ and $c'''(q) = 0$, so Assumption 1 is satisfied. In this special specification, the regulator's objective simplifies to $W = (1-3\alpha)\,\frac{(\Delta\theta)^2}{\Lambda^3}$, so W is concave on $[\bar{\Lambda}, 1]$ if $\alpha \geq 1/3$ and globally convex otherwise. Since full information revelation is optimal under global convexity, assumption 2 is unnecessary in this knife-edge quadratic case. We discuss the case when $\alpha < 1/3$ in Section 7.

4.2 Optimal Information Design

We now characterize the regulator's optimal information policy for $\alpha > 1/3$ (with $c''(q) > 0$ and $c'''(q) \geq 0$). Throughout, remember that the prior over types induced by the class prior μ_0 is

$$\Lambda_0 \;=\; \mu_0\lambda_A + (1-\mu_0)\lambda_B \;\in\; [\lambda_B, \lambda_A],$$

and, from Section 4.1, the interim objective $W(\Lambda)$ is affine on $[\lambda_B, \bar{\Lambda}]$ and concave on $(\bar{\Lambda}, \lambda_A]$, with strict concavity on $(\bar{\Lambda}, \lambda_A]$ whenever $\alpha > 1/3$ or $c''' > 0$.

Following Kamenica and Gentzkow 2011, the optimal information policy is obtained by 1) building the concavification of $W(\Lambda)$ for $\Lambda \in [\lambda_B, \lambda_A]$ and 2) evaluating this concavification at the prior $\Lambda_0 \in [\lambda_B, \lambda_A]$.

If $\lambda_B < \lambda_A < \bar{\Lambda}$ then $W(\Lambda)$ is linear decreasing in its entire domain $\Lambda \in [\lambda_B, \lambda_A]$. In this case, the concavification coincides with the function itself and all information policies yield the same value for the regulator.

If $\bar{\Lambda} < \lambda_B < \lambda_A$ then $W(\Lambda)$ is concave in its entire domain $\Lambda \in [\lambda_B, \lambda_A]$. In this case, the concavification again coincides with the function itself and the optimal information policy is no revelation.

We therefore focus henceforth on the interesting case where $\lambda_B < \bar{\Lambda} < \lambda_A$. This is the case where, if class membership were fully revealed, individuals of class ω_B would be offered a menu that excludes types θ_l, but individuals of class ω_A would be offered a menu that does not exclude types θ_l.

In this case, we show that the concavification of $W(\Lambda)$ consists of a straight line from the point $(\lambda_B, W(\lambda_B))$ to a point $(\hat{\Lambda}, W(\hat{\Lambda}))$ and then coincides with $W(\Lambda)$ for $\Lambda > \hat{\Lambda}$.

This implies an information policy a following family of one-parameter *partial revelation policies*, parameterized by $\hat{\Lambda}$. This family includes full revelation and full privacy as limits. We will then show that the optimal information policy is included in this family.

More precisely, fix a target $\hat{\Lambda} \in (\lambda_B, \lambda_A]$ and consider the signal space $\mathbb{S} = \{\text{“}B\text{”}, \text{pool}\}$. From Kamenica and Gentzkow 2011, we know that since the space of classes Ω has only two elements, only two messages are needed. The prior belief over classes is μ_0. Define the partial revelation policy $\mathsf{PRP}(\hat{\Lambda})$ as follows. First, if $\omega = \omega_A$, send the signal *pool* with probability 1. Second, if $\omega = \omega_B$, send the signal “B” with probability $1 - \gamma(\hat{\Lambda})$ and *pool* with probability $\gamma(\hat{\Lambda})$, where

$$\gamma(\hat{\Lambda}) \;=\; \frac{\mu_0(\lambda_A - \hat{\Lambda})}{(1-\mu_0)(\hat{\Lambda} - \lambda_B)} \;=\; 1 - \frac{\hat{\Lambda} - \Lambda_0}{(1-\mu_0)(\hat{\Lambda} - \lambda_B)}. \tag{11}$$

Upon sending the message “B”, the posterior is clearly $\Lambda = \lambda_B$. Upon sending the message “pool”, Bayes' rule gives the posterior over class memberships as

$$\hat{\mu} \;=\; \frac{\mu_0}{\mu_0 + (1-\mu_0)\gamma(\hat{\Lambda})},$$

and in turn this implies a posterior over types $\Lambda \;=\; \hat{\mu}\lambda_A + (1-\hat{\mu})\lambda_B \;=\; \hat{\Lambda}$.

Thus $\mathsf{PRP}(\hat{\Lambda})$ induces a two-point posterior on Λ with support $\{\lambda_B, \hat{\Lambda}\}$ and weights

$$x \;=\; \Pr\{\Lambda = \lambda_B\} \;=\; \frac{\hat{\Lambda} - \Lambda_0}{\hat{\Lambda} - \lambda_B}, \qquad 1 - x \;=\; \Pr\{\Lambda = \hat{\Lambda}\} \;=\; \frac{\Lambda_0 - \lambda_B}{\hat{\Lambda} - \lambda_B}. \tag{12}$$

These weights satisfy Bayes plausibility: $x\,\lambda_B + (1-x)\,\hat{\Lambda} = \Lambda_0$.

Notice that full information and zero information are a part of this family. If $\gamma(\hat{\Lambda}) = 1$, no information is revealed, and therefore full privacy obtains. If $\gamma(\hat{\Lambda}) = 0$, the signals separate classes A and B fully, and full revelation obtains (in this case, a signal of ”pool” would be fully revealing of class A). Notice also that these information policies convey information only about classes ω, not types θ which are the private information of individual.

We have defined a general partial revelation policy for an arbitrary target $\hat{\Lambda}$. Now, we we show that the optimal information policy belongs to this family and has $\hat{\Lambda} = \Lambda^\star$, which we define below.

Proposition 3. *Suppose Assumptions 2 and 1 hold. For any prior $\Lambda_0 \in [\lambda_B, \lambda_A]$:*

1. *If $\lambda_A \leq \bar{\Lambda}$, $W(\Lambda)$ is affine on $[\lambda_B, \lambda_A]$ and* all *information policies (in particular any* $\mathsf{PRP}(\hat{\Lambda})$*) deliver the same expected objective $W(\Lambda_0)$.*

2. *If $\lambda_B \geq \bar{\Lambda}$, $W(\Lambda)$ is (weakly) concave on $[\lambda_B, \lambda_A]$ and* full privacy *(the degenerate posterior at Λ_0) is optimal.*

3. *If $\lambda_B \leq \bar{\Lambda} < \lambda_A$, there exists a threshold $\Lambda^\star \in [\bar{\Lambda}, \lambda_A]$ such that:*

   - *if $\Lambda_0 < \Lambda^\star$, an optimal policy is* $\mathsf{PRP}(\Lambda^\star)$*, which induces posteriors $\{\lambda_B, \Lambda^\star\}$ with weights given by (12);*
   - *if $\Lambda_0 \geq \Lambda^\star$,* full privacy *is optimal.*

   Moreover, $\Lambda^\star$ is the unique solution to

$$\Lambda^\star = \arg \max_{\Lambda \in [\lambda_B, 1]} \frac{W(\Lambda) - W(\lambda_B)}{\Lambda - \lambda_B}. \tag{13}$$

To see the relevance of (13) to Proposition 3, observe that the regulator's optimal information structure lies in the family that corresponds to splittings of the prior belief Λ_0 into λ_B and some $\Lambda \geq \Lambda_0$. The regulator's payoff from a policy in this family can then be expressed as

$$\begin{aligned}
\mathbb{E}[W] &= \frac{\Lambda - \Lambda_0}{\Lambda - \lambda_B} W(\lambda_B) + \frac{\Lambda_0 - \lambda_B}{\Lambda - \lambda_B} W(\Lambda) \\
&= W(\lambda_B) + (\Lambda_0 - \lambda_B) \frac{W(\Lambda) - W(\lambda_B)}{\Lambda - \lambda_B},
\end{aligned} \tag{14}$$

so that the regulator is interested in maximizing the chord gradient described in (13). Indeed, when $\lambda_B < \bar{\Lambda}$ and $\lambda_A = 1$, it follows from Proposition 3 that the regulator's optimal payoff is given by the concavification

$$co(W)(\Lambda) = \begin{cases} W(\lambda_B) + \left(\frac{\Lambda - \lambda_B}{1 - \lambda_B}\right) w^\star, & \text{for } \lambda_B \leq \Lambda \leq \Lambda^\star \\ W(\Lambda), & \text{for } \Lambda^\star \leq \Lambda \leq 1, \end{cases}$$

where $w^\star$ is the value of the chord gradient in (13) at $\Lambda = \Lambda^\star$. Hence, we obtain a full characterization of the concavification. Moreover, if $\lambda_A < 1$ but $\lambda_A > \Lambda^\star$, the concavification remains the same as in the case with $\lambda_A = 1$. If instead $\lambda_A < \Lambda^\star$, the concavification coincides with full information.

As we discuss in the next section, both $\Lambda^\star$ and $w^\star$ depend in general on λ_B and α. It is useful to highlight the economic content of the policy in Case 3. Since $\lambda_B < \lambda_A$, class B has the smaller share of low types and therefore exhibits the largest downward distortions for the low type. The optimal partial policy *sometimes* reveals B (placing some mass at λ_B) but never reveals A; the remaining B-buyers are pooled with all A-buyers to produce a single posterior $\hat{\Lambda}$ that is chosen exactly at the point where the chord from λ_B to $\bar{\Lambda}$ is tangent to $W(\Lambda)$. As more B mass is drawn into the pool, Λ in the pool falls and the pool's distortion rises; $\hat{\Lambda}$ equates the marginal gain from carving out additional B mass (who then face the B-specific menu) with the marginal loss from worsening distortions in the pooled menu composed of all of class A and some of class B.

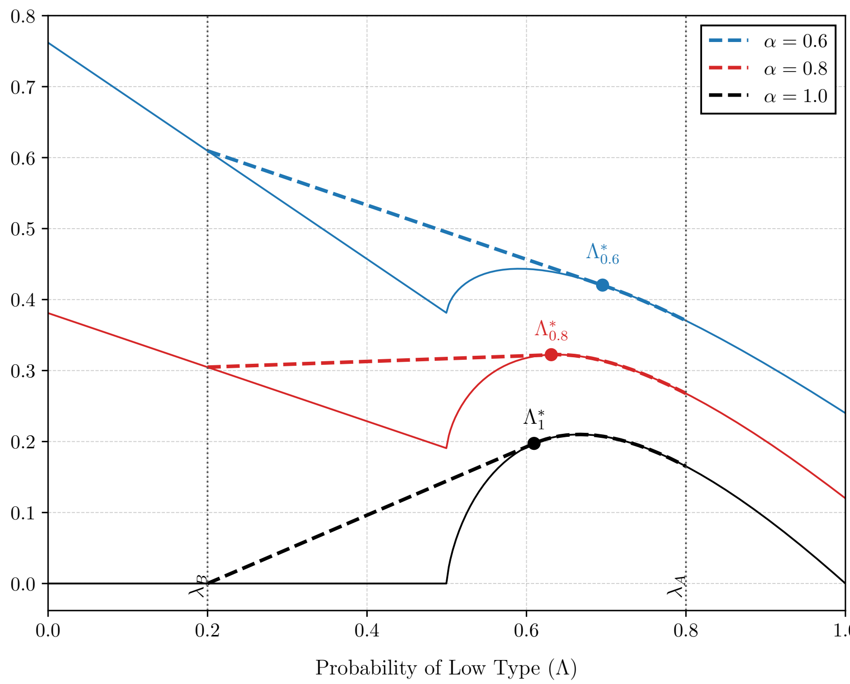


Figure 2: The regulator's objective function dependent on the interim belief Λ with cost function $c(q) = \frac{1}{3}q^3$. Dashed lines are concave closures for different relative welfare weights α.

4.3 Comparative Statics

In this section, we discuss how optimal information structures vary in the parameters α and λ_B, maintaining assumptions 2 and 1.

We treat the overall share of low types Λ_0 across both classes as the fundamental market characteristic that remains fixed. When we vary λ_A (the share of low types in class A), we simultaneously adjust λ_B to maintain the same aggregate Λ_0. This approach captures situations where the underlying composition of the consumer population stays constant, but different observable signals partition this population in different ways. For instance, various demographic characteristics—gender, age, location—represent alternative ways to divide the same consumer base into classes, each yielding different values of λ_A, λ_B while preserving the overall type distribution Λ_0.

We focus attention in this section on the non-trivial case where $\lambda_B < \bar{\Lambda} < \lambda_A$.

Lemma 4. *Under Assumptions 2 and 1, the function*

$$\frac{W(\Lambda) - W(\lambda_B)}{\Lambda - \lambda_B} \tag{15}$$

is strictly quasiconcave. Hence, there is a unique $\Lambda^\star$ satisfying (13), which solves the complementary slackness conditions

$$W'(\Lambda^\star) - \frac{W(\Lambda^\star) - W(\lambda_B)}{\Lambda^\star - \lambda_B} \geq 0 \tag{16}$$

$$\Lambda^\star - 1 \leq 0 \tag{17}$$

$$\left(W'(\Lambda^\star) - \frac{W(\Lambda^\star) - W(\lambda_B)}{\Lambda^\star - \lambda_B}\right)(\Lambda^\star - 1) = 0 \tag{18}$$

and satisfies $\Lambda^\star \in (\bar{\Lambda}, 1]$. Finally, when $\Lambda^\star$ satisfies (16) with equality, the left side of (16) has derivative $W''(\Lambda^\star) < 0$.

We now describe the comparative statics of $\Lambda^\star$ in the following two lemmas. Where convenient, we may write $\Lambda^\star(\lambda_B)$ or $\Lambda^\star(\alpha)$ to emphasize the relevant dependency.

Lemma 5. *(Comparative statics of $\Lambda^\star$ in λ_B) Fix $\alpha > \frac{1}{3}$. There exists a $\lambda'_B \in [0, \bar{\Lambda})$ (which may depend on α) such that $\Lambda^\star$ satisfies (16) with equality if and only if $\lambda_B \in [\lambda'_B, \bar{\Lambda})$. Moreover, $\Lambda^\star$ is a decreasing, differentiable function of λ_B on this range, with $\lim_{\lambda_B \to \bar{\Lambda}} \Lambda^\star = \bar{\Lambda}$.*

Figure 3 illustrates the comparative-statics intuition with respect to λ_B. When $\lambda'_B \in [\bar{\Lambda}, 1)$, the optimal information design features full privacy. In this region equation (16) is slack, so neither the information structure nor $\Lambda^\star$ responds to changes in λ_B.

At $\lambda'_B \in [0, \bar{\Lambda})$, equation (16) binds. Partial information revelation is optimal and the chord tangency at D corresponds to $\Lambda^\star = \Lambda'^\star$. At $\lambda''_B \in \left[\lambda'_B, \bar{\Lambda}\right)$, the chord gradient is shallower, shifting the tangency to point E, with corresponding $\Lambda^\star = \Lambda'^\star$, to the left of D. When $\lambda_B = 0$, the chord tangency occurs exactly at $\Lambda^\star = 1$.

Lemma 6. *(Comparative statics of $\Lambda^\star$ in α) Fix $\lambda_B < \bar{\Lambda}$. There exists a $\alpha' \in [\frac{1}{3}, \frac{1}{2}]$, with $\alpha' = \frac{1}{2}$ if $\lambda_B = 0$, such that $\Lambda^\star$ satisfies (16) with equality if and only if $\alpha > \alpha'$. Moreover, $\Lambda^\star$ is a decreasing, differentiable function of α on this range.*

The comparative statics of the regulator's optimal information structure follow straightforwardly from these lemmas. Indeed, the parameter λ_A simply enters as an additional constraint of the form $\Lambda^\star \leq \lambda_A$ in the problem described by (13). This constraint binds if and only if $\Lambda^\star > \lambda_A$, in which case (by the quasi-concavity of the objective) the constrained optimum is λ_A.

For general $\lambda_B < \bar{\Lambda} < \lambda_A$, the optimal information structure is therefore described by a Bayes plausible splitting to $\{\lambda_B, \min\{\Lambda^\star, \lambda_A\}\}$, where the probability of revealing class B is increasing in $\Lambda^\star$, with full revelation if and only if $\Lambda^\star \geq \lambda_A$.

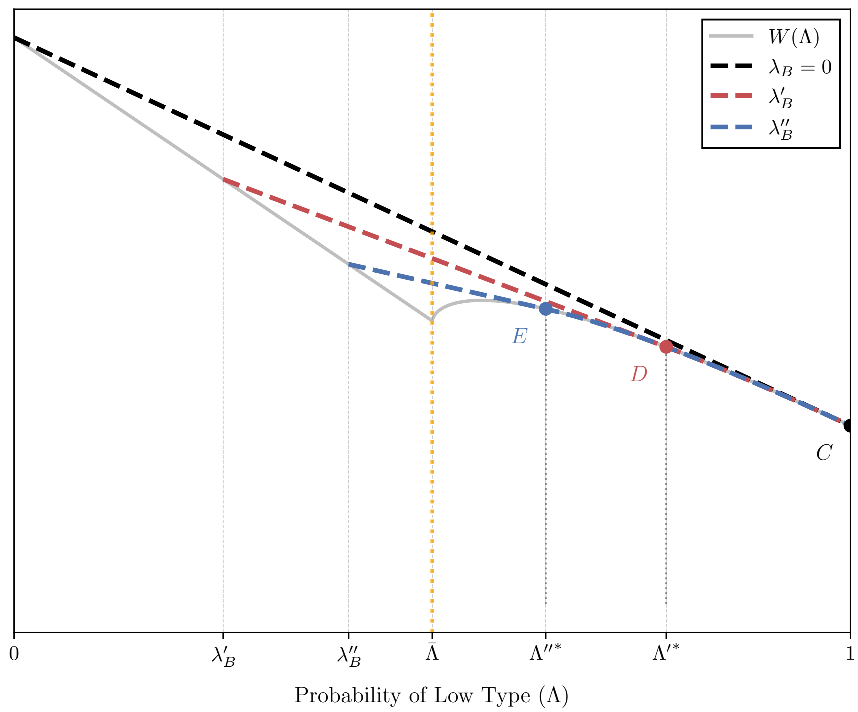


Figure 3: Graphical Intuition for Lemma 5 with $\alpha = 1/2$. Dashed lines are concave closures for different class compositions λ_B.

4.4 The Pareto frontier

In the comparative statics earlier, we examined how the optimal information structure varies locally in the weight assigned to consumer surplus. Using these results, we characterize the Pareto frontier in the (consumer surplus, producer surplus) plane generated by information policies. This is useful to quantify the trade-off faced by a regulator: the producer-surplus cost of delivering an additional unit of consumer surplus.

The results in the previous section provide a simple characterization of the consumer and producer surplus for any given α. When $\lambda_B < \bar{\Lambda}$ and $\alpha \geq \frac{1}{3}$, consumer surplus is:

$$CS(\alpha) \equiv \begin{cases} \mu_0 CS^{seg}(\lambda_A) & \text{if } \lambda_A < \Lambda^\star(\alpha) \\ \frac{\Lambda_0 - \lambda_B}{\Lambda^\star(\alpha) - \lambda_B} CS^{seg}(\Lambda^\star(\alpha)) & \text{if } \Lambda_0 < \Lambda^\star(\alpha) \leq \lambda_A \\ \Lambda_0 CS^{seg}(\Lambda_0) & \text{if } \Lambda^\star(\alpha) \leq \Lambda_0, \end{cases} \tag{19}$$

where CS^{seg} was defined in section 3.1. $CS(\alpha)$ therefore represents the ex ante expected consumer surplus across segments induced by the optimal information design of a regulator with weight α.

The producer surplus is defined analogously:

$$PS(\alpha) \equiv \begin{cases} \mu_0 \Pi^{seg}(\lambda_A) + (1 - \mu_0)\Pi^{seg}(\lambda_B) & \text{if} \lambda_A < \Lambda^\star(\alpha) \\ \frac{\Lambda_0 - \lambda_B}{\Lambda^\star(\alpha) - \lambda_B} \Pi^{seg}(\Lambda^\star(\alpha)) + \frac{\Lambda^\star(\alpha) - \Lambda_0}{\Lambda^\star(\alpha) - \lambda_B} \Pi^{seg}(\lambda_B) & \text{if } \Lambda_0 < \Lambda^\star(\alpha) \leq \lambda_A \\ \Lambda_0 \Pi^{seg}(\Lambda_0) & \text{if} \Lambda^\star(\alpha) \leq \Lambda_0. \end{cases} \tag{20}$$

Let Z be the set of all welfare vectors (CS, PS) that are achievable with some information design, and let

$$Z^* \equiv \{(CS, PS) \in Z | \nexists (CS', PS') \text{ } s.t. \text{ } CS' \geq CS, PS' \geq PS \text{ and } CS' + PS' > CS + PS\}$$

be the Pareto frontier. Note that $(CS(\alpha), PS(\alpha)) \in Z^*$ for $\alpha \geq \frac{1}{3}$. For $\alpha < \frac{1}{3}$, we have not characterized $CS(\alpha)$ and $PS(\alpha)$, and it is possible they are correspondences. We mildly abuse notation by continuing to denote it by $(CS(\alpha), PS(\alpha)))$ in what follows. Since we are ultimately interested in the $\alpha \geq \frac{1}{3}$ case, the following proposition takes $\lambda_A < \Lambda^\star\left(\frac{1}{3}\right)$, and shows that the entire Pareto frontier is characterized by points $(CS(\alpha), PS(\alpha))$ for $\alpha \geq \frac{1}{3}$.

Proposition 4. *Fix $\lambda_B < \bar{\Lambda}$ and assume $\bar{\Lambda} < \lambda_A < \Lambda^\star\left(\frac{1}{3}\right)$. Then $\exists \underline{\alpha} > \frac{1}{3}$ and $\overline{\alpha} \leq 1$ such that*

$$Z^* = \{(CS(\alpha), PS(\alpha))\}_{\alpha \in [\underline{\alpha}, \overline{\alpha}]}$$

If and only if $\lambda_A \leq \Lambda^\star(1)$, we can take $\underline{\alpha} = \overline{\alpha}$. Otherwise, we can choose $\underline{\alpha}$ and $\overline{\alpha}$ such that $\alpha \to (CS(\alpha), PS(\alpha))$ is a diffeomorphism on $[\underline{\alpha}, \overline{\alpha}]$, with $\overline{\alpha} < 1$ if and only if $\Lambda_0 > \Lambda^\star(1)$.

A few comments on Proposition 4. First, full information revelation is always firm optimal, and, except in knife-edge cases, there will be a neighborhood of 0 for which it remains so—that is, a whole range of α values which yield the same point on the Pareto frontier. For instance, in the case with 'trivial' class memberships (considered by Bergemann, Heumann, and Wang 2024 and Dosis and C. Rothschild 2024, and as implied by Lemma 6), full information revelation is optimal for all $\alpha \in \left[0, \frac{1}{2}\right]$, and $\Lambda^\star(\alpha) = 1$ in this range. The assumption that $\lambda_A < \Lambda^\star(\frac{1}{3})$ is thus often a reasonable one. Indeed, it follows from Lemma 5 that for any $\lambda_A \in \left[\bar{\Lambda}, 1\right]$ there exists a threshold $\lambda'_B \in \left(0, \bar{\Lambda}\right]$ (decreasing in λ_A) such that $\lambda_A < \Lambda^\star(\frac{1}{3})$ whenever $\lambda_B \leq \lambda'_B$. We make it here for expositional simplicity, as it ensures that full information is optimal at $\alpha = \frac{1}{3}$, and therefore for all lower α. This in turn ensures that the entire Pareto frontier is traced out for α for which we know W is concave. We discuss what can happen otherwise in section 7.

Second, when $\lambda_A \leq \Lambda^\star(1)$, full revelation is the *only* Pareto optimal point, as will be true, for instance, when λ_A is sufficiently close to $\bar{\Lambda}$. Otherwise, the consumer optimum will always involve some degree of privacy. Indeed, it follows from Lemmas 5 and 6 that there is a further threshold $\lambda''_B \in (0, \lambda'_B)$ (decreasing in λ_A) such that $\lambda_A > \Lambda^\star(1)$ when $\lambda_B > \lambda''_B$, and some privacy is optimal for sufficiently high α.

Third, only if $\Lambda_0 \geq \Lambda^\star(1)$ do consumers find *full* privacy to be optimal. In that case, there may be a range of α near $\alpha = 1$ for which full privacy is optimal, and we would then be able to trace out the entire frontier using a range of α values strictly interior to $[0, 1]$.

Fourth, because the Pareto frontier is differentiable, its slope at the point which is optimal for social planner with welfare weights α is equal to $-\frac{\alpha}{1-\alpha}$. If $\underline{\alpha} < \frac{1}{2}$, this means that the Pareto frontier will be shallower than 45-degrees; in this region, the "equity-efficiency tradeoff" of information is reversed relative to the case without classes: information regimes that are better for firms are *worse* for efficiency. As a simple corollary of Lemma 6, $\underline{\alpha} < \frac{1}{2}$ will always hold when $\lambda_B > 0$ and λ_A is sufficiently close to 1.

5 Extensions and Robustness

This section shows that the basic insights from the preceding do not hinge on the particulars of the Mussa-Rosen-like monopoly screening environment. We show that they apply in four distinct settings: a setting without screening (third-degree price discrimination only); a monopoly insurance setting; a competitive insurance setting; and an optimal income tax setting.

5.1 Only Third-Degree Price Discrimination

Consider instead a profit-maximizing seller who can only offer one product, so there is only third-degree price discrimination, as in Bergemann, Brooks, and Morris (2015). We use the same notation and class-level characteristics as in our main model and normalize the firm's cost to zero.

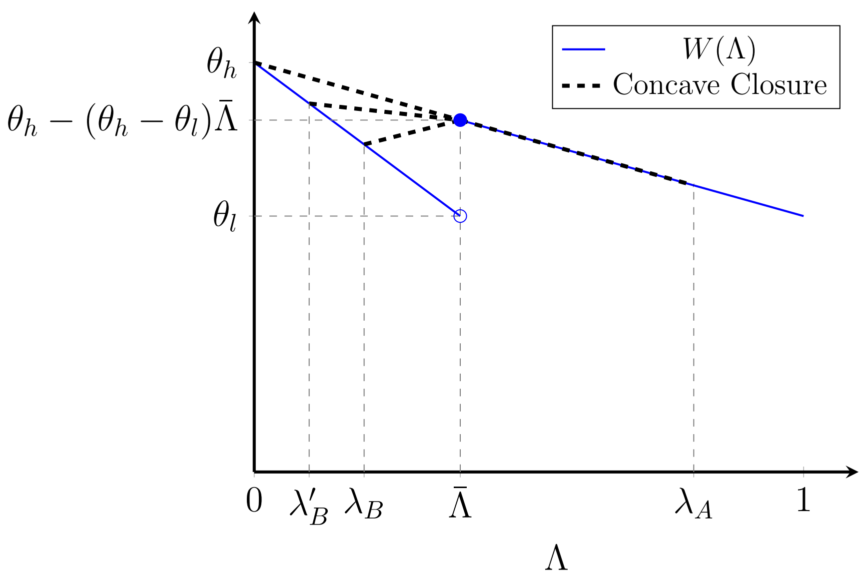


Figure 4: Interim welfare function and its concave closure in the case of only third-degree price discrimination. Dashed lines are concave closures for different class compositions λ_B.

Figure 4 illustrates the concave closure with a single-product seller under third-degree price discrimination. With only third-degree price discrimination, the exclusion threshold is $\bar{\Lambda} := 1 - \theta_l/\theta_h$. Note that in trivial cases, i.e., when $\lambda_B > \bar{\Lambda}$ or $\lambda_A < \bar{\Lambda}$, the concave closure coincides with the interim welfare function $W(\Lambda)$, which is linear. In this case, all information structures yield the same expected consumer and producer surplus.

Given a posterior low–type share Λ, the seller sets the revenue–maximizing price

$$p^*(\Lambda) = \begin{cases} \theta_h, & \Lambda < \bar{\Lambda}, \\ \theta_l, & \Lambda \geq \bar{\Lambda}, \end{cases}$$

yielding producer surplus $\Pi(\Lambda)$ and consumer surplus $U(\Lambda)$:

$$\Pi(\Lambda) = \begin{cases} \theta_h \cdot (1-\Lambda), & \Lambda < \bar{\Lambda}, \\ \theta_l, & \Lambda \geq \bar{\Lambda}, \end{cases} \qquad U(\Lambda) = \begin{cases} 0, & \Lambda < \bar{\Lambda}, \\ (\theta_h - \theta_l) \cdot (1-\Lambda), & \Lambda \geq \bar{\Lambda}. \end{cases} \tag{21}$$

Notice that $\bar{\Lambda}$ is exactly the threshold below which the low-type consumers are excluded. The interim social welfare is

$$\begin{aligned} W(\Lambda) := \Pi(\Lambda) + U(\Lambda) &= \begin{cases} \theta_h \cdot (1-\Lambda), & \Lambda < \bar{\Lambda}, \\ (\theta_h - \theta_l) \cdot (1-\Lambda) + \theta_l, & \Lambda \geq \bar{\Lambda}. \end{cases} \\ &= \begin{cases} -\theta_h \Lambda + \theta_h, & \Lambda < \bar{\Lambda}, \\ -(\theta_h - \theta_l) \cdot \Lambda + \theta_h, & \Lambda \geq \bar{\Lambda}. \end{cases} \end{aligned} \tag{22}$$

Our next result shows that the optimal information structure under only third-degree price discrimination has a similar structure as the one in Proposition 3, demonstrating the robustness of the insights in Section 4.

Proposition 5. *Under only third-degree price discrimination, we have*

1. *if $\lambda_B < \bar{\Lambda}$ and $\lambda_A > \bar{\Lambda}$:*
   (a) *for $\Lambda_0 \in [\lambda_B, \bar{\Lambda})$: an optimal information policy is* $\mathsf{PRP}(\bar{\Lambda})$, *which induces posteriors $\{\lambda_B, \bar{\Lambda}\}$;*
   (b) *for $\Lambda_0 \in [\bar{\Lambda}, \lambda_A]$: full privacy is optimal;*
2. *otherwise, all information structures are the same.*

Proof. This mainly follows from the (22) about the interim welfare function and Figure 4 about the characterization of the concave closure. $\square$

We also fully characterize the achievable region in this single-product setting, where the regulator (sender) can disclose only class identity. The feasible set either (i) degenerates to a single point or (ii) forms a triangle that is strictly contained within the BBM triangle (Bergemann, Brooks, and Morris 2015). Moreover, the frontier has slope everywhere at most 1 (i.e., less than 45°). See Appendix C.1 for details.

5.2 Monopolistic Insurance Markets

We next consider a classic two-type selection market a la J. E. Stiglitz 1977. There is a unit mass continuum of consumers who are considering purchasing an insurance contract. An insurance contract consists of the quantity of coverage $q \in [0, 1]$ of insurance it provides and its premium $t \geq 0$ (with $q = 0$ and $q = 1$ corresponding to no and full insurance, respectively.) Consumers differ only in their privately known type $i \in \{h, l\}$, with a mass $\Lambda_0 \in (0, 1)$ having type l.

The l-types have a lower willingness to pay for insurance coverage, but also impose a lower cost on firms when they buy insurance, when compared with h-types. Specifically, following Spence 1978, consumer preferences are quasilinear in premium: $u_i(q) - t$, and for ease of exposition, we assume $u_i = \theta_i \frac{1}{\gamma} q^\gamma$, with $\gamma < 0.5$. If a firm sells to type i, its profit is $t - \theta_i q$. Note that this preference structure implies a type-independent "first best" insurance quantity $q^{FB} = 1$ (i.e., full insurance). The key difference between this setting and that of Section 3 is the common value feature: firms' costs also depend on the buyer's type.

We first consider the monopolist's optimization problem given the firm's posterior beliefs Λ (i.e., the share of l types). As is standard, the monopolist chooses the expected-profit-maximizing incentive compatible and individually rational menu $\{(q_l, t_l), (q_h, t_h)\}$ of type-specific contracts. The individual rationality constraint for the h type is redundant, as is the incentive compatibility constraint for the l-type. We can thus write the monopoly equilibrium as the solution $(q_l(\Lambda), t_l(\Lambda), q_h^*, t_h(\Lambda))$ to the mathematical program:

$$\Pi(\Lambda) \equiv \max_{q_l, t_l, q_h, t_h} \Lambda(t_l - q_l\theta_l) + (1-\Lambda)(t_h - \theta_h q_h) \tag{23}$$

subject to incentive compatibility of type h

$$\frac{\theta_h}{\gamma} q_h - t_h \geq \frac{\theta_h}{\gamma} q_l^\gamma - t_l \tag{24}$$

and participation of type l

$$\theta_l \frac{1}{\gamma} q_l^\gamma - t_l \geq 0. \tag{25}$$

Standard arguments (J. E. Stiglitz 1977) yield a simple cutoff form for equilibrium. If

$$\Lambda \leq \bar{\Lambda} \equiv \frac{\theta_h - \theta_l}{\theta_h}, \tag{26}$$

then the l type is excluded, and the monopolist sells full insurance $q_h = 1$ to h types at the price θ_i/γ equal to their willingess to pay. For $\Lambda > \bar{\Lambda}$, the monopolist sells to both types. The l types obtain partial coverage at the level $q_l(\Lambda)$ solving

$$\theta_l q_l^{\gamma-1} - (1-\Lambda)\theta_h q_l^{\gamma-1} - \Lambda\theta_l = 0, \tag{27}$$

priced at their willingness to pay $\frac{\theta_l}{\gamma} q_l(\Lambda)^\gamma$. The h types continue to get full insurance, but now earn positive information rents $\frac{(\theta_h - \theta_l)}{\gamma} q_l^\gamma(\Lambda)$.

When (e.g.) $\alpha = \frac{1}{2}$, the social welfare generated per unit of individuals in a segment with firm beliefs Λ is:

$$\begin{aligned} W^M(\Lambda) &= \frac{1}{2} CS(\Lambda) + \frac{1}{2} PS(\Lambda) \qquad (28) \\ &= \frac{1}{2}\left(\Lambda\left(\frac{\theta_l}{\gamma} q_l(\Lambda)^\gamma - \theta_l q_l(\Lambda)\right) + (1-\Lambda)\left(\frac{\theta_h}{\gamma} - \theta_h\right)\right) \end{aligned}$$

if $\Lambda > \bar{\Lambda}$ and $\frac{1}{2}(1-\Lambda)\left(\frac{\theta_h}{\gamma} - \theta_h\right)$ for $\Lambda \leq \bar{\Lambda}$.

Per the following lemma, the properties of the welfare function in the monopoly insurance case perfectly mirror those described in Propositions 1 and 2 in the Mussa-Rosen case.

Lemma 7. *$W^M(\Lambda)$ is affine and strictly decreasing in Λ on $[0, \bar{\Lambda}]$, and is strictly concave on $[\bar{\Lambda}, 1]$.*

Optimal information policy therefore takes the exact same structure as described in Proposition 3. In particular, if $\lambda_B < \bar{\Lambda} < \lambda_A$, then there exists a threshold $\Lambda^\star$ such that optimal policy is partially revealing (splitting to λ_B and $\Lambda^\star$) of $\Lambda_0 < \Lambda^\star$, and calls for full privacy if $\Lambda_0 \geq \Lambda^\star$.

5.3 Competitive Insurance Markets

We now build on the previous section to study competitive insurance markets. The classical framework for this problem is that of M. Rothschild and J. Stiglitz 1976 and Wilson 1977.

In Rothschild–Stiglitz (RS) model, firms simultaneously offer menus of contracts and, by free entry and the absence of cross-subsidization, each individual contract must break even at the clientele it attracts: if type $i \in \{l, h\}$ selects (q_i, t_i), then $t_i = \theta_i q_i$. With single-crossing, the equilibrium (when it exists) is separating and independent of the type distribution Λ: the high-risk type receives full insurance at its actuarially fair premium, $(q_h, t_h) = (1, \theta_h)$, while the low-risk type is offered a distorted (partial-insurance) contract on its own zero-profit locus, $(q_l, t_l) = (q_l^{RS}, \theta_l q_l^{RS})$, where $q_l^{RS} \in (0, 1)$ is uniquely pinned down by the high-type IC condition that deters mimicking of the low-type contract:

$$\frac{\theta_h}{\gamma} - \theta_h \;=\; \frac{\theta_h}{\gamma}\,(q_l^{RS})^\gamma \;-\; \theta_l\, q_l^{RS} \quad\iff\quad \frac{\theta_h}{\gamma}\,(q_l^{RS})^\gamma \;-\; \theta_l\, q_l^{RS} \;-\; \theta_h\Big(1 - \frac{1}{\gamma}\Big) = 0.$$

Thus, RS delivers zero profit on each contract, full insurance for h, a downward distortion of q_l to satisfy h's IC, and an allocation that does not depend on Λ.

However, as is well known, the RS equilibrium does not exist for all values of Λ. A commonly used alternative equilibrium concept, which always exists in such markets for all values of Λ, is the so-called Miyazaki-Wilson-Spence MWS equilibrium (Miyazaki 1977; Wilson 1977; Spence 1978). Intuitively, this equilibrium is the allocation that results from maximizing the surplus of type l subject to firms breaking even (on average across all contracts offered) and the participation constraint of type h. This is the allocation that results from a game where firms are allowed to remove loss-making contracts. [10]

The MWS equilibrium coincides with the RS equilibrium for values of Λ for which the latter exists. For other values of Λ, the MWS equilibrium is given by the solution $\big(q_l^{MWS}(\Lambda), t_l^{MWS}(\Lambda), q_h^{MWS}(\Lambda), t_h^{MWS}(\Lambda)\big)$ to the mathematical program:

$$\max_{q_l, t_l, q_h, t_h} \theta_l \frac{1}{\gamma} q_l^\gamma - t_l \tag{29}$$

[10]More formally, to find a MWS equilibrium, one considers firms that compete by offering menus of contracts and are allowed to *withdraw* any contract that earns negative expected profits given the clientele induced by the prevailing set of offers. A firm may *introduce* a new contract (or menu) while anticipating that, once customers re-sort, all loss-making contracts—its own and rivals'—will subsequently be withdrawn. An equilibrium is a set of menus that is (i) closed under such withdrawals (no offered contract makes losses at the induced take-up), (ii) immune to profitable additions under foresight (no firm can add a contract that remains strictly profitable after the induced withdrawal step), and (iii) breaks even *on average across the contracts a firm offers*, allowing for cross-subsidization within a menu. See (Wilson 1977; Miyazaki 1977; Spence 1978); for game-theoretic foundations and existence under withdrawal dynamics, see, e.g., Mimra and Wambach (2011) and Netzer and Scheuer (2014).

subject to incentive compatibility 24 a break even constraint,

$$(1-\Lambda)\left(\theta_h q_h - t_h\right) + \Lambda\left(\theta_l q_l - t_l\right) \geq 0 \tag{30}$$

and a "outside option" constraint

$$\theta_h \frac{1}{\gamma} q_h^\gamma - t_h \geq \frac{\theta_h}{\gamma} - \theta_h \tag{31}$$

that h types must be at least as well off as they would be in a competitive market if they revealed their type.

Recall the definition of the exclusion threshold $\bar{\Lambda}$ in the context of monopoly insurance, from (26). It is straightforward to show (following Spence 1978) that there exists another cutoff denoted $\bar{\Lambda}^{MWS} > \bar{\Lambda}$ such that the outside option constraint 31 binds for $\Lambda \leq \bar{\Lambda}^{MWS}$. Then, the solution coincides with the Rothschild-Stigliz equilibrium, and is independent of Λ.

For $\Lambda > \bar{\Lambda}^{MWS}$, constraint 31 is slack and can be dropped. In this case, direct comparison reveals that the program defining the MWS equilibrium is mathematically dual to the program defining the monopoly equilibrium. In particular, the MWS competitive program maximizes L-type utility subject to a minimum profit constraint, while the monopoly program maximizes profits subject to a minimum L-type utility constraint. By quasi-linearity, this means that $q_l(\Lambda)$ (and $q_h(\Lambda) = 1$) and $t_h - t_l$ are the same in the monopoly and competitive equilibrium. The solutions differ only in the transfers t_l (and t_h). The premiums are pinned down by the zero-profit condition 30 in the competitive case (in contrast to being pinned down by the participation constraint 25 under monopoly.)

The MWS equilibrium (given beliefs Λ) can be understood as the allocation that obtains when the monopoly profits are fully competed away in the form of uniformly lower premiums. Let $W^{MWS}(\Lambda)$ be the total surplus that obtains in this competitive insurance market, under the MWS equilibrium, when the share of types l is Λ. When $\alpha = \frac{1}{2}$, transfers between producers and consumers do no affect the objective function. Therefore, $W^{MWS}(\Lambda)$ is identically equal to the monopoly's surplus $W^M(\Lambda)$ (from (28)) on $\Lambda \geq \bar{\Lambda}^{MWS}$, and hence also strictly concave there.

Moreover, because the equilibrium allocation (and hence the realized profit on a given individual of a given type) is independent of Λ for $\Lambda \leq \bar{\Lambda}^{MWS}$, $W^{MWS}(\Lambda)$ is again affine and decreasing on $[0, \bar{\Lambda}^{MWS}]$. In this region, equilibrium allocations are as in RS, which does not depend on the type distribution.

Optimal information policy therefore takes the same structure as described in Proposition 3. If $\lambda_B < \bar{\Lambda}^{MWS} < \lambda_A$, then there exists a threshold $\Lambda^\star$ such that optimal policy is partially revealing (splitting to λ_B and $\Lambda^\star$) if $\Lambda_0 < \Lambda^\star$, and calls for full privacy if $\Lambda_0 \geq \Lambda^\star$. The result is summarized in the following lemma.

Lemma 8. *$W^{MWS}(\Lambda)$ is affine and strictly decreasing in Λ on $[0, \bar{\Lambda}^{MWS}]$, and is strictly concave on $[\bar{\Lambda}^{MWS}, 1]$.*

5.4 Tax setting

In the insurance context of the two preceding subsections, we saw that the monopoly problem and the MWS competitive insurance equilibrium problem were mathematically dual to each other. In this section, we consider the dual to our baseline Mussa-Rosen model, and show that it can be interpreted as a model of optimal taxation. As in the insurance case, we focus for simplicity on the $\alpha = \frac{1}{2}$ special case.

In the Mussa-Rosen model, the monopolist maximizes its profits $\Pi = (1-\Lambda)(t_h - c(q_h)) + \Lambda(t_l - c(q_l))$ subject to incentive compatibility, a participation constraint $U(q_l, t_l, \theta_l) \geq 0$, and, implicitly $q_l \geq 0$ (the last constraint being what drives the "kink" in Figure 2).

The dual of this problem would consist of maximizing the utility $U(q_l, t_l, \theta_l) = \theta_l q_l - t_l$ subject to incentive compatibility and a zero profit constraint $\Pi \geq 0$ and $q_l \geq 0$.

By quasi-linearity, the solution to the dual features the same allocations q_l, q_h and the same difference in "prices" $t_h - t_l$ (and hence the same exclusion cutoff $\bar{\Lambda}$). The solution to the dual problem can therefore be found by starting from the solution of the primal problem and then"rebating" profits uniformly to consumers. Since transfers between consumers and producers do not affect the value of the objective function in the $\alpha = \frac{1}{2}$ case, the objective function for the dual is thus identical to the objective function in the primal. Therefore, optimal information design for the dual is the same as information design for the primal.

We now re-interpret variables to see that this dual problem is a reasonably standard Mirrleesian (Mirrlees 1971) optimal tax problem with two types a la J. E. Stiglitz (1982). Let $y = t$ be interpreted as "pre-tax earnings" and let $x = c(q)$ be interpreted as "consumption" or "after tax earnings", so that the total tax paid is $y - x$. Then profits can be re-interpreted as government surplus:

$$\Pi = (1 - \Lambda)(y_h - x_h) + \Lambda(y_l - x_l). \tag{32}$$

Similarly, utility can be written as

$$U_i = \theta_i \psi(x_i) - y_i, \tag{33}$$

where $\psi = c^{-1}$ is a concave function. This utility is cardinally equivalent to

$$U_i = \psi(x_i) - y_i/\theta_i, \tag{34}$$

so if we additionally interpret θ_i as a wage and hence y_i/θ_i as a labor supply, we have a standard two-type optimal tax problem with quasilinear-in-labor preferences as used, e.g., in Lollivier and Rochet (1983), Weymark (1987), and Boadway, Cuff, and Marchand (2000).

How can we interpret information design in a tax setting? The dual Mussa-Rosen problem is the *Rawlsian* optimal tax problem. When $\lambda_B > 0.5$, there is a natural interpretation of the information design problem: it corresponds to choosing *tax jurisdictions*. Once jurisdictions have been defined, policy will be determined by the median voter, which (by $\lambda_B > 0.5$) is the lower-wage worker—hence the Rawlsian tax policy within each jurisdiction. But the jurisdiction designer may care about welfare of both the high and low wage type—for instance, in the $\alpha = 1/2$ total surplus case where the two types are equally weighted.

Because the optimal tax problem is the Mussa-Rosen dual, and because the welfare function W is the same in the dual as the primal, all our results for optimal information design apply here. For instance, optimal design when $\Lambda_0 < \Lambda^\star$ is "partially revealing." It is optimal to create some homogeneous class-B jurisdictions, and some other jurisdiction that blends both classes.

6 Richer class and type spaces

In our main analysis, we made two simplifying assumptions: that the class membership distribution was binary, resulting in two potentially observable consumer classes, A and B, and that there are two underlying fundamental types. This section shows that neither of those simplifying assumptions are critical.

6.1 Binary types and continuous classes

First, nothing in our analysis relies on a binary distribution of class memberships. The main results extend readily to the case in which payoffs are as described in section 3 and types continue to follow a binary distribution, but where the class may be drawn from a richer distribution.

To be specific, let $\theta \in \{\theta_l, \theta_h\}$ and let $\Omega \subset \mathbb{R}$ be a (Borel) set of classes. For each class $\omega \in \Omega$, let $\lambda_\omega \in [0,1]$ denote the share of type-θ_l individuals in class ω. Let $\mu_0 \in \Delta(\Omega)$ be the prior over classes, and assume that the induced distribution of λ_ω under $\omega \sim \mu_0$ is atomless with full support $[\lambda_B, \lambda_A]$, where $0 \leq \lambda_B < \lambda_A \leq 1$.

Proposition 6 (Right-censorship with binary types and continuous classes)**.** *Under the setting of binary types and continuous classes, there exists an optimal information structure that is a* right-censorship policy *in λ_ω: there exists a cutoff $\Lambda^\star \in [\lambda_B, \lambda_A]$ such that*

- *for every class ω with $\lambda_\omega < \Lambda^\star$, the signal (essentially) reveals λ_ω;*
- *for every class ω with $\lambda_\omega \geq \Lambda^\star$, the same pooled signal is sent, inducing a common posterior share of low types*

$$\Lambda^{high} := \mathbb{E}[\lambda_\omega \mid s = \text{“high”}] = \mathbb{E}[\lambda_\omega \mid \lambda_\omega \geq \Lambda^\star].$$

Moreover: If $\lambda_A < \bar{\Lambda}$ or $\lambda_B > \bar{\Lambda}$, then $\Lambda^\star = \lambda_B$ is an optimal policy. Otherwise, the optimal threshold lies in $\left(\lambda_B, \bar{\Lambda}\right)$.

Proof. Given a signal s, the seller's posterior over types depends on s only through the posterior mean share of low types

$$\Lambda(s) := \mathbb{E}[\lambda_\omega \mid s].$$

Hence the continuation game and the regulator's objective depend on the information structure only via the induced distribution of posterior means $\Lambda(s)$, and the problem is a one-dimensional "unknown mean" persuasion problem in the sense of Dworczak and Martini (2019).

By Propositions 1 and 2 and Lemma 3, the interim welfare function $W(\Lambda)$ is S-shaped: affine on $[\lambda_B, \bar{\Lambda}]$ with a convex kink at $\bar{\Lambda}$, and strictly concave on $(\bar{\Lambda}, \lambda_A]$. For such S-shaped objectives that depend only on the posterior mean of a scalar state, it is well known (see, e.g., Kolotilin, Mylovanov, and Zapechelnyuk 2022) that there exists an optimal information structure that is an upper-censorship (right-censoring) policy: there is a cutoff $\Lambda^\star$ such that states below $\Lambda^\star$ are (essentially) fully separated, while all states above $\Lambda^\star$ are pooled into a single posterior mean.

Interpreting states as classes via the scalar state λ_ω yields the right-censorship form in λ_ω stated in the proposition. In the special cases $\lambda_A \leq \bar{\Lambda}$ or $\lambda_B \geq \bar{\Lambda}$, W is affine or concave on the support, so full pooling (the degenerate right-censoring with $\Lambda^\star = \lambda_B$) is optimal. $\square$

Right-censorship policies are the natural extension of the optimal information structures in Section 4 to the case of continuous class membership. When class membership is binary, we showed that the regulator's optimal signal pools some class-B consumers (those with low Λ) with the high-Λ class-A consumers. This pooling spares class-B consumers from a highly distorted contract, but it increases the distortion faced by the pooled class. Due to the S-shape of W, this trade-off is resolved by committing to reveal the class identity of a fraction of class-B consumers.

Likewise, with a continuum, the regulator would like to pool classes with lower Λ together with those with a higher Λ: in particular, this incentive is present on the concave portion of W, at the 'top' of the class distribution. On the other hand, since W is convex at the 'bottom' there is an incentive to reveal those classes with low values of Λ.

6.2 Continuous types and continuous classes

Our main analysis has so far assumed a binary type space. This raises the question of whether our findings are an artefact of this simplification. In general, moving beyond binary types can substantially complicate the regulator's information design problem, because the space of possible type distributions becomes high-dimensional. However, when heterogeneity across observable classes is *one-dimensional*, our approach extends directly.

We now consider a continuous type space together with a continuum of classes. Let θ take values in an interval $[\underline{\theta}, \overline{\theta}]$, and let $\Omega \subseteq [0,1]$ be the set of classes. For each class $\omega \in \Omega$, the distribution of types is a mixture of two benchmark distributions F_A and F_B on $[\underline{\theta}, \overline{\theta}]$:

$$F_\omega(\theta) \;=\; \omega F_A(\theta) + (1-\omega) F_B(\theta), \qquad \omega \in \Omega \subseteq [0,1]. \tag{35}$$

Intuitively, ω is an index of how "high-demand" the class is: higher ω shifts the type distribution toward F_A and away from F_B.[11] Let $\mu_0 \in \Delta(\Omega)$ be the prior over classes.

Given any information structure, a signal s induces a posterior $\mu(\cdot \mid s)$ over Ω. The induced distribution of types faced by the seller after observing s is then

$$F(\theta \mid s) \;=\; \int_\Omega F_\omega(\theta)\, \mu(d\omega \mid s) \;=\; \Big(\mathbb{E}[\omega \mid s]\Big) F_A(\theta) + \Big(1 - \mathbb{E}[\omega \mid s]\Big) F_B(\theta). \tag{36}$$

Thus, the entire type distribution after s is pinned down by the single scalar

$$m(s) := \mathbb{E}[\omega \mid s],$$

the posterior mean of the class index. As in the binary-type case, this one-dimensional structure allows us to reduce the regulator's problem to a moment persuasion problem in the sense of Dworczak and Martini (2019).

[11] In the example in the Appendix we assume F_A first-order stochastically dominates F_B, so higher ω corresponds to stochastically higher demand.

Let $W(F_\mu)$ be the regulator's interim payoff when the seller faces type distribution F_μ and optimally screens consumers. Then there exists a function $\widehat{W} : [0,1] \to \mathbb{R}$ such that

$$W(F_\mu) \;=\; \widehat{W}\big(m(\mu)\big)$$

for every posterior μ. In other words, the continuation game depends on the posterior μ only through its first moment $m(\mu)$.

Proposition 7 (Right-censorship with continuous types and classes)**.** *Under the setting of continuous types and classes, if $\widehat{W}$ is S-shaped on $[0,1]$, then there exists an optimal information structure that is a right-censorship policy in ω: there is a cutoff $\omega^\star \in \Omega$ such that*

- *for every class $\omega < \omega^\star$, the signal (essentially) reveals ω (full separation);*
- *for every class $\omega \geq \omega^\star$, the same pooled signal is sent, inducing a common posterior mean*
$$\mathbb{E}[\omega \mid s = \text{“high”}] \;=\; \mathbb{E}[\omega \mid \omega \geq \omega^\star].$$

Proof. Given any posterior $\mu \in \Delta(\Omega)$, we have by construction

$$F_\mu(\theta) = \int_\Omega F_\omega(\theta)\,\mu(d\omega) = \Big(\mathbb{E}_{\omega\sim\mu}[\omega]\Big)F_A(\theta) + \Big(1 - \mathbb{E}_{\omega\sim\mu}[\omega]\Big)F_B(\theta),$$

so the type distribution F_μ depends on μ only through the scalar $m(\mu) := \mathbb{E}_{\omega\sim\mu}[\omega]$. The seller's optimal menu and profit are determined by F_μ, and hence depend only on $m(\mu)$; likewise the regulator's interim payoff $W(F_\mu)$ depends on μ only via $m(\mu)$. Defining $\widehat{W}(m) := W(F_\mu)$ for any μ with $m(\mu) = m$ is therefore well-defined.

Under an information structure, each signal s induces a posterior $\mu(\cdot \mid s)$, and thus a posterior mean $m(s) := m(\mu(\cdot \mid s)) = \mathbb{E}[\omega \mid s]$. The regulator's expected payoff can be written as

$$\mathbb{E}\big[\,\widehat{W}\big(m(s)\big)\,\big],$$

subject to the Bayes plausibility constraint $\mathbb{E}[m(s)] = \mathbb{E}_{\omega\sim\mu_0}[\omega]$. This is exactly the one-dimensional “unknown mean” persuasion problem studied by Dworczak and Martini (2019).

If $\widehat{W}$ is S-shaped, Kolotilin, Mylovanov, and Zapechelnyuk (2022) show that there exists an optimal information structure that is an upper-censorship (right-censoring) policy: there is a cutoff $\omega^\star$ such that states with $\omega < \omega^\star$ are fully separated, while all states with $\omega \geq \omega^\star$ are pooled into a single posterior mean. Interpreting states as classes ω yields the right-censorship form described in the proposition. □

Proposition 7 shows that the binary-type assumption is not essential for our information design results. What matters is that heterogeneity across class memberships can be summarized by a *single* index ω that linearly mixes two benchmark type distributions F_B and F_A. In such environments the regulator faces exactly the same one-dimensional persuasion problem as in our baseline model, and whenever the resulting interim objective $\widehat{W}$ is S-shaped, the optimal information structure remains a right-censorship policy: low-ω classes are (essentially) revealed, while high-ω classes are pooled.

At the same time, characterizing the shape of $\widehat{W}$ as a general function of (F_B, F_A) is non-trivial. With continuous types, optimal menus may require ironing and other standard complications from the screening literature, and we do not attempt to provide general conditions under which $\widehat{W}$ is S-shaped. Instead, in the Appendix B, we analyze a tractable example in which F_A and F_B are uniform distributions (with F_A first-order stochastically dominating F_B), and show that in this case $\widehat{W}$ is indeed S-shaped. This confirms that our qualitative results are not an artefact of the binary-type environment in Section 3.

7 Boundaries

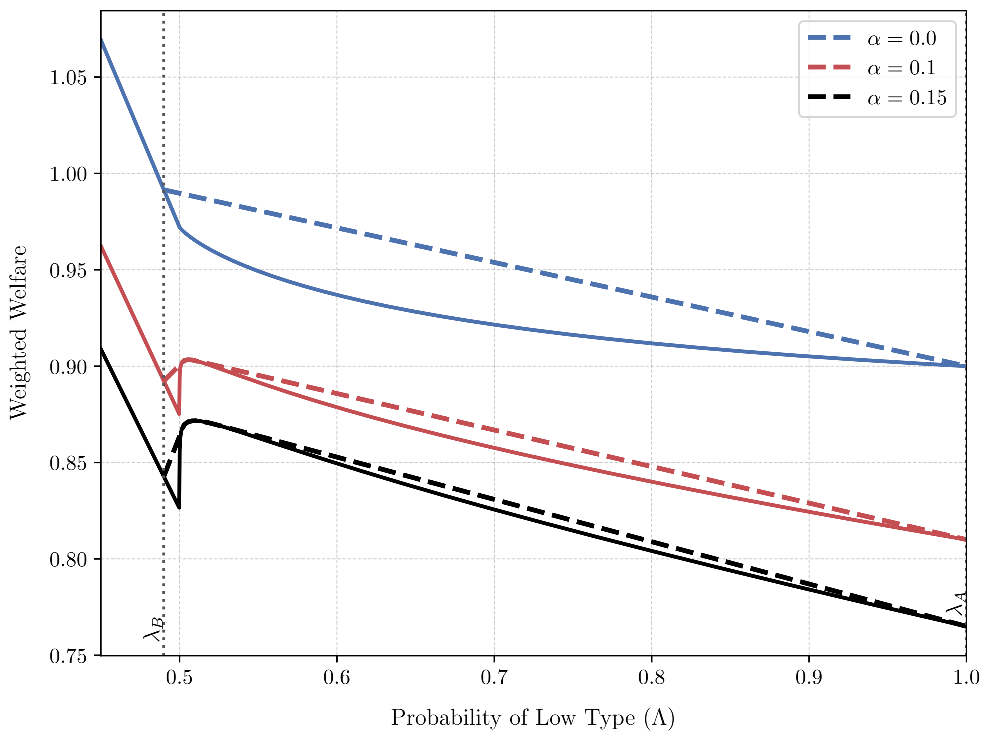


Figure 5: The regulator's objective function as a function of the interim belief Λ, given the cost function $c(q) = \frac{1}{10}q^{10}$ and varying values of $\alpha < \frac{1}{3}$. Dashed lines are concave closures for different α.

We have demonstrated that "probabilistic revelation of the higher value type" is a typical and robust feature of optimal information design. It is not universal, however. When $c''' > 0$ and $\alpha < \frac{1}{3}$, Proposition 2 does not hold in our two-type, two-class Mussa-Rosen baseline. There will typically be convex portions of the $W(\Lambda)$ curve in such cases. Examples are depicted in Figure 5. When $\alpha = 0$, as always, W is globally convex, and full revelation of the class is optimal for any classes (blue curves in Figure 5). Similarly, full revelation is optimal if $1 - \lambda_A$ and λ_B are both sufficiently close to 0. For $\alpha \in \left(0, \frac{1}{3}\right)$, however, W is concave near $\bar{\Lambda}$ (= 0.5 in this example) but switches (once) to convex for sufficiently high Λ (as in the orange and green curves in Figure 5). Then, if $\lambda_B < \bar{\Lambda}$ but sufficiently close to it, optimal information design will be to probabilistically reveal the A-class instead of the B-class whenever Λ_0 is sufficiently high.

Convex portions of W do not always imply a departure from the characterization in Proposition 2. In the monopoly insurance setting, for instance, convexity appears for low α "at the other end" (near $\bar{\Lambda}$) under typical parameterizations. In such cases, the typical case of probabilistic revelation of the higher value class will still obtain.

8 Conclusion

This paper has examined when a regulator who controls coarse class-level information should allow firms to condition on it in classic screening environments, and when it is optimal instead to limit disclosure. We focused on a seller that screens consumers through menus of quality–price contracts, while a planner observes only coarse class membership characteristics, which are associated with different distributions of underlying types. In contrast to standard information design benchmarks with direct signals of types, wherein full revelation maximizes total surplus (so any case for privacy relies on a preference for consumer surplus over producer surplus), we showed that once information is constrained to class memberships, informational frictions alone will frequently make privacy optimal even for a planner who primarily cares about total surplus.

Our main theoretical contribution is to characterize the structure of optimal information policies in this class membership setting. Under mild regularity conditions, the planner's optimal policy belongs to a simple family of partial pooling schemes: reveal class identity for a subset of consumers in the class that is relatively rich in high types, and pool the remaining members of that class with the low-type-intensive class. In a benchmark without screening, imperfectly informative classes both contract and flatten the surplus frontier, so that the policies that maximize consumer surplus and total surplus coincide. With screening, the same geometry that drives Mussa–Rosen-style exclusion generates a welfare function that is locally convex around exclusion thresholds and concave when both types trade. Given that the planner can only choose posteriors that lie between the class-level priors, optimal policies correspond to restricted concavification on this interval, which yields an interior degree of pooling in many cases. We further showed how the optimal amount of pooling depends on the distribution of class memberships and on the planner's weight on consumer surplus, and that partial revelation can improve both total surplus and consumer surplus relative to full disclosure, in sharp contrast to the standard price discrimination setting without classes.

We also demonstrated that the same mechanism operates across a range of canonical asymmetric information environments. In applications to monopolistic and competitive insurance, and optimal taxation, local convexity around exclusion-like thresholds and local concavity above those thresholds and thus imply a similar structure of privacy-protecting pooling policies. We showed how our analysis extends beyond binary types and categories to continuous type and class spaces, and how tools from constrained persuasion can be used to describe optimal policies in these richer settings. Therefore our general framework can help inform policy on insurance discrimination bans, occupation-based pricing, tax jurisdiction design, and digital platform data policies.

Several promising avenues for future research emerge from our analysis. First, extending the framework to dynamic settings where firms learn about class compositions over time could yield insights into the evolution of privacy regulations and market segmentation strategies. Second, incorporating strategic consumer responses, such as signaling or obfuscation of class membership, would enrich our understanding of equilibrium outcomes under different information regimes. Third, empirical work testing our theoretical predictions about the welfare effects of class-based information revelation could inform practical regulatory design, particularly in digital markets where class characteristics are increasingly observable. Finally, exploring the interaction between our information design approach and other regulatory instruments, such as direct restrictions on price discrimination or data protection laws, could provide a more comprehensive framework for understanding optimal regulation in markets with asymmetric information.

References

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris (2015). "The limits of price discrimination". In: *American Economic Review* 105.3, pp. 921–957.

Bergemann, Dirk, Tibor Heumann, and Michael C Wang (2024). "A unified approach to second and third degree price discrimination". In: *Cowles Foundation Discussion Paper No. 2420*.

Bergemann, Dirk and Stephen Morris (2016). "Information design, Bayesian persuasion, and Bayes correlated equilibrium". In: *American Economic Review* 106.5, pp. 586–591.

Boadway, Robin, Katherine Cuff, and Maurice Marchand (2000). "Optimal income taxation with quasi-linear preferences revisited". In: *Journal of Public Economic Theory* 2.4, pp. 435–460.

Competition and Markets Authority (2025). *CMA Annual Plan 2025 to 2026*. https://www.gov.uk/government/publications/cma-annual-plan-2025-to-2026/annual-plan-2025-to-2026. Accessed: 2025-10-10.

Dosis, Anastasios and Casey Rothschild (2024). "Screening and the Welfare Tradeoffs of Privacy". In: *SSRN 4706308*.

Doval, Laura and Vasiliki Skreta (2024). "Constrained information design". In: *Mathematics of Operations Research* 49.1, pp. 78–106.

Dworczak, Piotr and Giorgio Martini (2019). "The simple economics of optimal persuasion". In: *Journal of Political Economy* 127.5, pp. 1993–2048.

Einav, Liran and Amy Finkelstein (2011). "Selection in insurance markets: Theory and empirics in pictures". In: *Journal of Economic perspectives* 25.1, pp. 115–138.

Fainmesser, Itay P, Andrea Galeotti, and Ruslan Momot (2023). "Consumer profiling via information design". In: *Available at SSRN 4655468*.

Fallah, Alireza et al. (2024). "The Limits of Price Discrimination Under Privacy Constraints". In: *arXiv preprint arXiv:2402.08223*.

Farboodi, Maryam, Nima Haghpanah, and Ali Shourideh (2025). "Good Data and Bad Data: The Welfare Effects of Price Discrimination". In.

Federal Trade Commission (2025). *Mission of the Federal Trade Commission*. https://www.ftc.gov/about-ftc/mission. Accessed: 2025-10-10.

Galperti, Simone, Aleksandr Levkun, and Jacopo Perego (2024). "The value of data records". In: *Review of Economic Studies* 91.2, pp. 1007–1038.

Haghpanah, Nima and Ron Siegel (2022). "The limits of multiproduct price discrimination". In: *American Economic Review: Insights* 4.4, pp. 443–458.

— (2023). "Pareto-Improving Segmentation of Multiproduct Markets". In: *Journal of Political Economy* 131.6, pp. 1385–1617.

Kamenica, Emir and Matthew Gentzkow (2011). "Bayesian persuasion". In: *American Economic Review* 101.6, pp. 2590–2615.

Kolotilin, Anton, Timofiy Mylovanov, and Andriy Zapechelnyuk (2022). "Censorship as optimal persuasion". In: *Theoretical Economics* 17.2, pp. 561–585.

Le Treust, Maël and Tristan Tomala (2019). "Persuasion with limited communication capacity". In: *Journal of Economic Theory* 184, p. 104940.

Lipsey, Richard G and Kelvin Lancaster (1956). "The general theory of second best". In: *The review of economic studies* 24.1, pp. 11–32.

Lollivier, Stefan and Jean-Charles Rochet (1983). "Bunching and second-order conditions: A note on optimal tax theory". In: *Journal of Economic Theory* 31.2, pp. 392–400. ISSN: 0022-0531. DOI: https://doi.org/10.1016/0022-0531(83)90086-8. URL: https://www.sciencedirect.com/science/article/pii/0022053183900868.

Mimra, Wanda and Achim Wambach (2011). *A game-theoretic foundation for the Wilson equilibrium in competitive insurance markets with adverse selection*. Tech. rep. CESifo working paper.

Mirrlees, James A (1971). "An exploration in the theory of optimum income taxation". In: *The review of economic studies* 38.2, pp. 175–208.

Miyazaki, Hajime (1977). "The rat race and internal labor markets". In: *The Bell Journal of Economics* 8.2, pp. 394–418.

Mussa, Michael and Sherwin Rosen (1978). "Monopoly and product quality". In: *Journal of Economic theory* 18.2, pp. 301–317.

Netzer, Nick and Florian Scheuer (2014). "A game theoretic foundation of competitive equilibria with adverse selection". In: *International Economic Review* 55.2, pp. 399–422.

Rothschild, Michael and Joseph Stiglitz (1976). "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information". In: *The Quarterly Journal of Economics* 90.4, pp. 629–649.

Salanié, Bernard (2005). *The economics of contracts: a primer*. MIT press.

Spence, Michael (1978). "Product differentiation and performance in insurance markets". In: *Journal of Public Economics* 10.3, pp. 427–447.

Stiglitz, Joseph E (1977). "Monopoly, non-linear pricing and imperfect information: the insurance market". In: *The Review of Economic Studies* 44.3, pp. 407–430.

— (1982). "Self-selection and Pareto efficient taxation". In: *Journal of public economics* 17.2, pp. 213–240.

Weymark, John A (1987). "Comparative static properties of optimal nonlinear income taxes". In: *Econometrica: Journal of the Econometric Society*, pp. 1165–1185.

Wilson, Charles (1977). "A model of insurance markets with incomplete information". In: *Journal of Economic Theory* 16.2, pp. 167–207. ISSN: 10957235. DOI: 10.1016/0022-0531(77)90004-7.

APPENDIX

A 2x2 Model

A.1 Proof: Proposition 2

Proof. We first prove that $W(\Lambda)$ is concave on $\Lambda \in [\bar{\Lambda}, 1]$.

Throughout, write $\Delta\theta \equiv \theta_h - \theta_l > 0$, $q_l(\Lambda) \equiv q_l^*(\Lambda)$, and $q_h \equiv q_h^*$ for brevity.

We begin by decomposing $W(\Lambda)$. Recall that the second-best prices implied by the binding constraints are

$$t_l = \theta_l q_l, \qquad t_h = \theta_h q_h - \Delta\theta\, q_l.$$

Moreover, utilities are $U_l = \theta_l q_l - t_l = 0$;

$$U_h = \theta_h q_h - t_h = \theta_h q_h - (\theta_h q_h - \Delta\theta q_l) = \Delta\theta\, q_l.$$

Profits are $\pi_l = t_l - c(q_l) = \theta_l q_l - c(q_l)$; $\pi_h = t_h - c(q_h) = \theta_h q_h - \Delta\theta q_l - c(q_h)$.

Finally, the Planner's per-type payoff is $v(q, t, \theta) = \alpha U + (1-\alpha)\pi$, which implies

$$v_l = \alpha \cdot 0 + (1-\alpha)\big(\theta_l q_l - c(q_l)\big), \qquad v_h = \alpha \cdot (\Delta\theta\, q_l) + (1-\alpha)\big(\theta_h q_h - \Delta\theta q_l - c(q_h)\big).$$

$$\begin{aligned}
W(\Lambda) &= \Lambda\, v_l + (1-\Lambda)\, v_h \\
&= \Lambda\big[\alpha\, U_l + (1-\alpha)\, \pi_l\big] + (1-\Lambda)\big[\alpha\, U_h + (1-\alpha)\, \pi_h\big] \\
&= \Lambda\big[\alpha \cdot 0 + (1-\alpha)\big(\theta_l q_l - c(q_l)\big)\big] + (1-\Lambda)\big[\alpha(\Delta\theta\, q_l) + (1-\alpha)\big(\theta_h q_h - \Delta\theta\, q_l - c(q_h)\big)\big] \\
&= \alpha(1-\Lambda)\Delta\theta\, q_l + (1-\alpha)\Big\{\Lambda\big(\theta_l q_l - c(q_l)\big) + (1-\Lambda)\big(\theta_h q_h - \Delta\theta\, q_l - c(q_h)\big)\Big\}.
\end{aligned}$$

First, we derive q_l', q_l''. In this regime ($\Lambda > \bar{\Lambda}$), the interior FOC holds, i.e.

$$c'\big(q_l(\Lambda)\big) = \theta_l - \frac{1-\Lambda}{\Lambda}\,\Delta\theta = \theta_h - \frac{\Delta\theta}{\Lambda} \tag{37}$$

Implicit differentiation of (37) gives

$$c''\big(q_l(\Lambda)\big)\, q_l'(\Lambda) = \frac{\Delta\theta}{\Lambda^2} \quad \Rightarrow \quad q_l'(\Lambda) = \frac{\Delta\theta}{\Lambda^2\, c''\big(q_l(\Lambda)\big)} \;>\; 0.$$

Differentiating (37) again yields

$$c''\big(q_l\big)\, q_l''(\Lambda) + c'''\big(q_l\big)\, \big(q_l'(\Lambda)\big)^2 = -\frac{2\Delta\theta}{\Lambda^3}.$$

Hence

$$q_l''(\Lambda) = -\frac{2\Delta\theta}{\Lambda^3\, c''(q_l)} - \frac{c'''(q_l)}{c''(q_l)}\big(q_l'(\Lambda)\big)^2.$$

Since $c''' \geq 0$, the last term in the expression for $q_l''(\Lambda)$ is negative, so we can derive the following bound:

$$q_l''(\Lambda) \;\leq\; -\frac{2\Delta\theta}{\Lambda^3\, c''(q_l)}.$$

Now, using $q_l' = \Delta\theta/(\Lambda^2 c''(q_l))$, we can re-write this bound as

$$q_l''(\Lambda) \;\leq\; -\frac{2}{\Lambda}\, q_l'(\Lambda). \tag{38}$$

Now we derive $W''(\Lambda)$. To ease exposition, we will split this into consumer surplus (CS) and producer's surplus (PS). That is $W(\Lambda) = \alpha CS(\Lambda) + (1-\alpha)PS(\Lambda)$.

Since types θ_l always obtain a surplus of zero, consumer's surplus is

$$CS(\Lambda) = (1-\Lambda)\Delta\theta\, q_l(\Lambda).$$

Differentiating yields

$$CS'(\Lambda) = -\Delta\theta\, q_l + (1-\Lambda)\Delta\theta\, q_l'.$$

Differentiating again yields

$$CS''(\Lambda) = -2\Delta\theta\, q_l'(\Lambda) + (1-\Lambda)\Delta\theta\, q_l''(\Lambda). \tag{39}$$

Now we turn to Producer surplus, which is

$$PS(\Lambda) = \Lambda\big(\theta_l q_l - c(q_l)\big) + (1-\Lambda)\big(\theta_h q_h - \Delta\theta\, q_l - c(q_h)\big).$$

Producer's surplus is maximized with respect to q_h, q_l, so we can use the Envelope Theorem when differentiating it with respect to Λ. We obtain

$$PS'(\Lambda) = \big(\theta_l q_l - c(q_l) + \Delta\theta\, q_l\big) - \big(\theta_h q_h - c(q_h)\big).$$

where we have isolated the terms in q_h since these don't vary with Λ. Differentiating producer's surplus again yields

$$PS''(\Lambda) = \big(\theta_l - c'(q_l) + \Delta\theta\big)\, q_l'(\Lambda).$$

From (37), we obtained $\theta_l - c'(q_l) = \frac{1-\Lambda}{\Lambda}\Delta\theta$. Therefore,

$$\mathrm{PS}''(\Lambda) = \left(\frac{1-\Lambda}{\Lambda}\Delta\theta + \Delta\theta\right) q_l'(\Lambda) = \frac{\Delta\theta}{\Lambda}\, q_l'(\Lambda) \;>\; 0. \tag{40}$$

Now we combine these results to obtain $W''(\Lambda)$. Recall that $W = \alpha CS + (1-\alpha)PS$, therefore

$$W''(\Lambda) = \alpha\, \mathrm{CS}''(\Lambda) + (1-\alpha)\, \mathrm{PS}''(\Lambda).$$

Substituting in (39) and (40)) yields

$$W''(\Lambda) = \alpha\Big(-2\Delta\theta\, q_l' + (1-\Lambda)\Delta\theta\, q_l''\Big) + (1-\alpha)\Big(\frac{\Delta\theta}{\Lambda}\, q_l'\Big).$$

Factoring out the common terms yields

$$W''(\Lambda) = \Delta\theta\, q_l'(\Lambda)\left\{\frac{1-\alpha}{\Lambda} - 2\alpha + \alpha(1-\Lambda)\frac{q_l''(\Lambda)}{q_l'(\Lambda)}\right\}.$$

Now we apply the bound from (38)). Since $\frac{q_l''}{q_l'} \leq -\frac{2}{\Lambda}$, we obtain

$$W''(\Lambda) \leq \Delta\theta\, q_l'(\Lambda)\left\{\frac{1-\alpha}{\Lambda} - 2\alpha - \alpha(1-\Lambda)\frac{2}{\Lambda}\right\} = \Delta\theta\, q_l'(\Lambda)\,\frac{1-3\alpha}{\Lambda}.$$

Recall that $q_l' > 0$. Therefore, if $\alpha \geq 1/3$ the right-hand side is ≤ 0, with strict inequality when $\alpha > 1/3$ or when the inequality in (38) is strict somewhere (i.e., $c''' > 0$).

□

A.2 Proof: Lemma 3

Proof. First, consider PS^{seg} and observe that it is convex in Λ. As a result, it satisfies

$$\left(PS^{seg}\right)'(\Lambda_-) \leq \left(PS^{seg}\right)'(\Lambda_+)$$

for all Λ, including $\bar{\Lambda}$.

Next, consider $CS^{seg} = (1-\Lambda)\,(\theta_h - \theta_l)\, q_l^\star(\Lambda)$. Recall that to the left of $\bar{\Lambda}$, $q_l^\star(\Lambda) = 0$ and hence $CS^{seg}(\Lambda) = 0$. Therefore, $\left(CS^{seg}\right)'\left(\bar{\Lambda}_-\right) = 0$. On the other hand, direct calculation shows that the right-derivative is

$$\left(CS^{seg}\right)'\left(\bar{\Lambda}\right) = \left(1-\bar{\Lambda}\right)(\theta_h - \theta_l)\, q_l'\left(\bar{\Lambda}\right),$$

where from (4),

$$q_l'(\bar{\Lambda}) = \frac{(\theta_h - \theta_l)}{\bar{\Lambda}^2 c''(\bar{\Lambda})}$$

which is strictly positive under Assumption 1.

Finally under Assumption 2, $\alpha > 0$ and therefore

$$\begin{aligned} W'(\bar{\Lambda}_-) &= \alpha\left(CS^{seg}\right)'(\bar{\Lambda}_-) + (1-\alpha)\left(PS^{seg}\right)'(\bar{\Lambda}_-) \\ &< \alpha\left(CS^{seg}\right)'(\bar{\Lambda}_+) + (1-\alpha)\left(PS^{seg}\right)'(\bar{\Lambda}_+) = W'\left(\bar{\Lambda}_+\right). \end{aligned}$$

□

A.3 Proof: Proposition 3 (Optimal Information Design)

Proof. Here we derive the optimal information structure for the model with 2 types and 2 classes, assuming $\alpha > 1/3$.

The information design problem is to choose a distribution σ over $\Lambda \in [\lambda_B, \lambda_A]$ with mean Λ_0 to maximize $\mathbb{E}_\sigma[W(\Lambda)]$. Let $\widehat{W}(\cdot)$ be the concave envelope of $W(\cdot)$. The value to the regulator of this optimal information structure is the concave envelope evaluted at the prior: $\widehat{W}(\Lambda_0)$.

If $\lambda_A \leq \bar{\Lambda}$, by Proposition 1, W is affine on $[\lambda_B, \lambda_A]$. Therefore $\widehat{W} = W$ and any Bayes-plausible σ attains $W(\Lambda_0)$.

If $\lambda_B \geq \bar{\Lambda}$, by Proposition 2, W is (weakly) concave on $[\lambda_B, \lambda_A]$. Jensen's inequality gives $\mathbb{E}_\sigma[W(\Lambda)] \leq W(\mathbb{E}_\sigma[\Lambda]) = W(\Lambda_0)$, with equality for the degenerate σ at Λ_0 (full privacy), which is the optimal information policy.

We then turn to the third case: $\lambda_B \leq \bar{\Lambda} < \lambda_A$. Here W is affine on $[\lambda_B, \bar{\Lambda}]$ and concave on $(\bar{\Lambda}, \lambda_A]$.

We first describe the geometry of the concave envelope $\widehat{W}$. For any $y \in (\bar{\Lambda}, \lambda_A]$, let

$$L(\Lambda; y) \;=\; W(\lambda_B) + \frac{W(y) - W(\lambda_B)}{y - \lambda_B}\,(\Lambda - \lambda_B)$$

be the line through the points $(\lambda_B, W(\lambda_B))$ and $(y, W(y))$. Define the set

$$S \;=\; \Big\{ y \in [\bar{\Lambda}, \lambda_A] \;:\; L(\Lambda; y) \leq W(\Lambda) \text{ for all } \Lambda \in [\lambda_B, y] \Big\}.$$

Because W is affine on $[\lambda_B, \bar{\Lambda}]$ and concave on $[\bar{\Lambda}, \lambda_A]$, the set S is nonempty (it contains $\bar{\Lambda}$), closed, and connected. Let $\hat{\Lambda} = \sup S$. Two possibilities arise:

(a) $\hat{\Lambda} = \lambda_A$. Then $L(\Lambda; \lambda_A) \leq W(\Lambda)$ for all $\Lambda \in [\lambda_B, \lambda_A]$, so $\widehat{W} = W$ everywhere on $[\lambda_B, \lambda_A]$.

(b) $\hat{\Lambda} \in (\bar{\Lambda}, \lambda_A)$. By maximality, $L(\Lambda; \hat{\Lambda}) \leq W(\Lambda)$ on $[\lambda_B, \hat{\Lambda}]$ and $L(\Lambda; \hat{\Lambda}) = W(\Lambda)$ at $\Lambda = \lambda_B$ and $\Lambda = \hat{\Lambda}$; moreover, for any $\varepsilon > 0$, the line $L(\cdot; \hat{\Lambda} + \varepsilon)$ lies *strictly above* W at some point in $[\lambda_B, \hat{\Lambda}]$. Therefore $\widehat{W}$ coincides with the chord $L(\cdot; \hat{\Lambda})$ on $[\lambda_B, \hat{\Lambda}]$ and with W on $[\hat{\Lambda}, \lambda_A]$.

In words: the envelope is the upper hull that connects $(\lambda_B, W(\lambda_B))$ to a unique *tangency point* $(\hat{\Lambda}, W(\hat{\Lambda}))$ and then follows W thereafter. Under strict concavity on $(\bar{\Lambda}, \lambda_A]$ (which holds if $\alpha > 1/3$ or $c''' > 0$), the tangency point $\hat{\Lambda}$ is unique; in the knife-edge weakly concave case, one may have $\hat{\Lambda} = \lambda_B$ (which collapses back to (a)).

We have defined the concave envelope. We now use it to derive the optimal information regime.

If $\Lambda_0 \geq \hat{\Lambda}$, then $\widehat{W}(\Lambda_0) = W(\Lambda_0)$ and the degenerate posterior at Λ_0 (full privacy) is optimal.

If $\Lambda_0 < \hat{\Lambda}$, then $\widehat{W}(\Lambda_0)$ lies on the chord $L(\cdot; \hat{\Lambda})$. The unique Bayes-plausible two-point distribution that puts mass on $\{\lambda_B, \hat{\Lambda}\}$ with mean Λ_0 is given by (12). Under $\mathsf{PRP}(\hat{\Lambda})$, the posterior support is exactly $\{\lambda_B, \hat{\Lambda}\}$ with those weights (use (11) and Bayes' rule), so $\mathsf{PRP}(\hat{\Lambda})$ implements the envelope and is optimal.

Uniqueness of $\hat{\Lambda}$ in the strict-concavity case follows from the fact that a strictly concave function admits at most one point where the chord from $(\lambda_B, W(\lambda_B))$ is tangent.

□

A.4 Proof: Proposition 4

Proof. By standard information design logic, Z is a convex set, and the graph of Z^* is concave; so $(CS^*, PS^*) \in Z^*$ if and only if there exists an α for which $(CS^*, PS^*) \in (CS(\alpha), PS(\alpha))$, with $CS(\alpha)$ non-increasing and $PS(\alpha)$ non-decreasing in α.

Since $\Lambda^\star\left(\frac{1}{3}\right) > \lambda_A$, full revelation is optimal in the neighborhood of $\alpha = \frac{1}{3}$. Full revelation is always optimal when $\alpha = 0$. It follows from the monotonicity of $(CS(\alpha), PS(\alpha))$ that $(CS(\alpha), PS(\alpha)) = \left(CS\left(\frac{1}{3}\right), PS(\frac{1}{3})\right)$ for all $\alpha \in [1, \tilde{\alpha}]$ for some $\tilde{\alpha} > \frac{1}{3}$. Define $\underline{\alpha}$ as the maximal such $\tilde{\alpha}$ (which, by Lemma 6, is well defined). Then

$$Z^* = \{(CS(\alpha), PS(\alpha))\}_{\alpha \in [\underline{\alpha}, 1]}.$$

If $\lambda_A \leq \Lambda^\star(1)$, then full revelation is also optimal at $\alpha = 1$, and hence for all $\alpha \in [0, 1]$. Then we can take $\overline{\alpha} = \underline{\alpha}$ and the Pareto frontier is the single point corresponding to full information revelation.

If, on the other hand, $\lambda_A > \Lambda^\star(1)$, then $(CS(1), PS(1)) \neq \left(CS\left(\frac{1}{3}\right), PS(\frac{1}{3})\right)$, and Z^* is not a single point. In this case, define $\overline{\alpha}$ as the maximal α for which $\Lambda^\star(\alpha) \geq \Lambda_0$. By Lemma 6, $\overline{\alpha}$ is well-defined, and $\overline{\alpha} < 1$ iff $\Lambda_0 > \Lambda^\star(1)$. Then the optimal information policy is partially revealing—splitting to λ_B and $\Lambda^\star(\alpha)$—for all $[\underline{\alpha}, \overline{\alpha})$. By Lemma 6, the associated welfare $(CS(\alpha), PS(\alpha))$ is differentiable in α, and $\Lambda^\star(\alpha)$ is strictly decreasing on this range. An increase in α therefore results in mean-preserving contraction of the distribution of firm beliefs and thereby a strict reduction in PS. The mapping $\alpha \to (CS(\alpha), PS(\alpha))$ is therefore bijective on $[\underline{\alpha}, \overline{\alpha}]$. □

A.5 Proof: Comparative Statics

A.5.1 Proof of Lemma 4

Proof. It is easy to see that the objective function described in (13) is continuous and the constraint set compact, so that a maximum exists. It is convenient to begin by showing that this maximum is attained on $(\bar{\Lambda}, 1]$. Observe that, since q_l and hence consumer surplus

$$CS(\Lambda) = (1 - \Lambda) \cdot (\theta_h - \theta_l) \cdot q_l \tag{41}$$

is 0 for $\Lambda \leq \bar{\Lambda}$, we have

$$W(\Lambda) = (1 - \alpha) PS(\Lambda) = (1 - \alpha)(1 - \Lambda)(\theta_h q_h^* - c(q_h^*)),$$

where $PS(\Lambda)$ denotes the monopolist's profit. But since

$$PS(\Lambda) = \max_{q_l \leq q_h^*} \Lambda(\theta_l q_l - c(q_l)) + (1 - \Lambda)(\theta_h q_h - c(q_h) - (\theta_h - \theta_l) q_l) \tag{42}$$

is the upper envelope of functions linear in Λ, PS is convex. Moreover CS is bounded below by 0 and so for any $\Lambda \geq \bar{\Lambda}$ we have

$$\begin{aligned} W(\Lambda) &= \alpha CS(\Lambda) + (1 - \alpha) PS(\Lambda) \\ &\geq W(\lambda_B) + (1 - \alpha)(\Lambda - \bar{\Lambda})(\theta_h q_h - c(q_h)), \end{aligned}$$

where the inequality follows from $CS \geq 0$, the convexity of PS, and $\frac{dPS}{d\Lambda} = (\theta_h q_h - c(q_h))$ for $\Lambda < \bar{\Lambda}$. For $\Lambda \in (\bar{\Lambda}, 1]$, this inequality is strict, so that on rearranging, we have

$$\frac{W(\Lambda) - W(\lambda_B)}{\Lambda - \lambda_B} > \theta_h q_h - c(q_h) = W'(\lambda_B). \tag{43}$$

This shows that the max is attained on $\left(\bar{\Lambda}, 1\right]$.

We now prove that (15) is quasiconcave. To do so, it is sufficient to prove that (15) is single-peaked. Recall from Proposition 3 that W is strictly concave on $\left[\bar{\Lambda}, 1\right]$. Towards a contradiction, suppose that there are two distinct solutions $\Lambda^\star < \Lambda^{**}$ in the set of maximizers (13). Rearranging, both of these solutions must satisfy

$$W(\Lambda^\star) = W(\lambda_B) + w^\star(\Lambda^\star - \lambda_B)$$
$$W(\Lambda^{\star\star}) = W(\lambda_B) + w^\star(\Lambda^{\star\star} - \lambda_B),$$

where $w^\star$ is the maximal chord slope defined following Proposition 3. On the other hand, any other $\Lambda \in [\lambda_B, 1]$ must satisfy

$$W(\Lambda) \leq W(\lambda_B) + w^\star(\Lambda - \lambda_B)$$

That is, the points $(\Lambda^\star, W(\Lambda^\star)$ and $(\Lambda^{\star\star}, W(\Lambda^{\star\star}))$ lie on the line $W(\lambda_B) + w^\star(\Lambda - \lambda_B)$, while every other point $(\Lambda, W(\Lambda))$ for $\Lambda \in [\Lambda^\star, \Lambda^{\star\star}]$ lies weakly below the line - a contradiction to the strict concavity of W.

Since W is quasiconcave and takes its solution on $\Lambda \in \left(\bar{\Lambda}, 1\right]$ where (15) is easily verified to be differentiable, the standard first order and complementary slackness conditions are necessary and sufficient for identifying the maximizer $\Lambda^\star$. Noting that we have already proven that the constraint $\Lambda^\star \geq \bar{\Lambda}$ is slack, routine differentiation of (15) shows that these conditions are precisely (16)-(18).

Finally, we show that the second-order condition is strictly negative when (16) holds with equality. After a routine differentiation, the left side of (16) becomes

$$W''(\Lambda) - \frac{1}{\Lambda - \lambda_B}\left(W'(\Lambda) - \frac{W(\Lambda) - W(\lambda_B)}{\Lambda - \lambda_B}\right).$$

By hypothesis, the second term in this expression is 0 at $\Lambda = \Lambda^\star$. Moreover, the first is strictly negative, since W is strictly concave on $\left(\bar{\Lambda}, 1\right]$ and $\Lambda^\star > \bar{\Lambda}$. □

A.5.2 Proof of Lemma 5

Proof. We first prove existence of the threshold λ_B'. To do so, we show that if $\Lambda^\star$ satisfies (16) with equality for some λ_B, then the solution continues to do so for all $\tilde{\lambda}_B > \lambda_B$. A simple closedness argument (left to the reader) will then establish that $\bar{\Lambda} = \inf\{\lambda_B : \Lambda^\star(\lambda_B)$ solves (16) with equality$\}$.

Suppose $\Lambda^\star(\lambda_B)$ satisfies (16) with equality.

$$W'\left(\Lambda^\star(\lambda_B)\right) = \frac{W(\Lambda^\star(\lambda_B) - W(\lambda_B))}{\Lambda^\star(\lambda_B) - \lambda_B}.$$

By strict quasiconcavity of (15), this implies

$$W'(1) < \frac{W(1) - W(\lambda_B)}{1 - \lambda_B}.$$

Now, differentiating the right side of this inequality with respect to λ_B gives

$$\frac{W(1) - W(\lambda_B) - (1 - \lambda_B)W'(\lambda_B)}{(1 - \lambda_B)^2} \tag{44}$$

which is positive by (43). Hence, for any $\tilde{\lambda}_B > \lambda_B$, $\Lambda^\star(\tilde{\lambda}_B)$ cannot satisfy (16) as an inequality, proving the claim.

Next, we show that $\lambda'_B < \bar{\Lambda}$. To see this, observe that by strict concavity of W on $\left[\bar{\Lambda}, 1\right]$ we have

$$W'(1) < \frac{W(1) - W(\bar{\Lambda})}{1 - \bar{\Lambda}}$$

By continuity, there exists an $\varepsilon > 0$ such that

$$W'(1) < \frac{W(1) - W(\lambda_B)}{1 - \lambda_B}$$

for all $\lambda_B \in \left[\bar{\Lambda} - \varepsilon, \bar{\Lambda}\right]$, and hence, the unique maximizer $\Lambda^\star$ in (13) satisfies (16) with equality for all such λ_B.

To show that $\Lambda^\star$ is decreasing on $\left[\lambda'_B, \bar{\Lambda}\right]$, suppose $\lambda'_B < \bar{\Lambda}$ and fix $\lambda_B \in \left[\lambda'_B, \bar{\Lambda}\right]$. $\Lambda^\star$ solves

$$W'\left(\Lambda^\star(\lambda_B)\right) = \frac{W(\Lambda^\star(\lambda_B) - W(\lambda_B))}{\Lambda^\star(\lambda_B) - \lambda_B}).$$

Differentiating this condition partially with respect to λ_B, the right side changes by

$$-\frac{1}{(\Lambda^\star - \lambda_B)^2}\left(W'(\lambda_B - \frac{W(\Lambda^\star) - W(\lambda_B)}{\Lambda^\star - \lambda_B})\right) > 0,$$

which follows again from (43). Thus, for any $\tilde{\lambda}_B \in (\lambda_B, \bar{\Lambda})$, we have

$$W'\left(\Lambda^\star(\lambda_B)\right) < \frac{W(\Lambda^\star(\lambda_B) - W(\tilde{\lambda}_B))}{\Lambda^\star(\lambda_B) - \tilde{\lambda}_B})$$

and hence by quasi-concavity, the solution must satisfy $\Lambda^\star(\tilde{\lambda}_B) < \Lambda^\star(\lambda_B)$. To see that $\Lambda^\star$ is differentiable on $\lambda_B > \lambda'_B$, note that the first-order condition (16) holds with equality on $\lambda_B > \lambda'_B$ and moreover by Lemma 4, the second-order condition is strictly negative. Hence, the Implicit Function Theorem applies, so that $\Lambda^\star(\alpha)$ is differentiable.

Finally, we show that $\lim_{\lambda_B \to \bar{\Lambda}} \Lambda^\star = \bar{\Lambda}$. For the sake of a contradiction, suppose not. Then, since $\Lambda^\star$ is decreasing in λ_B it must converge to some limit $\underline{\Lambda} > \bar{\Lambda}$. Noting that W' and W are continuous, we may pass limits through (16) (which holds with equality for all λ_B sufficiently close to $\bar{\Lambda}$) to conclude

$$W'(\underline{\Lambda}) = \frac{W(\underline{\Lambda}) - W(\bar{\Lambda})}{\underline{\Lambda} - \bar{\Lambda}} := \underline{w}.$$

By strict concavity of W, the right side of this expression is strictly less than $\tilde{w} := \frac{W(\Lambda)-W(\bar{\Lambda})}{\Lambda-\bar{\Lambda}}$, where we take $\Lambda = \frac{\underline{\Lambda}+\bar{\Lambda}}{2}$. Let $\Delta = \tilde{w} - \underline{w} > 0$. By continuity, there exists a $\varepsilon > 0$ such that for all $\lambda_B \in \left[\bar{\Lambda} - \varepsilon, \bar{\Lambda}\right]$ and $\Lambda^\star \in [\underline{\Lambda}, \underline{\Lambda} + \varepsilon]$, we have

$$\frac{W(\Lambda^\star) - W(\lambda_B)}{\Lambda^\star - \lambda_B} < \underline{w} + \frac{\Delta}{2}$$

and yet

$$\frac{W(\Lambda) - W(\lambda_B)}{\Lambda - \lambda_B} > \tilde{w} - \frac{\Delta}{2},$$

which implies

$$\frac{W(\Lambda) - W(\lambda_B)}{\Lambda - \lambda_B} > \frac{W(\Lambda^\star) - W(\lambda_B)}{\Lambda^\star - \lambda_B},$$

contradicting the optimality of $\Lambda^\star$.

□

A.5.3 Proof of Lemma 6

Proof. Note that

$$W(\Lambda) = \alpha CS(\Lambda) + (1-\alpha)PS(\Lambda)$$

where CS and PS are defined respectively by (41), (42). We begin by showing that if $\Lambda^\star$ solves (16) with equality at some $\alpha < 1$, then it also does so for all $\tilde{\alpha} > \alpha$. If so, then at α we have, on rearrangement of (16) at the boundary $\Lambda^\star(\alpha)$

$$-\alpha\left(CS'(\Lambda^\star) - \frac{CS(\Lambda^\star) - CS(\lambda_B)}{\Lambda^\star - \lambda_B}\right) = (1-\alpha)\left(PS'(\Lambda^\star) - \frac{PS(\Lambda^\star) - PS(\lambda_B)}{\Lambda^\star - \lambda_B}\right)$$

Since PS is convex (and strictly so, for $\Lambda \geq \bar{\Lambda}$), the left side of this equation is strictly positive. Thus, we must have

$$CS'(\Lambda^\star) - \frac{CS(\Lambda^\star) - CS(\lambda_B)}{\Lambda^\star - \lambda_B} < 0,$$

which implies for $\tilde{\alpha} > \alpha$ that

$$W'(\Lambda^\star(\alpha)) < \frac{W(\Lambda^\star(\alpha)) - W(\lambda_B)}{\Lambda^\star(\alpha) - \lambda_B}.$$

By strict quasi-concavity, this implies that $\Lambda^\star(\tilde{\alpha}) < \Lambda^\star(\alpha)$, and therefore that $\Lambda^\star(\tilde{\alpha})$ solves (16) with equality at $\tilde{\alpha}$, which establishes the existence of the requisite threshold α'.

To see that $\alpha' \leq \frac{1}{2}$, it is convenient to begin by showing that $\alpha' = \frac{1}{2}$ when $\lambda_B = 0$. When $\alpha = \frac{1}{2}$, the regulator's payoff satisfies

$$W(\Lambda) = \frac{1}{2}\left[(1-\Lambda)(\theta_h q_h^\star - c(q_h^\star)) + \Lambda(\theta_l q_l^\star(\Lambda) - c(q_l^\star(\Lambda)))\right].$$

We show that $\Lambda^\star = 1$ satisfies (16) with equality when $\lambda_B = 0$. An argument similar to the comparative static above (but for $\tilde{\alpha} < \alpha$) then implies that for $\tilde{\alpha} < \alpha$, $\Lambda^\star(\tilde{\alpha})$ must satisfy (16) with strict inequality. Hence, $\alpha' = \frac{1}{2}$.

To see this, direct evaluation gives

$$\frac{W(1) - W(0)}{1 - 0} = \frac{1}{2}\left[(\theta_l q_l^\star(\Lambda) - c(q_l^\star(\Lambda))) - (\theta_h q_h^\star - c(q_h^\star))\right]$$

and, after taking the derivative,

$$W'(1) = \frac{1}{2}\left[(\theta_l q_l^\star(\Lambda) - c(q_l^\star(\Lambda))) - (\theta_h q_h^\star - c(q_h^\star))\right],$$

so that $\Lambda^\star = 1$ does satisfy (16) with equality. To extend the result to $\lambda_B \in [0, \bar{\Lambda})$, note from Lemma 5 that $\Lambda^\star(\lambda_B, \alpha)$ is decreasing in λ_B. Hence, $\Lambda^\star$ satisfies (16) with equality at $\alpha = \frac{1}{2}$ for all $\lambda_B \in [0, \bar{\Lambda})$, from which it follows that $\alpha' < \frac{1}{2}$.

For the comparative statics, notice that the argument given above already establishes that $\Lambda^\star(\alpha)$ is decreasing on $[\alpha', 1]$. The proof of differentiability is essentially identical to that in the proof of Lemma 5, and hence omitted.

□

B Uniform Example for Continuous Types and Groups

This appendix provides a simple parametric example illustrating Proposition 7 of the S-shaped interim welfare function. We take two benchmark type distributions F_A, F_B on $[0,1]$ and consider the one-dimensional family of mixtures

$$F_m \;=\; mF_A + (1-m)F_B, \qquad m \in [0,1],$$

where m is the posterior mean of the class index as in Section 6.2. For each m, the seller faces F_m and the regulator's interim welfare is $\widehat{W}(m)$, so by Proposition 7 the shape of $\widehat{W}$ fully determines the optimal information structure.

We specialize to the case in which the benchmark distributions are uniform with the same upper bound:

$$F_A = \mathcal{U}[0,1], \qquad F_B = \mathcal{U}[\theta_0, 1], \quad \text{with } \theta_0 \in [0,1).$$

Thus for any $m \in [0,1]$ the mixture distribution over types has c.d.f.

$$\Lambda_m(\theta) = mF_A(\theta) + (1-m)F_B(\theta),$$

and density $\Lambda_m'(\theta)$.

Because F_A and F_B share the same upper bound, it can be shown that the regularity condition always holds, i.e., for every $m \in [0,1]$, the virtual type

$$\theta \;-\; \frac{1 - \Lambda_m(\theta)}{\Lambda_m'(\theta)}$$

is non-decreasing in $\theta \in \Theta$. Hence the seller's optimal allocation for distribution Λ_m is given by the standard Myerson rule:

$$q^*(\theta \mid m) = \max\Big\{0,\ \theta - \frac{1-\Lambda_m(\theta)}{\Lambda_m'(\theta)}\Big\}. \tag{45}$$

We take quadratic costs $c(q) = \frac{1}{2}q^2$ and equal weight on consumer and producer surplus, so total interim welfare is

$$\widehat{W}(m) := \mathbb{E}_{\theta\sim\Lambda_m}\left[\theta\, q^*(\theta \mid m) - \frac{\big(q^*(\theta \mid m)\big)^2}{2}\right].$$

The next result shows that, in this uniform example, $\widehat{W}(m)$ is S-shaped. For simplicity, consider $\theta_0 \in (3/4, 1)$.

Proposition 8 (Uniform example: S-shaped welfare)**.** *Let $F_A = \mathcal{U}[0,1]$ and $F_B = \mathcal{U}[\theta_0, 1]$ with $\theta_0 \in (0,1)$, and define $\widehat{W}(m)$ as above. Then:*

1. *For $m \in \big[0, \frac{1}{2\theta_0}\big]$, $\widehat{W}(m)$ is affine in m (in fact linear and decreasing).*
2. *For $m \in \big(\frac{1}{2\theta_0}, \frac{3}{4\theta_0}\big)$, $\widehat{W}(m)$ is strictly convex: $\widehat{W}''(m) > 0$.*
3. *For $m \in \big(\frac{3}{4\theta_0}, 1\big]$, $\widehat{W}(m)$ is strictly concave: $\widehat{W}''(m) < 0$.*

In particular, $\widehat{W}(m)$ is S-shaped: affine for small m, then strictly convex, then strictly concave.

Proof. A direct computation of the mixture c.d.f. and density gives

$$\Lambda_m(\theta) = \begin{cases} m\theta, & \theta \in [0,\theta_0], \\ m\theta + (1-m)\dfrac{\theta-\theta_0}{1-\theta_0}, & \theta \in (\theta_0, 1], \end{cases} \qquad \Lambda_m'(\theta) = \begin{cases} m, & \theta \in [0,\theta_0], \\ \dfrac{1-\theta_0 m}{1-\theta_0}, & \theta \in (\theta_0, 1]. \end{cases}$$

Substituting into (45) yields the virtual value

$$v_m(\theta) := \theta - \frac{1-\Lambda_m(\theta)}{\Lambda_m'(\theta)} = \begin{cases} 2\theta - \dfrac{1}{m}, & \theta \in [0,\theta_0], \\ 2\theta - 1, & \theta \in (\theta_0, 1], \end{cases}$$

and hence $q^*(\theta \mid m) = \max\{0, v_m(\theta)\}$.

A simple threshold argument then shows:

- if $m \le \frac{1}{2\theta_0}$, the zero-quantity region covers $[0,\theta_0]$, and trade occurs only for $\theta \ge \frac{1}{2}$; in this case $\widehat{W}(m)$ reduces to

$$\widehat{W}(m) = \frac{\theta_0}{2}\big(1 - \theta_0 m\big),$$

which is linear (and decreasing) in m;

- if $m > \frac{1}{2\theta_0}$, a positive-quantity region appears on $[\frac{1}{2m}, \theta_0]$, and a straightforward integration yields

$$\widehat{W}(m) = \frac{1}{2}\left[-\frac{\theta_0}{m} + \frac{1}{4m^2} - \theta_0^2 m + \theta_0^2 + \theta_0\right].$$

Differentiating twice on the second region gives

$$\widehat{W}''(m) = \frac{1}{2}\left[-\frac{2\theta_0}{m^3} + \frac{3}{2m^4}\right] = \frac{1}{2m^4}\Big(\frac{3}{2m} - 2\theta_0\Big).$$

Thus the sign of $\widehat{W}''(m)$ coincides with the sign of $\frac{3}{2m} - 2\theta_0$. It follows that

$$\widehat{W}''(m) > 0 \iff m < \frac{3}{4\theta_0}, \qquad \widehat{W}''(m) < 0 \iff m > \frac{3}{4\theta_0}.$$

Combining this with the linear region for $m \le \frac{1}{2\theta_0}$ yields items (1)–(3) and hence the S-shape of $\widehat{W}(m)$. $\square$

C Extensions and Robustness

C.1 Only Third-Degree Price Discrimination

We expand the analysis in Section 5.1 to characterize the implementable region and the Pareto frontier. Recall that an *information structure* is a distribution τ over posteriors $\Lambda \in [\lambda_B, \lambda_A]$ satisfying Bayes–plausibility $\mathbb{E}_\tau[\Lambda] = \Lambda_0$. Its outcomes are

$$\big(u(\tau), \pi(\tau)\big) = \big(\mathbb{E}_\tau[U(\Lambda)],\ \mathbb{E}_\tau[\Pi(\Lambda)]\big).$$

Lemma 9 (Two–posterior sufficiency)**.** *For any Bayes–plausible information structure τ on $[\lambda_B, \lambda_A]$ (with $\mathbb{E}_\tau[\Lambda] = \Lambda_0$), let*

$$S_L := \{\Lambda : \Lambda \le \bar{\Lambda}\}, \qquad S_R := \{\Lambda : \Lambda > \bar{\Lambda}\}, \qquad t := \tau(S_L) \in [0,1].$$

Define the conditional means (whenever the corresponding set has positive mass)

$$\Lambda_L := \mathbb{E}_\tau[\Lambda \mid S_L] \in [\lambda_B, \bar{\Lambda}], \qquad \Lambda_R := \mathbb{E}_\tau[\Lambda \mid S_R] \in (\bar{\Lambda}, \lambda_A].$$

With the conventions that if $t = 1$ (resp. $t = 0$) then $\Lambda_L := \Lambda_0$ (resp. $\Lambda_R := \Lambda_0$), consider the two–point distribution

$$\tau' \ := \ t\,\delta_{\Lambda_L} \ + \ (1-t)\,\delta_{\Lambda_R}.$$

Then:

1. *τ' is Bayes–plausible: $\mathbb{E}_{\tau'}[\Lambda] = \Lambda_0$.*
2. *τ' reproduces outcomes: $(u(\tau'), \pi(\tau')) = (u(\tau), \pi(\tau))$.*

Consequently, the set of feasible outcome pairs $\{(u(\tau), \pi(\tau)) : \tau$ Bayes–plausible$\}$ is generated (without loss of generality) by information structures τ' supported on at most two posteriors.

Proof of Lemma 9. Our proof is divided into several steps.

Step 1 (Piecewise affinity). By (21), on S_L we have $U(\Lambda) = 0$ and $\Pi(\Lambda) = \theta_h(1-\Lambda)$, both affine in Λ. On S_R we have $U(\Lambda) = (\theta_h - \theta_l)(1-\Lambda)$ and $\Pi(\Lambda) = \theta_l$, again affine in Λ.

Step 2 (Two–posterior reduction). Let $t := \tau(S_L)$. When $t \in (0,1)$, define $\Lambda_L := \mathbb{E}_\tau[\Lambda \mid S_L]$ and $\Lambda_R := \mathbb{E}_\tau[\Lambda \mid S_R]$. By piecewise affinity and the law of iterated expectations,

$$\begin{aligned}
\mathbb{E}_\tau[U(\Lambda)] &= t\,\mathbb{E}_\tau[U(\Lambda) \mid S_L] + (1-t)\,\mathbb{E}_\tau[U(\Lambda) \mid S_R] \\
&= t\,U(\Lambda_L) + (1-t)\,U(\Lambda_R) \;=\; \mathbb{E}_{\tau'}[U(\Lambda)], \\
\mathbb{E}_\tau[\Pi(\Lambda)] &= t\,\mathbb{E}_\tau[\Pi(\Lambda) \mid S_L] + (1-t)\,\mathbb{E}_\tau[\Pi(\Lambda) \mid S_R] \\
&= t\,\Pi(\Lambda_L) + (1-t)\,\Pi(\Lambda_R) \;=\; \mathbb{E}_{\tau'}[\Pi(\Lambda)].
\end{aligned}$$

Moreover, Bayes–plausibility is preserved:

$$\mathbb{E}_{\tau'}[\Lambda] \;=\; t\,\Lambda_L + (1-t)\,\Lambda_R \;=\; \mathbb{E}_\tau[\Lambda] \;=\; \Lambda_0.$$

Step 3 (Degenerate and boundary cases). If $t = 0$ (all mass on S_R), set $\Lambda_R := \Lambda_0$ and take $\tau' = \delta_{\Lambda_R}$; τ' trivially preserves (U, Π) and the mean. If $t = 1$ (all mass on S_L), set $\Lambda_L := \Lambda_0$ and take $\tau' = \delta_{\Lambda_L}$. If τ places mass at the boundary $\Lambda = \bar{\Lambda}$, any fixed tie–breaking leads to an affine specification on each side; absorb the boundary mass into either S_L or S_R consistently with that tie–breaking and repeat the argument.

Combining the steps yields the claim. □

In the nontrivial case $\lambda_B < \bar{\Lambda} < \lambda_A$, define

$$\alpha := \frac{\lambda_A - \Lambda_0}{\lambda_A - \lambda_B}, \qquad t_B := \frac{\lambda_A - \Lambda_0}{\lambda_A - \bar{\Lambda}}, \qquad s_A := \frac{\bar{\Lambda} - \Lambda_0}{\bar{\Lambda} - \lambda_B}.$$

We use four benchmark outcomes (ex–ante quantities in lowercase (u, π)):

$$P^{\mathrm{ND}} = \begin{cases} \big(0,\ \theta_h(1-\Lambda_0)\big), & \Lambda_0 < \bar{\Lambda}, \\ \big((\theta_h - \theta_l)(1-\Lambda_0),\ \theta_l\big), & \Lambda_0 \ge \bar{\Lambda}, \end{cases} \qquad \text{(no information)}$$

$$P^{BA} = \Big(\ (1-\alpha)\,(\theta_h - \theta_l)\,(1-\lambda_A),\ \ \alpha\,\theta_h(1-\lambda_B) + (1-\alpha)\,\theta_l\ \Big)\ ,$$
(full class-membership revelation)

$$P^{\bar{\Lambda}A} = \Big(\ t_B\,(\theta_h - \theta_l)\,(1-\bar{\Lambda}) + (1-t_B)\,(\theta_h - \theta_l)\,(1-\lambda_A),\ \ \theta_l\ \Big), \qquad \text{(support } \{\,\bar{\Lambda},\ \lambda_A\,\}\text{)}$$

$$P^{B\bar{\Lambda}} = \Big(\ (1-s_A)\,(\theta_h - \theta_l)\,(1-\bar{\Lambda}),\ \ \theta_h\big[\,s_A(1-\lambda_B) + (1-s_A)\,(1-\bar{\Lambda})\,\big]\ \Big).$$
(support $\{\,\lambda_B,\ \bar{\Lambda}\,\}$)

Proof. We write interim (posterior) quantities with capitals (U, Π) and ex–ante expectations with lowercase (u, π).

(i) P^{ND} *(no information).* Degenerate posterior at Λ_0. If $\Lambda_0 < \bar{\Lambda}$ then $p^* = \theta_h$, so $(u, \pi) = (0, \theta_h(1-\Lambda_0))$. If $\Lambda_0 \ge \bar{\Lambda}$ then (by the boundary convention when equality holds) we evaluate at the right piece: $(u, \pi) = ((\theta_h - \theta_l)(1-\Lambda_0), \theta_l)$.

(ii) P^{BA} *(full class membership revelation).* Support $\{\lambda_B, \lambda_A\}$ with weights $\alpha, 1-\alpha$, where Bayes–plausibility $\alpha\lambda_B + (1-\alpha)\lambda_A = \Lambda_0$ yields $\alpha = (\lambda_A - \Lambda_0)/(\lambda_A - \lambda_B)$. At $\lambda_B < \bar{\Lambda}$: $(U, \Pi) = (0, \theta_h(1-\lambda_B))$. At $\lambda_A > \bar{\Lambda}$: $(U, \Pi) = ((\theta_h - \theta_l)(1-\lambda_A), \theta_l)$. Averaging gives the displayed P^{BA}.

(iii) $P^{\bar{\Lambda}A}$ *(support $\{\bar{\Lambda}, \lambda_A\}$).* Place weight t_B on $\bar{\Lambda}$ and $1-t_B$ on λ_A, where $t_B\bar{\Lambda}+(1-t_B)\lambda_A = \Lambda_0$ so $t_B = \frac{\lambda_A - \Lambda_0}{\lambda_A - \bar{\Lambda}}$. Using the boundary convention at $\bar{\Lambda}$ and that $\lambda_A > \bar{\Lambda}$ implies price θ_l:

$$u = t_B(\theta_h - \theta_l)(1-\bar{\Lambda}) + (1-t_B)(\theta_h - \theta_l)(1-\lambda_A), \qquad \pi = \theta_l,$$

which yields $P^{\bar{\Lambda}A}$.

(a) *Implementation (assume $\Lambda_0 \geq \bar{\Lambda}$).* Use a binary signal $s \in \{L, R\}$ with

$$\Pr[L \mid B] = 1, \qquad \Pr[L \mid A] = \frac{\mu_0}{1-\mu_0} \cdot \frac{\bar{\Lambda} - \lambda_B}{\lambda_A - \bar{\Lambda}} \in [0, 1].$$

Then R occurs only under A and hence reveals A (posterior λ_A); L pools to posterior $\bar{\Lambda}$ with probability t_B, implementing $P^{\bar{\Lambda}A}$ exactly.

(iv) $P^{B\bar{\Lambda}}$ *(support $\{\lambda_B, \bar{\Lambda}\}$).* Place weight s_A on λ_B and $1-s_A$ on $\bar{\Lambda}$, where $s_A\lambda_B + (1-s_A)\bar{\Lambda} = \Lambda_0$, i.e., $s_A = \frac{\bar{\Lambda} - \Lambda_0}{\bar{\Lambda} - \lambda_B} \in [0, 1]$. At λ_B: $(U, \Pi) = (0, \theta_h(1-\lambda_B))$; at $\bar{\Lambda}$: $(U, \Pi) = ((\theta_h - \theta_l)(1-\bar{\Lambda}), \theta_l)$. Averaging,

$$u = (1-s_A)(\theta_h - \theta_l)(1-\bar{\Lambda})$$
$$\pi = s_A\,\theta_h(1-\lambda_B) + (1-s_A)\,\theta_l = \theta_h\Big[s_A\big((1-\lambda_B) - (1-\bar{\Lambda})\big) + (1-\bar{\Lambda})\Big] = \theta_h(1-\Lambda_0).$$

This is the displayed $P^{B\bar{\Lambda}}$.

(a) *Implementation (assume $\Lambda_0 \leq \bar{\Lambda}$).* Use $s \in \{L, R\}$ with

$$\Pr[L \mid A] = 0, \qquad \Pr[L \mid B] = \frac{\bar{\Lambda} - \Lambda_0}{(\bar{\Lambda} - \lambda_B)\mu_0} \in [0, 1].$$

Then L reveals B (posterior λ_B), while R pools to posterior $\bar{\Lambda}$ with probability $1-s_A$, implementing $P^{B\bar{\Lambda}}$ exactly.

□

Let $\mathcal{F} := \{(u(\tau), \pi(\tau)) : \ \tau \text{ Bayes–plausible on } [\lambda_B, \lambda_A]\}$ denote the set of all achievable outcome pairs.

Proposition 9 (Achievable Region)**.** *We have*

(a) Trivial cases. If $\lambda_A \leq \bar{\Lambda}$ (resp. $\lambda_B \geq \bar{\Lambda}$), the seller always posts θ_h (resp. θ_l) at every posterior, hence the achievable set is the singleton

$$\mathcal{F} = \{P^{\mathrm{ND}}\} = \begin{cases} \{(0, \theta_h(1-\Lambda_0))\}, & \lambda_A \leq \bar{\Lambda}, \\ \{((\theta_h - \theta_l)(1-\Lambda_0), \theta_l)\}, & \lambda_B \geq \bar{\Lambda}. \end{cases}$$

(b) Nontrivial case $\lambda_B < \bar{\Lambda} < \lambda_A$.

$$\textit{(i) If } \Lambda_0 < \bar{\Lambda}: \quad \mathcal{F} = \mathsf{convhull}\{\, P^{\mathrm{ND}},\ P^{B\bar{\Lambda}},\ P^{BA}\,\}.$$
$$\textit{(ii) If } \Lambda_0 \geq \bar{\Lambda}: \quad \mathcal{F} = \mathsf{convhull}\{\, P^{\mathrm{ND}},\ P^{\bar{\Lambda}A},\ P^{BA}\,\}.$$

Proof of Proposition 9. By definition of $p^*(\Lambda)$, the interim pair $(U(\Lambda), \Pi(\Lambda))$ is

$$\big(U(\Lambda), \Pi(\Lambda)\big) = \begin{cases} \big(0,\ \theta_h(1-\Lambda)\big), & \Lambda < \bar{\Lambda} \quad \text{(left region } S_L\text{)}, \\ \big((\theta_h - \theta_l)(1-\Lambda),\ \theta_l\big), & \Lambda \geq \bar{\Lambda} \quad \text{(right region } S_R\text{)}, \end{cases}$$

and on each side of the threshold the map $\Lambda \mapsto (U(\Lambda), \Pi(\Lambda))$ is *affine.*

For any Bayes–plausible τ with mean Λ_0, let $t := \tau(S_L)$ be the weight on the left region, and let

$$\Lambda_L := \mathbb{E}[\Lambda \mid S_L] \in [\lambda_B, \bar{\Lambda}), \qquad \Lambda_R := \mathbb{E}[\Lambda \mid S_R] \in [\bar{\Lambda}, \lambda_A]$$

be the conditional means (defined whenever the corresponding region has positive mass; in boundary cases we set Λ_L or Λ_R equal to Λ_0 as in Lemma 9). Then the *two-point* information structure

$$\tau' := t\,\delta_{\Lambda_L} + (1-t)\,\delta_{\Lambda_R},$$

where, for any posterior mean λ, δ_λ denotes the degenerate distribution (Dirac measure) that assigns probability one to λ (a point mass at λ), preserves the mean ($t\Lambda_L + (1-t)\Lambda_R = \Lambda_0$) and reproduces outcomes:

$$(u(\tau), \pi(\tau)) = (u(\tau'), \pi(\tau')).$$

Therefore, *it suffices to analyze two–posterior experiments.* For such a two–point structure we have explicitly

$$\big(u, \pi\big) = \Big((1-t)\,(\theta_h - \theta_l)\,(1-\Lambda_R),\ \ t\,\theta_h(1-\Lambda_L) + (1-t)\,\theta_l\Big), \tag{46}$$

with the Bayes weight

$$t = \frac{\Lambda_R - \Lambda_0}{\Lambda_R - \Lambda_L} \in [0,1], \qquad \Lambda_L \in [\lambda_B, \bar{\Lambda}),\ \ \Lambda_R \in [\bar{\Lambda}, \lambda_A]. \tag{47}$$

We now characterize $\mathcal{F}$ by cases.

(a) Trivial cases.

If $\lambda_A \leq \bar{\Lambda}$, *all* posteriors lie in S_L; the seller always posts θ_h, hence $U(\Lambda) = 0$ and $\Pi(\Lambda) = \theta_h(1-\Lambda)$ pointwise. Taking expectations under any τ gives

$$(u(\tau), \pi(\tau)) = (0, \theta_h(1 - \mathbb{E}[\Lambda])) = (0, \theta_h(1-\Lambda_0)) =: P^{\mathrm{ND}}.$$

Thus $\mathcal{F} = \{P^{\mathrm{ND}}\}$. Similarly, if $\lambda_B \geq \bar{\Lambda}$, *all* posteriors lie in S_R; the seller always posts θ_l, hence

$$(u(\tau), \pi(\tau)) = ((\theta_h - \theta_l)(1-\Lambda_0), \theta_l) =: P^{\mathrm{ND}},$$

so again $\mathcal{F}$ is a singleton. This proves part (a).

From now on assume the *nontrivial* configuration $\lambda_B < \bar{\Lambda} < \lambda_A$, where both sides of the pricing threshold are reachable.

(b.i) Nontrivial case and $\Lambda_0 < \bar{\Lambda}$ (prior on the left). Under no information the seller posts θ_h, so $P^{\mathrm{ND}} = (0, \theta_h(1 - \Lambda_0))$.

Step 1: For fixed Λ_R, feasible outcomes trace a line segment from P^{ND} when varying Λ_L.

Fix any $\Lambda_R \in [\bar{\Lambda}, \lambda_A]$ and let Λ_L vary in $[\lambda_B, \Lambda_0]$. By (47), $t(\Lambda_L) = \frac{\Lambda_R - \Lambda_0}{\Lambda_R - \Lambda_L}$ increases continuously from $t(\Lambda_0) = 1$ (when $\Lambda_L = \Lambda_0$) down to $t(\lambda_B) = \frac{\Lambda_R - \Lambda_0}{\Lambda_R - \lambda_B} \in (0, 1)$ (when $\Lambda_L = \lambda_B$). Since (46) is affine in t, the image of $\Lambda_L \in [\lambda_B, \Lambda_0]$ is the straight line segment with endpoints:

$$\underbrace{\Big((1-1)(\theta_h - \theta_l)(1 - \Lambda_R),\ \theta_h(1 - \Lambda_0)\Big)}_{= P^{\mathrm{ND}}}$$

$$R(\Lambda_R) := \Big((1 - \tilde{t})\,(\theta_h - \theta_l)\,(1 - \Lambda_R),\quad \tilde{t}\,\theta_h(1 - \lambda_B) + (1 - \tilde{t})\,\theta_l\Big),$$

where $\tilde{t} := \frac{\Lambda_R - \Lambda_0}{\Lambda_R - \lambda_B}$. Thus, *for each fixed* Λ_R, the feasible points form the segment $[\,P^{\mathrm{ND}},\ R(\Lambda_R)\,]$. Taking the union over all Λ_R gives

$$\mathcal{F} = \bigcup_{\Lambda_R \in [\bar{\Lambda}, \lambda_A]} [\,P^{\mathrm{ND}},\ R(\Lambda_R)\,]. \tag{48}$$

Step 2: The "upper edge" $\{R(\Lambda_R)\}$ is exactly the line segment $[\,P^{B\bar{\Lambda}},\ P^{BA}\,]$.

Evaluate $R(\Lambda_R)$ at the two extremes, using the boundary convention at $\Lambda_R = \bar{\Lambda}$:

$$R(\bar{\Lambda}) = \Big((1 - s_A)\,(\theta_h - \theta_l)\,(1 - \bar{\Lambda}),\quad s_A\,\theta_h(1 - \lambda_B) + (1 - s_A)\,\theta_l\Big) = P^{B\bar{\Lambda}},$$

$$R(\lambda_A) = \Big((1 - \alpha)\,(\theta_h - \theta_l)\,(1 - \lambda_A),\quad \alpha\,\theta_h(1 - \lambda_B) + (1 - \alpha)\,\theta_l\Big) = P^{BA}.$$

We now show that as Λ_R varies, $R(\Lambda_R)$ sweeps *only* this segment.

Consider the two simple two–posterior information structures:

$$\tau^{B\bar{\Lambda}} :\ \{\lambda_B,\ \bar{\Lambda}\} \text{ with left mass } s_A, \qquad \tau^{BA} :\ \{\lambda_B,\ \lambda_A\} \text{ with left mass } \alpha.$$

For any $\eta \in [0, 1]$, consider their mixture $\sigma_\eta := \eta\,\tau^{B\bar{\Lambda}} + (1-\eta)\,\tau^{BA}$. Its support is $\{\lambda_B,\ \bar{\Lambda},\ \lambda_A\}$, and its *right* mass equals $\eta(1 - s_A) + (1 - \eta)(1 - \alpha)$. The *conditional mean on the right region* is

$$\bar{\Lambda}_R(\eta) \ = \ \frac{\eta(1 - s_A)\bar{\Lambda} + (1 - \eta)(1 - \alpha)\lambda_A}{\eta(1 - s_A) + (1 - \eta)(1 - \alpha)} \ \in \ [\,\bar{\Lambda}, \lambda_A\,],$$

which varies continuously and *monotonically* from $\bar{\Lambda}$ (when $\eta = 1$) to λ_A (when $\eta = 0$).

By the two–posterior lemma (collapse conditional laws to their means), σ_η can be replaced by the two–point experiment $\{\lambda_B, \bar{\Lambda}_R(\eta)\}$ with left mass $\tilde{t}(\bar{\Lambda}_R(\eta)) = \frac{\bar{\Lambda}_R(\eta) - \Lambda_0}{\bar{\Lambda}_R(\eta) - \lambda_B}$, which delivers exactly $R(\bar{\Lambda}_R(\eta))$ by (46). On the other hand, outcomes are *linear* in the information structure, so the outcome of σ_η is

$$\eta\,P^{B\bar{\Lambda}} + (1 - \eta)\,P^{BA}.$$

Therefore, $R(\bar\Lambda_R(\eta)) \in [\, P^{B\bar\Lambda},\, P^{BA}\,]$ for every η; as η runs from 1 to 0, $\bar\Lambda_R(\eta)$ runs over the whole interval $[\bar\Lambda, \lambda_A]$, hence

$$\{R(\Lambda_R) : \Lambda_R \in [\bar\Lambda, \lambda_A]\} = [\, P^{B\bar\Lambda},\, P^{BA}\,]. \tag{49}$$

Step 3: Conclusion for $\Lambda_0 \le \bar\Lambda$. Combining (48) and (49), and using the cone identity $\bigcup_{y\in[a,b]}[x, y] = \mathsf{convhull}\{x, a, b\}$, we obtain

$$\mathcal{F} = \mathsf{convhull}\{\, P^{\mathrm{ND}},\, P^{B\bar\Lambda},\, P^{BA}\,\}.$$

(b.ii) Nontrivial case and $\Lambda_0 \ge \bar\Lambda$ (prior on the right). Under no information the seller posts θ_l, so $P^{\mathrm{ND}} = ((\theta_h - \theta_l)(1 - \Lambda_0), \theta_l)$.

Step 1: For fixed Λ_L, feasible outcomes trace a ray from P^{ND} when varying Λ_R.

Fix any $\Lambda_L \in [\lambda_B, \bar\Lambda)$ and let Λ_R vary in $[\Lambda_0, \lambda_A]$. By (47), $t(\Lambda_R) = \frac{\Lambda_R - \Lambda_0}{\Lambda_R - \Lambda_L}$ increases from 0 (at $\Lambda_R = \Lambda_0$) to $t_A(\Lambda_L) := \frac{\lambda_A - \Lambda_0}{\lambda_A - \Lambda_L}$ (at $\Lambda_R = \lambda_A$). Because (46) is affine in t, the image is the segment with endpoints

$$\underbrace{\Big((\theta_h - \theta_l)(1 - \Lambda_0),\ \theta_l\Big)}_{= P^{\mathrm{ND}}} \quad \text{and}$$

$$Q(\Lambda_L) := \Big((1 - t_A(\Lambda_L))(\theta_h - \theta_l)(1 - \lambda_A),\ \ t_A(\Lambda_L)\,\theta_h(1 - \Lambda_L) + (1 - t_A(\Lambda_L))\,\theta_l\Big).$$

Taking the union over all Λ_L we get

$$\mathcal{F} = \bigcup_{\Lambda_L \in [\lambda_B, \bar\Lambda)} [\, P^{\mathrm{ND}},\, Q(\Lambda_L)\,] \ \cup\ [\, P^{\mathrm{ND}},\, P^{\bar\Lambda A}\,], \tag{50}$$

where the additional segment $[P^{\mathrm{ND}}, P^{\bar\Lambda A}]$ corresponds to the *exact* boundary experiment with support $\{\bar\Lambda, \lambda_A\}$ (both posteriors in S_R), weight $t_B = \frac{\lambda_A - \Lambda_0}{\lambda_A - \bar\Lambda}$ on $\bar\Lambda$, and outcome $P^{\bar\Lambda A}$.

Step 2: The "upper edge" is the segment $[\, P^{BA},\, P^{\bar\Lambda A}\,]$.

At $\Lambda_L = \lambda_B$ we obtain

$$Q(\lambda_B) = \Big((1 - \alpha)(\theta_h - \theta_l)(1 - \lambda_A),\ \ \alpha\,\theta_h(1 - \lambda_B) + (1 - \alpha)\,\theta_l\Big) = P^{BA}.$$

The other endpoint $P^{\bar\Lambda A}$ is attained *without limits* by the two–point support $\{\bar\Lambda, \lambda_A\}$ with the stated Bayes weight t_B; by our boundary convention, (U, Π) at $\bar\Lambda$ equals $((\theta_h - \theta_l)(1 - \bar\Lambda), \theta_l)$. Therefore all upper endpoints in (50) lie on the straight segment $[\, P^{BA},\, P^{\bar\Lambda A}\,]$.

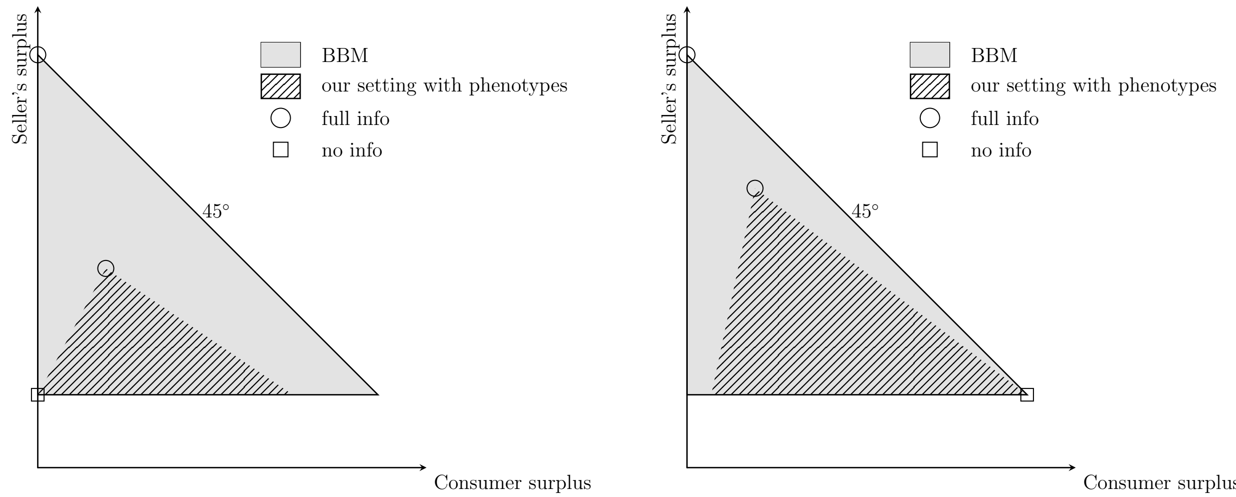
Step 3: Conclusion for $\Lambda_0 > \bar\Lambda$.

Combining (50) with Step 2 and using again the cone identity,

$$\mathcal{F} = \mathsf{convhull}\{\, P^{\mathrm{ND}},\, P^{\bar\Lambda A},\, P^{BA}\,\}.$$

This completes the proof. □

Proposition 10 (Pareto frontier and slopes)**.** *Assume the non-trivial case $\lambda_B < \bar\Lambda < \lambda_A$.*



(a) $\Lambda_0 < \bar{\Lambda}$ (b) $\Lambda_0 \geq \bar{\Lambda}$

Figure 6: Feasible Region in Third-degree Price Discrimination under the Non-trivial Case $\lambda_B < \bar{\Lambda} < \lambda_A$

(i) If $\Lambda_0 < \bar{\Lambda}$, the Pareto frontier is the line segment $\overline{P^{B\bar{\Lambda}}P^{BA}}$.

(ii) If $\Lambda_0 \geq \bar{\Lambda}$, the Pareto frontier is the line segment $\overline{P^{\mathrm{ND}}P^{BA}}$.

In both cases, with (u, π)–space (horizontal u, vertical π), the slope of the frontier is

$$\frac{d\pi}{du} = -\frac{\bar{\Lambda} - \lambda_B}{\bar{\Lambda}\,(1 - \lambda_B)} < 0,$$

with $|\frac{d\pi}{du}| \leq 1$.

Proof of Proposition 10. By Proposition 9, in the nontrivial case the achievable region is a triangle (including its interior). Hence, the Pareto frontier is a line segment. It therefore suffices to identify the information structures that maximize the seller's expected utility and the consumers' expected surplus. Obviously, full information maximizes the seller's expected utility, so one endpoint of this line segment is P^{BA}.

We now identify the other endpoint. The interim consumer-surplus function $U(\Lambda)$, defined in Equation (21), is constant on $[\lambda_B, \bar{\Lambda})$, exhibits a jump at $\Lambda = \bar{\Lambda}$, and is linearly decreasing on $(\bar{\Lambda}, , \lambda_A]$. By an argument analogous to Proposition 5, when $\Lambda_0 < \bar{\Lambda}$, the consumer-surplus–maximizing information structure has support $\{\lambda_B, \bar{\Lambda}\}$ and induces the outcome $P^{B\bar{\Lambda}}$; when $\Lambda_0 \geq \bar{\Lambda}$, no information is optimal, which induces P^{ND}. This completes the proof.

We now compute the slopes in the two cases.

(i) Case $\Lambda_0 < \bar{\Lambda}$: slope of $\overline{P^{B\bar{\Lambda}}P^{BA}}$.

Recall

$$\alpha := \frac{\lambda_A - \Lambda_0}{\lambda_A - \lambda_B}, \qquad s_A := \frac{\bar{\Lambda} - \Lambda_0}{\bar{\Lambda} - \lambda_B},$$

and the ex-ante coordinates (using $\theta_l = \theta_h(1-\bar\Lambda)$, $\theta_h - \theta_l = \theta_h\bar\Lambda$)

$$P^{BA} = \big(u^{BA}, \pi^{BA}\big) = \Big((1-\alpha)\,\theta_h\bar\Lambda\,(1-\lambda_A),\quad \theta_h\big[\alpha(1-\lambda_B) + (1-\alpha)(1-\bar\Lambda)\big]\Big),$$
$$P^{B\bar\Lambda} = \big(u^{B\bar\Lambda}, \pi^{B\bar\Lambda}\big) = \Big((1-s_A)\,\theta_h\bar\Lambda\,(1-\bar\Lambda),\quad \theta_h\big[s_A(1-\lambda_B) + (1-s_A)(1-\bar\Lambda)\big]\Big).$$

Hence,

$$\pi^{BA} - \pi^{B\bar\Lambda} = \theta_h\,(\alpha - s_A)\big(\bar\Lambda - \lambda_B\big),$$

and

$$u^{BA} - u^{B\bar\Lambda} = \theta_h\bar\Lambda\Big((1-\alpha)(1-\lambda_A) - (1-\bar\Lambda)(1-s_A)\Big).$$

A direct algebraic simplification gives

$$\alpha - s_A = \frac{(\Lambda_0 - \lambda_B)\,(\lambda_A - \bar\Lambda)}{(\lambda_A - \lambda_B)\,\big(\bar\Lambda - \lambda_B\big)}, \qquad (1-\alpha)(1-\lambda_A) - (1-\bar\Lambda)(1-s_A) = -\,\frac{(\Lambda_0 - \lambda_B)\,(1-\lambda_B)\,(\lambda_A - \bar\Lambda)}{(\lambda_A - \lambda_B)\,\big(\bar\Lambda - \lambda_B\big)}.$$

Taking the ratio (the common factors cancel) yields

$$\frac{\pi^{BA} - \pi^{B\bar\Lambda}}{u^{BA} - u^{B\bar\Lambda}} = -\,\frac{\bar\Lambda - \lambda_B}{\bar\Lambda\,(1-\lambda_B)}.$$

(ii) Case $\Lambda_0 \geq \bar\Lambda$: slope of $\overline{P^{\mathrm{ND}}P^{BA}}$.

Now $P^{\mathrm{ND}} = \big((\theta_h - \theta_l)(1-\Lambda_0),\ \theta_l\big) = \big(\theta_h\bar\Lambda(1-\Lambda_0),\ \theta_h(1-\bar\Lambda)\big)$, while P^{BA} is as above. Using Bayes–plausibility $1 - \Lambda_0 = \alpha(1-\lambda_B) + (1-\alpha)(1-\lambda_A)$,

$$\pi^{BA} - \pi^{\mathrm{ND}} = \theta_h\Big(\alpha(1-\lambda_B) + (1-\alpha)(1-\bar\Lambda) - (1-\bar\Lambda)\Big) = \theta_h\,\alpha\big(\bar\Lambda - \lambda_B\big),$$
$$u^{BA} - u^{\mathrm{ND}} = \theta_h\bar\Lambda\Big((1-\alpha)(1-\lambda_A) - \big[\alpha(1-\lambda_B) + (1-\alpha)(1-\lambda_A)\big]\Big) = -\,\theta_h\bar\Lambda\,\alpha(1-\lambda_B).$$

Therefore

$$\frac{\pi^{BA} - \pi^{\mathrm{ND}}}{u^{BA} - u^{\mathrm{ND}}} = -\,\frac{\bar\Lambda - \lambda_B}{\bar\Lambda\,(1-\lambda_B)}.$$

Summary. In either case,

$$\left|\frac{d\pi}{du}\right| = \frac{\bar\Lambda - \lambda_B}{\bar\Lambda\,(1-\lambda_B)} \;\leq\; 1,$$

since

$$\bar\Lambda(1-\lambda_B) \;-\; \big(\bar\Lambda - \lambda_B\big) = (1-\bar\Lambda)\,\lambda_B \;\geq\; 0.$$

As $(1-\bar\Lambda) \in (0,1)$, equality holds if and only if $\lambda_B = 0$. This completes the proof. □

C.2 Insurance

C.2.1 Proof of Proposition 7

We now prove Proposition 7. We show that, in an monopoply insurance market, when $\alpha = 1/2$, the objective function (social surplus) is linear for low values of Λ and concave for high values of Λ.

Without loss of generality, we can normalize $\theta_l = 1$. Then, differentiating 28 twice gives:

$$2W''^M(\Lambda) = 2\left(q_l^{\gamma-1} - 1\right) q_l' + \Lambda\left((\gamma-1)q_l^{\gamma-2}\right)(q_l')^2 + \Lambda\left(q_l^{\gamma-1} - 1\right) q_l''. \tag{51}$$

Differentiating 27 twice gives:

$$q_l' = \frac{q_l\left(1 - \theta_h q_l^{\gamma-1}\right)}{(\gamma-1)\Lambda} \quad \text{and} \quad q_l'' = \frac{q_l'}{\Lambda(\gamma-1)}\left[2 - \gamma - \gamma\theta_h q_l^{\gamma-1}.\right] \tag{52}$$

Plugging these into 51 and re-arranging yields:

$$2W''^M(\Lambda) = \frac{q_l(\Lambda)\left(\theta_h q_l(\Lambda)^{\gamma-1} - 1\right)^2}{(1-\gamma)^2\Lambda}\left[(2\gamma-1)q_l(\Lambda)^{\gamma-1} - \gamma,\right] \tag{53}$$

which is negative whenever $2\gamma - 1 < 0$, i.e., whenever $\gamma < 0.5$