The Production of Information to Price Discriminate

Willy Lefez (Humboldt Universität)

Discussion Paper No. 535

July 02, 2025

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, Ludwig Maximilian University of Munich, 80539 Munich, Germany
info@rationality-and-competition.de

The Production of Information to Price Discriminate*

Willy Lefez†

July 2, 2025

Abstract

We study price discrimination by a monopolistic seller that endogenously produces a market segmentation at a cost, and question the efficiency of the production of market segmentations led by private incentives. We show that the efficient market segmentation gives all the gains in total surplus to the buyer, and the seller profit stays at the uniform profit level. Our result suggests that the private production of information by sellers to price discriminate is significantly inefficient.

Keywords: Price Discrimination, Cost of Information, Production of Information.

JEL Classification: D42, D83, L12.

*We are grateful to Roland Strausz, Helmut Bester, Andreas Asseyer, Dilip Ravidran, Tobias Gamp, Vincent Meisner, Volker Nocke, Nicolas Schutz, Balázs Szentes for helpful discussions as well as participants to the ZEW conference on ICT and workshop participants at Humboldt university for their questions and comments. This research was supported by the PRIVDIMA advanced ERC grant. This article is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 853123). Support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) is gratefully acknowledged.

†Humboldt University. Email: willy.lefez@hu-berlin.de

1 Introduction

Firms do not simply acquire information, they produce information. They transform data, a digital resource yet unfit for direct use, into insights that help decision-making. Examples abound: KPMG, a large consulting firm, argues that "Data is the great hidden resource that flows, largely untapped, through major organizations" and offers clients to "convert data into insights."[1] Meta transforms its massive amount of user data into insights to help businesses target and reach a personalized audience segment.[2] Many firms' core business model is to produce information. Palantir, for instance, helps institutions and businesses to secure, store, compile, and analyze their data to improve decision-making.[3] Insights is what this industry refers to as a "deep understanding of data" that "helps organizations make better decisions."[4] Transforming data into insights is costly and becoming one of the major investments of firms.[5]

We study the production of information where insights are market segmentations used to price discriminate. We consider a seller that endogenously produces at a cost a market segmentation. Think of a seller acquiring data about potential buyers, such as zip codes, demographics, cookies, or the browser history of past purchases, then using it to segment the market into consumer groups with different estimates about their value for the good. In each market segment, the seller sets a price given the estimated distribution of buyer values, i.e., the demand. The produced market segmentation is the input for the seller's price discrimination strategy.

Taking into account that producing information is costly, this paper questions the efficiency of the production of information to price discriminate. We argue why the endogenous choice of market segmentations is crucial to our understanding of the welfare consequences of price discrimination.

As shown in the seminal work of Bergemann et al. (2015) (BBM thereafter), how a monopolistic seller is informed about buyers entirely pins down the surplus division between buyers and the seller. Any surplus division that ensures the seller no less

[1] KPGM.

[2] Insights to Go.

[3] Foundry is one of Palentir's personalized data solution.

[4] See google's course on insights.

[5] See Report on data analytics industry.

profit than uniform pricing, non-negative buyer surplus, and no more total surplus than the efficient level is achieved by price discrimination under some information about buyers. In other words, how the seller produces information determines the surplus outcome of price discrimination.

We know that if the production of information is unfeasible or too costly, then we have the notoriously inefficient monopoly uniform-pricing outcome. The Coase theorem fails because of information asymmetries between the seller and buyers. So by reducing information asymmetries in the market, the production of information is socially valuable. On the other hand, if the production of information is free, we know the seller acquires all the information to first-degree price discriminate resulting in an efficient outcome. Therefore, any distortions must take root in costly production frictions.

In the context of price discrimination one cannot model information in a reduced form, e.g., as a quality dimension like accuracy, precision, or variance. Indeed, as apparent in BBM, comparative statics on welfare does not boil down to an issue of too little or too much information or to an issue of a too coarse or too refined market segmentation: Capturing distortions requires a full-fledged information model. The production of information in this context is out of the scope of standard firms' production models.

We compare the profit-maximizing production of information to the welfare-maximizing production of information, where information is used by the seller to price discriminate.[6] We assume a flexible production that allows the seller or the planner to acquire any signal at a cost, reflecting an industry that provides ever more tailored and personalized data solutions. We assume both face the same cost, excluding externalities such as privacy costs that could worsen distortions.

Our main result shows that, under various market environments, the welfare-maximizing information gives all the welfare gains to the buyer, nothing goes to the seller. That is, the planner solution coincides with the buyer-optimal one and the seller profit stays at the uniform level.

To increase trade efficiency, the planner produces a market segmentation that induces lower prices, but does not over-produce to save on costs. Hence, it stops

[6] By welfare we mean the sum of the buyer surplus and the seller profit.

producing information exactly when the seller switches to a lower price. As a result, the efficiently produced market segmentation makes the seller indifferent between multiple prices in all but one segment. This key property of efficient segmentations holds for all continuous and strictly Blackwell monotone costs of information. We then show this property implies the efficient segmentation gives all welfare gains to buyers in various market conditions.

In the case where buyers have binary values, the implication is immediate: with no further restriction the efficient segmentation gives all the welfare gains to buyers. In the case where buyers have finitely many values, we show our main result holds for a restricted set of prior distributions of buyer values. Finally, in the case where buyers have three values and with a symmetry and differentiability assumption on the production cost of information, we show our main result holds for unrestricted prior distributions of buyers values.

Figure 1 represents our main result on the BBM surplus triangle. The red

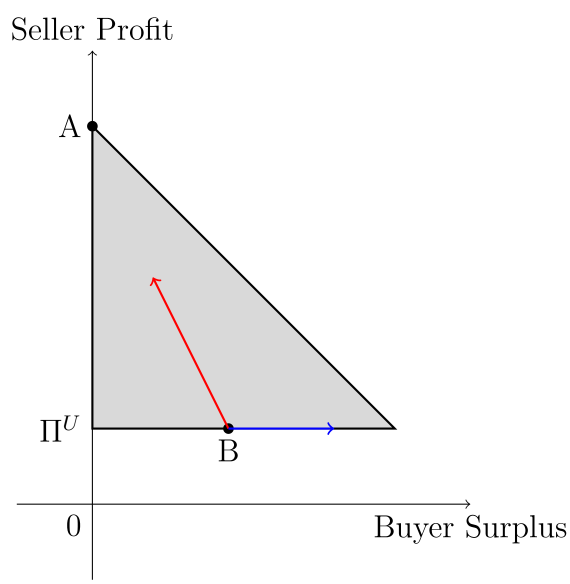

Figure 1: Seller vs Planner Segmentations on the BBM Surplus Triangle

arrow illustrates the profit-maximizing segmentation: Starting from point B, the uniform pricing or no-information point, the seller acquires information to ideally reach point A, the 1st degree price discrimination or full information point, if information is cheap but may stop before otherwise.[7] The blue arrow captures the welfare-maximizing segmentation: Starting from point B the planner acquires information that only benefits buyers, hence staying on the bottom edge of the surplus triangle.

We consider a monopolistic seller that, in the first stage, produces a posterior belief distribution about buyers' values at a cost. A posterior belief distribution corresponds to a market segmentation that distributes buyers in different segments with different value distributions. In the second stage, the seller sets a price in each segment to maximize profit. We contrast with a planner that produces information for the seller to maximize total surplus.

The profit-maximizing production of information problem is formally equivalent to a rational inattention problem (Sims 2003 and Caplin & Dean 2013).[8] The welfare-maximizing production of information is a Bayesian persuasion problem with costly information acquisition (Gentzkow & Kamenica 2014).[9]

We view the production cost of information as a technical cost to transform data into insight. It accounts for each step of the production process (collecting, compiling, analyzing data) and it accounts for production factor's price. We assume that the seller rationalizes production: If two production plans generate the same price discrimination strategies, the seller produces the cheapest one. We show that the resulting rationalized production cost is Blackwell monotone: More information (more spread out beliefs) costs more.

In the first part of the paper (section 3), we present our main result in the binary buyer value case. While without production cost of information, the seller acquires full information and perfectly price discriminate buyers, resulting in an efficient outcome; with costly production of information, we show that efficient segmentations never benefits seller: all the surplus gained goes to buyers.

In section 4, we establish the key property that at an efficient market segmentation the seller is indifferent between multiple prices in all segments except the highest-price-inducing one. Proposition 3 extends the our main result to the case

[7] Proposition 8 characterizes the seller's trade-off.

[8] The problem of incurring attention costs to decode a signal and pick the right action is the same as the problem of acquiring costly information about buyers values to set the right price.

[9] The receiver (seller) is misaligned with the sender (planner).

where buyers have finitely many types for a restricted set of prior buyer value distributions.[10] We then build an example with an asymmetric cost of producing information for which starting from a prior out of proposition 3 region, the planner produces a segmentation that gives profit to the seller.

Section 5 restricts attention to differentiable costs of information. We show that, in contrast to the efficient segmentation, at profit-maximizing market segmentation, only one price is optimal for the seller in all segments. We then consider *symmetric* information costs where spreading is equally costly in all directions. If buyers have three types we extend our main result to all prior distribution of buyer values.

Section 6 presents first-order characterizations of the planner and seller solutions used for section 5's results, which holds even for non-posterior separable costs

Related Literature. Our paper relates to the literature sparked by Bergemann et al. (2015) studies price discrimination under any market segmentation induced by some information about buyers. See, e.g., Ali et al. (2020), Hidir & Vellodi (2021), Ichihashi & Smolin (2022), Elliott et al. (2022), Barreto et al. (2022), Haghpanah & Siegel (2022), and De Cornière et al. (2024). Ravid et al. (2022) considers how buyers should costly learn before trading with a monopolist.[11] The closest paper to ours is Tekdir (2024) that also studies endogeneous market segmentation acquisition for the seller in the case of entropy costs of information. They exhibit a non-monotonic relationship in consumer surplus and welfare when information is made more costly. Our paper compares the seller problem with the planner for all strictly Blackwell monotone on costs of information. Devine & Munoz-Garcia (2018) studies second-degree price discrimination with costly information acquisition.

The seller problem is a rational inattention (Sims 2003, Sims 2010, Caplin & Dean 2013, Caplin 2016, and Maćkowiak et al. 2018). How an agent incurs attention costs to learn an uncertain state of nature is formally related to how a firm incurs a cost to transform data into insights. Matějka & McKay (2015) and Pomatto et al. (2023) derive first-order characterizations for optimal learning.

The social planner problem of producing a surplus-maximizing posterior distri-

[10] Finitely many values is BBM's main set-up.

[11] See also Roesler & Szentes (2017) for this problem with no learning cost. Mensch & Ravid (2022) and Thereze (2022) also study price discrimination with costly buyer learning.

bution for the firm is an application of costly Bayesian persuasion (Kamenica & Gentzkow 2011 and Gentzkow & Kamenica 2014). The social planner (sender) has misaligned incentives compared with the firm (receiver). We specifically exploit using duality techniques introduced by Dworczak & Kolotilin (2024) and Dworczak & Martini (2019).

We contribute to both literatures in that we extend the characterization of solutions to the case where costs are not necessarily posterior separable, using the concept of Bregman divergence.

The next section presents the model. Section 3 resents the binary value case. Section 4 presents results for the finite value case. Section 5 introduces variational techniques. Section 6 compares social versus private information production under differentiable production costs of information. Section 7 concludes.

2 Model

Players and Market. There is a single seller with no cost to produce the good facing a unit-mass of unit-demand buyers with positive private values $v \in \{v_1, ..., v_N\} = \Omega$. There is a commonly known fully supported prior distribution $\rho_0 \in \Delta(\Omega)$ of buyer types in the market.[12] The v coordinate of vector ρ_0, denoted ρ_0^v, corresponds to the proportion of type v buyers. The buyer value distribution represents market demand; the number of buyers purchasing at price p is given by:

$$D_0(p) = \sum_{v \geq p} \rho_0^v.$$

The best seller profit without price discrimination is:

$$\pi(\rho_0) = \max_{p \in \mathbb{R}} pD_0(p) = \max_{p \in \mathbb{R}} p \sum_{v \geq p} \rho_0^v.$$

[12] If the seller has a per unit cost say κ, then set new values as $\tilde{v} = v - \kappa$ and remove negative $\tilde{v}$ from $\tilde{\Omega}$ (the seller never sets a price below marginal cost in this model) to map back to our setup. If ρ_0 does not have full support, i.e., $\rho_0^{v'} = 0$ for some v', simply remove v' from the set of buyers values Ω.

First Stage. The seller produces a posterior belief distribution of buyers value $\tau \in \Delta(\Delta(\Omega))$ with mean ρ_0 at cost $C(\tau)$. We assume that the cost is continuous and that producing no information costs 0.[13] The next subsection discusses further assumptions and interpretation of the production cost of information.

Think of τ as a segmentation of the initial market ρ_0 that distributes into each segment ρ a mass $\tau(d\rho)$ of buyers. Where $\rho \in \Delta(\Omega)$ is the distribution of buyer types in the market segment and so captures the segment's demand $D_\rho(p) = \sum_{p \leq v} \rho^v$.

The typical example of the students and non-students market segmentation is captured by a distribution τ with binary support. The weak segment ρ_l reflects students' demand (type distribution), where $\tau(d\rho_l)$ is the proportion of students in the population. The strong segment ρ_h reflects the non-students' demand (type distribution) where $\tau(d\rho_h) = 1 - \tau(d\rho_l)$ is the proportion of non-students in the population. Given the overall proportion ρ_0 of buyer types, the segmentation is feasible if and only if $\tau(d\rho_l)\rho_l + \tau(d\rho_h)\rho_h = \rho_0$. This set-up captures all market segmentations, see Bergemann et al. (2015).

Second Stage. For each market segment ρ supported by τ, the seller sets a price $p \in \mathbb{R}$. Buyers in each market segment observe the seller price and decide whether to buy the good.

Payoffs. Each buyer purchases the seller good if their private value is above their market segment price, otherwise trade does not happen and the buyer and seller get 0 payoff. The seller profit from trading with a type v buyer at price p is $u_s(p, v) = p\mathbf{1}_{\{v \geq p\}}$ and the corresponding buyer surplus is $u_b(p, v) = (v - p)\mathbf{1}_{\{v \geq p\}}$. So trading is efficient and generates surplus v.[14]

We focus on profit-maximizing and welfare-maximizing market segmentations net of information production cost given that the seller sets optimal prices in each market segment.

[13]Continuity ensures existence of solutions to the maximization problems. No information is the distribution δ_{ρ_0} that draws the prior with probability 1, so $C(\delta_{\rho_0}) = 0$.

[14]We have normalized Ω to exclude buyers types below marginal cost.

The Seller's Problem. Let π map beliefs into profits under optimal pricing:

$$\pi(\rho) := \max_{p \in \mathbb{R}} p \sum_{p \leq v} \rho^v = \max_{p \in \mathbb{R}} p D_\rho(p).$$

The firm's profit-maximizing production of information problem is:

$$\max_{\tau \in \Delta(\Delta(\Omega))} \int_{\Delta(\Omega)} \pi(\rho)\tau(d\rho) - C(\tau), \text{ s.t. } \int_{\Delta(\Omega)} \rho\tau(d\rho) = \rho_0. \tag{$\mathcal{P}_1$}$$

Profit-maximizing prices are in $\{v_1, ..., v_N\}$.[15] We denote by P_i the set of posteriors where price v_i is profit-maximizing:

$$P_i := \left\{ \rho \in \Delta(\Omega);\ v_i \in \arg\max_p p \sum_{p \leq v} \rho^v \right\}$$

The Planner's Problem. If many prices are optimal at a posterior ρ, we assume the seller sets the lowest one and we denote it by $p(\rho)$.[16] Let w map beliefs into the resulting welfare:

$$w(\rho) = \sum_{v \geq p(\rho)} v\rho^v.$$

The welfare optimal production of information problem is:

$$\max_{\tau \in \Delta(\Delta(\Omega))} \int_{\Delta(\Omega)} w(\rho)\tau(d\rho) - C(\tau), \text{ s.t. } \int_{\Delta(\Omega)} \rho\tau(d\rho) = \rho_0. \tag{$\mathcal{P}_2$}$$

2.1 Rationalized Production Costs

We view the production cost of information as a technical cost to transform data into insights. It accounts for each step of the production process like purchasing, collecting, compiling, storing, cleaning, and analyzing data. It accounts for produc-

[15] A price in between two buyer values $v_i < p < v_{i+1}$ is suboptimal as the price $p' = v_{i+1}$ yields the same amount of sales for a larger mark-up.

[16] We assume the seller (receiver) plays the planner (sender) preferred equilibrium. This is a standard assumption in Bayesian persuasion to ensure the existence of solutions in the planner problem.

tion factors prices such as energy, and computing technology, and wages like data scientists' labor. It accounts for suppliers' prices if part of the production is outsourced.

We assume that the seller rationalizes production: If one information production plan leads to the same price discrimination strategies as another, then the seller produces the cheapest one. We argue below why rationalized production costs are Blackwell monotone, that is non-decreasing in the mean preserving spread (MPS) order.

From the Blackwell theorem, we know that a decision maker's set of strategies expands if and only if it gains Blackwell information.[17] Without information, the seller can only set prices independent of v. If instead it has full information on v, it can perfectly correlate prices with v (first degree price-discriminate buyers), but it can also set prices independent of v or with arbitrary correlation. With partial information on v, it can correlate but not perfectly prices with v.

Let $\tilde{C} : X \to \mathbb{R}_+$ be the technical production cost of information that maps the choice of a production plan $x \in X$ into its costs, and let $\mu : X \to \Delta(\Delta(\Omega))$ maps production plans with the posterior belief distributions they output.

Definition 1. *A rationalized production cost of information τ is the cost of the cheapest production plan that achieves price discrimination strategies feasible under τ:*

$$C(\tau) := \inf_{x \in X} \left\{ \tilde{C}(x);\ \mu(x)\ \mathit{MPS}\ \tau \right\}$$

This definition makes apparent that the rationalized production cost is Blackwell monotone, because the set of production plans that output no less information than τ contains all production plans that output no less information than any MPS of τ.

Lemma 1. *A rationalized production cost of information is Blackwell monotone.*

We therefore view Blackwell monotonicity as a consequence of the rationalization of a technical production cost. This approach captures sellers that are experienced

[17] See Le Cam (1996) or de Oliveira (2018). The distribution τ is a mean preserving spread of a distribution τ' if and only if the seller can replicate all price discrimination strategies feasible under τ' with information τ.

with the data industry, that have experimented with multiple information production strategies, and that intend to minimize their production cost.

We reinforce the previous property by assuming that there are no free information gains.

Assumption 1. The production cost of information is strictly Blackwell monotone: For all τ MPS of τ' such that $\tau \neq \tau'$ then $C(\tau) > C(\tau')$.

We will add more assumptions in the variational approach section of the paper. The next section studies the private and social production of information problem for $N = 2$.

3 Strong Misalignment for the Two Type Case

Consider the case where $\Omega = \{v_l, v_h\}$, with $v_l < v_h$. A belief ρ refers to the probability of the buyer being of a high type $P(v = v_h)$. Accordingly, a posterior belief distribution τ supports a (measurable) subset of $[0, 1]$.

3.1 Private and Social Marginal Values of Information

The value of information measures the gain in payoff for a decision maker of an information τ compared to no information. The value of information for the seller and planner are:

$$\Pi(\tau) = \int_0^1 \pi(\rho)\tau(d\rho) - \pi(\rho_0)$$
$$W(\tau) = \int_0^1 w(\rho)\tau(d\rho) - w(\rho_0)$$

These values are linear (posterior separable) in τ as both decision makers are expected surplus maximizers.

The seller has two dominant pricing strategies. If $\rho \leq \frac{v_l}{v_h}$, the best price v_l to trade with both types leaving the low type without surplus. Otherwise if $\rho \geq \frac{v_l}{v_h}$ the best price is v_h to trade only with the high types leaving them with no surplus. We assume the seller sets a low price at the cutoff $\frac{v_l}{v_h}$. The marginal values of information[18] π and w are:

$$\pi(\rho) = \begin{cases} v_l & \text{if } \rho \leq \frac{v_l}{v_h} \\ \rho v_h & \text{if } \rho \geq \frac{v_l}{v_h} \end{cases}$$

$$w(\rho) = \begin{cases} (1-\rho)v_l + \rho v_h & \text{if } \rho \leq \frac{v_l}{v_h} \\ \rho v_h & \text{if } \rho > \frac{v_l}{v_h} \end{cases}$$

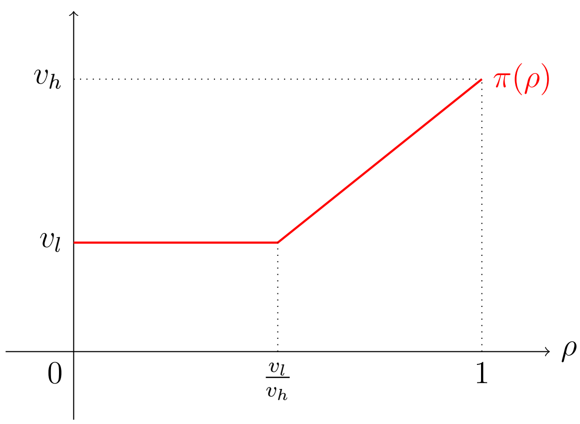

Figure 2: Marginal Profit

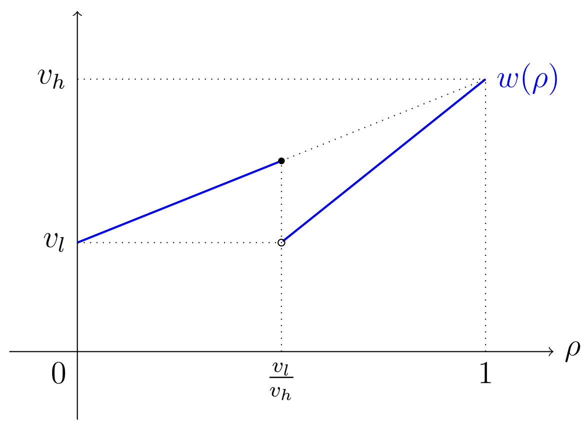

Figure 3: Marginal Welfare

Both profit and welfare are piecewise linear: Information is payoff relevant only if it affects decisions. Spreading or contracting beliefs on an interval where the same price is played has no effect profit or welfare, so π or w are linear on such intervals. Piece-wise linearity hints at the standard property that coarse posterior belief distribution are without loss or optimal with costly information production.

Definition 2. *A posterior belief distribution τ is coarse if it supports one posterior per price played.*

The marginal profit is convex, which stems from the Blackwell theorem, more spread out beliefs yields higher profit. In contrast, the marginal welfare is non-convex, which reflects what we know about the welfare effect of price discrimination: Allowing price discrimination, i.e., the use of information, may or may not increase

[18] We use the term marginal value of information because π and w are the derivatives in τ of the seller and planner values of information.

welfare.

Discontinuity in marginal social value reflects the incentive misalignment between seller and planner.[19] Buyer surplus drops for a small belief change around $\frac{v_l}{v_h}$ that makes the seller switch from a low to a high price. Moreover, it seems suboptimal for the seller to produce $\rho = \frac{v_l}{v_h}$. This posterior makes the seller indifferent between the two prices, but the purpose of information is to help decision making. More formally, the marginal profit forms a kink around $\frac{v_l}{v_h}$ so that any belief spread out of this posterior is valuable. We call this posterior indecisive.

Definition 3. *A posterior ρ is indecisive if multiple prices are optimal at ρ. Otherwise, it is decisive if only one price is optimal.*

3.2 No Information Cost Benchmark

We know that without costs of information the seller acquires full information to perfectly price discriminate buyers. The resulting allocation is efficient, and the seller extracts the entire surplus.

The seller and planner solutions concavify respectively marginal profit and welfare (Kamenica & Gentzkow 2011 and Caplin & Dean 2013).

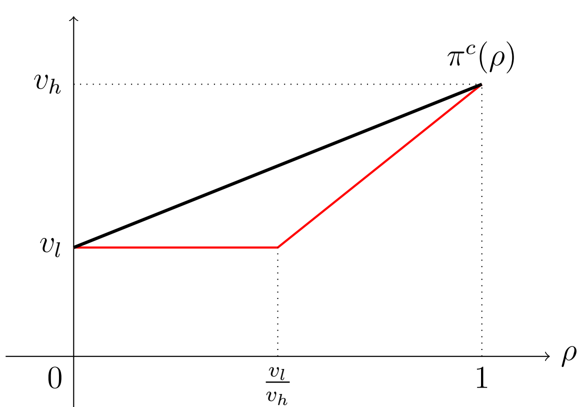

Figure 4: Concavification of Marginal Profit

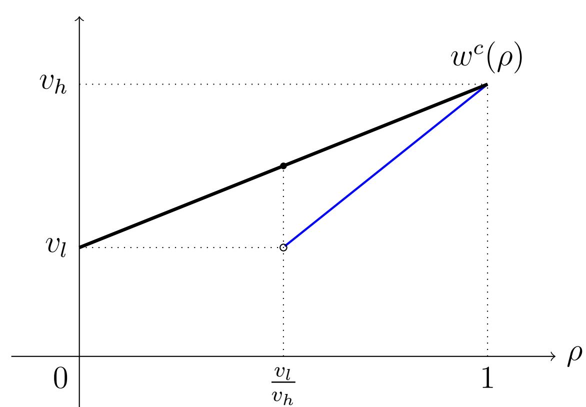

Figure 5: Concavifiction of Marginal Welfare

The concavifications of marginal profit and welfare are equal: $\pi^c = w^c$, reflecting

[19]Upper semi-continuity follows from our assumption that the receiver breaks ties in favor of the sender, ensuring the existence of solutions.

that first degree price discrimination is efficient. The distribution that concavifies profit involves the extreme point beliefs 0 and 1 (because the profit is convex). That is, full information is required to perfectly price discriminate.

The first linear piece of welfare on $[0, \frac{v_l}{v_h}]$ is aligned with the welfare value at $\rho = 1$, meaning that for all $\rho \in [0, \frac{v_l}{v_h}]$, $w(\rho) = w^c(\rho)$. That is, when at the prior the seller sets a low price trade is efficient, no information is needed. Unlike for profit, many distributions concavify w, provided they induce price v_h only at $\rho = 1$.[20]

All in all, there is no distortion: The solution to the seller problem is a solution to the planner problem. Without costs, private information acquisition is efficient. With costly information, however, the only distribution that concavifies profit is also the most expensive to produce, but the planner can pick cheaper distributions to reach efficiency.

3.3 Full Misalignment with Costly Information

Full information is now no longer an efficient solution as it is the costliest to produce. Surprisingly, we show the efficient solution is now buyer optimal and generates no profit to the seller: All the welfare gained is passed to buyers.

We first prove an intuitive lemma that optimal distributions are coarse.

Lemma 2. *Solutions to the seller and planner problem are coarse.*

Proof. Standard result in the literature, see the appendices for completeness. □

Solutions either support only $\{\rho_0\}$ (no information) or have binary support $\{\rho_l, \rho_h\}$ with posterior ρ_l inducing price v_l and posterior ρ_h inducing price v_h. Next, we establish a robust distortion between the supports of the seller and planner solutions.

Proposition 1. *Assume some information is produced at the solution.*

1. *The seller's solution only supports decisive posteriors.*

2. *The planner's solution supports an indecisive low posterior, i.e., $\rho_l = \frac{v_l}{v_h}$.*

[20]These efficient information structures allocate differently the total surplus to the seller and buyers, and form the hypotenuse of the BBM surplus triangle.

Proof. See the appendices. □

If the prior induces a low price, the planner does not produce information as trade is already efficient. If the prior induces a high price, then the cheapest belief spread that induces a low price supports the indecisive posterior $\frac{v_l}{v_h}$. That is, there is no benefit to spread below $\frac{v_l}{v_h}$ as it already induces a low price. Whereas for the seller, producing the indecisive posterior $\frac{v_l}{v_h}$ is never optimal as it makes the seller indifferent between the two prices.

Proposition 1 has strong implications for the allocation of the total surplus at the planner solution.

Proposition 2. *The seller profit under the planner solution stays at the uniform profit level: All the surplus gained from the information produced is passed on to buyers. So that, the welfare-optimal information production coincides with the buyer-optimal one.*

Proof. See the appendices. □

If the prior is in the low price inducing region then no information is produced, trade is efficient and there is no price discrimination. If, instead, the prior is in the high price inducing region, the planner induces low prices by spreading the indecisive posterior. The seller is indifferent at the low posterior and therefore has the same profit as when setting the high price uniformly. In any cases, the seller profit stays at the uniform profit level. Despite the planner maximizing consumer surplus and profit, to economize on information costs it produces a distribution that only benefits the buyer, making the planner's solutions and buyer optimal solutions coincide.

Unlike in the free information benchmark where there is no misalignment between the planner and the seller as the concavification of welfare and profit are equal, the starkest misalignment obtains under costly information. For any strictly Blackwell monotone cost, the efficient solution is buyer optimal.

3.4 Example with Euclidean Costs

Assume an Euclidean production cost of information for a $\kappa > 0$:

$$C(\tau) = \kappa \int_0^1 \|\rho - \rho_0\|^2 \tau(d\rho)$$

Total cost of τ is obtained by integrating how far each posterior is drawn from the prior. This cost function is posterior separable, continuous and no information costs 0. It is also strictly Blackwell monotone, because the marginal cost $\rho \mapsto \kappa\|\rho - \rho_0\|^2$ is strictly convex (see lemma 5).

Let $h_i = \pi$ for $i = 1$ and $h_i = w$ for $i = 2$, the optimal production of information problem rewrites:

$$\max_{\tau \in \Delta[0,1]} \int_0^1 (h_i(\rho) - \kappa\|\rho - \rho_0\|^2) d\tau(\rho), \text{ s.t } \int_0^1 \rho d\tau(\rho) = \rho_0 \qquad (\mathcal{P}_i)$$

The concavification argument applies to the map $h_i(\rho) - \kappa\|\rho - \rho_0\|^2$.

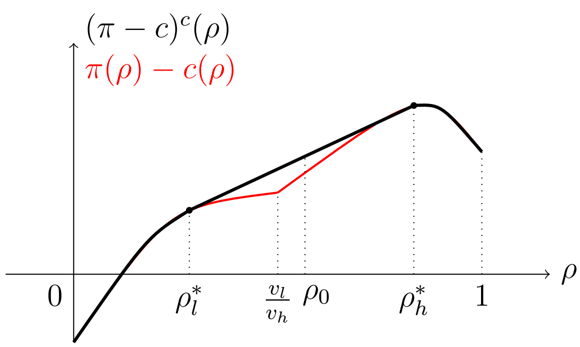

Figure 6: Concavification of Marginal Profit net of Cost

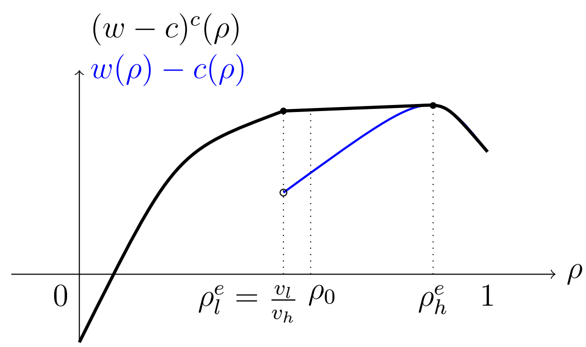

Figure 7: Concavification of Marginal Welfare net of Cost

Figure 6 and 7 show the seller and planner solution for a prior in the high-price region. The difference between the concavified and non-concavified value at the prior is the gain in profit (resp. welfare) net of the production cost.

If $\rho_0 \leq \frac{v_l}{v_h}$, $(w - c)^c(\rho) = w(\rho) - c(\rho)$, that is the planner solution is to produce no information: trade is already efficient in markets where a low price is played. But for some $\rho_0 \leq \frac{v_l}{v_h}$, the seller produces information, it segments the market to mark-up high type buyers in the high segment ρ_h^*. If $\rho_0 > \frac{v_l}{v_h}$, both the seller and the planner produce information. However, the planner stops spreading the low belief at $\frac{v_l}{v_h}$ making the seller indifferent in the low segment (proposition 1).

The Seller Solution. We compute (ρ_l^*, ρ_h^*) using proposition 8 and corollary 1 necessary and sufficient conditions:

$$\begin{cases} v_h - \kappa(\rho_h^* - (1-\rho_h^*)) = -\kappa(\rho_l^* - (1-\rho_l^*)) & \text{(Smooth Pasting)} \\ \rho_h^* v_h - v_l = \kappa(\rho_h^* - \rho_l^*)^2 & \text{(Bdiv)} \\ \tau(d\rho_h^*) = \frac{\rho_0 - \rho_l^*}{\rho_h^* - \rho_l^*} & \text{(Feasibility)} \end{cases}$$

$$\iff \begin{cases} \rho_h^* = \frac{v_l}{v_h} + \frac{v_h}{4\kappa} \\ \rho_l^* = \frac{v_l}{v_h} - \frac{v_h}{4\kappa} \\ \tau(d\rho_h^*) = \frac{1}{2} + \frac{2\kappa}{v_h}(\rho_0 - \frac{v_l}{v_h}) \end{cases}$$

The first condition ensures that marginally spreading or contracting ρ_l^* and ρ_h^* is not profitable. The second condition equates the Bregman divergences of marginal profit and marginal cost. We have interior solutions for $\kappa > \max\left\{\frac{v_h^2}{4(v_h - v_l)}, \frac{v_h^2}{4v_l}\right\}$, otherwise $\rho_l^* = 0$ or $\rho_h^* = 1$. Furthermore, we need $-\frac{v_h}{4\kappa} < \rho_0 - \frac{v_l}{v_h} < \frac{v_h}{4\kappa}$ otherwise no information production is optimal.

The Planner Solution. There is no production of information if $\rho_0 \leq \frac{v_l}{v_h}$, as trade is already efficient. We focus on $\rho_0 > \frac{v_l}{v_h}$. Proposition 1 (or 3 for the general case) implies that ρ_l is indecisive so equal to $\frac{v_l}{v_h}$. If interior, ρ_h satisfies the Bregman divergence condition (proposition 9).

$$\begin{cases} \rho_l^e = \frac{v_l}{v_h} & \text{(Indecisiveness)} \\ (1-\rho_l^e) v_l = \kappa(\rho_l^e - \rho_h^e)^2 & \text{(Bdiv)} \\ \tau(d\rho_h^e) = \frac{\rho_0 - \rho_l^e}{\rho_h^e - \rho_l^e} & \text{(Feasibility)} \end{cases} \iff \begin{cases} \rho_l^e = \frac{v_l}{v_h} \\ \rho_h^e = \frac{v_l}{v_h} + \left((1 - \frac{v_l}{v_h})\frac{v_l}{\kappa}\right)^{\frac{1}{2}} \\ \tau(d\rho_h^e) = \frac{\rho_0 - \frac{v_l}{v_h}}{\left((1-\frac{v_l}{v_h})\frac{v_l}{\kappa}\right)^{\frac{1}{2}}} \end{cases}$$

The second condition equates the Bregman divergences of marginal welfare and marginal cost (proposition 9). The prior must be such that: $0 < \rho_0 - \frac{v_l}{v_h} < \left((1 - \frac{v_l}{v_h})\frac{v_l}{\kappa}\right)^{\frac{1}{2}}$ otherwise the solution is no production. The solution is interior for

$\kappa > \frac{v_l v_h}{v_h - v_l}$, otherwise $\rho_h = 1$.

Comparison. If the prior induces a low price, then no information production is efficient. The seller produces information and therefore reduces welfare for $\rho_0 \in \left[\frac{v_l}{v_h} - \frac{v_h}{4\kappa}, \frac{v_l}{v_h}\right]$. If the prior induces a high price, then producing information increases welfare. It allows for low-type consumers that are excluded under the uniform price, to trade with positive probability. Furthermore, if $\rho_l^* > 0$, buyer surplus increases as some high-type consumers are charged a low price which also implies that welfare increases net of production costs.

4 Extension to Finite Buyer Values

This section studies the general model with $v \in \{v_1, ..., v_N\}$. Lemma 2 directly extends to this case. We take for granted that profit- and welfare-maximizing distributions are coarse.

4.1 Marginal Values of Information

The seller's marginal value of information is the solution of a maximization problem: taking the best decision at posterior ρ. Lemma 3 presents an envelope result:[21]

Lemma 3. *1. π is continuous, convex, and piecewise linear.*

2. The subdifferentials of π are convex hulls of state-vector payoffs evaluated at optimal prices:[22]

$$\partial\pi(\rho) = conv\left\{u_s(v_i, .) \in \mathbb{R}^N;\ \rho \in P_i\right\}$$

Proof. Rochet (1987) proves a related statement in the context of multidimensional screening.[23] See the appendices for completeness. □

[21] Lemma 3 is a multidimensional envelop theorem that cannot be deduced from Milgrom & Segal (2002) theorem.

[22] Where $\mathbf{1}_i$ is a vector with coordinates 1 for coordinates weakly above i and 0 otherwise.

[23] See also Daskalakis et al. (2017). The first claim is routinely applied in the Bayesian persuasion and information theory literatures.

Convexity of the marginal value v captures that spreading beliefs increases profit (the Blackwell theorem). Piece-wise linearity reflects that spreading or contracting posteriors on a subset where the same price is played has no effect on profit. Information matters only if it affects decision-making.

The second part of the proposition is an envelope result: One can ignore the effect of re-adjusting the seller's optimal price when subdifferentiating optimal profit with respect to posteriors. Lemma 3 shows that π is differentiable at ρ if and only if ρ is decisive. At indecisive posteriors π forms a kink.

There is no envelope result for the planner case, as the seller has misaligned incentives. The next lemma collects useful properties of the planner's marginal value of information.

Lemma 4. *The marginal welfare $w(\rho)$ is piece-wise linear and upper semi-continuous. It is differentiable on the interior of P_i, the corresponding gradient is:*

$$w'(\rho) = u_b(v_i, .) + u_s(v_i, .)$$

Proof. See appendices. □

Piece-wise linearity reflects that information matters only if it affects prices. Non-convexity means allowing price discrimination may or may not increase welfare. The marginal value is discontinuous at posteriors where the seller switches prices. Upper semi-continuity follows from the assumption that the seller sets the planner-preferred price if indifferent.

4.2 Planner and Buyer Alignment

This subsection extends proposition 1 to the N buyer value case and corollary 1 for restricted prior distributions. Proposition 2's statement about the seller extends with additional smoothness assumption on the cost (see section 5).

Proposition 3. *At the planner solution, all supported posteriors, except the highest-price-inducing one, are indecisive.*

Proof. See the appendices. □

Consider shifting a low-price-inducing posterior ρ_l towards a higher-price-inducing posterior ρ_h, i.e. $\rho_l' = \rho_l + \varepsilon(\rho_h - \rho_l)$. Maintaining a mean of ρ_0 implies that ρ_l' must be drawn slightly more often than ρ_l and ρ_h slightly less often. If ρ_l is decisive, there is an $\varepsilon > 0$ for which ρ_l' induces the same low price than ρ_l. In that case, this variation induces low prices more often: Welfare goes up. On the other hand, this variation is a mean preserving contraction, and is therefore cheaper to produce. So low-price-inducing posteriors cannot be decisive at the planner solution.

The planner spreads posteriors to induce lower prices, but does not waste resources over-spreading. Consequently, all low-price-inducing posteriors make the seller indifferent between multiple prices. This feature is unappealing for the seller that would produce information to know what price to set. With differentiable costs, proposition 5 shows that all seller-produced posteriors are decisive.

Proposition 3 sheds an interesting light on the third-degree price discrimination literature, in which it is commonly assumed that each segment has a profit maximizing price characterized by a first-order condition. Such assumption excludes market segmentations featuring *indecisive* demands, where many prices are optimal. However, these segmentations have better welfare properties. That is, assuming log-concavity of demand tilts welfare analysis against the use of price discrimination.

In the two type case, this result implies that the planner solution gives all welfare gains to the buyer and nothing to the seller. In the N type case, we show the same implication holds for a specific set of initial markets.

Proposition 4. *Suppose either of these two:*

1. $\rho_0 \in P_1$.
2. $\rho_0 \in P_i \setminus conv\{\rho \in P_j \; ; \; j \neq i\}$ *for* $i > 1$, *and the profit ranking of prices at* ρ_0 *satisfies* $v_1 \succeq v_2 \succeq \ldots \succeq v_{i-1}$.

then the planner solution gives all the welfare gains to the buyers and nothing to the seller.

Proof. See the appendices. □

Figure 8's grey areas are the initial markets where the planner solution gives all the welfare gains to the buyers. In these areas, the planner and buyer solutions

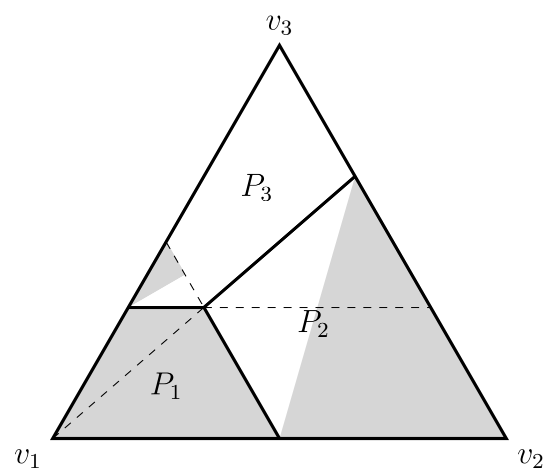

Figure 8: Simplex with $\{v_1, v_2, v_3\} = \{1, 2, 3\}$

coincide. The seller obtains part of the welfare gains at the planner solution if at a low posterior, the seller is not indifferent with the highest induced price. The seller would then be strictly worse-off by setting the same price at both posteriors.

4.3 A Counter Example for N=3

We now present an example to provide intuition for why our main result for binary values does not directly generalize to the finite value case.

The figure below represents the $N = 3$ simplex with $v_1 = 1, v_2 = 2, v_3 = 3$ and a prior $\rho_0 = (\frac{3}{10}, \frac{1}{5}, \frac{1}{2})$. We have $\rho_0 \in P_3$, so the uniform monopoly price is 3. Let the matrix M_L be the projection matrix on the linear subspace L in blue on the figure. So $I_3 - M_L$ is the projection matrix on the corresponding orthogonal space.[24]

Consider the following cost function for $\epsilon, B > 0$:

$$C(\tau) = \int \left(\epsilon(\rho - \rho_0)^T M_L(\rho - \rho_0) + B(\rho - \rho_0)^T (I_3 - M_L)(\rho - \rho_0)\right) d\tau(\rho)$$

Lemma 5 (see next section) ensures C is Blackwell monotone.[25]

For ϵ small and B large, it is cheap to spread beliefs along the linear subspace

[24] Where I_3 is the 3 by 3 identity matrix.

[25] Indeed, M_L and $I_3 - M_L$ are symmetric and semi positive definite as projection matrix. So C's gradient: $\rho \mapsto (\rho - \rho_0)^T (\epsilon M_L + B(I_3 - M_L))(\rho - \rho_0)$ is convex.

L but expensive to spread in other directions. In this environment, the spread (ρ_1^e, ρ_3^e) is efficient and is the planner solution for ϵ sufficiently small and B sufficiently large. However, the monopolist gains profit under this spread compared to the uniform level. Indeed, the belief supporting a low price of 1 is indecisive, but the monopolist is indifferent between price v_1 and v_2 and not with price v_3 the uniform monopoly price.

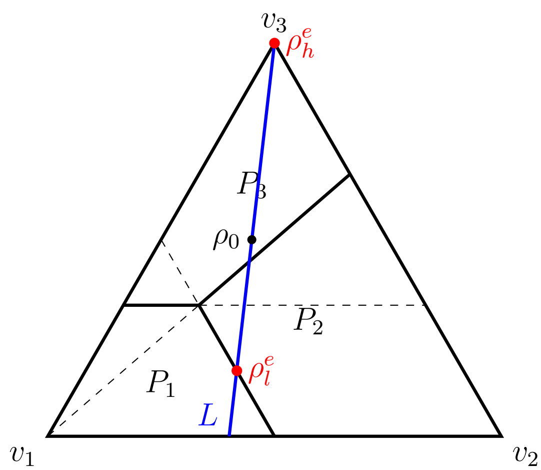

Figure 9: Simplex with $\{v_1, v_2, v_3\} = \{1, 2, 3\}$

The next section serves two purposes. We show that under differentiable costs where spreading is equally costly in all direction, then proposition 4 extends to all starting market ρ_0 for $N = 3$. In other words, raising the seller at the efficient segmentation is justified by asymmetries in the production cost of information. Furthermore, under differentiable costs we show the seller produces only decisive posteriors, exhibiting a robust support distortion.

5 Differentiable and Symmetric Costs

5.1 Marginal Costs of Producing Information

Assumption 2. The cost of producing information is Fréchet differentiable.

The corresponding marginal of C at τ, c_τ, is a continuous map from $\Delta(\Omega)$ to $\mathbb{R}$.[26] The marginal cost of producing information reflects the cheapest shift of production plans to output extra information.

Consider the seller shifting its production from a distribution τ to a distribution $\tau + \epsilon(\tau' - \tau)$ for a small $\epsilon > 0$. The marginal cost c_τ approximates the change in information production cost as follows:[27]

$$C(\tau + \epsilon(\tau' - \tau)) - C(\tau) = \epsilon \int_{\Delta(\Omega)} c_\tau(\rho)(\tau' - \tau)(d\rho) + o(\epsilon).$$

For a posterior ρ, $c_\tau(\rho)$ is per unit of mass production cost change. To compute the total change in production cost, one aggregates (integrates) per posterior $c_\tau(\rho)$ weighted by the size of each change $\epsilon(\tau' - \tau)(d\rho)$. That is, marginal changes in the cost of producing information have a posterior separable interpretation.[28]

Suppose c_τ is convex, then a marginal cost change from τ towards a more informative (a mean preserving spread) τ' is positive. So if the marginal cost is convex at all τ the production cost of information marginally increases in the mean preserving spread order. The next proposition shows that this property is necessary and sufficient for the production cost of information to be Blackwell monotone.

Lemma 5. *Assume C is differentiable.*

1. *C is Blackwell monotone if and only if c_τ is convex at all τ.*

2. *If c_τ is strictly convex at all τ, then C is strictly Blackwell monotone.*

Proof. Ravid et al. (2022) proves 1. in the space of cumulative distributions.[29] See the appendices for completeness. □

[26] See Riesz-Markov representation theorem. The space of continuous functions is endowed with the sup norm $\|.\|_\infty$, and the space of measures with the corresponding dual norm $\|\tau\|_{TV} = \sup_{f, \|f\|_\infty \leq 1} \left\{ \int_{\Delta(\Omega)} f(\rho)\tau(d\rho) \right\}$.

[27] The map $o(\epsilon)$ is such that $\lim_{\epsilon \to 0} \frac{o(\epsilon)}{\epsilon} = 0$.

[28] This is because a) marginal cost changes correspond to changes in the total cost's local linear approximation, and b) linear costs of information are posterior-separable. Therefore, marginal cost changes inherit the posterior separable property.

[29] Ely et al. (2015) and Gentzkow & Kamenica (2014) discuss this claim for linear costs of information.

The first part of the proposition is a first-order characterization of Blackwell monotonicity. So differentiable rationalized production cost of information have convex marginal costs. We assume strict convexity of marginal costs to recover assumption 1:

Assumption 3. The marginal cost of information c_τ is strictly convex for all τ.

We make an additional smoothness assumption:

Assumption 4: For all τ, the marginal cost c_τ is differentiable.

Marginal costs are strictly convex, so almost everywhere differentiable. Assumption 4 removes a few kinks. Differentiability makes c_τ approximately linear around any beliefs, so that the cost of small spreads is vanishingly small.

5.2 Robust Support Distortion

We now extend proposition 1 to the finite buyer value case and establish for all differentiable costs of information a distortion between the seller-produced and the planner-produced market segmentations.

Proposition 5. *Assume some information is produced at the seller and planner solutions.*

1. *All supported profit-maximizing posteriors are decisive.*
2. *(Restatement of Proposition 3) All supported welfare-maximizing posteriors except the one inducing the highest price are indecisive.*

Proof. See appendices □

The seller produces information to improve its decision making, so intuitively it seems suboptimal to produce posteriors where two or more prices are optimal. Formally, this statement hinges on the differentiability of the marginal cost (assumption 4).[30] Differentiability means the marginal cost is approximately linear at all points and therefore small posterior spreads are approximately costless. In contrast, a

[30] If the marginal cost is not differentiable and has kinks, then the same statement holds if the kinks of the marginal values are sharper than the marginal cost's ones.

small posterior spread around an indecisive posterior, i.e. where the marginal value has a kink, yields non-vanishing gains. Spreading out of a kink is always locally profitable net of costs, and so produced posteriors are decisive.

The logic is radically different for the social planner. Indecisive posteriors induces the seller to set prices lower than the prior-induced one using minimal spreads, thereby economizing production costs.

This distortion is robust to regulations that do not ban private information production. Indeed, if a planner provides to the seller a market segmentation that features indecisive posteriors, then the seller will always produce additional information to spread out of indecisive posteriors and make all posteriors decisive.

5.3 Planner and Buyer Alignment

We assume in this subsection that $N = 3$. We further make a symmetry assumption: it is equally costly to spread in all directions at the margin:

Assumption 5. For all τ there is a $\kappa_\tau > 0$ such that:

$$c_\tau(\rho) = \kappa_\tau \|\rho - \rho_0\|^2$$

For instance, the Euclidiean cost of information satisfy this assumption for $\kappa_\tau = \kappa$. A quadratic Euclidean cost of information satisfies this assumption as well:[31]

$$C(\tau) = \frac{1}{2}\left(\kappa \int_{\Delta(\Omega)} \|\rho - \rho_0\|^2 d\tau(\rho)\right)^2, \text{ as marginal costs are:}$$
$$c_\tau(\rho) = \underbrace{\kappa \int_{\Delta(\Omega)} \|z - \rho_0\|^2 d\tau(z)}_{\equiv \kappa_\tau} \|\rho - \rho_0\|^2$$

A cost of information such that for a map f is: $C(\tau) = f\left(\int_{\Delta(\Omega)} \|\rho - \rho_0\|^2 d\tau(\rho)\right)$ satisfies assumption 5. For such costs, we extend our main results for all starting market.

Proposition 6. *Assume that $2v_2 \geq v_3$, then at the planner solution the seller*

[31] Remark that this cost function is not posterior separable.

profit stays at the uniform profit level: All the surplus gained from the information produced is passed on to buyers. So, the welfare-optimal information production coincides with the buyer-optimal one.

Proof. See the appendices. □

The proof shows on a case-by-case basis that segmentations providing the same total welfare but additional profit to the seller are costlier to produce than ones that do not give additional profit.[32]

6 Variational Approach

This section presents the first-order characterizations of profit- and welfare-maximizing segmentations used to derive section 5's results. To make our first-order approach sufficient, we assume global convexity on the total cost:[33]

Assumption 6. The cost of producing information C is convex.

6.1 First-Order Characterizations

Let $h_i = \pi$ for $i = 1$ and $h_i = w$ for $i = 2$, the optimal production of information problem is:

$$\max_{\tau \in \Delta(\Omega)} \int_{\Delta(\Omega)} h_i(\rho) d\tau(\rho) - C(\tau), \text{ s.t } \int_{\Delta(\Omega)} \rho d\tau(\rho) = \rho_0 \qquad (\mathcal{P}_i)$$

The next proposition extends known characterizations of solutions for C linear to the case where the cost is not linear but differentiable (and convex for sufficiency):

[32] Our intuition is that this result extends to any N. However, the number of cases to check explodes, and we did not find an efficient method to replicate the $N = 3$ proof.

[33] Without assumption 6, all following result remain necessary. Ravid et al. (2022) makes the same assumption (in the space of cumulative distributions). The rational inattention literature considers costs that are posterior-separable (linear) and so convex.

Proposition 7. *The following statements are equivalent for $i \in \{1, 2\}$:*

1. *τ is solution to $(\mathcal{P}_i)$.*
2. *τ concavifies $h_i - c_\tau$ at ρ_0.*
3. *τ is feasible and $\exists \lambda$ s.t. supp$\{\tau\} \subset \arg\max_{\rho \in \Delta(\Omega)} \left\{ h_i(\rho) - c_\tau(\rho) - \sum_{v \in \Omega} \lambda^v (\rho^v - \rho_0^v) \right\}$.*

Proof. See the appendices. □

The equivalence between 2. and 3. follows from Dworczak & Kolotilin (2024) results. The equivalence between 1. and 2. generalizes the concavification characterization (Caplin & Dean 2013, Gentzkow & Kamenica 2014) to costs that are not necessarily posterior separable.

To interpret this 3., consider that the seller and planner have a limited posterior budget mass of 1 they can produce. Given that fixed budget, they produce the posteriors that generate the highest value per unit of mass net of the production cost and an opportunity cost captured by λ. The opportunity cost reflects the extra cost of compensating for the mean when producing a posterior.

To be supported (produced), a posterior belief must solve a sub-maximization problem. For $i \in \{1, 2\}$ there is a λ_i such that for all supported ρ:

$$\max_{\rho \in \Delta(\Omega)} h_i(\rho) - c_\tau(\rho) - \sum_{v \in \Omega} \lambda_i^v \rho^v \qquad (SP_i)$$

We study this problem in the next subsection to characterize and compare the seller and planner solutions.

6.2 Characterization with Bregman Divergences

The concept of Bregman divergence central to our characterization of solutions.[34]

Definition 4. *The Bregman divergence of a convex function f captures how fast*

[34] Frankel & Kamenica (2019) uses the Bregman divergence and shows that the value of news pieces that are coupled with measures of uncertainty are Bregman divergences of the latter. In our paper, the Bregman divergence comes up as an optimality condition.

the function curves away from its tangent from ρ_2 to ρ_1:

$$Bf(\rho_1, \rho_2) = f(\rho_1) - f(\rho_2) - f'(\rho_2) \cdot (\rho_1 - \rho_2)$$

Geometrically $Bf(\rho_1, \rho_2)$ is the difference between function values at ρ_1 to its ρ_2 tangent evaluated at ρ_1. Think of the tangent at ρ_1 as a linear benchmark where it is costless to spread belifs around ρ_1. The Bregman divergence measures the cost of spreading beliefs compared to that linear (costless) benchmark.

Profit-Maximizing Information Production. The next proposition shows that supported posteriors solve is an equality between the Bregman divergence of the marginal profit and the marginal cost.

Proposition 8. *A feasible distribution τ is a solution to $(\mathcal{P}_1)$ if and only if*

$$\begin{aligned}
&1.\forall \rho_1, \rho_2 \in supp\{\tau\}: \quad \sum_{v \geq p_1} p_1 \rho_1^v - \sum_{v \geq p_2} p_2 \rho_1^v = Bc_\tau(\rho_1, \rho_2) \\
&2.\forall \rho_3 \in \Delta(\Omega),\ \forall \rho_2 \in supp\{\tau\}: \sum_{v \geq p_3} p_3 \rho_3^v - \sum_{v \geq p_2} p_2 \rho_3^v \leq Bc_\tau(\rho_3, \rho_2)
\end{aligned}$$

Proof. See the appendices. □

Condition 1. is the counterpart of marginal value equals marginal cost in our problem. The value of taking different actions ($a(\rho_1)$ instead of $a(\rho_2)$), that is, the value of spreading beliefs equals the costs of spreading beliefs. The Bregman divergence is the measure of convexity that captures the cost of spreading.

Corollary 1. *Consider τ a solution to $(\mathcal{P}_1)$. Then, for any two posteriors $\rho_1, \rho_2 \in supp\{\tau\}$ inducing prices $p_1 > p_2$ and for all $v \in \Omega$ one has:*

$$\frac{\partial c_\tau}{\partial \rho^v}(\rho_1) - \frac{\partial c_\tau}{\partial \rho^v}(\rho_2) + \frac{\partial c_\tau}{\partial \rho^{\underline{v}}}(\rho_2) - \frac{\partial c_\tau}{\partial \rho^{\underline{v}}}(\rho_1) = p_1 \mathbf{1}_{\{v \geq p_1\}} - p_2 \mathbf{1}_{\{v \geq p_2,\ p_2 \neq \underline{v}\}}$$

Proof. See the appendices. □

The smooth pasting condition means that for all supported posteriors the marginal values net of marginal costs have the same direction on $\Delta(\Omega)$. If smooth pasting fails for two posteriors, then marginally contracting in or spreading out these two posteriors increases marginal value net of marginal cost.

Welfare-Maximizing Information Production. Because the seller indifferent in many segments, welfare is discontinuous on the efficient support, making first-order characterizations tricky. Yet, if the highest-price-inducing posterior is decisive, then welfare is differentiable around this posterior and we obtain a characterization for the planner solution using Bregman divergences.

Proposition 9. *Consider a feasible τ. Assume the highest price inducing posterior ρ_k is decisive then τ is a solution to $(\mathcal{P}_2)$ if and only if*

$$1.\ \forall \rho_1 \in supp\{\tau\} : \sum_{p_1 \leq v < p_k} v\rho_1^v = Bc_\tau(\rho_1, \rho_k)$$

$$2.\ \forall \rho_2 \in \Delta(\Omega) : \sum_{p_2 \leq v < p_k} v\rho_2^v \leq Bc_\tau(\rho_2, \rho_k)$$

Proof. See the appendices. □

Solutions the euclidean example apply proposition 8 and 9 conditions.

7 Conclusion

We question the efficiency of private information production to price discriminate. We show that a privately optimal signal distributes consumers in decisive segment where one price is optimal. There are non-vanishing incentives to acquire more information when a decision-maker is indifferent between multiple actions. In contrast, to economize on production costs, the efficient market segmentation features all but one indecisive segments. For various market conditions and large classes of production costs of information, we show this property implies all welfare gains from efficiently produced market segmentation go to the buyer, and the seller's profit stays at the uniform level.

Unlike the inconclusive price discrimination welfare result under general but exogenous market segmentations, we show stark distortions emerge if one assumes the production of information is endogenous and costly. Our result warns against the strong inefficiencies of privately produced information to price discriminate.

Appendices

Proof of Lemma 2

Suppose not, and consider a solution τ that supports $\rho_1 < \rho_2$ such that $[\rho_1, \rho_2] \subset [0, \frac{v_l}{v_h}]$ or $\subset (\frac{v_l}{v_h}, 1]$. Consider another distribution τ' that contracts posteriors ρ_1 and ρ_2 into $\rho_3 = \frac{1}{\tau(d\rho_1)+\tau(d\rho_2)}\left(\tau(d\rho_1)\rho_1 + \tau(d\rho_2)\rho_2\right)$ drawn with probability $\tau(d\rho_1) + \tau(d\rho_2)$. Distribution τ is a mean preserving spread of τ', and so $C(\tau) > C(\tau')$. However, τ' yields the same welfare and profit as τ as w and π are linear on $[\rho_1, \rho_2]$, so τ cannot be a solution. Therefore, solutions are coarse. $\square$

Proof of Proposition 1

1. Suppose instead the solution supports $\rho_l < \rho_h$ with $\rho_l = \frac{v_l}{v_h}$. Since π is linear on $[\rho_l, \rho_h]$ contracting posteriors strictly reduces cost without reducing profit, a contradiction. Same argument for $\rho_h = \frac{v_l}{v_h}$.

2. First off, if $\rho_0 \leq \frac{v_l}{v_h}$, no information production generates the efficient surplus, therefore this is the solution. Hence, if there is production of information at the solution then $\rho_0 > \frac{v_l}{v_h}$. Suppose instead τ supports $\rho_l < \frac{v_l}{v_h} < \rho_h$. Consider a τ' that supports $\frac{v_l}{v_h}, \rho_h$ instead. Feasibility implies $\tau(d\rho_h) = \frac{\rho_0-\rho_l}{\rho_h-\rho_l}$ and $\tau'(d\rho_h) = \frac{\rho_0-\frac{v_l}{v_h}}{\rho_h-\frac{v_l}{v_h}}$. Consider the following conditional distribution:

$$G(B|\rho) = \begin{cases} \delta_\rho(B) & \text{if } \rho \neq \frac{v_l}{v_h} \\ \frac{\frac{v_l}{v_h}-\rho_l}{\rho_h-\rho_l}\delta_{\rho_l}(B) + \frac{\rho_h-\frac{v_l}{v_h}}{\rho_h-\rho_l}\delta_{\rho_h}(B) & \text{if } \rho = \frac{v_l}{v_h} \end{cases}$$

Remark that for all $\rho \in [0,1]$ $\int_0^1 D(x|\rho)dx = \rho$ and that for all $B \in \mathcal{B}[0,1]$ $\int_B G(x|\rho)\tau'(d\rho) = \tau(B)$. Therefore τ is a dilation of τ' and so τ is a MPS of

τ'.[35] Moreover, the difference in welfare between τ' and τ is:

$$\begin{aligned}&\int_0^1 w(\rho)\tau'(d\rho) - \int_0^1 w(\rho)\tau(d\rho)\\ =&\frac{\rho_h-\rho_0}{\rho_h-\frac{v_l}{v_h}}((1-\frac{v_l}{v_h})v_l+\frac{v_l}{v_h}v_h)+\frac{\rho_0-\frac{v_l}{v_h}}{\rho_h-\frac{v_l}{v_h}}\rho_h v_h-\frac{\rho_h-\rho_0}{\rho_h-\rho_l}((1-\rho_l)v_l+\rho_l v_h)-\frac{\rho_0-\rho_l}{\rho_h-\rho_l}\rho_h v_h\\ =&(\rho_h-\rho_0)\frac{(1-\rho_h)(\frac{v_l}{v_h}-\rho_l)}{(\rho_h-\rho_l)(\rho_h-\frac{v_l}{v_h})}v_l \geq 0\end{aligned}$$

With equality iff $\rho_h = 1$. Therefore τ' generates weakly higher welfare and costs strictly less to produce so τ cannot be a solution. $\square$

Proof of Proposition 2

Remark that the value of information is greater for the planner than for the buyer, that is for all τ we have:

$$\int_0^1 (w(\rho)-\pi(\rho))\tau(d\rho) - (w(\rho_0)-\pi(\rho_0)) \leq \int_0^1 w(\rho)\tau(d\rho) - w(\rho_0)$$

Therefore, if the planner solution is no information production which implies that for all τ the welfare generated is smaller than the production cost:

$$\int_0^1 w(\rho)\tau(d\rho) \leq C(\tau)$$

Then so is the case for buyer surplus: $\int_0^1 (w(\rho)-\pi(\rho))\tau(d\rho)+\pi(\rho_0) \leq C(\tau)$. And so the solution to buyer optimal production of information is no information production.

Assume instead there is information production at the planner solution (implying $\rho_0 > \frac{v_l}{v_h}$). By lemma 2 the support is binary and by proposition 4 the low posterior is $\rho_l = \frac{v_l}{v_h}$. The planner problem boils down to:

$$\max_{\rho_h\in(\rho_0,1]} \frac{\rho_h-\rho_0}{\rho_h-\frac{v_l}{v_h}} w\left(\frac{v_l}{v_h}\right) + \frac{\rho_0-\frac{v_l}{v_h}}{\rho_h-\frac{v_l}{v_h}} w(\rho_h) - C(\tau)$$

[35] See Le Cam (1996).

Because profit is linear on $[\frac{v_l}{v_h}, \rho_h]$, we have that $(1-\tau(d\rho_h))\pi\left(\frac{v_l}{v_h}\right) + \tau(d\rho_h)\pi(\rho_h) = \rho_0$. Therefore the planner problem rewrites:

$$\max_{\rho_h \in (\rho_0,1]} \frac{\rho_h - \rho_0}{\rho_h - \frac{v_l}{v_h}} \left(w\left(\frac{v_l}{v_h}\right) - \pi\left(\frac{v_l}{v_h}\right)\right) + \frac{\rho_0 - \frac{v_l}{v_h}}{\rho_h - \frac{v_l}{v_h}} (w(\rho_h) - \pi(\rho_h)) + \pi(\rho_0) - C(\tau)$$

Which is the buyer optimal information production problem. $\square$

Proof of Lemma 3

1. v is defined on the convex set $\Delta(\Omega)$. The epigraph of π is:

$$\begin{aligned} epi(\pi(\rho)) &= \left\{ (\rho, r) \in \Delta(\Omega) \times \mathbb{R};\ \max_{p\in\Omega} \sum_{i=1}^{N} u_s(p, v_i)\rho^{v_i} \leq r \right\} \\ &= \bigcap_{p\in\Omega} \left\{ (\rho, r) \in \Delta(\Omega) \times \mathbb{R};\ \sum_{i=1}^{N} u_s(p, v_i)\rho^{v_i} \leq r \right\} \end{aligned}$$

$\left\{(\rho, r) \in \Delta(\Omega) \times \mathbb{R};\ \sum_{i=1}^{N} u_s(p, v_i)\rho^{v_i} \leq r\right\}$ is a half space and is closed and convex for all $a \in A$. Intersections of closed and convex sets are closed and convex, therefore π is convex and continuous.

Recall P_i is a finite intersection of halfspaces, so it is convex and closed. Moreover, $P_i \subset \Delta(\Omega)$, so it is compact. In addition, $\cup_{v_i\in\Omega} P_i = \Delta(\Omega)$ since the decision problem has a solution for all ρ, so there is some v_i where $P_i \neq \emptyset$. By construction $\pi(\rho) = \sum_{v\in\Omega} u_s(v_i, v)\rho^v \iff \rho \in P_i$, and π is linear on each non-empty P_i. Therefore π is linear on all (non empty) piece P_i where the same price is optimal.

2. This is a direct application of Danskin's theorem, see Rochet (1987). $\square$

Proof of Lemma 4

For all $\rho \in P_i$, $w(\rho) = \sum_{v\in\Omega} v\mathbf{1}_{\{v_i\leq v\}}\rho^v$ is linear in ρ. Therefore, w linear on P_i and continuous on $int(P_i)$, but discontinuous at the borders of P_i when the seller switches prices. Upper semicontinuity follows from the seller setting the planner-preferred price if indifferent.

Proof of Proposition 3

Assume the planner solution is $\tau \neq \delta_{\rho_0}$, and let $\rho_k, \rho_i \in supp\{\tau\}$ that induces prices $v_k > v_i$ respectively. Construct a distribution of τ' from τ, that transfers $\epsilon > 0$ mass from ρ_k to ρ_i and shifts ρ_i towards ρ_k, $\rho_i' = \rho_i + \gamma(\rho_k - \rho_i)$ to maintain the mean at ρ_0. That is:

$$\tau'(d\rho) = \begin{cases} 0 & \text{if } \rho = \rho_i \\ \tau(d\rho_i) + \epsilon & \text{if } \rho = \rho_i' \\ \tau(d\rho_k) - \epsilon & \text{if } \rho = \rho_k \\ \tau(d\rho) & \text{otherwise} \end{cases}$$

And τ' has mean ρ_0 if:

$$\begin{aligned}
&\int_{\Delta(\Omega)} \rho\tau'(d\rho) = \rho_0 \\
\iff &\int_{\Delta(\Omega)} \rho\tau(d\rho) + \tau(d\rho_i)\gamma(\rho_k - \rho_i) + \epsilon(\rho_i + \gamma(\rho_k - \rho_i) - \rho_k) = \rho_0 \\
\iff &\gamma(\rho_k - \rho_i)(\tau(d\rho_i) + \epsilon) = \epsilon(\rho_k - \rho_i) \\
\iff &\gamma = \frac{\epsilon}{\tau(d\rho_i) + \epsilon}
\end{aligned}$$

We now argue τ MPS τ'. Indeed, define the conditional distribution:

$$G(d\rho|z) = \begin{cases} \delta_z & \text{if } z \neq \rho_i + \frac{\epsilon}{\tau(d\rho_i)+\epsilon}(\rho_k - \rho_i) \\ \frac{\epsilon}{\tau(d\rho_i)+\epsilon}\delta_{\rho_k} + \frac{\tau(d\rho_i)}{\tau(d\rho_i)+\epsilon}\delta_{\rho_i} & \text{if } z = \rho_i + \frac{\epsilon}{\tau(d\rho_i)+\epsilon}(\rho_k - \rho_i) \end{cases}$$

And remark that:

$$\int_{\Delta(\Omega)} \rho G(d\rho|z) = z \quad \forall z$$
$$\tau(d\rho) = \int_{\Delta(\Omega)} G(d\rho|z)\tau'(dz) \quad \forall \rho$$

Therefore, τ is a MPS of τ'. Consequently, $C(\tau') < C(\tau)$.

The variation in welfare from τ to τ' is:

$$\int w(\rho)\tau'(d\rho) - \int w(\rho)\tau(d\rho) = (\tau(d\rho_i)+\epsilon)(w(\rho_i+\gamma(\rho_k-\rho_i))) - \tau(d\rho_i)w(\rho_i) - \epsilon w(\rho_k)$$

Assume that welfare is differentiable at ρ_i in direction $\rho_k - \rho_i$. Then there is a small ϵ such that at $\rho_i + \frac{\epsilon}{\tau(d\rho_i)+\epsilon}(\rho_k - \rho_i)$ induces price v_i. The variation in welfare becomes:

$$(\tau(d\rho_i)+\epsilon)w(\rho_i) + \epsilon \sum_{v \geq v_i} v(\rho_k^v - \rho_i^v) - \tau(d\rho_i)w(\rho_i) - \epsilon \sum_{v \geq v_k} v\rho_k^v = \epsilon \sum_{v_i \leq v < v_k} v\rho_k^v \geq 0$$

Therefore, at all posteriors welfare must not be differentiable towards higher posteriors. Which implies that all posteriors except the highest price-inducing one are indecisive. $\square$

Proof of Proposition 4

We first prove a useful lemma on welfare discontinuities. We introduce the following notation. For a subset of indices $I \subset \{1, ..., N\}$, we denote by ρ_I a posterior where all prices v_i with $i \in I$ are optimal, i.e. $\rho_I \in \cap_{i \in I} P_i$. We further more denote by s the smallest element of I.

Lemma 6. *1. Welfare is differentiable at a posterior ρ_I in direction $\rho - \rho_I$ if and only if p_s is weakly more profitable than p_i for all $i \in I$ at ρ.*

2. Welfare is differentiable at a posterior ρ_I in direction $\rho_I - \rho$ if and only if p_s is weakly less profitable than p_i for all $i \in I$ at ρ.

Proof. 1. Welfare is differentiable in direction $\rho - \rho_I$ at ρ_I if and only if there is an $\epsilon > 0$ such that $\rho_I + \epsilon(\rho - \rho_I) \in P_s$. That is, for some $\epsilon > 0$ and for all j:

$$\sum_{p_s \leq v} p_s[(1-\epsilon)\rho_I^v + \epsilon\rho^v] \geq \sum_{p_j \leq v} p_j[(1-\epsilon)\rho_I^v + \epsilon\rho^v]$$

Because for $j \notin I$ the inequality is strict, such small ϵ exists regardless of ρ. However for $j \in I$ the inequality boils down to: $\sum_{p_s \leq v} p_s\rho^v \geq \sum_{p_j \leq v} p_j\rho^v$ That is we must have p_s more profitable than p_j for all $j \in I$ at ρ.

2. Welfare is differentiable in direction $\rho - \rho_I$ at ρ_I if and only if there is an $\epsilon > 0$ such that $\rho_I + \epsilon(\rho_I - \rho) \in P_s$. For strict inequalities the argument is the same however for $j \in I$ the inequality boils down to:

$$\sum_{p_s \leq v} p_s \rho^v \leq \sum_{p_j \leq v} p_j \rho^v$$

That is we must have p_s weakly less profitable than p_j for all $j \in I$ at ρ. □

We now prove the proposition. First, if $\rho_0 \in P_1$ then trade is efficient without production of information. The seller payoff is the uniform profit.

Under 2., consider the planner solution's support $\{\rho_1, ..., \rho_k\}$ with $\rho_0 \in P_i$.

Claim 1: ρ_k induces the uniform price.
Proof: Suppose ρ_k induces a strictly lower price than ρ_0. Because it is the highest price inducing posteriors that implies all other posteriors induce prices strictly lower than ρ_0. But that imply $\rho_0 \in conv\{P_j; j < i\}$ a contradiction. Suppose, ρ_k induces a strictly higher price than ρ_0. All other posteriors average out at some ρ_{-k} such that ρ_k, ρ_0 and ρ_{-k} are aligned. If a posterior is in P_i, it is indecisive (not the highest one), so $\rho_{-k} \in conv\{P_j;\ j \neq i\}$. Therefore, $\rho_0 \in conv\{P_j;\ j \neq i\}$, a contradiction.

Claim 1 implies $\rho_k \in P_i$, applying lemma 6, implies that all other posteriors are indecisve with v_k, which is the uniform price. Therefore the seller has the same profit as if it were setting v_k in all segments. □

Proof of Lemma 5

1. ($\Leftarrow$) Assume c_τ is convex at all τ. Consider τ' a mean preserving spread of τ, and the smooth curve $\gamma\ :\ \epsilon \in [0,1] \mapsto \epsilon\tau' + (1-\epsilon)\tau$.
For each ϵ, $\frac{\partial \gamma}{\partial \epsilon}$ is the measure $\tau' - \tau$. Applying the gradient theorem:

$$C(\tau') - C(\tau) = \int_0^1 \left(\int_{\Delta(\Omega)} c_{\gamma(\epsilon)}(\rho)\tau'(d\rho) - \int_{\Delta(\Omega)} c_{\gamma(\epsilon)}(\rho)\tau(d\rho) \right) d\epsilon$$

The integrand is non-negative for all ϵ since τ' is a mean preserving spread of τ and $c_{\gamma(\epsilon)}$ is convex for all $\epsilon \in [0,1]$. Therefore $C(\tau') \geq C(\tau)$.

($\Rightarrow$) Assume C is Blackwell monotone. Consider τ' a mean preserving spread of τ. For all $\epsilon \in [0,1]$, the measure $\epsilon\tau' + (1-\epsilon)\tau$ is a mean preserving spread of τ. Indeed, for all convex and continuous map f the quantity:

$$\begin{aligned}&\int_{\Delta(\Omega)} f(\rho)(\epsilon\tau' + (1-\epsilon)\tau)(d\rho) - \int_{\Delta(\Omega)} f(\rho)\tau(d\rho)\\ =&\epsilon\left(\int_{\Delta(\Omega)} f(\rho)\tau'(d\rho) - \int_{\Delta(\Omega)} f(\rho)\tau(d\rho)\right) \geq 0\end{aligned}$$

is positive since τ' is a mean preserving spread of τ.
C is differentiable at τ:[36]

$$C(\epsilon\tau' + (1-\epsilon)\tau) = C(\tau) + \epsilon \int_{\Delta(\Omega)} c_\tau(\rho)(\tau' - \tau)(d\rho) + o(\epsilon)$$

Since $\epsilon\tau' + (1-\epsilon)\tau$ is a mean preserving spread of τ and C is Blackwell increasing then:

$$\begin{aligned}&\epsilon \int_{\Delta(\Omega)} c_\tau(\rho)(\tau' - \tau)(d\rho) + o(\epsilon) = C(\epsilon\tau' + (1-\epsilon)\tau) - C(\tau) \geq 0\\ \implies &\int_{\Delta(\Omega)} c_\tau(\rho)(\tau' - \tau)(d\rho) + \frac{o(\epsilon)}{\epsilon} \geq 0.\end{aligned}$$

This condition holds for all τ' MPS τ and all $\epsilon > 0$. Letting $\epsilon \to 0$:

$$\int_{\Delta(\Omega)} c_\tau(\rho)(\tau' - \tau)(d\rho) \geq 0, \text{ for all } \tau' \text{ mean preserving spreads of } \tau. \tag{MPS}$$

We show the (MPS) condition implies c_τ is convex. Consider for $\epsilon > 0$ and $\rho = \alpha\rho_1 + (1-\alpha)\rho_2$ with $\alpha \in [0,1]$ a τ' such that:

$$\tau' = \tau - \epsilon\left(\delta_\rho + \alpha\delta_{\rho_1} + (1-\alpha)\delta_{\rho_2}\right)$$

[36] Where $o(\epsilon)$ is such that $\lim_{\epsilon\mapsto 0} \frac{o(\epsilon)}{\epsilon} = 0$

τ' spreads mass from ρ to ρ_1 and ρ_2 preserving mean and total mass of 1. Thus τ' is a mean preserving spread of τ. This holds even if $\rho \notin supp\{\tau\}$, since C's differentiability is defined over some open set $U \supset \Delta(\Delta(\Omega))$ with non-empty interior.[37] Therefore, there is a small enough ϵ such that $\tau' \in U$. Applying (MPS) to τ' yields:

$$\alpha c_\tau(\rho_1) + (1-\alpha)c_\tau(\rho_2) \geq c_\tau(\alpha\rho_1 + (1-\alpha)\rho_2)$$

Which holds for all $\rho_1, \rho_2 \in \Delta(\Omega)$ and $\alpha \in [0,1]$. Therefore, c_τ is convex and which concludes the proof of 1.

2. Assume c_τ is strictly convex at all τ. As before, consider τ' a mean preserving spread of $\tau \neq \tau'$, and the smooth curve γ:

$$\gamma \ : \ \epsilon \in [0,1] \mapsto \epsilon\tau' + (1-\epsilon)\tau$$

For each ϵ, $\frac{\partial\gamma}{\partial\epsilon}$ is the measure $\tau' - \tau$. Applying the gradient theorem:

$$C(\tau') - C(\tau) = \int_0^1 \left(\int_{\Delta(\Omega)} c_{\gamma(\epsilon)}(\rho)\tau'(d\rho) - \int_{\Delta(\Omega)} c_{\gamma(\epsilon)}(\rho)\tau(d\rho) \right) d\epsilon$$

By Le Cam (1996), because τ' is a MPS of τ, there exists a markov kernel G with $\int_{\Delta(\Omega)} zG(dz|\rho) = \rho$ for all ρ such that:

$$\forall B \in \mathcal{B}[\Delta(\Omega)], \ \tau'(B) = \int_{\Delta(\Omega)} G(B|\rho)\tau(d\rho)$$

Plugging this formulation in the above condition one has:

$$\begin{aligned} C(\tau') - C(\tau) &= \int_0^1 \left(\int_{\Delta(\Omega)} \int_{\Delta(\Omega)} c_{\gamma(\epsilon)}(z)G(dz|\rho)\tau(d\rho) - \int_{\Delta(\Omega)} c_{\gamma(\epsilon)}(\rho)\tau(d\rho) \right) d\epsilon \\ &= \int_0^1 \int_{\Delta(\Omega)} \int_{\Delta(\Omega)} (c_{\gamma(\epsilon)}(z) - c_{\gamma(\epsilon)}(\rho))G(dz|\rho)\tau(d\rho)d\epsilon \end{aligned}$$

[37] Fréchet differentiability has bite only on full dimensional open sets.

A function c on $\Delta(\Omega)$ is strictly convex iff for all $x_0 \in \Delta(\Omega)$ there exists $u \in \mathbb{R}^n$ s.t:

$$c(z) - c(x_0) > u \cdot (z - x_0) \quad \forall z \in \Delta(\Omega),\ z \neq x_0$$

Since $\tau' \neq \tau$, there is a subset S with $\tau(S) > 0$ for which for all $\rho \in S$, $G(dz|\rho) \neq \delta_\rho(dz)$. For these ρ apply the characterization of strict convexity for $c_{\gamma(\epsilon)}$ at $x_0 = \rho$ and integrate over w.r.t $G(.|\rho)$. That is, for all ρ, $\exists u_\rho$:

$$\begin{aligned} & c_{\gamma(\epsilon)}(z) - c_{\gamma(\epsilon)}(\rho) > u_\rho \cdot (z - x_0) \\ \implies & \int_{\Delta(\Omega)} (c_{\gamma(\epsilon)}(z) - c_{\gamma(\epsilon)}(\rho)) G(dz|\rho) > \int_{\Delta(\Omega)} u_\rho \cdot (z - \rho) G(dz|\rho) = 0 \end{aligned}$$

The last equality follows from the fact linear operators commute and that $\int_{\Delta(\Omega)} zG(dz|\rho) = \rho$. Since it holds for all ρ in S with $\tau(S) > 0$ for all ϵ the integrand is strictly positive:

$$\int_{\Delta(\Omega)} \int_{\Delta(\Omega)} (c_{\gamma(\epsilon)}(z) - c_{\gamma(\epsilon)}(\rho)) G(dz|\rho) \tau(d\rho) > 0$$

Which holds for all $\epsilon \in [0, 1]$, therefore: $C(\tau') > C(\tau)$. □

Proof of Proposition 7

1. $\Rightarrow$ 2. Assume τ is a solution of $(\mathcal{P}_i)$. Consider a feasible τ', an $\epsilon > 0$ and the following quantity:

$$\frac{H_i(\tau + \epsilon(\tau' - \tau)) - H_i(\tau)}{\epsilon} = \int_{\Delta(\Omega)} h_i(\rho)(\tau' - \tau)(d\rho) - \frac{C(\tau + \epsilon(\tau' - \tau)) - C(\tau)}{\epsilon}$$

Because C is differentiable, and τ is a solution one has:

$$\int_{\Delta(\Omega)} (h_i(\rho) - c_\tau)(\tau' - \tau)(d\rho) - o(\epsilon) \leq 0$$

Which holds for all positive ϵ, therefore for all feasible τ':

$$\int_{\Delta(\Omega)} (h_i(\rho) - c_\tau)\tau'(d\rho) \leq \int_{\Delta(\Omega)} (h_i(\rho) - c_\tau)\tau(d\rho)$$

Therefore τ concavifies $h_i - c_\tau$ at ρ_0.

1. $\Leftarrow$ 2. Assume τ concavifies $h_i - c_\tau$ at ρ_0, then for all feasible τ':

$$\int_{\Delta(\Omega)} (h_i(\rho) - c_\tau(\rho))(\tau' - \tau)(d\rho) \leq 0$$

Since H_i is concave (C is convex) then for all feasible τ':

$$H_i(\tau') - H_i(\tau) \leq \int_{\Delta(\Omega)} (h_i(\rho) - c_\tau(\rho))(\tau' - \tau)(d\rho) \leq 0$$

And so τ is solution to $(\mathcal{P}_i)$.

Consequently, solutions of $(\mathcal{P}_i)$ solves the following problem:

$$(h_i - c_\tau)^c(\rho_0) = \sup_{\tau \in \Delta(\Omega)} \left\{ \int_{\Delta(\Omega)} (h_i(\rho) - c_\tau(\rho))\tau(d\rho);\ \int \rho\tau(d\rho) = \rho_0 \right\}$$

To prove 2. $\iff$ 3. we use Dworczak & Kolotilin (2024) that characterizes primal and dual solutions of this problem.[38]

Lemma 7. *Corollary 1, Dworczak & Kolotilin (2022)*
A feasible τ^* and a feasible λ are primal and dual solutions if and only if for all $\rho \in supp\{\tau\}$:

$$h_i(\rho) - c_{\tau^*}(\rho) = \sum_{v \in \Omega} \lambda^v \rho^v$$

Where the set of feasible $\lambda \in \mathbb{R}^N$ is: $F = \left\{ h_i(\rho) - c_\tau(\rho) \leq \sum_{v\in\Omega} \lambda^v \rho^v,\ \forall \rho \in \Delta(\Omega) \right\}$. Equivalently, a feasible τ^* is a solution to $(\mathcal{P}_i)$ if and only if there is a $\lambda \in \mathbb{R}^{|\Omega|}$ such that:

$$\forall \rho \in supp\{\tau^*\}\ ,\ \rho \in \arg\max_{\rho' \in \Delta(\Omega)} \left\{ h_i(\rho) - c_{\tau^*}(\rho) - \sum_{v\in\Omega} \lambda^v \rho^v \right\}.$$

[38] The authors study a more general version in which Ω is not necessarily finite. They argue their result with Ω finite can be recovered from Rockafellar (1970).

Remark that the max is achieved as the objective is upper semicontinuous (lemma 3 and 4) and $\Delta(\Omega)$ is compact. $\square$

Proof of Proposition 5

Consider τ a solution to the seller's problem which supports ρ. Proposition 7 implies ρ is a solution to (SP). Directional derivatives of the objective at ρ are negative. All derivatives in direction $d \in \mathbb{R}^{|\Omega|}$ with $\sum_{v\in\Omega} d^v = 0$ are feasible and exists since π is convex. So for direction d and $-d$ one has:

$$\begin{cases} \pi'(\rho; d) - c'_\tau(\rho) \cdot d \leq \lambda \cdot d \\ \pi'(\rho; -d) + c'_\tau(\rho) \cdot d \leq -\lambda \cdot d \end{cases}$$

Therefore at a solution ρ one has:

$$\begin{aligned} &\pi'(\rho; d) - c'_\tau(\rho) \cdot d \leq \lambda \cdot d \leq -\pi'(\rho; -d) - c'_\tau(\rho) \cdot d \\ \implies &\pi'(\rho; d) \leq -\pi'(\rho; -d) \end{aligned}$$

However because π is convex then $\pi'(\rho; d) \geq -\pi'(\rho; -d)$, So $\pi'(\rho; d) = -\pi'(\rho; -d)$. Because this holds for any d with $\sum_{v\in\Omega} d^v = 0$, ρ is in the relative interior of a linear part of π, so π is differentiable at ρ. $\square$

Proof of Proposition 6

We start with a useful lemma.

Lemma 8. *Assume τ is a solution and consider $\rho_1, \rho_2 \in supp\{\tau\}$ inducing prices $p_1 < p_2$. For all ρ such that w is differentiable in direction $\rho - \rho_1$ at ρ_1 and indirection $\rho_2 - \rho$ at ρ_2 then we have that:*

$$\sum_{p_1 \leq v < p_2} v\rho^v + Bc_\tau(\rho, \rho_1) - Bc_\tau(\rho, \rho_2) \leq 0$$

Proof. Because $\rho_1, \rho_2 \in supp\{\tau\}$ proposition 7 implies there is a λ such that:

$$\begin{aligned} & w(\rho_1) - c_\tau(\rho_1) - \lambda \cdot \rho_1 = w(\rho_2) - c_\tau(\rho_2) - \lambda \cdot \rho_2 \\ \implies & w(\rho_1) - w(\rho) + c_\tau(\rho) - c_\tau(\rho_1) - \lambda \cdot (\rho_1 - \rho) = w(\rho_2) - w(\rho) + c_\tau(\rho) - c_\tau(\rho_2) - \lambda \cdot (\rho_2 - \rho) \end{aligned}$$

Furthermore, as welfare is differentiable in direction $\rho - \rho_1$ at ρ_1 and indirection $\rho_2 - \rho$ at ρ_2 we have that:

$$\begin{aligned} w'(\rho_1) \cdot (\rho_1 - \rho) - c'_\tau(\rho_1) \cdot (\rho_1 - \rho) \geq \lambda \cdot (\rho_1 - \rho) \\ w'(\rho_2) \cdot (\rho_2 - \rho) - c'_\tau(\rho_2) \cdot (\rho_2 - \rho) \leq \lambda \cdot (\rho_2 - \rho) \end{aligned}$$

Combining:

$$\begin{aligned} & w(\rho_1) - w(\rho) + c_\tau(\rho) - c_\tau(\rho_1) - w'(\rho_1) \cdot (\rho_1 - \rho) + c'_\tau(\rho_1) \cdot (\rho_1 - \rho) \\ \leq & w(\rho_2) - w(\rho) + c_\tau(\rho) - c_\tau(\rho_2) - w'(\rho_2) \cdot (\rho_2 - \rho) + c'_\tau(\rho_2) \cdot (\rho_2 - \rho) \\ \iff & - Bw(\rho, \rho_1) + Bc_\tau(\rho, \rho_1) \leq -Bw(\rho, \rho_2) + Bc_\tau(\rho, \rho_2) \\ \iff & \sum_{p_1 \leq v < p_2} v\rho^v + Bc_\tau(\rho, \rho_1) - Bc_\tau(\rho, \rho_2) \leq 0 \end{aligned}$$

□

Low price-inducing posteriors must be indecisive (proposition 3). Seller has a higher profit than uniform level if low prices are not indecisive other higher supported price. With $N = 3$ this happens if (a) $\rho_1 \in P_1 \cap P_2 \setminus P_3$ and $\rho_3 \in P_3 \setminus (P_2 \cup P_1)$, or if (b) $\rho_1 \in P_1 \cap P_3 \setminus P_2$ and $\rho_2 \in P_2 \setminus (P_3 \cup P_1)$. We show that in both cases there is a variation increasing welfare net of costs.

Case a). Assume that $\rho_3, \rho_1 \in supp\{\tau\}$ such that $\rho_3 \in P_3 \setminus (P_2 \cup P_1)$, $\rho_1 \in P_1 \cap P_2 \setminus P_3$. And consider a $\rho_2 \in P_2 \cap P_3$ such that $\rho_1 + \alpha(\rho_3 - \rho_2) \in P_1 \cap P_2$ for some $\alpha > 0$.
Claim 1. $\rho_2^1 = \rho_3^1$.
Proof. $\rho_1 \in P_1 \cap P_2$ so: $v_1 = v_2(1 - \rho_1^1) \iff v_2\rho_1^1 = v_2 - v_1$.

Likewise $\rho_1 + \alpha(\rho_3 - \rho_2) \in P_1 \cap P_2$:

$$\begin{aligned}
& v_1 = v_2(1 - \rho_1^1 + \alpha(\rho_2^1 - \rho_3^1)) \\
\iff & v_2\rho_1^1 + v_2\alpha(\rho_2^1 - \rho_3^1) = v_2 - v_1 \\
\iff & v_2\alpha(\rho_2^1 - \rho_3^1) = 0 \iff \rho_2^1 = \rho_3^1
\end{aligned} \qquad \square$$

Claim 2. $\rho_2 \in \Delta(\Omega)$.
Proof. let $\rho_2^1 = \rho_3^1$, and consider x s.t. $\rho_2^3 = \rho_3^3 - x$ and $\rho_2^2 = \rho_3^2 + x$. Because $\rho_2 \in P_2 \cap P_3$:

$$(\rho_3^3 - x)v_3 = v_2(\rho_3^3 + \rho_3^2) \iff x = \rho_3^3 - \frac{v_2}{v_3}(\rho_3^3 + \rho_3^2) \in (0, \rho_3^3)$$

And so $\rho_2^1 = \rho_3^1$, $\rho_2^2 = (1 - v_2/v_3)(\rho_2^2 + \rho_2^3)$ and $\rho_2^3 = \frac{v_2}{v_3}(\rho_3^3 + \rho_3^2)$. $\square$

Claim 3. $\rho_3, \rho_1 \in supp\{\tau\} \implies v_2\rho_3^2 \leq (c_\tau'(\rho_1) - c_\tau'(\rho_2)) \cdot (\rho_3 - \rho_2) - Bc_\tau(\rho_3, \rho_2)$
Proof. ρ_1, $\rho_1 + \alpha(\rho_3 - \rho_2) \in P_1 \cap P_2$, therefore welfare is differentiable at ρ_1 in direction $\rho_3 - \rho_2$. By proposition 7, for the dual solution λ we have that:

$$(w'(\rho_1) - c_\tau'(\rho_1)) \cdot (\rho_3 - \rho_2) \leq \lambda \cdot (\rho_3 - \rho_2)$$

$\rho_3 \in supp\{\tau\}$ and $\rho_2 \notin supp\{\tau\}$, by proposition 7:

$$\begin{aligned}
& w(\rho_3) - c_\tau(\rho_3) - \lambda \cdot \rho_3 \geq w(\rho_2) - c_\tau(\rho_2) - \lambda \cdot \rho_2 \\
\iff & w(\rho_3) - c_\tau(\rho_3) - (w(\rho_2) - c_\tau(\rho_2)) \geq \lambda \cdot (\rho_3 - \rho_2)
\end{aligned}$$

Combining yields:

$$\begin{aligned}
& w'(\rho_1) \cdot (\rho_3 - \rho_2) - w(\rho_3) + w(\rho_2) \leq c_\tau'(\rho_1) \cdot (\rho_3 - \rho_2) - c_\tau(\rho_3) + c_\tau(\rho_2) \\
\iff & v_2\rho_3^2 + v_1(\rho_3^1 - \rho_2^1)\rho_2^3 \leq (c_\tau'(\rho_1) - c_\tau'(\rho_2)) \cdot (\rho_3 - \rho_2) - c_\tau(\rho_3) + c_\tau(\rho_2) + c_\tau'(\rho_2) \cdot (\rho_3 - \rho_2) \\
\iff & v_2\rho_3^2 \leq (c_\tau'(\rho_1) - c_\tau'(\rho_2)) \cdot (\rho_3 - \rho_2) - Bc_\tau(\rho_3, \rho_2)
\end{aligned} \qquad \square$$

Also, $v_2\rho_2^2 \leq Bc_\tau(\rho_2, \rho_3) \implies$, using above:

$$\begin{aligned}
&v_2\rho_3^2 \leq c'_\tau(\rho_1) \cdot (\rho_3 - \rho_2) - c_\tau(\rho_3) + c_\tau(\rho_2) \\
\iff &v_2\rho_3^2 \leq (c'_\tau(\rho_1) - c'_\tau(\rho_3)) \cdot (\rho_3 - \rho_2) + Bc_\tau(\rho_2, \rho_3)
\end{aligned}$$

Claim 4 Assume that $C_\tau(\rho) = \kappa_\tau \|\rho - \rho_0\|^2$, with $\kappa_\tau > 0$. Then

$$(c'_\tau(\rho_1) - c'_\tau(\rho_2)) \cdot (\rho_3 - \rho_2) - Bc_\tau(\rho_3, \rho_2) < 0$$

Proof. We have that $Bc_\tau(\rho_3, \rho_2) = \kappa_\tau \|\rho_3 - \rho_2\|^2$ and that $C'_\tau(\rho) = 2\kappa_\tau(\rho - \rho_0)$. So:

$$\begin{aligned}
&2\kappa_\tau(\rho_1 - \rho_2) \cdot (\rho_3 - \rho_2) - \kappa_\tau(\rho_3 - \rho_2) \cdot (\rho_3 - \rho_2) \\
\propto &\left(\rho_1 - \frac{\rho_2 + \rho_3}{2}\right) \cdot (\rho_3 - \rho_2) \\
= &(\rho_1^3 - \rho_3^3 + \frac{x}{2})x - (\rho_1^2 - \rho_3^2 - \frac{x}{2})x \\
\propto &\rho_1^3 - \rho_1^2 + \rho_3^2 - \rho_3^3 + x = \rho_1^3 - \rho_1^2 + \rho_3^2 - \frac{v_2}{v_3}(\rho_3^3 + \rho_3^2)
\end{aligned}$$

We know price v_2 is strictly better than price v_3 at ρ_1. Therefore:

$$\rho_1^3 - \rho_1^2 < \rho_1^3(1 - \frac{v_3 - v_2}{v_2}) = \rho_1^3 \frac{2v_2 - v_3}{v_2}$$

Furthermore, because price v_3 is strictly than v_2 at ρ_3:

$$\rho_3^2 \frac{v_3 - v_2}{v_3} - \frac{v_2}{v_3}\rho_3^3 < \rho_3^3 \left(\frac{v_3 - v_2}{v_2}\frac{v_3 - v_2}{v_3} - \frac{v_2}{v_3}\right) = \rho_3^3 \frac{v_3^2 - 2v_2v_3}{v_2v_3} = \rho_3^3 \frac{v_3 - 2v_2}{v_2}$$

Because price v_1 strictly worse than price v_2 at ρ_3, then one has $\rho_1^3 < \frac{v_1}{v_3} < \rho_3^3$. Therefore: $(\rho_1^3 - \rho_3^3)\frac{2v_2 - v_3}{v_2} < 0 \iff 2v_2 > v_3$. □

That is, whenever $2v_2 > v_3$, the variation saves cost, and, therefore, the planner solution gives no profit to the seller.

Case b). $\rho_1 \in P_1 \cap P_3 \setminus P_2$ and $\rho_2 \in P_2 \setminus (P_3 \cup P_1)$. $\rho_1 \in P_1 \cap P_3 \setminus P_2$ implies that $\rho_1^3 = \frac{v_1}{v_3}$ and that $(1 - \rho_1^1) < \rho_1^3 \frac{v_3}{v_2} = \frac{v_1}{v_2}$. Using lemma 8 on $\rho = (0, 1 - \frac{v_1}{v_3}, \frac{v_1}{v_3})$, it must be that:

$$Bc_\tau(\rho,\rho_1) \leq Bc_\tau(\rho,\rho_2) \iff \|\rho-\rho_1\|^2 \leq \|\rho-\rho_2\|^2$$

We will show that $\|\rho-\rho_1\|^2 > \|\rho-\rho_2\|^2$ for all feasible ρ_1,ρ_2. $\|\rho-\rho_1\|^2$ can be arbitrarily close to $\|\rho-\tilde{\rho}\|^2$ where $\tilde{\rho} = (1-\frac{v_1}{v_2}, \frac{v_1}{v_2}-\frac{v_1}{v_3}, \frac{v_1}{v_3})$ is the posterior that makes the seller indifferent between all three prices.

We show that $\max_{\rho_2}\|\rho-\rho_2\|^2 = \|\rho-\tilde{\rho}\|^2$ for $\rho_2 \in P_2$ and for which price v_3 is preferred to v_1. This problem maximizes a convex function over a polytope. Therefore, the maximum is achieved at one of the corners of the polytope. It has three corners: $\rho, \tilde{\rho}$ and $\rho_{2,3} = (0, 1-\frac{v_2}{v_3}, \frac{v_2}{v_3})$. We compare the distance of our two candidates $\tilde{\rho}$.

$$\|\rho-\rho_{2,3}\|^2 = (\frac{v_2}{v_3}-\frac{v_1}{v_3})^2 + (\frac{v_1}{v_3}-\frac{v_2}{v_3})^2 = 2(\frac{v_2-v_1}{v_3})^2$$
$$\|\rho-\tilde{\rho}\|^2 = \left(1-\frac{v_1}{v_2}\right)^2 + \left(1-\frac{v_1}{v_3}-\frac{v_1}{v_2}+\frac{v_1}{v_3}\right)^2 = 2\left(\frac{v_2-v_1}{v_2}\right)^2$$

Because $v_2 < v_3$, $\|\rho-\rho_{2,3}\|^2 < \|\rho-\tilde{\rho}\|^2$. Hence, $\tilde{\rho}$ is the maximizer. □

Proof of Proposition 8

($\Rightarrow$) If ρ is solution to (SP), proposition 7 shows π is differentiable at ρ, and by optimality feasible directional derivatives are negative:

$$(\pi'(\rho) - c_\tau'(\rho)) \cdot d \leq \lambda \cdot d$$

Because d and $-d$ are feasible directions one has:

$$\forall\ d \text{ s.t. } \sum_v d^v = 0,\ (\pi'(\rho) - c_\tau'(\rho)) \cdot d = \lambda \cdot d$$

Now pick ρ that solves (SP). For all $\rho' \in \Delta(\Omega)$ (remark $\sum_{v\in\Omega}(\rho'^v - \rho^v) = 0$):

$$\begin{aligned}
&\pi(\rho) - c_\tau(\rho) - \lambda \cdot \rho \geq \pi(\rho') - c_\tau(\rho') - \lambda \cdot \rho' \\
\Longrightarrow\ &\pi(\rho) - c_\tau(\rho) \geq \pi(\rho') - c_\tau(\rho') - (\pi'(\rho) - c'_\tau(\rho)) \cdot (\rho' - \rho) \\
\Longrightarrow\ &c_\tau(\rho') - c_\tau(\rho) - c'_\tau(\rho) \cdot (\rho' - \rho) \geq \pi(\rho') - \pi(\rho) - \pi'(\rho) \cdot (\rho' - \rho) \\
\Longleftrightarrow\ &Bc_\tau(\rho', \rho) \geq B\pi(\rho', \rho)
\end{aligned}$$

And if ρ_1 and ρ_2 both solves (SP) then:

$$\pi(\rho_1) - c_\tau(\rho_1) - \lambda \cdot \rho_1 = \pi(\rho_2) - c_\tau(\rho_2) - \lambda \cdot \rho_2 \iff Bc_\tau(\rho_2, \rho_1) = B\pi(\rho_2, \rho_1)$$

Using lemma 3 then $B\pi(\rho', \rho) = \sum_{v\in\Omega} \left(u_s(a(\rho'), v) - u_s(a(\rho), v)\right) \rho'^v$. In particular for ρ and ρ' both solutions:

$$\sum_{v\in\Omega} \left(u_s(a(\rho'), v) - u_s(a(\rho), v)\right) \rho'^v = Bc_\tau(\rho', \rho)$$

($\Leftarrow$) Consider a feasible distribution τ such $\forall \rho \in supp\{\tau\}$ and $\forall \rho' \in \Delta(\Omega)$:

$$B\pi(\rho', \rho) \leq Bc_\tau(\rho', \rho)$$

And $\forall \rho_1, \rho_2 \in supp\{\tau\}$: $B\pi(\rho_2, \rho_1) = Bc_\tau(\rho_2, \rho_1)$. Then, for a $\rho_1 \in supp\{\tau\}$:

$$\begin{aligned}
&\rho_1 \in \arg\max_{\rho' \in \Delta(\Omega)} \{B\pi(\rho', \rho_1) - Bc_\tau(\rho', \rho_1)\} \\
\Longleftrightarrow\ &\rho_1 \in \arg\max_{\rho' \in \Delta(\Omega)} \{\pi(\rho') - c_\tau(\rho') - (\pi'(\rho_1) - c'_\tau(\rho_1)) \cdot \rho'\}
\end{aligned}$$

Because at $\rho' = \rho_1$ the upperbound 0 of the problem is achieved. The equivalence follows from constant terms not affecting the argmax. Also any $\rho_2 \in supp\{\tau\}$ solves this problem. So picking $\lambda = \pi'(\rho_1) - c'_\tau(\rho_1)$ yields proposition 7 statement. $\square$

Proof of Corollary 1

Consider a solution τ. Proposition 8 is equivalent to:

$$\forall \rho \in supp\{\tau\},\ supp\{\tau\} \subset \arg\max_{\rho' \in \Delta(\Omega)} \{B\pi(\rho', \rho) - Bc_\tau(\rho', \rho)\}$$

For any two supported posteriors ρ_1, ρ_2, the first order condition implies that for all d such that $\sum_v d^v = 0$:

$$(\pi'(\rho_1) - c'_\tau(\rho_1)) \cdot d = (\pi'(\rho_2) - c'_\tau(\rho_2)) \cdot d$$

Or equivalently, $\forall \rho_1, \rho_2 \in supp\{\tau\}$, there is a $\varphi_{1,2} \in \mathbb{R}$ such that:

$$\pi'(\rho_1) - c'_\tau(\rho_1) = \pi'(\rho_2) - c'_\tau(\rho_2) + \varphi_{1,2} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

So gradients of marginal value minus marginal costs at supported posteriors are equal up to a translation.

To simplify the notations, let supported posteriors ρ_1, ρ_2 lead to prices p_1, p_2 and let $\underline{v}$ be the lowest state in Ω. Consider the case where $p_1 > p_2$ remark that:

$$\frac{\partial \pi}{\partial \rho^v}(\rho_1) - \frac{\partial \pi}{\partial \rho^{\underline{v}}}(\rho_1) = p_1 \mathbf{1}_{\{v \geq p_1\}}$$
$$\frac{\partial \pi}{\partial \rho^v}(\rho_2) - \frac{\partial \pi}{\partial \rho^{\underline{v}}}(\rho_2) = p_2 \mathbf{1}_{\{v \geq p_2,\ p_2 \neq \underline{v}\}}$$

Using the equality of gradients up to a translation one has that for all supported posteriors ρ_1 and ρ_2 inducing prices $p_1 > p_2$ and for all $v \in \Omega$:

$$\frac{\partial c_\tau}{\partial \rho^v}(\rho_1) - \frac{\partial c_\tau}{\partial \rho^v}(\rho_2) - \frac{\partial c_\tau}{\partial \rho^{\underline{v}}}(\rho_1) + \frac{\partial c_\tau}{\partial \rho^{\underline{v}}}(\rho_2) = p_1 \mathbf{1}_{\{v \geq p_1\}} - p_2 \mathbf{1}_{\{v \geq p_2,\ p_2 \neq \underline{v}\}} \qquad \square$$

Proof of Proposition 9

($\Rightarrow$) Assume τ is a solution to ($\mathcal{P}_2$), so that each $\rho \in supp\{\tau\}$ solves (SP2). In particular, for ρ_k the objective of (SP2) is differentiable and so $\forall d$ such that $\sum_{v \in \Omega} d^v = 0$:

$(w'(\rho) - c'_\tau(\rho)) \cdot d = \sum_{v \in \Omega} \lambda^v d^v$. Hence, $\forall \rho_j \in \Delta(\Omega)$:

$$
\begin{aligned}
& w(\rho_j) - c_\tau(\rho_j) - \lambda \cdot (\rho_j - \rho_k) \leq w(\rho_k) - c_\tau(\rho_k) \\
\Longrightarrow\; & w(\rho_j) - w(\rho_k) - w'(\rho_k) \cdot (\rho_j - \rho_k) \leq c_\tau(\rho_j) - c_\tau(\rho_k) - c'_\tau(\rho_k) \cdot (\rho_j - \rho_k) \\
\Longleftrightarrow\; & \sum_{p_j \leq v < p_k} v \rho_j^v \leq B c_\tau(\rho_j, \rho_k)
\end{aligned}
$$

And with equality if $\rho_j \in supp\{\tau\}$ as it solves (SP2) as well.
($\Leftarrow$) Assume τ is such that:

1. $\forall \rho_1 \in supp\{\tau\} : \displaystyle\sum_{p_1 \leq v < p_k} v \rho_1^v = B c_\tau(\rho_1, \rho_k)$
2. $\forall \rho_2 \in \Delta(\Omega) : \displaystyle\sum_{p_2 \leq v < p_k} v \rho_2^v \leq B c_\tau(\rho_2, \rho_k)$

Therefore we have:

$$
\begin{aligned}
supp\{\tau\} \subset & \arg\max_{\rho_i \in \Delta(\Omega)} \left\{ \sum_{p_i \leq v < p_k} v \rho_i^v - B c_\tau(\rho_i, \rho_k) \right\} \\
= & \arg\max_{\rho \in \Delta(\Omega)} \{ w(\rho) - w(\rho_k) - w'(\rho_k) \cdot (\rho - \rho_k) - c_\tau(\rho) + c_\tau(\rho_k) + c'_\tau(\rho_k) \cdot (\rho - \rho_k) \} \\
= & \arg\max_{\rho \in \Delta(\Omega)} \{ w(\rho) - c_\tau(\rho) - (w'(\rho_k) - c'_\tau(\rho_k)) \cdot \rho + \text{ constant terms } \}
\end{aligned}
$$

Therefore, all $\rho \in supp\{\tau\}$ solve SP_2 for $\lambda = w'(\rho_k) - c'_\tau(\rho_k)$, and so τ solves P_2. $\square$

.1 Working case of Entropy mutual information

$$
\begin{aligned}
C(\tau) &= \int_\Delta H(\rho_0) - H(\rho) d\tau(\rho) \\
&= \int_{\Delta(\Omega)} \sum_{i=1}^{n} (\rho^i \log(\rho^i) - \rho_0^i \log(\rho_0^i)) d\tau(\rho)
\end{aligned}
$$

$$B_{-H}(\rho_1, \rho_2) = D_{KL}(\rho_1 \| \rho_2) = \sum_i \rho_1^i \log \left(\frac{\rho_1^i}{\rho_2^i} \right)$$

References

Ali, S. N., Lewis, G. & Vasserman, S. (2020), Voluntary disclosure and personalized pricing, *in* 'Proceedings of the 21st ACM Conference on Economics and Computation', pp. 537–538.

Barreto, D. M., Ghersengorin, A. & Augias, V. (2022), 'Price discrimination with redistributive concerns'.

Bergemann, D., Brooks, B. & Morris, S. (2015), 'The limits of price discrimination', *American Economic Review* **105**(3), 921–57.

Caplin, A. (2016), 'Measuring and modeling attention', *Annual Review of Economics* **8**, 379–403.

Caplin, A. & Dean, M. (2013), Behavioral implications of rational inattention with shannon entropy, Technical report, National Bureau of Economic Research.

Daskalakis, C., Deckelbaum, A. & Tzamos, C. (2017), 'Strong duality for a multiple-good monopolist', *Econometrica* **85**(3), 735–767.

De Cornière, A., Mantovani, A. & Shekhar, S. (2024), 'Third-degree price discrimination in two-sided markets', *Management Science* .

de Oliveira, H. (2018), 'Blackwell's informativeness theorem using diagrams', *Games and Economic Behavior* **109**, 126–131.

Devine, B. R. & Munoz-Garcia, F. (2018), Nonlinear pricing with costly information acquisition, Technical report, Working Paper.

Dworczak, P. & Kolotilin, A. (2024), 'The persuasion duality', *Theoretical Economics* **19**(4), 1701–1755.

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

Elliott, M., Galeotti, A., Koh, A. & Li, W. (2022), 'Matching and information design in marketplaces', *Available at SSRN* .

Ely, J., Frankel, A. & Kamenica, E. (2015), 'Suspense and surprise', *Journal of Political Economy* **123**(1), 215–260.

Frankel, A. & Kamenica, E. (2019), 'Quantifying information and uncertainty', *American Economic Review* **109**(10), 3650–80.

Gentzkow, M. & Kamenica, E. (2014), 'Costly persuasion', *American Economic Review* **104**(5), 457–62.

Haghpanah, N. & Siegel, R. (2022), 'The limits of multiproduct price discrimination', *American Economic Review: Insights* **4**(4), 443–58.

Hidir, S. & Vellodi, N. (2021), 'Privacy, personalization, and price discrimination', *Journal of the European Economic Association* **19**(2), 1342–1363.

Ichihashi, S. & Smolin, A. (2022), 'Data collection by an informed seller', *arXiv preprint arXiv:2204.08723* .

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

Le Cam, L. (1996), 'Comparison of experiments: A short review', *Lecture Notes-Monograph Series* pp. 127–138.

Maćkowiak, B., Matějka, F. & Wiederholt, M. (2018), 'Survey: Rational inattention, a disciplined behavioral model', *Available at SSRN 3266436* .

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

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

Milgrom, P. & Segal, I. (2002), 'Envelope theorems for arbitrary choice sets', *Econometrica* **70**(2), 583–601.

Pomatto, L., Strack, P. & Tamuz, O. (2023), 'The cost of information: The case of constant marginal costs', *American Economic Review* **113**(5), 1360–1393.

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

Rochet, J.-C. (1987), 'A necessary and sufficient condition for rationalizability in a quasi-linear context', *Journal of mathematical Economics* **16**(2), 191–200.

Rockafellar, R. T. (1970), *Convex analysis*, Vol. 18, Princeton university press.

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

Sims, C. A. (2003), 'Implications of rational inattention', *Journal of monetary Economics* **50**(3), 665–690.

Sims, C. A. (2010), Rational inattention and monetary economics, *in* 'Handbook of monetary economics', Vol. 3, Elsevier, pp. 155–181.

Tekdir, I. (2024), 'Third degree price discrimination under costly information acquisition', *arXiv preprint arXiv:2406.06026* .

Thereze, J. (2022), Screening costly information, Technical report, mimeo, Princeton.