VERY CREDIBLE AUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Online platforms routinely ask users to reveal private information to improve "targeting": which product to recommend, which advertisement to show, or which offer to display. At the same time, revealed data can be used for more aggressive price discrimination, leaving users worse off. This paper studies a minimal disclosure-and-pricing game in which *neither side can commit*: the user cannot commit to a willingness-to-pay claim and the platform cannot commit to a pre-announced pricing rule. We call mechanisms implementable under such constraints *very credible*. We first show a one-dimensional impossibility: with a single product and a single buyer, any very credible mechanism is outcome-equivalent to posted monopoly pricing and uninformative communication. We then show that multidimensional preferences create room for positive-sum disclosure: with multiple products but limited attention (the platform can display only one offer), users can truthfully reveal *which* product they prefer without enabling stronger price discrimination. A simple two-product example yields strictly higher seller revenue and strictly higher buyer surplus relative to no disclosure.

1 INTRODUCTION AND MOTIVATION

A conventional wisdom is that one should not reveal private information to an online platform (as much). For example, this can come from the uncertainty or ambiguity on what the data will be used in the future: Will it be used for training a better predictive model? Will it be accessible to others to whom we do not wish to share? Will it be prone to some data attack due to security flaws?

For a more economics example, we can consider the case where the platform is trying to sell a product to us as well as possibly personalized price. If it only knows that our valuation $V \sim \text{Unif}([0, 1])$, then the price selected by the platform, which is the monopoly price, will be $\frac{1}{2}$, and the expected payoff for us is $\frac{1}{4}$. If the platform knows the realization of our private valuation, then the platform can conveniently set the personalized price to be just below (or at) V, and extracting all the surplus from us. The expected payoff for us can be infinitesimally small.

However, in marketing, the effect of data disclosure can be the opposite (Neumann et al., 2024). Its rationale is that more information will let the platform knows more about you, so the platform can be better at its recommendation. This phenomenon is called "targeting".

The two effects of adversarial pricing and targeting make it difficult to analyze the welfare effect of information disclosure. In a general case, where the information is observed directly by the platform, this will highly depend the discernment power as well as accuracy of its data processing. A better information often leads to higher total surplus, but we cannot predict the platform surplus gain before observing such data processing structure.

This also prompts the movement for the user to provide information voluntarily instead. If the platform were trusted by the user to use such information only for targeting, then the user will be willing to provide an accurate fine-grained data. The question is then rest on the platform that are not trusted. In this paper, we will study the scenario where voluntary report is still preferable.

From credible to very credible. The mechanism design literature typically assumes an auctioneer or platform can commit to a rule of the game. Credible auctions (Akbarpour & Li, 2020) study the more realistic possibility that the seller cannot commit, especially when bids are not verifiable by bidders. The key insight is that some auction formats (e.g., first-price auctions without reserve) align the seller's incentives with the announced rule.

In many consumer-platform settings the commitment problem is more severe:

- The platform cannot credibly commit to a pricing policy after it learns user information.
- The user cannot credibly commit to a willingness-to-pay cap after seeing an offer.

In such environments, "mechanisms" are effectively whatever can arise from strategic interaction in real time. This motivates our central object: *very credible mechanisms*—outcomes implementable without commitment on either side.

Main message. In one-dimensional environments (a single good), very credible mechanisms have little bite: voluntary disclosure collapses to monopoly pricing. In multidimensional environments (multiple goods, limited attention), users can sometimes disclose *which* good they prefer in a way that benefits both sides, because this information improves targeting without tightening the platform's inference about willingness to pay for the chosen good.

2 Model: Disclosure and Pricing without Commitment

A platform (seller) can present at most one offer at a time. A customer (buyer) has private values over a set of products.

2.1 Environment

Let $I = \{1, \dots, k\}$ index products. The buyer's value vector is $V = (V_1, \dots, V_k) \in \mathcal{V} \subseteq \mathbb{R}_+^k$ drawn from a common prior $\Pi \in (\mathcal{V})$. The platform observes Π but not the realization V.

The interaction is a three-stage game:

1. **Message.** The buyer sends a (cheap-talk) message $m \in \mathcal{M}$.
2. **Offer.** After observing m, the platform chooses an offer $(i, p) \in I \times \mathbb{R}_+$, or chooses to make no offer (denoted).
3. **Acceptance.** After observing the offer (i, p), the buyer accepts or rejects. If it accepts, trade occurs and payoffs are

$$u_B = V_i - p, \qquad u_S = p.$$

   If it rejects or the platform makes no offer, payoffs are 0.

2.2 Very credible mechanisms

The "mechanism" here is not a commitment device; it is a *self-enforcing* pattern of communication and offers.

Definition 1 (Very credible mechanism)**.** Fix an environment (I, Π). A *very credible mechanism* is an outcome implemented by a sequential equilibrium of the disclosure-and-pricing game described above. Equivalently, it is a tuple (μ, σ, a) consisting of

- a buyer messaging strategy $\mu : \mathcal{V} \to (\mathcal{M})$,
- a platform offer strategy $\sigma : \mathcal{M} \to ((I \times \mathbb{R}_+) \cup \{\})$,
- a buyer acceptance rule $a : \mathcal{V} \times (I \times \mathbb{R}_+) \to \{0, 1\}$,

together with beliefs, such that (μ, σ, a) is sequentially rational and beliefs are consistent (in the standard sense of sequential equilibrium).

Remark 1. The definition intentionally imposes *no* commitment. The buyer's message is non-binding cheap talk. The platform can deviate after seeing the message and choose any offer. Thus, "design" in this paper is closer to identifying *which outcomes are implementable without commitment* rather than designing an optimal commitment mechanism.

3 One-Dimensional Benchmark: Why Disclosure Collapses

We begin with the case $k = 1$ (one product) to isolate the fundamental tension between disclosure and price discrimination.

Let $V \in [0, 1]$ have continuous cdf F with full support and let the platform's cost be zero.

3.1 A basic observation

In this one-good environment, the buyer's message can only influence the *price* the platform offers. Since the buyer's utility from trade is $V - p$, every type weakly prefers lower prices. This creates a "race to the bottom" in message choice: whenever one on-path message induces a lower price than another, higher types want to mimic it.

The next theorem formalizes the resulting collapse.

Theorem 1 (No informative very credible mechanisms with one good)**.** *Consider the one-product environment. In any pure-strategy sequential equilibrium in which trade occurs with positive probability, the platform's on-path behavior is outcome-equivalent to posting a single price $p^\star$ and ignoring messages. Moreover, $p^\star$ is a monopoly price for the prior, i.e.,*

$$p^\star \in \arg\max_{p \in [0,1]} \; p\,(1 - F(p)).$$

The proof is straightforward (see the appendix A), since if any type gets a lower price, user will lie and report that price.

4 Multidimensional Preferences: Targeting without Stronger Pricing

The impossibility in Theorem 1 hinges on a one-dimensional tradeoff: the only action is a price, and lower prices are always preferred by all buyer types. With multiple products, the platform's action has a *targeting* component (which product to offer) in addition to pricing. A buyer may be willing to disclose information that improves targeting if that information does not (materially) increase the platform's ability to price discriminate.

We illustrate this with a simple two-product example in which truthful preference disclosure is very credible.

4.1 A two-product example

There are two products $I = \{A, B\}$. The buyer's value vector $V = (V_A, V_B)$ is drawn from the distribution

$$\Pi = \tfrac{1}{2}\Big(\mathrm{Unif}[0, \tfrac{1}{2}] \times \mathrm{Unif}[\tfrac{1}{2}, 1]\Big) + \tfrac{1}{2}\Big(\mathrm{Unif}[\tfrac{1}{2}, 1] \times \mathrm{Unif}[0, \tfrac{1}{2}]\Big).$$

Equivalently, one product is "high" (uniform on $[\frac{1}{2}, 1]$) and the other is "low" (uniform on $[0, \frac{1}{2}]$), with equal probability for which product is high.

The platform can display only one offer (i, p) at a time.

4.2 Benchmark: no disclosure

If the buyer sends an uninformative message (or if the platform ignores messages), then by symmetry the platform randomizes between A and B and posts the monopoly price for the marginal distribution of each coordinate. Each coordinate is $\mathrm{Unif}[0, 1]$, so the monopoly price is $\frac{1}{2}$. Trade occurs with probability $\frac{1}{2}$ and expected revenue is $\frac{1}{4}$.

4.3 A very credible targeting mechanism

Consider the following protocol:

> **Message space:** $\mathcal{M} = \{A, B\}$. The buyer reports which product has the higher realized value. After report $m \in \{A, B\}$, the platform offers product m at price $\frac{1}{2}$.

Theorem 2 (Very credible preference disclosure)**.** *In the two-product environment above, there exists a sequential equilibrium in which (i) the buyer truthfully reports the higher-valued product, and (ii) the platform offers the reported product at price $\frac{1}{2}$. This equilibrium yields expected revenue $\frac{1}{2}$ and expected buyer surplus $\frac{1}{4}$, strictly exceeding the no-disclosure benchmark.*

Proof. Fix buyer strategy: report A if $V_A \geq V_B$ and report B otherwise. Fix platform strategy: after message $m \in \{A, B\}$ offer $(m, \frac{1}{2})$.

Buyer optimality. Under Π, the reported product is always the "high" product with value in $[\frac{1}{2}, 1]$. Accepting the offer at price $\frac{1}{2}$ yields utility $V_m - \frac{1}{2} \geq 0$, while rejecting yields 0. If the buyer lies and reports the low product, it faces price $\frac{1}{2}$ for a value in $[0, \frac{1}{2}]$ and (weakly) prefers to reject, obtaining 0. Hence truthful reporting and acceptance are optimal.

Platform optimality. Given message m, the posterior over V_m is $\mathrm{Unif}[\frac{1}{2}, 1]$. For this distribution, revenue from price $p \in [0, 1]$ is $p \cdot \mathbb{P}(V_m \geq p)$, maximized at $p = \frac{1}{2}$. Offering the other product yields strictly lower optimal revenue because it is $\mathrm{Unif}[0, \frac{1}{2}]$ conditional on the message. Thus, offering $(m, \frac{1}{2})$ is a best response.

Payoffs. Revenue is always $\frac{1}{2}$ because trade occurs with probability 1. Buyer surplus is $\mathbb{E}\left[V_m - \frac{1}{2}\right]$ where $V_m \sim \mathrm{Unif}[\frac{1}{2}, 1]$, hence $\frac{3}{4} - \frac{1}{2} = \frac{1}{4}$. □

Remark 2. The key feature is that the buyer discloses information about *which good is a better match*, not about *how much* it is willing to pay. The platform's optimal price remains $\frac{1}{2}$ after disclosure, so lying does not create a profitable "cheaper price" deviation for high types.

5 Discussion and Open Questions

The example suggests a general design principle for very credible mechanisms:

> *Voluntary disclosure is most plausible when it improves targeting but does not sharpen the platform's inference about willingness to pay for the targeted offer.*

In one-dimensional environments, any informative disclosure directly sharpens the platform's inference about willingness to pay, inviting price discrimination and destroying incentive compatibility (Theorem 1). In multidimensional environments, disclosure can be "orthogonal" to willingness to pay for the eventual offer (as in Theorem 2).

A natural open question is to characterize distributions Π for which such beneficial and very credible disclosure exists. For example, if Π concentrates on a one-dimensional manifold (e.g., $V_B = \lambda V_A$ almost surely), then "preference" information may be redundant and disclosure might again collapse to the one-dimensional benchmark.

6 Conclusion

We proposed "very credible mechanisms" as a way to reason about data disclosure in settings where neither consumers nor platforms can commit. In a single-good environment, the scope for beneficial voluntary disclosure collapses: equilibrium outcomes are equivalent to posted monopoly pricing. With multiple goods and limited attention, users can sometimes truthfully disclose coarse preference information that improves targeting and benefits both sides. Understanding when such positive-sum disclosure exists—and what it implies for privacy and platform policy—is an interesting direction for future work.

REFERENCES

Mohammad Akbarpour and Shengwu Li. Credible auctions: A trilemma. *Econometrica*, 88(2): 425–467, 2020.

Nico Neumann, Catherine E Tucker, Levi Kaplan, Alan Mislove, and Piotr Sapiezynski. Data deserts and black boxes: the impact of socio-economic status on consumer profiling. *Management Science*, 70(11):8003–8029, 2024.

A PROOF FOR THEOREM 1

Proof. Let $p(m)$ denote the (deterministic) equilibrium price after message m. Suppose two messages m_1, m_2 are sent on path and induce prices $p(m_1) < p(m_2)$. Consider any type $v \geq p(m_2)$ who buys after m_2. That type strictly prefers to send m_1 and buy at the lower price $p(m_1)$, contradicting optimality of the message strategy. Hence, among types who trade with positive probability, there can be at most one on-path price.

Therefore, any equilibrium with trade is outcome-equivalent to a single posted price $p^\star$ (possibly accompanied by additional off-path or non-trading messages). Given that the equilibrium message is uninformative on the trading path, the platform's posterior is the prior, and sequential rationality implies it chooses a monopoly price for the prior. □