Privacy Preserving Auctions*

Ran Eilat†, Kfir Eliaz,‡ and Xiaosheng Mu§

Abstract

In many auction settings the auctioneer must disclose the identity of the winner and the price he pays. We characterize the auction that minimizes the winner's privacy loss among those that maximize total surplus or the seller's revenue, and are strategy-proof. Privacy loss is measured with respect to what an outside observer learns from the disclosed price, and is quantified by the mutual information between the price and the winner's willingness to pay. When only interim individual-rationality is required, the most privacy preserving auction involves stochastic ex-post payments. Under ex-post individual rationality, it is the second-price auction with deterministic payments.

Keywords: Bayesian privacy, Mutual information, Auction theory

*Financial support from BSF grant 2022294 is gratefully acknowledged.

†Ben Gurion University of the Negev; eilatr@bgu.ac.il

‡School of Economics, Tel-Aviv University and David Eccles School of Business, the University of Utah; kfire@tauex.ac.il

§Department of Economics, Princeton University; xmu@princeton.edu

1 Introduction

In many auction markets, it is common practice for the auctioneer to disclose the identity of the winning bidder and the price he paid. For instance, this transparency is prevalent in many public procurement auctions worldwide. In the U.S., cities such as New York, Chicago, and Philadelphia make the contract amount and winning bidder publicly accessible.[1] Likewise, the U.S. Department of the Treasury publishes the names of winners and the sale prices for auctions of seized property.[2] Another example is that of prominent auction houses, which have built their reputation on their capacity to secure high sale prices, and also frequently disseminate the results of their auctions. While some may not reveal the identity of the winner, this information often finds its way into the media.[3,4]

This disclosure of information may raise concerns for potential bidders. For instance, a bidder may fear that the disseminated information could potentially be leveraged against him in subsequent auctions. Additionally, a bidder who wins a contract through a procurement auction will often need to negotiate with subcontractors. Knowledge of his true value for winning the contract may undermine his bargaining position in these negotiations. Other buyers may be concerned that winning an auction and paying either an excessively high or low price could expose them to criticism from third parties (e.g. managers, clients, or the general public).

This leads to the question: Given the necessity of disclosing the winner's identity and payment (e.g., due to regulatory requirements or as an anti-corruption measure), which type of auction minimizes the winner's privacy loss while still accomplishing the auctioneer's primary objective, which can be either efficiency or revenue maximization? In this paper, we take a first step towards addressing this question.

Measuring privacy loss. To investigate the question, we employ the Bayesian approach to measuring privacy loss, as proposed in Eilat, Eliaz and Mu (2021). The cornerstone of this approach is the idea that privacy loss is a *relative* notion: How much new information is effectively learned about the winner's willingness to pay ("type") from observing his payment should be measured relative to what was previously known about the winner.

In the context of this paper, consider an "outsider" who observes the winner's identity

[1] See open-contracting.org for a list of worldwide databases of public procurement auction results.

[2] See www.treasury.gov/auctions/treasury/rp/bidresults.shtml.

[3] See, e.g., thecollector.com and artnews.com.

[4] The disclosure of the winner's identity and payment is often justified as a compromise between full transparency (i.e., disclosing all participants and their bids) and complete opacity in an auction. Such a compromise is warranted because full transparency may facilitate collusion among bidders, while complete opacity may create opportunities for corruption.

and payment. In accordance with the Bayesian tradition of mechanism design, the outsider is assumed to have a prior belief regarding the winner's type. When she observes the actual payment, the outsider updates her beliefs about the winner's type. The Bayesian approach to privacy loss quantifies the expected change in the outsider's equilibrium beliefs triggered by the new information. A greater change indicates a more significant privacy loss.

There are several equivalent methods for computing privacy loss. Throughout most of the analysis we employ the following representation: Privacy loss is defined as the *mutual information* between the random variable representing the winner's type and the random variable representing the payment. This representation quantifies the reduction in uncertainty about the former random variable caused by observing a realization of the latter. In Section 3 we discuss two equivalent methods for computing the same quantity.

The merits of the Bayesian approach to privacy loss are discussed in Eilat, Eliaz and Mu (2021). Here, we mention only that Bayesian privacy loss does not require making any assumptions about the exact nature of future strategic interactions. This allows us to rank auctions according to the "amount of information" they divulge about the "sensitive" variable (the winner's type), without committing to the exact manner in which this information will be used in the future.

Preview of the model and main results. We consider a pure private values environment with risk-neutral buyers whose participation in the auction is voluntary. A single item is offered for sale. When the auction ends, two pieces of information are disclosed to an outside observer: the winner's identity and payment.[5] For most of our analysis, we focus on the class of *efficient* mechanisms with a dominant-strategy equilibrium. Because payments can in principle be stochastic, this class contains many mechanisms (see Section 4.2). Within this class, we seek the mechanism that minimizes the winner's privacy loss. We subsequently demonstrate that our main findings remain applicable when the designer's objective is revenue maximization.

Our focus on dominant-strategy mechanisms is motivated by several considerations. First, this assumption makes the analysis more tractable. Specifically, we rely on this assumption in Step 3 of the proof for Theorem 1. Second, dominant-strategy mechanisms are considered to be desirable in practical applications. This is because they simplify the strategic reasoning for bidders and exhibit robustness in the sense that equilibrium outcomes do not rely on bidders' high-order beliefs. Finally, the class of dominant-strategy

[5]We assume that no information is disclosed about the losing bidders – neither their identity, nor their bids, are revealed. In light of this, we are concerned only with the privacy loss of the winner.

mechanisms provides a natural candidate – the second-price auction (SPA) – that can serve as a benchmark for assessing the most privacy-preserving mechanism, as it does not directly disclose the winner's willingness to pay. In contrast, other prevalent auction formats that are not strategy proof, such as the first-price and the Dutch auctions, reveal all information about the winner, and hence are the *worst* mechanisms in terms of the winner's privacy.

We demonstrate that, in certain cases, introducing post-bid randomization to the winner's payment can reduce the mechanism's privacy loss. From an outsider's perspective, this randomization sometimes weakens the connection between the winner's willingness to pay and his final payment. In particular, if the price that the winner can pay is allowed to be arbitrarily high (as long as interim individual rationality is satisfied), it is possible to achieve efficiency or revenue maximization with privacy loss converging to zero. This is accomplished through a mechanism in which, for any bid profile, the winner's payment is stochastic and takes the form of a lottery between 0 and some large payment K, with probabilities determined such that both incentive compatibility and interim individual rationality are satisfied.

However, in many real-life scenarios, requiring buyers to make arbitrarily high payments is impractical. Therefore, we examine two types of restrictions on the realized payments. First, we consider the case where there is a uniform upper bound on the payments, namely that all buyer types are limited to paying no more than some constant K. We show that the mechanism described above, in which the payment is stochastic and the winner pays either 0 or K, minimizes the winner's privacy loss among all efficient (or revenue-maximizing) mechanisms that involve payments at most K.

Next, we examine a scenario in which the maximum amount that a winner can pay is type dependent. Specifically, we assume that voluntary participation is applied *ex-post* and therefore a buyer's payment cannot exceed his true valuation. Perhaps surprisingly, our main result in this case is completely opposite to the case with a uniform price cap: stochastic payments do not prove effective in reducing the winner's privacy loss when ex-post individual rationality is required. In fact, we show that under mild conditions, a well-known mechanism with *deterministic* payments minimizes the winner's privacy loss: the second-price auction (with the addition of a reserve price if revenue maximization is the objective).

The proof for this result comprises three steps. First, we establish a general lemma that characterizes the lower bound on the mutual information between two ordered random variables with given marginal distributions. Next, we verify that the joint distribution of winner types and payments under the second-price auction indeed achieves the mutual information lower bound identified in the lemma. Finally, we note that any other dominant-strategy mechanism induces a payment distribution that constitutes a mean-preserving spread of the distribution of the second highest type among the bidders. We then show that any mean-preserving spread payment distribution can only increase the aforementioned lower bound.

2 Related Literature

Eilat, Eliaz and Mu (2021) introduced the notion of Bayesian privacy in mechanisms, but studied privacy loss with respect to the *designer* of a mechanism that has access to the participants' actions. It analyzes a monopolistic seller who faces one buyer and seeks to design the profit maximizing mechanism subject to some exogenous cap on privacy loss, which is measured by the mutual information between the buyer's action and type. In contrast, this paper is concerned with the privacy loss from the perspective of an *outsider* who observes only the outcome of the mechanism, where the outcome is the winner's identity and payment. Additionally, this paper solves a different problem: find the mechanism with the minimal privacy loss among all those that maximize some objective function.

Our paper is related to the literature on auction design that takes into account the inference that will be made about the winner after the auction. A recent paper by Dworczak (2020) studies the problem of a seller, whose payoff depends not only on the outcome of the mechanism, but also on the outcome in an aftermarket. The paper represents the aftermarket via the seller's payoff that depends both on the winner's type and on the posterior belief about this type. Given an aftermarket, the seller's problem is to design both an allocation rule and a disclosure rule to maximize his payoff. The paper restricts attention to a class of allocation rules that are dominant-strategy implementable via "cutoff mechanisms," where the winner has to outbid a random threshold that does not depend on his bid. The seller may disclose any information about the realization of the random cutoff.

There are three key differences between our framework and that of Dworczak (2020). First, our approach is context independent in the sense that it does not require specifying the exact payments for the seller (or the buyers) in the aftermarket. Second, our seller cares about posterior beliefs in a *lexicographic* manner: Among the mechanisms that meet some objective, he chooses the one that preserves the most privacy about the winner's type. Finally, in contrast to Dworczak (2020), our seller *must* disclose the price the winner paid. If our seller had the option to not disclose any information, he would choose it.[6]

[6] Dworczak (2020) gives sufficient conditions on the seller's payoff function for which it is optimal to conduct an SPA and disclose the price paid by the winner. However, our result that the SPA solves the seller's

A related literature studies the effect of disclosure policies on particular post-auction interaction between the bidders and third parties. Calzolari and Pavan (2006*b*) study the optimal disclosure of information between an upstream and a downstream principal who contract sequentially with the same agent. Calzolari and Pavan (2006*a*) consider the case of a monopolist who designs an allocation rule and a disclosure policy to maximize revenue, taking into account that the winning bidder may resell the object. Molnár and Virág (2008) consider a seller who designs an auction and a flexible disclosure rule to maximize expected revenue, taking into account that the winner's payoff depends both on the value he derives from the good and on the posterior belief about his value, given the information disclosed by the seller. They give sufficient conditions on the winner's payoff under which the seller discloses all or no information about bidders' types.

Instead of jointly designing the selling mechanism and the disclosure policy, other authors (some notable examples include Goeree (2003), Das Varma (2003), Katzman and Rhodes-Kropf (2008) and Giovannoni and Makris (2014)) compared different auction formats and different disclosure rules on revenue when the auction was followed by some form of competition, or when the winner cares about the posterior belief formed about his type. Similarly, Bergemann and Hörner (2018) analyze Markov-perfect equilibria of infinitely repeated first-price auctions, and compare the effect on revenue and efficiency of different disclosure rules.

In the computer science literature, a popular approach to measuring privacy in mechanisms uses the notion of "differential privacy", which was introduced by Dwork et al. (2006) (see the surveys by Pai and Roth (2013) and Heffetz and Ligett (2014)). The key difference from our approach is that differential privacy is non-Bayesian. Because it does not incorporate a prior belief, it is not concerned with what new information is learned, relative to what an outside observer knew or believed before the mechanism was executed. Furthermore, as long as the environment is prior-free, maximizing ex-ante expected revenue or welfare is not a well-defined problem. If we were to allow a prior in defining the objective, but measured privacy loss using differential privacy, we would not be able to meet the objective since it is very sensitive to the buyer's reports.

A second approach in computer science applies cryptographic tools to ensure that the bidder-bid relation is kept private. See the early paper by Naor, Pinkas and Sumner (1999) and the recent survey in Alvarez and Nojoumian (2020). However, this approach is not applicable in our setting, where there is an *exogenous requirement* to publicly reveal the identity of the winner and the price paid.

problem is obtained *only* when we impose ex-post individual rationality, a restriction which is orthogonal to the condition identified in Dworczak (2020).

3 Model

A seller owns one unit of an indivisible good, whose value to him is normalized to zero. There are n potential risk-neutral buyers. The willingness to pay (i.e. "type") of buyer $i \in N = \{1,...,n\}$, denoted θ_i, is privately and independently drawn from a distribution F over $[\underline{\theta},\overline{\theta}]$ with $\overline{\theta} > \underline{\theta} \geq 0$. We restrict attention to distributions with strictly positive and continuously differentiable densities f over the interval $[\underline{\theta},\overline{\theta}]$, which exhibit a monotone hazard rate – i.e., the ratio $(1-F(\theta))/f(\theta)$ is decreasing in θ.

The seller designs a mechanism M whose outcome is an allocation of the good, which could remain with the seller, and a profile of possibly stochastic payments. To streamline the exposition, we assume that M is a simultaneous move mechanism, which is without loss of generality as explained below. The seller considers only mechanisms that have dominant-strategy equilibria, and where only buyers make payments to the seller. Participation in the mechanism is voluntary, and a buyer who opts out gets a payoff of zero. We assume that the seller can commit to the details of the mechanism.[7]

If one of the buyers wins the good, then the winner's identity and his payment are publicly disclosed. Until Section 4.3, we focus on "efficient" mechanisms in which the good is always allocated to a buyer with the highest type. This restriction simplifies the definitions below by ensuring that the winner is always well-defined.

Given a mechanism M that always allocates the good and a dominant-strategy equilibrium (DSE) σ in this mechanism, let P^σ and W^σ denote the random variables that represent the winner's payment and winner's type induced by σ and F, and let G^σ denote their joint probability distribution. Let G^σ_P and G^σ_W denote the marginal distributions of P^σ and W^σ, respectively, while $G^\sigma_{W|P}$ denotes the conditional distribution of W^σ given P^σ. An outsider, who observes the winner's identity and payment (i.e. the realization of P^σ), updates his beliefs about the winner's type (the value of W^σ).

Definition 1 (Privacy loss) *The privacy loss associated with a mechanism M that always allocates the good and a DSE σ is the mutual information between the induced random variables W^σ (winner's type) and P^σ(winner's payment):*

$$MI\left(W^\sigma,P^\sigma\right) = D_{KL}\left(G^\sigma || G^\sigma_W \otimes G^\sigma_P\right) \tag{MI}$$

where D_{KL} is the Kullback-Leibler (KL) divergence, and $\otimes$ denotes the product distribution.

[7] E.g., he cannot ignore bids, engage in shill bidding, or change randomization probabilities. This (standard) assumption can be justified by ethical guidelines or legal constraints, or by reputational considerations of third parties, such as accounting firms, who oftentimes conduct the auction in practice.

Equivalent representations of privacy loss. Mutual information of two random variables is an information-theoretic concept that measures the mutual dependence between the two variables. It can be computed in several equivalent ways, one of which is shown on the right-hand side of Eq. (MI). Another method to compute the same quantity is given by:

$$MI\left(W^{\sigma}, P^{\sigma}\right) = \mathbb{E}_{P^{\sigma}}\left[D_{KL}\left(G^{\sigma}_{W|P} || G^{\sigma}_{W}\right)\right].$$

This representation emphasizes that the privacy loss is equal to the expected KL divergence from the posterior belief of the winner's type after observing the payment to the prior belief, with expectations taken with respect to the realized payment. Symmetrically, we also have

$$MI\left(W^{\sigma}, P^{\sigma}\right) = \mathbb{E}_{W^{\sigma}}\left[D_{KL}\left(G^{\sigma}_{P|W} || G^{\sigma}_{P}\right)\right].$$

This is useful for computation, as we can often express the payment in terms of the winner type without going through Bayesian updating.

Another equivalent representation of the mutual information is the following:

$$MI(W^{\sigma}, P^{\sigma}) = H(W^{\sigma}) - \mathbb{E}_{P^{\sigma}}[H(W^{\sigma}|P^{\sigma})],$$

where $H(\cdot)$ is the Shannon entropy of a distribution. Here, privacy loss is computed as the expected entropy reduction in the belief about winner type. Because the entropy $H(W^{\sigma})$ is constant across all efficient mechanisms, this representation suggests that minimizing privacy loss is equivalent to maximizing expected residual uncertainty about winner type.

Example 1. To illustrate the definition, suppose there are two buyers whose valuations are distributed uniformly on $[0,1]$. Suppose further that the seller uses an SPA, which admits a DSE σ in which both buyers bid their value. Before the auction is carried out, the prior is that the winner's type w is the highest of two independent draws from a uniform distribution. Hence, $G^{\sigma}_{W}(w) = w^2$ is the prior CDF of winner type, with a density of $2w$. In an SPA, the realized payment p is the value of the loser, which is the lowest of two independent draws from a uniform distribution. Therefore, $G^{\sigma}_{P}(p) = 1-(1-p)^2$, with a density of $2(1-p)$. The joint distribution $G^{\sigma}(w,p)$ is uniform over the triangle $0 \le p \le w \le 1$, with a density of 2. Plugging these into the KL-divergence formula, we have:

$$D_{KL}\left(G^{\sigma} || G^{\sigma}_{W} \otimes G^{\sigma}_{P}\right) = \int_0^1 \int_0^w 2\log\frac{2}{2w\cdot 2(1-p)}\,dp\,dw = 1 - \log 2.$$

□

The seller's objective is to design a mechanism with a DSE that maximizes the total expected surplus, such that there is no other mechanism with a DSE that achieves the same objectives but with lower privacy loss. Later, we will explain how our analysis extends to the case of revenue maximization.

Formally, let $\mathscr{M}$ denote the class of all pairs (M,σ), where M is a normal-form mechanism, and σ is a DSE in M in which each buyer's interim expected payoff is non-negative (i.e., interim individual rationality is satisfied). Let $V(M,\sigma)$ denote the expected social surplus in the DSE σ of M. The seller's problem is then given by:

$$\begin{aligned} \inf_{M,\sigma} \quad & D_{KL}\left(G^{\sigma}||G_W^{\sigma}\otimes G_P^{\sigma}\right) && \text{(Seller's problem)} \\ \text{s.t.} \quad & (M,\sigma)\in \arg\max_{(M',\sigma')\in\mathscr{M}} V\left(M',\sigma'\right) \end{aligned}$$

A *direct revelation mechanism* is a normal-form mechanism in which bidders report their types. Formally, a direct revelation mechanism is a tuple $M = \langle q, t_1 \dots t_n\rangle$, where $q:[\underline{\theta},\overline{\theta}]^n \to \Delta(I)$ is an allocation function that maps a profile of reports to a lottery over who gets the good (with I being the set of all players including the seller), and $t_i:[\underline{\theta},\overline{\theta}]^n \to \Delta(\mathbb{R}_+)$ maps the profile of reports to a *potentially stochastic* payment of buyer i (i.e., after the type profile is reported the payment can still be stochastic). Let $q_i(\theta)$ be the probability that the good is assigned to buyer i according to the distribution $q(\theta)$, and let $T_i(\theta) = \mathbb{E}[t_i(\theta)]$ be the expected *ex-post* payment of buyer i, where the expectation is taken with respect to the distribution of payments implied by $t_i(\theta)$. Thus, the expected utility of buyer i when the realized profile of types is θ is given by $u_i(\theta) = q_i(\theta)\cdot\theta_i - T_i(\theta)$.

It is without loss of generality to restrict attention to direct revelation mechanisms where truth-telling is a DSE. This is because privacy loss is calculated solely based on what an outsider observes, and not directly influenced by the players' reports to the designer. By exactly the same arguments that lead to the standard revelation principle, we obtain the following result:

Observation 1 (Revelation principle) *For any mechanism with a dominant strategy equilibrium, there exists a direct revelation mechanism in which truth-telling is a dominant strategy, such that the two equilibria induce the same stochastic mapping from type profiles to outcomes, and thus induce the same privacy loss.*

In light of this, in the remainder of the paper we will focus on direct revelation mechanisms with truthful DSE. To ease notation, we will omit the superscript σ.

As discussed above, there exists an essentially unique allocation $q(\theta)$ that characterizes an efficient mechanism:[8]

$$q_i(\theta) = \begin{cases} 1, & \text{if } \theta_i = \max_{1\le j\le n} \theta_j \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Next, by standard arguments, dominant-strategy incentive-compatibility (DSIC) requires the expected ex-post transfers to satisfy the following equation for every buyer i and every type profile (θ_i, θ_{-i}):

$$T_i(\theta_i, \theta_{-i}) = -u_i\left(\underline{\theta}, \theta_{-i}\right) - \int_{\underline{\theta}}^{\theta_i} q_i\left(\hat{\theta}, \theta_{-i}\right) d\hat{\theta} + q_i(\theta_i, \theta_{-i}) \cdot \theta_i. \tag{DSIC}$$

We then observe that $u_i(\underline{\theta}, \theta_{-i}) \le 0$ because only buyers make payments to the seller, and because the good is never allocated to type $\underline{\theta}$ by Eq. (1). But interim individual rationality requires $\mathbb{E}_{\theta_{-i}}\left[u_i(\underline{\theta}, \theta_{-i})\right] \ge 0$, so $u_i\left(\underline{\theta}, \theta_{-i}\right) = 0$ for any θ_{-i}.

By plugging Eq. (1) into Eq. (DSIC), we obtain:

$$T_i(\theta_i, \theta_{-i}) = \begin{cases} \max\{\theta_{-i}\}, & \text{if } \theta_i \ge \max\{\theta_{-i}\} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Thus, the designer's problem reduces to the following: Among stochastic ex-post payment functions $t_1 \dots t_n$ that satisfy Eq. (2), find those that minimize the mutual information between winner's type and payment.

4 Characterization

A key factor in characterizing the solution to the seller's problem is the maximum price that a bidder may be required to pay. In light of this, we will explore two natural cases. First, we will assume that the highest price cannot exceed an exogenous cap which is uniform across bidders regardless of their type. For example, this is the case when all bidders face a budget constraint that is identical for all types.

Next, we will consider the case where the price cap is type-dependent and cannot exceed the bidder's willingness to pay. Under this specification, bidders must agree to pay the realized price, i.e. we impose the stronger constraint of ex-post individual rationality. We show that while stochastic payments prove beneficial with the exogenous uniform price cap, the same does not hold when ex-post individual rationality is required.

[8] I.e., the allocation is unique up to zero measure type profiles.

4.1 Privacy with Uniform Price Caps

Given a positive real number K, define a $K-$capped mechanism to be a mechanism in which no buyer pays more than K in any realization of his payment. Namely, the upper bound of the support of $t_i(\theta)$ is smaller than K for all i and for all θ.

For any $K \geq \overline{\theta}$, we say that a mechanism $M = \langle q, t_1 \dots t_n \rangle$ is *a* $\{0,K\}$*-mechanism* if, for any profile of reports $\theta \in [\underline{\theta}, \overline{\theta}]^n$ and every buyer i, the distribution of $t_i(\theta)$ is supported on $\{0,K\}$. Notice that any K-capped mechanism $\langle q, t_1 \dots t_n \rangle$ can be transformed into a $\{0,K\}$-*mechanism* as follows. For any type profile θ, we keep the same allocation function and modify the stochastic ex-post payment function of each buyer i to be a lottery with support $\{0,K\}$ whose mean is equal to $T_i(\theta)$ of the original mechanism. This transformation does not affect the expected payment of any buyer at any type profile and thus maintains both incentive compatibility and efficiency. Since DSIC and efficiency pin down $T_i(\theta)$, there exists a unique efficient mechanism that is also a $\{0,K\}$-mechanism.

The result below shows that the efficient $\{0,K\}$-mechanism minimizes privacy loss among all efficient K-capped mechanisms, and it is an essentially unique minimizer.

Proposition 1 *For any $K \geq \overline{\theta}$, the efficient* $\{0,K\}$*-mechanism minimizes privacy loss among all efficient K-capped mechanisms. Moreover, if any efficient K-capped mechanism achieves minimal privacy loss, then for every buyer i, the realized payment $t_i(\theta)$ is supported on* $\{0,K\}$ *for almost every type profile θ such that $\theta_i = \max_{1 \leq j \leq n} \theta_j$.*

Proof: Given an efficient K-capped mechanism M, we can view the efficient $\{0,K\}$-mechanism as the following transformation of M: For any profile of reports θ, any buyer i and any payment p in the support of $t_i(\theta)$, replace the payment p by a lottery that induces the payment K with probability p/K and the payment 0 with remaining probability. This results in an efficient $\{0,K\}$-mechanism, which must be the unique one discussed above.

Denote by P_M and $P_{\{0,K\}}$ the random variables that represent the winner's payments in M and the $\{0,K\}$-mechanism, respectively. The above transformation allows us to represent $P_{\{0,K\}}$ as a random variable that only depends on P_M. In particular, conditional on P_M, the random variable $P_{\{0,K\}}$ is *conditionally independent* from the winner's type W. Therefore, by the Date Processing Inequality, we have

$$MI(W, P_M) \geq MI\left(W, P_{\{0,K\}}\right).$$

This proves that the efficient $\{0,K\}$ mechanism minimizes privacy loss.

To show that it is essentially the unique minimizer, note that the Data Processing Inequality holds equal only if P_M is also conditionally independent from W conditional on

$P_{\{0,K\}}$. Below we show that this can only be the case if P_M is supported on $\{0,K\}$, which will imply the result.

Let $\overline{H}$ denote the distribution of P_M conditional on $P_{\{0,K\}} = K$, while $\underline{H}$ denotes its distribution conditional on $P_{\{0,K\}} = 0$. Let $\overline{\mu}$ and $\underline{\mu}$ denote the expectation of $\overline{H}$ and $\underline{H}$, respectively. Given any winner type W, let $\alpha(W)$ denote the conditional probability that $P_{\{0,K\}} = K$. We have the following conditional expectation:

$$\mathbb{E}\left[P_{\{0,K\}} \mid W\right] = \alpha(W) \cdot K.$$

On the other hand, assuming conditional independence from W, the random variable P_M will have $\alpha(W)$ probability to follow the distribution $\overline{H}$ and remaining $1-\alpha(W)$ probability to follow the distribution $\underline{H}$. Therefore,

$$\mathbb{E}[P_M \mid W] = \alpha(W) \cdot \overline{\mu} + (1-\alpha(W)) \cdot \underline{\mu} = \alpha(W) \cdot \left(\overline{\mu} - \underline{\mu}\right) + \underline{\mu}.$$

Recall that the transformation from P_M to $P_{\{0,K\}}$ does not change expected ex-post payments at any type profile. So the above two conditional expectations $\mathbb{E}\left[P_{\{0,K\}} \mid W\right]$ and $\mathbb{E}[P_M \mid W]$ must be equal for every value W of the winner type. Because $\alpha(W)$ is not constant in W,[9] this equality implies $\underline{\mu} = 0$ and $\overline{\mu} = K$. Since the support of P_M is contained in the interval $[0,K]$, the distribution $\overline{H}$ must be the point-mass at K while $\underline{H}$ is the point-mass at 0. This completes the proof that P_M is supported on $\{0,K\}$. ■

The next result shows that with a sufficiently large price cap K, the seller can make the privacy loss arbitrarily small.

Proposition 2 *For any $\varepsilon > 0$ there exists $K(\varepsilon) > 0$ such that the efficient $\{0,K(\varepsilon)\}$-mechanism achieves privacy loss smaller than ε.*

Proof. With $T_i(\theta)$ given by Eq. (2), let $\tau(\theta_i) = \mathbb{E}_{\theta_{-i}}[T_i(\theta)]$ denote the interim expected payment of buyer type θ_i in any efficient mechanism. Then, in the unique efficient $\{0,K\}$-mechanism, winner type W would pay K with probability $\tau(W)/K$ and pay zero with the remaining probability. Averaging across W (according to the distribution of the winner's type, G_W), the unconditional probability that the winner pays K is $\mathbb{E}[\tau(W)]/K$, and 0 with

[9] $\alpha(W) \cdot K$ is the expected winner payment conditional on winner type W, which is the conditional expectation of the second highest type. When W is close to $\underline{\theta}$, the second highest type is also close to $\underline{\theta}$. But if W is bounded away from $\underline{\theta}$, then the second highest type also has a positive conditional probability of being bounded away from $\underline{\theta}$, making its expectation bounded away from $\underline{\theta}$ as well.

the remaining probability. Hence the privacy loss in the efficient $\{0, K\}$-mechanism is given by:

$$\begin{aligned}
D_{KL}(G||G_W \otimes G_P) &= \mathbb{E}_W \left[D_{KL} \left(G_{P|W} || G_P \right) \right] \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\tau(w)}{K} \log \frac{\tau(w)/K}{\mathbb{E}[\tau(W)]/K} + \left(1 - \frac{\tau(w)}{K}\right) \log \frac{1 - \tau(w)/K}{1 - \mathbb{E}[\tau(W)]/K} \right) dG_W(w) \qquad (3)
\end{aligned}$$

The integrand on the right-hand side of Eq. (3) is bounded above by

$$\frac{\overline{\theta}}{K} \log \frac{\overline{\theta}}{\mathbb{E}[\tau(W)]} + \log \frac{1}{1 - \overline{\theta}/K} = O(1/K).$$

Thus, as $K \to \infty$ the integral converges to zero. ∎

Remark. The auctions characterized in Propositions 1 and 2 involve rather extreme stochastic payment schemes. These are not meant to be interpreted as descriptive or normative statements about privacy-preserving auctions. Instead, their purpose is to highlight the potential role of randomization in reducing privacy loss in auctions. This stands in stark contrast to our result in the next section where randomness is not useful in alleviating privacy concerns in the presence of ex-post individual rationality.

4.2 Privacy with Ex-post Individual Rationality

In this section we consider the case where the mechanism must satisfy ex-post individual rationality (EPIR). Namely, the winner's payment cannot exceed his valuation. In contrast to our previous results, we now show that in this case, the most privacy-preserving auction uses a deterministic pricing rule: the winner simply pays the second-highest bid.

Theorem 1 *The standard SPA with deterministic payments minimizes the privacy loss among all efficient, DSIC and ex-post individually rational mechanisms.*[10]

Before we proceed to the proof, it is worth noting that the restriction to ex-post individually-rational dominant-strategy mechanisms still leaves the door open to a wide variety of auctions. Namely, although conditional on winning the *expected* payment of the winner must be independent of the winner's type and be equal to the second-highest bid, this payment can potentially be stochastic ex-post (i.e., after all bids have been submitted). Therefore, the *distribution* of prices that the winner pays *can* vary with the profile of bids,

[10] It follows from the proof below that randomized payments strictly increase the privacy loss when the hazard rate $\frac{1-F(\theta)}{f(\theta)}$ is strictly decreasing in θ.

including the winner's bid. The distribution of prices only needs to adhere to the following conditions: (i) its mean has to be equal to the second highest value, and (ii) its support must be bounded above by the winner's value. A variety of stochastic price schedules satisfy these conditions, as we show below.

A simple example, in the spirit of Proposition 1, is the following. Given a profile of bids where b_1 is the highest bid and b_2 is the second highest, the mechanism can determine the winner's price by randomizing between 0 and b_1 with probabilities $1-b_2/b_1$ and b_2/b_1, respectively. Consequently, for every profile of bids, the winner pays the second highest bid on average. More complex continuous distributions that satisfy the required properties can also be devised. An example is when the winner's price is drawn from a scaled Beta distribution with parameters $\alpha = b_2$ and $\beta = b_1 - b_2$ that is supported on $[0, b_1]$, whose mean is precisely b_2. Theorem 1 proves that all these examples will generate a higher loss of privacy compared to the deterministic SPA.

Proof of Theorem 1. The proof proceeds in three steps. First, given two distributions X and Y on $\mathbb{R}$, where X is non-atomic and $X(s) \le Y(s) \quad \forall s$, we derive a lower bound on the mutual information between any two jointly distributed random variables with marginal distributions X and Y. Applied to our setting, this result gives us a lower bound on the mutual information between winner's type and payment, given these two variables' marginal distributions. Next, we show that the SPA induces a joint distribution of winner type and payment that achieves the above mutual information lower bound, given its marginal distributions. Finally, we show that with the marginal distribution of winner type pinned down by the prior F (due to efficiency), any other marginal distribution of payment increases the mutual information lower bound compared to the one induced by the SPA.

Step 1. We begin by deriving the lower bound on the mutual information between two ordered random variables with given marginal distributions. A key observation for this result is that the joint density that attains this lower bound has a multiplicative form. To illustrate this observation in a simpler setup, consider the following discrete example in which the optimal joint distribution can be characterizes using a standard Lagrangian method.

Example 2. Suppose that $\mathscr{X}$ and $\mathscr{Y}$ are two discrete random variables, jointly distributed on $\{1,2,3\} \times \{1,2,3\}$, where $\mathscr{X} \ge \mathscr{Y}$ with probability 1. Denote the probability mass functions of the two random variables by g_1 and g_2, respectively. Table (1a) provides an example. To find the joint distribution λ that minimizes the mutual information between the two

	$x=1$	$x=2$	$x=3$	$g_2(y)$
$y=3$			$\lambda(3,3)$	0.1
$y=2$		$\lambda(2,2)$	$\lambda(3,2)$	0.3
$y=1$	$\lambda(1,1)$	$\lambda(2,1)$	$\lambda(3,1)$	0.6
$g_1(x)$	0.1	0.4	0.5	

(a)

	$x=1$	$x=2$	$x=3$	$g_2(y)$
$y=3$			0.1	0.1
$y=2$		0.15	0.15	0.3
$y=1$	0.1	0.25	0.25	0.6
$g_1(x)$	0.1	0.4	0.5	

(b)

Table 1: Parameters for Example 2. (a) Given random variables $\mathscr{X}$ and $\mathscr{Y}$ that satisfy $\mathscr{X} \geq \mathscr{Y}$, with probability mass functions $g_1(\cdot)$ and $g_2(\cdot)$, respectively, find the joint distribution $\lambda(x,y)$ on $1 \leq y \leq x \leq 3$ that minimizes the mutual information between the two random variables (b) The MI-minimizing joint distribution.

random variables, we solve:

$$\min_{\lambda} \sum_{x=1}^{3} \sum_{y=1}^{x} \lambda(x,y) \cdot \log\left(\frac{\lambda(x,y)}{g_1(x) \cdot g_2(y)}\right)$$

$$s.t. \ \sum_{y=1}^{x} \lambda(x,y) = g_1(x) \ \forall x, \ \text{ and } \ \sum_{x=y}^{3} \lambda(x,y) = g_2(y) \ \forall y$$

Differentiating the associated Lagrangian, we obtain the following first-order conditions:

$$\lambda^*(x,y) = h_1(x) \times h_2(y) \times 1_{y \leq x} \qquad \forall x, y \in \{1,2,3\}$$

where $h_1(x) = e^{\alpha(x)+\log(g_1(x))}$ and $h_2(y) = e^{\beta(y)+\log(g_2(y))-1}$, and $\alpha(x)$ and $\beta(y)$ are the Lagrange multipliers associated with the marginal constraints. Given the first-order conditions and the marginal constraints, the parameters of the example yield the solution $h_1(1) = \frac{1}{6}, h_1(2) = \frac{5}{12}, h_1(3) = \frac{5}{12}$ and $h_2(1) = \frac{15}{25}, h_2(2) = \frac{9}{25}, h_2(3) = \frac{6}{25}$.[11] This solution is described in Table (1b), where, for example, $\lambda(2,2) = h_1(2) \times h_2(2) = 0.15$. □

The following result extends the illustration in Example 2 to any pair of random variables X and Y, where X is non-atomic.[12]. The proof of this lemma is presented in Section 5 below.

[11]There may be multiple solutions for h_1 and h_2, but all solutions yield the same product.

[12]This result generalizes the bivariate case of Theorem 5.4 in Butucea et al. (2018) to environments where the marginal distributions may not admit densities. This generalization is important for our application, as the payment distribution is endogenously chosen and may not have a density (for example see Section 4.3 below). See also the independent work by Arnold, Molchanov and Ziegel (2020).

Lemma 1 *Let X and Y be two Borel probability measures on $\mathbb{R}$, and with an abuse of notation let $X(s), Y(s)$ also denote their CDFs. Assume X is non-atomic (i.e. $X(s)$ is continuous in s) and $X(s) \leq Y(s)$ for all $s \in \mathbb{R}$.*

Define $\mathscr{M}(X,Y)$ to be the set of joint distributions λ of two random variables $\mathscr{X}$ and $\mathscr{Y}$ with marginal distributions X and Y, and satisfying $\mathscr{X} \geq \mathscr{Y}$ with λ- probability 1. Then, with the convention $\log \frac{1}{0} = \infty$, it holds that

$$\inf_{\lambda \in \mathscr{M}(X,Y)} D_{KL}(\lambda \,||\, X \otimes Y) = -1 + \int_{\mathbb{R}} \log \frac{1}{Y(s) - X(s)} \, dX(s). \tag{4}$$

The infimum above is achieved as minimum whenever the RHS of Eq. (4) is finite, in which case the unique minimizer λ^* is the joint distribution defined by

$$\frac{d\lambda^*}{d(X \otimes Y)}(x,y) = \frac{1}{Y(x) - X(x)} \cdot e^{-\int_y^x \frac{1}{Y(s)-X(s)} \, dX(s)} \cdot 1_{Y(x) > X(x)} \cdot 1_{y \leq x}. \tag{5}$$

That is, the Radon-Nikodym derivative of λ^* with respect to the product measure $X \otimes Y$ is zero if either $Y(x) = X(x)$ or $y > x$. Otherwise this density is $\frac{1}{Y(x)-X(x)} \cdot e^{-\int_y^x \frac{1}{Y(s)-X(s)} \, dX(s)}$.

Moreover, if the RHS of Eq. (4) is finite and if there exists $\hat{\lambda} \in \mathscr{M}(X,Y)$ such that $\frac{d\hat{\lambda}}{d(X \otimes Y)}(x,y) = h_1(x) \cdot h_2(y)$ for a pair of functions h_1, h_2 that are positive and bounded away from zero, then $\hat{\lambda} = \lambda^*$ as described above.

Applying Lemma 1 to our setup, we can let $X = G_W$ be the marginal distribution of winner type, and $Y = G_P$ be the marginal distribution of payment in an efficient, DSIC, EPIR mechanism. The RHS of Eq. (4) provides a lower bound on privacy loss, given by:

$$-1 + \int_{\mathbb{R}} \log \frac{1}{G_P(s) - G_W(s)} \, dG_W(s). \tag{6}$$

We emphasize that the condition $\mathscr{X} \geq \mathscr{Y}$ with λ-probability 1 is crucial for the lemma; otherwise $\lambda = X \otimes Y$ could lead to zero mutual information. In our setup, this ranking condition corresponds to the winner's type always exceeding his payment, as required by ex-post individual rationality.

Step 2. We now show that the joint distribution of winner type and payment under the SPA achieves the mutual information lower bound in Eq. (4), given its marginal distributions. Note that $X = G_W = F^n$ is the marginal distribution of winner type, with density $g_W(s) = nf(s)F(s)^{n-1}$. Denote the CDF of the second highest type out of n independent draws from F by $G_L(s) = F(s)^n + n(1-F(s))F(s)^{n-1}$. Then $Y = G_L$ is the marginal distribution of payment, with density $g_L(s) = n(n-1)f(s)(1-F(s))F(s)^{n-2}$.

Under the SPA, the joint distribution $\hat{\lambda}$ of winner type and payment is the joint distribution of the highest and second highest types among n independent draws from F. This joint distribution has density $d\hat{\lambda}(w,p) = nf(w)\cdot(n-1)f(p)F(p)^{n-2}$. Therefore,

$$\frac{d\hat{\lambda}}{d(G_W \otimes G_L)}(w,p) = \frac{1}{nF(w)^{n-1}(1-F(p))} \quad \forall\ \underline{\theta} \le p < w \le \overline{\theta}.$$

Define $h_1(w) = \frac{1}{nF(w)^{n-1}} \ge \frac{1}{n}$ and $h_2(p) = \frac{1}{1-F(p)} \ge 1$. The last part of Lemma 1 shows that $\hat{\lambda}$ minimizes mutual information given its marginals. It is the unique minimizer as $\int \log \frac{1}{G_L(s)-G_W(s)}\, dG_W(s) = \int \log \frac{1}{n(1-F(s))F(s)^{n-1}}\, dF(s)^n = \int_0^1 \log \frac{1}{n(1-x)x^{n-1}}\, dx^n$ is finite.[13]

Remark. Intuitively, while the highest and second highest types are not independently distributed, their joint distribution can be obtained by conditioning a product distribution on the "triangular region" that one of them is always larger than the other. Lemma 1 ensures that whenever the joint distribution of two ordered random variables has such a property, this joint distribution minimizes mutual information given the marginals. This feature was also illustrated in Example 2.

Step 3. Under DSIC, the winner's expected payment at any type profile is the second highest type. Thus for any DSIC mechanism, G_P is a *mean-preserving spread* of G_L, and due to EPIR, $G_P(\overline{\theta}) = 1 = G_L(\overline{\theta})$. From this we will show that G_L minimizes the RHS of Eq. (6) among all possible G_P, which will prove Theorem 1.

From the mean-preserving spread property, we have $\int_{-\infty}^{t} G_P(s)\, ds \ge \int_{-\infty}^{t} G_L(s)\, ds$ for every $t \in \mathbb{R}$. Equality holds for $t \ge \overline{\theta}$ because $G_P(\overline{\theta}) = 1 = G_L(\overline{\theta})$. We then obtain:

$$\int_t^{\overline{\theta}} G_P(s)\, ds \le \int_t^{\overline{\theta}} G_L(s)\, ds \quad \text{for every } t \in [\underline{\theta}, \overline{\theta}].$$

For any two real numbers $a, b \ge 0$, we have $\log \frac{b}{a} \le \frac{b}{a} - 1$ and $a \log \frac{a}{b} \ge a - b$. Thus

$$\int_t^{\overline{\theta}} (G_L(s) - G_W(s)) \cdot \log \frac{G_L(s) - G_W(s)}{G_P(s) - G_W(s)}\, ds \ge \int_t^{\overline{\theta}} (G_L(s) - G_P(s))\, ds \ge 0.$$

Rearranging yields the following inequality:

$$\int_t^{\overline{\theta}} (G_L(s) - G_W(s)) \cdot \log \frac{1}{G_P(s) - G_W(s)}\, ds \ge \int_t^{\overline{\theta}} (G_L(s) - G_W(s)) \cdot \log \frac{1}{G_L(s) - G_W(s)}\, ds. \quad (7)$$

[13] $\int_0^1 \log \frac{1}{1-x}\, dx^n \le \int_0^1 n \log \frac{1}{1-x}\, dx = n$, and $\int_0^1 \log \frac{1}{x^{n-1}}\, dx^n \le \int_0^1 \frac{1}{x^{n-1}}\, dx^n = n$ as well.

Integrating across t, we obtain that for any function $c(t) \geq 0$:

$$\int_{\underline{\theta}}^{\overline{\theta}} c(t)\left(\int_{t}^{\overline{\theta}} (G_L(s) - G_W(s)) \cdot \log \frac{1}{G_P(s) - G_W(s)}\, ds\right) dt$$
$$\geq \int_{\underline{\theta}}^{\overline{\theta}} c(t)\left(\int_{t}^{\overline{\theta}} (G_L(s) - G_W(s)) \cdot \log \frac{1}{G_L(s) - G_W(s)}\, ds\right) dt.$$

Let $C(s) = \int_{\underline{\theta}}^{s} c(t)\, dt$. By changing the order of integration, the above yields:

$$\int_{\underline{\theta}}^{\overline{\theta}} C(s)(G_L(s) - G_W(s)) \cdot \log \frac{1}{G_P(s) - G_W(s)}\, ds \geq \int_{\underline{\theta}}^{\overline{\theta}} C(s)(G_L(s) - G_W(s)) \cdot \log \frac{1}{G_L(s) - G_W(s)}\, ds. \tag{8}$$

We now choose $c(t) = \partial\left(\frac{f(t)}{1-F(t)}\right)/\partial t$ for $\underline{\theta} \leq t < \overline{\theta}$, which is well-defined and continuous because f is continuously differentiable. By the monotone hazard rate assumption, $c(t)$ is non-negative. Thus, for this choice of $c(t)$, Eq. (8) holds with $C(s) = \int_{\underline{\theta}}^{s} c(t)\, dt = \frac{f(s)}{1-F(s)} - f(\underline{\theta})$. By adding Eq. (8) and $f(\underline{\theta})$ multiples of Eq. (7) for $t = \underline{\theta}$, we obtain:

$$\int_{\underline{\theta}}^{\overline{\theta}} \frac{f(s)}{1-F(s)}(G_L(s) - G_W(s)) \cdot \log \frac{1}{G_P(s) - G_W(s)}\, ds \tag{9}$$
$$\geq \int_{\underline{\theta}}^{\overline{\theta}} \frac{f(s)}{1-F(s)}(G_L(s) - G_W(s)) \cdot \log \frac{1}{G_L(s) - G_W(s)}\, ds.$$

Using the expressions for the distributions G_W and G_L, we have:

$$\frac{f(s)}{1-F(s)}(G_L(s) - G_W(s)) = \frac{f(s)}{1-F(s)} \cdot n(1-F(s))F(s)^{n-1} = nf(s)F(s)^{n-1} = g_W(s).$$

Hence, Eq. (9) implies that:

$$\int_{\underline{\theta}}^{\overline{\theta}} \log \frac{1}{G_P(s) - G_W(s)}\, dG_W(s) \geq \int_{\underline{\theta}}^{\overline{\theta}} \log \frac{1}{G_L(s) - G_W(s)}\, dG_W(s). \tag{10}$$

Namely, among all possible payment distributions G_P, G_L minimizes the mutual information lower bound in Eq. (6). ■

4.3 Privacy and Revenue Maximization

In the above analysis, we assumed that the designer's objective is to achieve efficiency. We now demonstrate that our main results extend to the case of revenue maximization. By Myerson (1981), the essentially unique allocation for a revenue-maximizing mechanism can be implemented by an SPA with a reserve price r, where r maximizes $r(1-F(r))$ and is unique due to monotone hazard rate. Thus, Eq. (1) is modified so that $q_i(\theta)=1$ if and only if $\theta_i = \max_{1\le j\le n}\theta_j \ge r$. Under DSIC, the expected ex-post transfers are also the same as in the SPA with reserve price r, given by $T_i(\theta_i,\theta_{-i}) = \max\{\max\{\theta_{-i}\}, r\}$ in case $q_i(\theta)=1$.

A revenue-maximizing designer who cares about privacy seeks to minimize privacy loss among all stochastic ex-post payment functions that average to the above expected payments. Note however that we need to extend the previous definition of privacy loss to the current setting, because the winner is not always defined (in particular when all buyers have value less than r). We propose the following extension of Definition 1: For any mechanism M and DSE σ, let W^σ denote the random variable of winner type *conditional on the event $\mathscr{E}^\sigma$ that the good is allocated*. Similarly let P^σ denote the random variable of winner payment conditional on the same event $\mathscr{E}^\sigma$. Then

Definition 2 (Privacy loss in the general case) *The privacy loss associated with a mechanism M and a DSE σ is the mutual information between the conditional random variables W^σ and P^σ, multiplied by the probability that the winner exists:*

$$\mathbb{P}(\mathscr{E}^\sigma)\cdot MI\left(W^\sigma, P^\sigma\right)$$

This coincides with Definition 1 when the mechanism always allocates the good, but provides a natural generalization to cases where the good is sometimes withheld.

Under this definition, we can again show that randomized payments do not help preserve privacy once ex-post individual rationality is required:

Theorem 2 *The standard SPA with an optimal reserve price and deterministic payments minimizes the privacy loss among all revenue-maximizing, DSIC and ex-post individually rational mechanisms.*

Proof. We follow the previous proof of Theorem 1 and point out the modifications. Step 1 is unchanged. In Step 2, we consider the joint distribution $\tilde{\lambda}$ of winner type W and payment P (conditional on existence of a winner) that is induced by the SPA with reserve price r. The key observation is that for any $p \ge r$,

$$\mathbb{P}[W \le w \mid P = p] = \frac{F(w)-F(p)}{1-F(p)}. \tag{11}$$

To see this, suppose that t_1 is the highest type, which means $t_1 \ge \max_{i>1} t_i$ and also $t_1 \ge r$ because the winner exists. Thus $t_1 \ge P$ and $\mathbb{P}[W \le w \mid P = p] = \mathbb{P}[t_1 \le w \mid t_1 \ge P = p]$, which is further equal to $\mathbb{P}[t_1 \le w \mid t_1 \ge p]$ by the independence across types.

From Eq. (11) we obtain that under the SPA with reserve price r, the conditional density of W given $P = p$ is simply $\frac{f(w)}{1-F(p)}$ for $w \geq p$. The marginal distribution of W is $G_W(w) = \frac{F(w)^n - F(r)^n}{1-F(r)^n}$, so the unconditional density of W is $\frac{nf(w)F(w)^{n-1}}{1-F(r)^n}$ for $w \geq r$. Dividing the conditional density by the unconditional density, we arrive at the Radon-Nikodym derivative of the joint distribution of (W,P) with respect to the product of their marginals:

$$\frac{d\hat{\lambda}}{d(G_W \otimes G_L)}(w,p) = \frac{1-F(r)^n}{nF(w)^{n-1}(1-F(p))} \quad \forall\ r \leq p < w \leq \overline{\theta}.$$

With $h_1(w) = \frac{1}{nF(w)^{n-1}}$ and $h_2(p) = \frac{1-F(r)^n}{1-F(p)}$, we can apply the last part of Lemma 1 to conclude that $\hat{\lambda}$ minimizes mutual information given its marginals.[14]

As for Step 3, note that with a reserve price r, $G_W(s) = \frac{F(s)^n - F(r)^n}{1-F(r)^n} \cdot 1_{s \geq r}$ is the CDF of winner type conditional on existence of a winner, and $G_L(s) = \frac{F(s)^n - F(r)^n + n(1-F(s))F(s)^{n-1}}{1-F(r)^n} \cdot 1_{s \geq r}$ is the conditional CDF of payment (G_L has a mass point at r). We want to show that whenever G_P is a mean-preserving spread of G_L, it holds that

$$\int_r^{\overline{\theta}} \log \frac{1}{G_P(s) - G_W(s)}\, dG_W(s) \geq \int_r^{\overline{\theta}} \log \frac{1}{G_L(s) - G_W(s)}\, dG_W(s). \tag{12}$$

The proof is essentially the same as before, since we still have $\frac{f(s)}{1-F(s)}(G_L(s) - G_W(s)) = \frac{nf(s)F(s)^{n-1}}{1-F(r)^n} = g_W(s)$ in the final calculation. Intuitively, the reserve price r affects G_W and G_L by the same linear transformation. The integrals in Eq. (12) also change linearly. ■

5 Proof of Lemma 1

We begin with introducing some notation. For any interval I, we will write $\int_I u(s)\, dX(s)$ for the Lebesgue integral of a measurable function $u(s)$ with respect to the measure X. Sometimes we also write $\int_a^b u(s)\, dX(s)$, even though we still have in mind the Lebesgue integral unless otherwise specified – since X is non-atomic, whether or not the endpoints a and b are included in the range of integration does not matter. Likewise, $\int_I u(s)\, dY(s)$ is the integral of u with respect to Y over the interval I, but we will not write $\int_a^b u(s)\, dY(s)$ since the endpoints may matter.

For a bivariate function $v(x,y)$, we denote by

$$\int_{y \leq x} v(x,y)\, dX(x)\, dY(y)$$

[14] As r maximizes $r(1-F(r))$, $F(r) < 1$ must hold and so $h_2(p)$ is bounded away from zero. In addition, $\hat{\lambda}$ is the unique minimizer because $\int \log \frac{1}{G_L(s) - G_W(s)}\, dG_W(s)$ is finite like before. We omit the calculation.

the integral of $v(x,y)$ with respect to the product measure $X \otimes Y$ over the region $y \le x$. When the Fubini-Tonelli Theorem applies, this integral can be rewritten as a double integral. Our notations will distinguish these different forms of integration.

The following lemma characterizes when there exists a "ranked" joint distribution with given marginals X and Y:

Lemma 2 *$\mathcal{M}(X,Y) \neq \emptyset$ if and only if $Y(s) \ge X(s)$ for every $s \in \mathbb{R}$.*

Proof of Lemma 2. If $\mathscr{X}$ and $\mathscr{Y}$ are two random variables that satisfy $\mathscr{X} \ge \mathscr{Y}$ almost surely, then the distribution of $\mathscr{X}$ first-order stochastically dominates the distribution of $\mathscr{Y}$. This implies $Y(s) \ge X(s)$ for every s. Conversely, by the well-known "coupling" characterization, if X first-order stochastically dominates Y, then there exist random variables $\mathscr{X}$ and $\mathscr{Y}$ with marginal distributions X and Y respectively, and satisfy $\mathscr{X} \ge \mathscr{Y}$ almost surely. For example, we can choose t to be a $Unif[0,1]$ random variable, and let $\mathscr{X} = X^{-1}(t) = \min\{z : X(z) \ge t\}$ and $\mathscr{Y} = Y^{-1}(t) = \min\{z : Y(z) \ge t\}$. ∎

5.1 Preliminary Results

As can be seen from the definition of λ^* in Eq. (5), the points s where the CDFs $Y(s)$ and $X(s)$ coincide are special. In this section we prove some preliminary results about these points.

For any $s \in \mathbb{R}$, let $Y_-(s) = \lim_{t<s, t\to s} Y(t)$ be the Y-measure of $(-\infty, s)$. Note that

1. while $Y(s)$ is right-continuous in s, $Y_-(s)$ is left-continuous;
2. we do not define X_- because X is assumed to be non-atomic;
3. $Y(s) \ge Y_-(s) \ge X_-(s) = X(s)$ holds for every s.

We then define the following sets:

$$A = \{s \in \mathbb{R} : Y(s) = X(s)\};$$
$$\overline{A} = \{s \in \mathbb{R} : Y_-(s) = X(s)\}.$$

Lemma 3 *$\overline{A}$ is a closed set that contains A.*

Proof of Lemma 3. Since $Y(s) \ge Y_-(s) \ge X(s)$, any $s \in A$ necessarily also belongs to $\overline{A}$. Thus $\overline{A}$ contains A. To see that $\overline{A}$ is closed, consider any sequence $s_n \in \overline{A}$ that converges to some $s \in \mathbb{R}$. Without loss we can assume s_n is monotone in n. If s_n increases in n, then by left-continuity $Y_-(s_n) = X(s_n)$ implies $Y(s) = X(s)$ and $s \in \overline{A}$. If instead s_n decreases in

n, then $Y_-(s) \le \lim_n Y_-(s_n) = \lim_n X(s_n) = X(s)$. But we discussed above that $Y_-(s) \ge X(s)$, so equality holds and s again belongs to $\overline{A}$. ■

Since $\overline{A}$ is a closed set, its complement $\overline{A}^c$ is open. This complement can then be written as the union of at most countably many disjoint open intervals $I_1, I_2, \ldots$, which we fix in the sequel. Let us write $I_k = (a_k, b_k)$,[15] and note that a_k, b_k must both belong to $\overline{A}$; otherwise they belong to another open interval I_m, which would intersect with I_k. We now consider two possibilities. If $a_k \in A$ then we define $\hat{I}_k = I_k = (a_k, b_k)$, and if $a_k \in \overline{A} \backslash A$ we define $\hat{I}_k = [a_k, b_k)$.

Lemma 4 *A^c is the union of the disjoint intervals $\hat{I}_k$.*

Proof of Lemma 4. Clearly these intervals are disjoint. Moreover, by construction, if $s \in \hat{I}_k$ then either $s \in (a_k, b_k) \subset \overline{A}^c \subset A^c$ or $s = a_k \in \overline{A} \backslash A \subset A^c$. Either way s belongs to A^c.

Conversely, if $s \in A^c$ then there are two cases. One case is if $s \in \overline{A}^c$, in which case s belongs to some $I_k \subset \hat{I}_k$. The remaining case is if $s \in \overline{A} \backslash A$, so that $Y(s) > Y_-(s) = X(s)$. Thus for t slightly larger than s, $Y(s) > X(t)$ also holds and we thus have $Y_-(t) > X(t)$. It follows that any such t belongs to $\overline{A}^c$. All these t must belong to a single open interval I_k, and thus $s = a_k$ belongs to $\hat{I}_k$ by construction. ■

The next result relates the measure of the set A under X and under Y.

Lemma 5 *The Y-measure of A is equal to the X-measure of A.*[16]

Proof of Lemma 5. Note that the Y-measure of A^c is the total Y-measure of $\hat{I}_k$ summing across k. For each k, the Y-measure of $\hat{I}_k$ is $Y_-(b_k) - Y(a_k)$ if $a_k \in A$ and $Y_-(b_k) - Y_-(a_k)$ if $a_k \in \overline{A} \backslash A$. In both cases the measure equals $Y_-(b_k) - Y_-(a_k)$ since $a_k \in A$ would imply $Y(a_k) = Y_-(a_k)$.

Thus, from $a_k, b_k \in \overline{A}$ we know that the Y-measure of $\hat{I}_k$ is $X(b_k) - X(a_k)$ for every k, which is equal to the X-measure of $\hat{I}_k$ (recall X is non-atomic). Summing across k implies that the Y-measure of A^c is equal to the X-measure of A^c. Taking the complement then yields the lemma. ■

---

[15] Here we allow for the possibility that $a_k = -\infty$ and/or $b_k = \infty$. The subsequent analysis applies to these special cases with minimal changes.

[16] However, the Y-measure of $\overline{A}$ may be bigger than its X-measure. For example if X is uniform on $[0,1]$ and Y is the point-mass at 0, then $\overline{A} = (-\infty, 0] \cup [1, \infty)$ and it has X-measure zero but Y-measure one. In this example $A = (-\infty, 0) \cup [1, \infty)$, which does have Y-measure zero.

5.2 Proof of Lemma 1 When $X(A) > 0$

The following result shows that if the X-measure of the set A is strictly positive, then *every* joint distribution $\lambda \in \mathcal{M}(X,Y)$ is *not* absolutely continuous with respect to $X \otimes Y$. In these cases the KL-divergence $D(\lambda \,||\, X \otimes Y)$ is always infinite, and Lemma 1 holds because $-1 + \int_{\mathbb{R}} \log \frac{1}{Y(s)-X(s)} \, dX(s) \geq -1 + \int_A \log \frac{1}{Y(s)-X(s)} \, dX(s) = \infty$, where the last equality holds by the assumption that $Y(s) - X(s) = 0$ for a positive X-measure of points s.

Lemma 6 *If A has positive X-measure, then every $\lambda \in \mathcal{M}(X,Y)$ is not absolutely continuous with respect to $X \otimes Y$.*

Proof of Lemma 6. Choose any $\lambda \in \mathcal{M}(X,Y)$. Consider any point $s \in \overline{A}$, such that $Y_{-}(s) = X(s)$. Thus λ assigns the same measure to the region $y < s$ as to the region $x < s$. But by assumption λ is supported on $x \geq y$, so we also have $\lambda(y < s) = \lambda(y \leq x < s)$, which implies $\lambda(y < s \leq x) = 0$. In words, for any $s \in \overline{A}$, λ assigns zero measure to those pairs (y,x) with $y < s \leq x$.

We use this to show that λ assigns zero measure to the set $\overline{S} = \{(x,y) : \, x \in \overline{A} \text{ and } y < x\}$. Indeed, for any rational number $r \in \mathbb{R}$, we can let $s_r \in \overline{A}$ be the point that is closest to r (which exists because $\overline{A}$ is closed). Then define $S_r = \{(x,y) : \, y < s_r \leq x\}$, which we know has λ-measure zero. Thus the union of S_r across rational numbers r also has measure zero. This union covers $\overline{S}$ because for any $x > y$ with $x \in \overline{A}$, we can choose a rational number $r \in (\frac{x+y}{2}, x)$. Then the closest point s_r satisfies $|s_r - r| \leq |x - r|$, which implies $s_r \in (y, x]$ and so $(x,y) \in S_r$. Hence $\cup_r S_r$ covers $\overline{S}$, which must have λ-measure zero.

In particular, the subset $S = \{(x,y) : \, x \in A \text{ and } y < x\}$ also has λ-measure zero. Since λ has marginal X on the x-dimension, we know that the λ-measure of $T = \{(x,y) : \, x \in A \text{ and } y \leq x\}$ is the X-measure of A. Thus the set difference

$$T \backslash S = \{(x,y) : \, x = y \in A\}$$

has λ-measure equal to $X(A) > 0$. But this set $T \backslash S$ is part of the 45-degree line, which has measure zero according to $X \otimes Y$.[17] Hence λ is not absolutely continuous with respect to $X \otimes Y$. ∎

[17] For each y, the X-measure of those x such that $x = y$ is zero because X is non-atomic. The overall measure of the 45-degree line is thus also zero by Tonelli's Theorem.

5.3 Support of λ^*

From now on we assume the set A has X-measure zero. In this section we study properties of the joint distribution λ^*, whose density with respect to $X \otimes Y$ is

$$h^*(x,y) = \frac{1}{Y(x)-X(x)} \cdot e^{-\int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)} \cdot 1_{Y(x)>X(x)} \cdot 1_{y\le x}. \tag{13}$$

While h^* is defined for any $y \le x$, the following result shows that it is supported on those pairs (x,y) such that x and y belong to the same interval $\hat{I}_k$ for some k, where the intervals $\hat{I}_k$ were defined previously in Lemma 4.

Lemma 7 *Suppose $y \in A^c$ (i.e. $Y(y) > X(y)$), and let k be the unique index such that $y \in \hat{I}_k$. Then for $x \ge y$, $\int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)$ is finite if and only if $x \in \hat{I}_k$. Consequently, $h^*(x,y)$ as defined in (13) is strictly positive if and only if $x \ge y$ and $x \in \hat{I}_k$.*

Proof of Lemma 7. The second statement follows immediately from the first, since for $x \in \hat{I}_k \subset A^c$ it holds that $Y(x)-X(x) > 0$. To prove the statement about $\int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)$, recall $\hat{I}_k = [a_k, b_k)$ or (a_k, b_k). Then because $b_k \in \overline{A}$, we have $Y_-(b_k) = X(b_k)$. Thus

$$\int_{[y,b_k)} \frac{1}{Y(s)-X(s)}\, dX(s) \ge \int_{[y,b_k)} \frac{1}{Y_-(b_k)-X(s)}\, dX(s) = \log \frac{Y_-(b_k)-X(y)}{Y_-(b_k)-X(b_k)} = \infty,$$

where the penultimate equality uses the substitution $z = X(s)$, and the last equality uses $Y_-(b_k) \ge Y(y) > X(y)$. It follows that $\int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)$ is infinite whenever $x \ge b_k$.

As for $x \in [y, b_k)$, $\int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)$ is finite because the integrand $\frac{1}{Y(s)-X(s)}$ is bounded from above on the compact interval $[y,x]$. To see why, suppose for contradiction that there exists a sequence $s_n \in [y,x]$ with $Y(s_n) - X(s_n) \to 0$. Passing to a subsequence, we may assume s_n is monotone in n and has a limit $s \in [y,x]$. If s_n decreases in n, then $Y(s_n) - X(s_n) \to 0$ implies $Y(s) = X(s)$ by right-continuity, but this contradicts $s \in [y,x] \subset \hat{I}_k \subset A^c$. If s_n increases in n, then $Y(s_n) - X(s_n) \to 0$ implies $s > y$ and $Y_-(s) = X(s)$. But this contradicts $s \in (y,x] \subset I_k \subset \overline{A}^c$. ■

5.4 λ^* Belongs to $\mathcal{M}(X,Y)$ When $X(A) = 0$

We now apply Lemma 7 to show the following result:

Lemma 8 *If A has X-measure zero, then $\lambda^* \in \mathcal{M}(X,Y)$.*

Proof of Lemma 8. By construction λ^* is supported on $y \le x$, so we just need to check λ^* has marginals X and Y. Consider *any* joint distribution $\lambda \in \mathcal{M}(X,Y)$ that is absolutely continuous with respect to the product measure $X \otimes Y$. Let $h(x,y)$ be the density $\frac{d\lambda}{d(X\otimes Y)}$, with $h(x,y) = 0$ whenever $y > x$. Then the marginal requirements on λ equivalently translate into

$$\int_{\mathbb{R}} h(x,y)\, dX(x) = 1 \quad \text{for } Y\text{-almost every } y; \tag{14}$$

$$\int_{\mathbb{R}} h(x,y)\, dY(y) = 1 \quad \text{for } X\text{-almost every } x. \tag{15}$$

When A has X-measure zero and therefore also Y-measure zero by Lemma 5, these equalities for $h = h^*$ are proved in the following two lemmata. ■

Lemma 9 *h^* defined in Eq. (13) satisfies Eq. (14) for every $y \in A^c$ (i.e. $Y(y) > X(y)$).*

Lemma 10 *h^* defined in Eq. (13) satisfies Eq. (15) for every $x \in A^c$ (i.e. $Y(x) > X(x)$).*

Proof of Lemma 9. Fix any y with $Y(y) > X(y)$, and suppose $y \in \hat{I}_k = [a_k, b_k)$ or (a_k, b_k). Then thanks to Lemma 7,

$$\int_{\mathbb{R}} h^*(x,y)\, dX(x) = \int_{[y,b_k)} h^*(x,y)\, dX(x) = \int_{[y,b_k)} \frac{1}{Y(x)-X(x)} \cdot e^{-\int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)}\, dX(x).$$

For this fixed y, let $\alpha(x) = \int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)$ for $x \geq y$. Then as shown in Lemma 7, $\alpha(x)$ is finite for $x \in [y, b_k)$ and approaches ∞ as $x \to b_k$. Moreover, $\alpha(x)$ is increasing and continuous on the interval $[y, b_k)$, where continuity follows from the Dominated Convergence Theorem and X being non-atomic.

Since the function $\alpha(x)$ is equal to 0 at $x = y$ and increases continuously for $x < b_k$, we can view it as defining a non-atomic measure (also called α) on $[y, b_k)$. Directly from the definition $\alpha(x) = \int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)$, we see that α is absolutely continuous with respect to X, with density function $\frac{d\alpha}{dX}(s) = \frac{1}{Y(s)-X(s)}$ on this interval (this density is finite since $s \in \hat{I}_k \subset A^c$). It follows that

$$\begin{aligned}\int_{\mathbb{R}} h^*(x,y)\, dX(x) &= \int_{[y,b_k)} \frac{1}{Y(x)-X(x)} \cdot e^{-\alpha(x)}\, dX(x) \\ &= \int_{[y,b_k)} e^{-\alpha(x)}\, d\alpha(x) = \int_0^\infty e^{-z}\, dz = 1.\end{aligned}$$

The penultimate equality crucially uses $\lim_{x<b_k, x\to b_k} \alpha(x) = \alpha(b_k) = \infty$ when making the substitution $z = \alpha(x)$. This proves the lemma. ■

Proof of Lemma 10. Fix any x with $Y(x) > X(x)$, and suppose $x \in \hat{I}_k = [a_k, b_k)$ or (a_k, b_k). Then for $h = h^*$, the equality in (15) reduces to

$$\int_{(-\infty,x]} e^{-\int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)}\, dY(y) = Y(x) - X(x).$$

By Lemma 7, we can restrict the range of integration to $[a_k, x]$ or $(a_k, x]$. In fact we can always assume the range of integration is $[a_k, x]$, because $\hat{I}_k = (a_k, b_k)$ would imply $a_k \in A \subset \overline{A}$, and thus Y does not have an atom at a_k. In this case including the point a_k in the range of integration does not affect the integral on the LHS above.

For this fixed x, let $\beta(y) = \int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)$ for $y \le x$. By Lemma 7, the function $\beta(y)$ is finite for $y \in (a_k, x] \subset \hat{I}_k$, and it is thus continuous on this interval by the Dominated Convergence Theorem. Although $\beta(a_k)$ could be infinite (in case $a_k \notin \hat{I}_k$), the function β is still right-continuous at a_k by the Monotone Convergence Theorem. Thus $\beta(y)$ is decreasing and continuous on the closed interval $[a_k, x]$.

We need to show that $\int_{[a_k,x]} e^{-\beta(y)}\, dY(y) = Y(x) - X(x)$. Let $g(y) = e^{-\beta(y)}$, then g is increasing and continuous for $y \in [a_k, x]$ with $g(x) = 1$.[18] It remains to show that

$$\int_{[a_k,x]} g(y)\, dY(y) = Y(x) - X(x). \tag{16}$$

If $a_k = x$, then the LHS above is simply $Y(\{x\})$ (the mass of Y at x) because $g(x) = 1$. In this case the above equality holds because $x = a_k \in \overline{A}$ implies $Y_-(x) = X(x)$, and thus $Y(\{x\}) = Y(x) - X(x)$.

Below we consider $a_k < x$. Note that we still have $Y_-(a_k) = X(a_k)$. We prove (16) by approximating the LHS integral by the integrals of increasing step functions. Specifically, consider any partition of the interval $[a_k, x]$ into disjoint intervals $[y_0, y_1] \cup (y_1, y_2] \cup \cdots (y_{n-1}, y_n]$ with $a_k = y_0 < y_1 < \cdots < y_n = x$. For each such partition, define two functions $\underline{g}(y)$ and $\overline{g}(y)$ such that for each $y \in (y_{i-1}, y_i]$, $\underline{g}(y) = g(y_{i-1})$ whereas $\overline{g}(y) = g(y_i)$. Naturally, we also let $\underline{g}(y_0) = g(y_0)$ and $\overline{g}(y_0) = g(y_1)$.

Since g is an increasing function, we have $\underline{g} \le g \le \overline{g}$ point-wise for any partition. Moreover, since g is continuous on the interval $[a_k, x]$, the functions $\underline{g}, \overline{g}$ converge point-wise to g as the partition becomes finer and finer. Thus, by the Dominated Convergence Theorem (which applies since $\underline{g}, \overline{g}$ are uniformly bounded between 0 and 1), we have that $\int_{[a_k,x]} g(y)\, dY(y)$ is the *common limit* of the integrals $\int_{[a_k,x]} \underline{g}(y)\, dY(y)$ and $\int_{[a_k,x]} \overline{g}(y)\, dY(y)$, as the partition becomes arbitrarily fine. Thus, to show (16), it suffices to show the following inequality for every partition:

$$\int_{[a_k,x]} \underline{g}(y)\, dY(y) \le Y(x) - X(x) \le \int_{[a_k,x]} \overline{g}(y)\, dY(y).$$

Using the fact that $\underline{g}$ and $\overline{g}$ are simple functions, we can rewrite their integrals as finite

[18] In case $a_k = -\infty$, we define $\beta(-\infty) = \int_{-\infty}^x \frac{1}{Y(s)-X(s)}\, dX(s)$ and $g(-\infty) = e^{-\beta(-\infty)}$ accordingly. The subsequent arguments also apply to this case.

sums. The above inequalities then become

$$g(y_0)\cdot(Y(y_1)-Y_-(y_0))+\sum_{i=1}^{n-1} g(y_i)\cdot(Y(y_{i+1})-Y(y_i))\le Y(y_n)-X(y_n);$$

$$g(y_1)\cdot(Y(y_1)-Y_-(y_0))+\sum_{i=1}^{n-1} g(y_{i+1})\cdot(Y(y_{i+1})-Y(y_i))\ge Y(y_n)-X(y_n).$$

For the first inequality, we prove by induction that

$$g(y_0)\cdot(Y(y_1)-Y_-(y_0))+\sum_{i=1}^{m-1} g(y_i)\cdot(Y(y_{i+1})-Y(y_i))\le g(y_m)\cdot(Y(y_m)-X(y_m)). \tag{17}$$

The base case $m=1$ says $g(y_0)\cdot(Y(y_1)-Y_-(y_0))\le g(y_1)\cdot(Y(y_1)-X(y_1))$. Since $y_0=a_k$ and $Y_-(y_0)=X(y_0)$, it suffices to show for any $y_0<y_1$:

$$g(y_0)\cdot(Y(y_1)-X(y_0))\le g(y_1)\cdot(Y(y_1)-X(y_1)).$$

This holds trivially if $g(y_0)=0$ or $Y(y_1)-X(y_0)=0$. Otherwise

$$\log\frac{g(y_1)}{g(y_0)}=\beta(y_0)-\beta(y_1)=\int_{y_0}^{y_1}\frac{1}{Y(s)-X(s)}\,dX(s)\ge\int_{y_0}^{y_1}\frac{1}{Y(y_1)-X(s)}\,dX(s)=\log\frac{Y(y_1)-X(y_0)}{Y(y_1)-X(y_1)},$$

as we desire to show.[19] As for the induction step in (17) from m to $m+1$, we need to verify that $g(y_m)(Y(y_m)-X(y_m))+g(y_m)(Y(y_{m+1})-Y(y_m))\le g(y_{m+1})(Y(y_{m+1})-X(y_{m+1}))$. This reduces to

$$g(y_m)\cdot(Y(y_{m+1})-X(y_m))\le g(y_{m+1})\cdot(Y(y_{m+1})-X(y_{m+1})),$$

which can be proved in exactly the same way as above (where we showed this for $m=0$).

The above analysis dealt with the lower bound $\underline{g}$. As for $\overline{g}$, we will similarly show by induction that

$$g(y_1)\cdot(Y(y_1)-Y_-(y_0))+\sum_{i=1}^{m-1} g(y_{i+1})\cdot(Y(y_{i+1})-Y(y_i))\ge g(y_m)\cdot(Y(y_m)-X(y_m)). \tag{18}$$

The base case $m=1$ holds because $Y_-(y_0)=X(y_0)\le X(y_1)$. For the induction step, we need to verify $g(y_m)(Y(y_m)-X(y_m))+g(y_{m+1})(Y(y_{m+1})-Y(y_m))\ge g(y_{m+1})(Y(y_{m+1})-X(y_{m+1}))$,

[19] The final equality here follows by viewing the integral as a Riemann-Stieltjes integral, and making the substitution $z=X(s)$.

which is equivalent to

$$g(y_m)\cdot(Y(y_m)-X(y_m)) \geq g(y_{m+1})\cdot(Y(y_m)-X(y_{m+1})).$$

This clearly holds if $Y(y_m) \leq X(y_{m+1})$, so we assume $Y(y_m) > X(y_{m+1})$. We then have[20]

$$\log\frac{g(y_{m+1})}{g(y_m)} = \int_{y_m}^{y_{m+1}} \frac{1}{Y(s)-X(s)}\,dX(s) \leq \int_{y_m}^{y_{m+1}} \frac{1}{Y(y_m)-X(s)}\,dX(s) = \log\frac{Y(y_m)-X(y_m)}{Y(y_m)-X(y_{m+1})}.$$

This proves the induction step and implies (18).

Therefore (16) holds and the lemma is proved. ■

5.5 λ^* Minimizes Mutual Information When $X(A)=0$

By Lemma 8 we know that $\lambda^* \in \mathcal{M}(X,Y)$. In this section we show $D_{KL}(\lambda \,||\, X\otimes Y) \geq D_{KL}(\lambda^* \,||\, X\otimes Y)$ for any $\lambda \in \mathcal{M}(X,Y)$. We introduce the following result, which ensures that the support of λ is a subset of the support of λ^*.

Lemma 11 *Suppose A has X-measure zero. Then every $\lambda \in \mathcal{M}(X,Y)$ is supported on those points (x,y) with $y \leq x$ and $y,x \in \hat{I}_k$ for the same index k.*

Proof of Lemma 11. First of all, λ is supported on $A^c \times A^c$ because it has marginals X and Y, which assign zero measure to A. Thus we can restrict attention to $x,y \in A^c = \cup_k \hat{I}_k$. Recall that $\hat{I}_k = [a_k,b_k)$ or (a_k,b_k). In either case the left end-point a_k belongs to $\overline{A}$, so that $Y_-(a_k) = X(a_k)$. Thus, just as we showed in the proof of Lemma 6, λ must assign zero measure to the set $S_k = \{(x,y): \ y < a_k \leq x\}$. Since the number of indices k is at most countable, the union of the sets S_k also has λ-measure zero. Note that if $y \leq x$ and y,x belong to $\hat{I}_j$ and $\hat{I}_k$ respectively (with $j \neq k$), then $(x,y) \in S_k$. Thus the union of S_k covers all such points (x,y). Taking the relative complement of this union in $A^c \times A^c$ implies that λ is only supported on the remaining points where x and y do belong to the same $\hat{I}_k$. ■

We now show that the KL-divergence from any $\lambda \in \mathcal{M}(X,Y)$ to $X\otimes Y$ can be decomposed as the sum of the KL-divergence from λ to λ^* and the KL-divergence from λ^* to $X\otimes Y$, so that λ^* uniquely minimizes the KL-divergence. This "triangle equality" does not in general hold, but it holds here because the density of λ^* has a *multiplicatively separable* form, a property that we study further in the next section.

[20] Note that $g(y_m) = e^{-\beta(y_m)} > 0$ for any $y_m > y_0 = a_k$.

Lemma 12 *Suppose A has X-measure zero. Then for every $\lambda \in \mathcal{M}(X,Y)$, it holds that*

$$D_{KL}(\lambda \,||\, X \otimes Y) = D_{KL}(\lambda \,||\, \lambda^*) + K(X,Y),$$

where $K(X,Y) = -1 + \int_{\mathbb{R}} \log \frac{1}{Y(s)-X(s)}\, dX(s)$. Consequently, $D_{KL}(\lambda^* \,||\, X \otimes Y) = K(X,Y) \leq D_{KL}(\lambda \,||\, X \otimes Y)$, and when $K(X,Y) < \infty$ equality holds if and only if $\lambda = \lambda^*$.

Proof of Lemma 12. If λ is not absolutely continuous with respect to $X \otimes Y$, then because λ^* is absolutely continuous with respect to $X \otimes Y$, λ is also not absolutely continuous with respect to λ^*. In this case both $D_{KL}(\lambda \,||\, X \otimes Y)$ and $D_{KL}(\lambda \,||\, \lambda^*)$ are infinite, and the lemma holds.

Suppose instead that λ is absolutely continuous with respect to $X \otimes Y$, admitting a density $h(x,y)$. Then from Lemma 11, it is without loss (up to sets that have measure zero under $X \otimes Y$) to assume $h(x,y) > 0$ only if they belong to the same $\hat{I}_k$ and $y \leq x$. For notational ease, we let T_k denote the "triangular region" associated with $\hat{I}_k$:

$$T_k = \{(x,y) : \ y \leq x \text{ and } y, x \in \hat{I}_k\}.$$

Then h is strictly positive only on $\cup_k T_k$. We also recall from Lemma 7 that the density h^* associated with λ^* is strictly positive on and only on $\cup_k T_k$.

We can write the mutual information induced by λ as follows:

$$\begin{aligned} D_{KL}(\lambda \,||\, X \otimes Y) &= \int_{\mathbb{R}^2} h(x,y) \log h(x,y)\, dX(x)\, dY(y) \\ &= \int_{\cup_k T_k} h(x,y) \log h(x,y)\, dX(x)\, dY(y) \\ &= \int_{\cup_k T_k} h(x,y) \log \frac{h(x,y)}{h^*(x,y)}\, dX(x)\, dY(y) + \int_{\cup_k T_k} h(x,y) \log h^*(x,y)\, dX(x)\, dY(y) \\ &= D_{KL}(\lambda \,||\, \lambda^*) + \sum_k \int_{T_k} h(x,y) \log h^*(x,y)\, dX(x)\, dY(y). \end{aligned} \tag{19}$$

In this derivation one may be worried about absolute integrability affecting the equality between the second line and the third line. This turns out to not be an issue because in the third line, the first integrand $h(x,y)\log \frac{h(x,y)}{h^*(x,y)}$ is bounded below by $h(x,y) - h^*(x,y)$, so the *negative part* of $h(x,y)\log \frac{h(x,y)}{h^*(x,y)}$ is absolute integrable. Meanwhile, as shown below, the second integrand $h(x,y)\log h^*(x,y)$ is bounded below by $-h(x,y)\int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)$, which is also absolute integrable with integral 1.

We now compute $\int_{T_k} h(x,y) \log h^*(x,y)\, dX(x)\, dY(y)$ for each k. Recall that for $x, y \in \hat{I}_k$,

$h^*(x,y) = \frac{1}{Y(x)-X(x)} \cdot e^{-\int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)}$. Thus $\log h^*(x,y) = \log \frac{1}{Y(x)-X(x)} - \int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)$, and it follows that

$$\begin{aligned}&\int_{T_k} h(x,y) \log h^*(x,y)\, dX(x)\, dY(y)\\ &= \int_{T_k} h(x,y) \log\left(\frac{1}{Y(x)-X(x)}\right) dX(x)\, dY(y) \;-\; \int_{T_k} h(x,y) \cdot \left(\int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)\right) dX(x)\, dY(y).\end{aligned} \tag{20}$$

To simplify the first term on the RHS above, we recall that h is the density of $\lambda \in \mathscr{M}(X,Y)$, and thus satisfies the marginal requirements (14) and (15). In particular, (15) gives $\int_{\mathbb{R}} h(x,y)\, dY(y) = 1$ for X-almost every x, and thus $\int_{\hat{I}_k} h(x,y)\, dY(y) = 1$ for X-almost every $x \in \hat{I}_k$. Applying Tonelli's Theorem, we thus have

$$\begin{aligned}\int_{T_k} h(x,y) \log\left(\frac{1}{Y(x)-X(x)}\right) dX(x)\, dY(y) &= \int_{\hat{I}_k} \log \frac{1}{Y(x)-X(x)} \cdot \left(\int_{\hat{I}_k} h(x,y)\, dY(y)\right) dX(x)\\ &= \int_{\hat{I}_k} \log \frac{1}{Y(x)-X(x)}\, dX(x).\end{aligned} \tag{21}$$

As for the second term on the RHS of (20), we have[21]

$$\begin{aligned}&\int_{T_k} h(x,y) \cdot \left(\int_{[y,x)} \frac{1}{Y(s)-X(s)}\, dX(s)\right) dX(x)\, dY(y)\\ &= \int_{y,s,x \in \hat{I}_k :\ y \le s < x} h(x,y) \frac{1}{Y(s)-X(s)}\, dX(s)\, dX(x)\, dY(y)\\ &= \int_{s \in \hat{I}_k} \frac{1}{Y(s)-X(s)} \cdot \left(\int_{y,x \in \hat{I}_k :\ y \le s, x > s} h(x,y)\, dX(x)\, dY(y)\right) dX(s).\end{aligned} \tag{22}$$

Now observe that the integral $\int_{y,x \in \hat{I}_k :\ y \le s < x} h(x,y)\, dX(x)\, dY(y)$ is simply the measure that λ assigns to the region $\{(y,x) \in T_k :\ y \le s < x\}$. Since the different $\hat{I}_k$ are disjoint, we see from Lemma 11 that the λ measure of this region is just equal to the λ-measure of the larger region $\{(y,x) :\ y \le s < x\}$, which is just $Y(s) - X(s)$. Hence, plugging in the RHS of (22), we obtain

$$\begin{aligned}&\int_{T_k} h(x,y) \cdot \left(\int_{[y,x)} \frac{1}{Y(s)-X(s)}\, dX(s)\right) dX(x)\, dY(y)\\ &= \int_{s \in \hat{I}_k} \frac{1}{Y(s)-X(s)} \cdot (Y(s) - X(s))\, dX(s) = \int_{s \in \hat{I}_k} 1\, dX(s) = X(\hat{I}_k).\end{aligned} \tag{23}$$

[21] We can write $\int_y^x \frac{1}{Y(s)-X(s)}\, dX(s)$ as $\int_{[y,x)} \frac{1}{Y(s)-X(s)}\, dX(s)$ because X is non-atomic.

If we now plug (21) and (23) into (20) and then back into (19), we arrive at

$$\begin{aligned}
D_{KL}(\lambda \,||\, X\otimes Y) &= D_{KL}(\lambda \,||\, \lambda^*) + \sum_k \left(\int_{\hat{I}_k} \log \frac{1}{Y(x)-X(x)} \, dX(x) - X(\hat{I}_k) \right) \\
&= D_{KL}(\lambda \,||\, \lambda^*) + \int_{A^c} \log \frac{1}{Y(x)-X(x)} \, dX(x) - X(A^c) \\
&= D_{KL}(\lambda \,||\, \lambda^*) + \int_{\mathbb{R}} \log \frac{1}{Y(x)-X(x)} \, dX(x) - 1 \\
&= D_{KL}(\lambda \,||\, \lambda^*) + K(X,Y),
\end{aligned} \tag{24}$$

where the penultimate equality uses $X(A^c) = 1$. This completes the proof. ■

5.6 Multiplicatively Separable Density Must be λ^*

It remains to prove the last paragraph in the statement of Lemma 1. To do this we show the following analogue of Lemma 12:

Lemma 13 *If $\hat{\lambda} \in \mathcal{M}(X,Y)$ satisfies $\frac{d\hat{\lambda}}{d(X\otimes Y)}(x,y) = h_1(x)\cdot h_2(y)$ for a pair of functions h_1, h_2 that are positive and bounded away from zero, then for every $\lambda \in \mathcal{M}(X,Y)$ it holds that*

$$D_{KL}(\lambda \,||\, X\otimes Y) = D_{KL}(\lambda \,||\, \hat{\lambda}) + D_{KL}(\hat{\lambda} \,||\, X\otimes Y).$$

Lemma 13 immediately implies that $\hat{\lambda}$ minimizes mutual information whenever the minimum is achieved. Thus $\hat{\lambda} = \lambda^*$ whenever the RHS of Eq. (4) is finite.

Proof of Lemma 13. Like before, it is without loss to assume λ admits density h with respect to $X\otimes Y$; otherwise both sides of the desired equality are infinite. We then have

$$\begin{aligned}
D_{KL}(\lambda \,||\, X\otimes Y) &= \int_{\mathbb{R}^2} h(x,y) \log h(x,y) \, dX(x)\, dY(y) \\
&= \int_{\mathbb{R}^2} h(x,y) \log \frac{h(x,y)}{h_1(x)h_2(y)} \, dX(x)\, dY(y) + \int_{\mathbb{R}^2} h(x,y) \log(h_1(x)h_2(y)) \, dX(x)\, dY(y) \\
&= D_{KL}(\lambda \,||\, \hat{\lambda}) + \int_{\mathbb{R}^2} h(x,y) \log h_1(x) \, dX(x)\, dY(y) + \int_{\mathbb{R}^2} h(x,y) \log h_2(y) \, dX(x)\, dY(y).
\end{aligned} \tag{25}$$

Here we made use of the assumption that $h_1(x)$ and $h_2(y)$ are bounded away from zero, which ensures that the negative parts of $h(x,y)\log h_1(x)$ and $h(x,y)\log h_2(y)$ are absolutely integrable.

Since λ has marginals X and Y, we have $\int h(x,y) \; dY(y) = 1$ for X-almost every x and $\int h(x,y) \; dX(x) = 1$ for Y-almost every y. So by the Fubini-

Tonelli Theorem, $\int_{\mathbb{R}^2} h(x,y)\log h_1(x)\ dX(x)\ dY(y) = \int \log h_1(x)\ dX(x)$, and similarly $\int_{\mathbb{R}^2} h(x,y)\log h_2(y)\ dX(x)\ dY(y) = \int \log h_1(x)\ dX(x) = \int \log h_2(y)\ dY(y)$. Plugging these into Eq. (25), we obtain

$$D_{KL}(\lambda \,||\, X \otimes Y) = D_{KL}(\lambda \,||\, \hat{\lambda}) + \int \log h_1(x)\ dX(x) + \int \log h_2(y)\ dY(y).$$

Since this equality holds in particular for $\lambda = \hat{\lambda}$, we obtain $D_{KL}(\hat{\lambda} \,||\, X \otimes Y) = \int \log h_1(x)\ dX(x) + \int \log h_2(y)\ dY(y)$. Therefore it follows that

$$D_{KL}(\lambda \,||\, X \otimes Y) = D_{KL}(\lambda \,||\, \hat{\lambda}) + D_{KL}(\hat{\lambda} \,||\, X \otimes Y),$$

as we desire to show. ■

6 Conclusion

This paper takes a first step at exploring the implication of Bayesian privacy concerns in designing efficient and optimal auctions. Because in many environments the identify of the auction winner and the price she paid are made public, we focus on minimizing the privacy loss to the winner. We propose to quantify this loss by the mutual information between the winner's type and her payment. Our main result suggests a justification for standard second-price auctions in the sense that they are the most privacy-preserving auctions among all ex-post individually-rational dominant-strategy mechanisms.

Our proof technique relies on a new result on the lower bound of the mutual information between two ordered random variables. We use this result, in conjunction with the feature of dominant strategy mechanism that the winner's payment is a mean-preserving spread of the second highest bid, to establish our main theorem. A natural follow-up to our results in to study ex-post individually-rational privacy preserving auctions under other solution concepts. Since this would necessitate different proof techniques, it is left as an open question for future research.

References

Alvarez, Ramiro, and Mehrdad Nojoumian. 2020. "Comprehensive survey on privacy-preserving protocols for sealed-bid auctions." *Comput. Secur.*, 88.

Arnold, Sebastian, Ilya Molchanov, and Johanna F. Ziegel. 2020. "Bivariate distributions with ordered marginals." *Journal of Multivariate Analysis*, 177: 104585.

Bergemann, Dirk, and Johannes Hörner. 2018. "Should First-Price Auctions Be Transparent?" *American Economic Journal: Microeconomics*, 10(3): 177–218.

Butucea, Cristina, Jean-François Delmas, Anne Dutfoy, and Richard Fischer. 2018. "Maximum entropy distribution of order statistics with given marginals." *Bernoulli*, 24(1): 115 – 155.

Calzolari, Giacomo, and Alessandro Pavan. 2006*a*. "Monopoly with Resale." *The RAND Journal of Economics*, 37(2): 362–375.

Calzolari, Giacomo, and Alessandro Pavan. 2006*b*. "On the optimality of privacy in sequential contracting." *Journal of Economic Theory*, 130(1): 168–204.

Das Varma, Gopal. 2003. "Bidding for a process innovation under alternative modes of competition." *International Journal of Industrial Organization*, 21(1): 15–37.

Dworczak, Piotr. 2020. "Mechanism Design With Aftermarkets: Cutoff Mechanisms." *Econometrica*, 88(6): 2629–2661.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. "Calibrating Noise to Sensitivity in Private Data Analysis." In *Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science.* , ed. Halevi S. and Rabin T. Springer, Berlin Heidelberg.

Eilat, Ran, Kfir Eliaz, and Xiaosheng Mu. 2021. "Bayesian Privacy." *Theoretical Economics*, 16: 1557–1603.

Giovannoni, Francesco, and Miltiadis Makris. 2014. "Reputational Bidding." *International Economic Review*, 55(3): 693–710.

Goeree, Jacob K. 2003. "Bidding for the future: signaling in auctions with an aftermarket." *Journal of Economic Theory*, 108(2): 345–364.

Heffetz, Ori, and Katrina Ligett. 2014. "Privacy and Data-Based Research." *Journal of Economic Perspectives*, 28(2): 75–98.

Katzman, Brett E., and Matthew Rhodes-Kropf. 2008. "The Consequences of Information Revealed in Auctions." *Applied Economics Research Bulletin*, 2: 53–87.

Molnár, József, and Gábor Virág. 2008. "Revenue maximizing auctions with market interaction and signaling." *Economics Letters*, 99(2): 360–363.

Myerson, Roger B. 1981. "Optimal Auction Design." *Mathematics of Operations Research*, 6(1): 58–73.

Naor, Moni, Benny Pinkas, and Reuban Sumner. 1999. "Privacy preserving auctions and mechanism design." 129–139. ACM.

Pai, Mallesh M., and Aaron Roth. 2013. "Privacy and Mechanism Design." *SIGecom*

Exch., 12(1): 8–29.