

Privacy-Preserving Auctions^{*}

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

Abstract

In many auction settings, the auctioneer must disclose the winner's identity and payment. We characterize the auction that minimizes the winner's privacy loss among those that maximize total surplus or seller revenue and are strategy-proof. Privacy loss is defined by what an outside observer learns from the disclosed payment, measured as the mutual information between the payment and the winner's valuation. Under ex-post individual rationality and a monotone hazard rate, privacy loss is minimized by the second-price auction with deterministic payments. When only interim individual rationality is required, the most privacy-preserving auction uses stochastic ex-post payments.

Keywords: Bayesian privacy, Mutual information, Auction theory

^{*}Financial support from BSF grant 2022294 is gratefully acknowledged.

[†]School of Economics, Tel-Aviv University; eilatr@tauex.tau.ac.il

[‡]School of Economics, Tel-Aviv University and King's College London Business School; 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 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).

In light of the above, this paper addresses the following 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?

Measuring privacy loss. To investigate the question, we employ the Bayesian approach to measuring privacy loss, as proposed in [Eilat, Eliaz and Mu \(2021\)](#).

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

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

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

⁴The 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. For instance, a recent OECD report on the implementation of its Recommendation on Public Procurement ([OECD, 2025](#)) encourages transparency by saying that “transparency can help mitigate risks of favouritism, corruption, and inefficiency by making procurement processes more open and accountable”. On the other hand, it warns against the potential hazards of transparency by saying that in some cases it may “facilitate collusion among bidders, especially in markets with limited suppliers”. See also [Cramton and Schwartz \(2000\)](#) for a discussion of the risks of full transparency.

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 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 Bayesian privacy loss. For most of the analysis, we use 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 three equivalent methods for computing the same quantity.

Preview of the model. We consider a pure private values environment (where values are drawn from a distribution that satisfies the monotone hazard rate condition) 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. As is common in many “real-world” settings (as in the examples mentioned above), 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.

For most of our analysis, we focus on the class of *efficient* mechanisms with a dominant-strategy equilibrium under an *ex-post* participation constraint – that is, after the auction concludes and the price is announced, the winner retains the option to decline completing the deal (we discuss the case of interim participation constraint in Section 5). Because payments can in principle be stochastic, this class contains many mechanisms. 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. We view this objective as a *conservative approach* for privacy preservation

in the sense that the designer’s first priority is to maximize either surplus or revenue, and his secondary desideratum is to minimize privacy loss.⁵

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 2 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.⁶ Finally, the class of dominant-strategy 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.⁷

Preview of the main results. In principle, introducing randomness into the winner’s payment can obfuscate the information revealed about his willingness to pay. For example, suppose that the realized second-highest value is θ_2 . Dominant-strategy incentive compatibility requires the winner’s expected payment to be θ_2 . This requirement can be implemented by any stochastic pricing rule with mean θ_2 . An ex-post participation constraint imposes an additional restriction: the realized price cannot exceed the winner’s value, θ_1 . Thus, a stochastic pricing rule can satisfy both restrictions if it draws the price from a distribution supported on $[0, \theta_1]$ with mean θ_2 .

Our first main result shows that, under the assumptions of our model, stochastic payments are *ineffective* in reducing the winner’s privacy loss. Theorem 1 establishes that the standard second-price auction, which uses a deterministic pricing rule, minimizes the winner’s privacy loss among all efficient, dominant-strategy incentive compatible and ex-post individually rational mechanisms. The-

⁵An alternative approach to modeling the designer’s goal would be to consider an objective that is a weighted sum of surplus (or revenue) and privacy loss. However, this approach would require a different set of solution techniques, and is left as an open question for future research.

⁶While dominant-strategy incentive compatibility and ex-post individual rationality do not make use of the prior distribution, the revenue-maximizing designer (in Section 4.4) is fully Bayesian. In our model, so is the outside observer who gleans information from the outcome of the mechanism.

⁷The auction examples in the opening paragraph highlight the practice of disclosing both the winners and the prices they paid. Nonetheless, the auction formats used in those settings are not always strategy-proof.

orem 2 extends this finding to the revenue-maximization setting, with the addition of an optimal reserve price.

To prove Theorem 1, we first establish a general lemma characterizing a tight lower bound on the mutual information between two *ordered* random variables with given marginal distributions (Lemma 1). We then verify that the joint distribution of winner types and payments induced by the second-price auction attains the mutual-information lower bound identified in the lemma. Finally, we observe that any other dominant-strategy mechanism induces a payment distribution that is a mean-preserving spread of the distribution of the second-highest bidder type, and we show that any mean-preserving spread can only increase the lower bound. The proof of Theorem 2 follows the same outline.

An ex-post participation constraint can be interpreted as a type-dependent price cap. Alternatively, one may relax participation to hold only in the interim and instead impose price caps that are uniform across types – that is, all buyer types are constrained to pay no more than a fixed constant K . We show that the conclusion is reversed in this case: a stochastic payment mechanism minimizes the winner’s privacy loss. This mechanism asks for a randomized payment of either 0 or K , with probabilities chosen to satisfy incentive compatibility. If payments are allowed to be arbitrarily high (as $K \rightarrow \infty$), then efficiency or revenue maximization can be achieved with privacy loss converging to zero, while maintaining incentive compatibility and interim individual rationality.

We emphasize that such stochastic pricing rules are not intended as descriptive or normative models of privacy-preserving auction design: we neither claim that such mechanisms are used in real-world auctions nor advocate for their adoption. Rather, we present them to demonstrate a possibility for randomization to enhance Bayesian privacy, so long as interim individual rationality is acceptable.

Organization. The paper is organized as follows. In Section 2 we review the related literature. In Section 3 we present the framework and define our privacy notion. Section 4 presents our main results, Theorem 1 and Theorem 2, which establish that the second-price auction (with a reserve price) minimizes privacy loss among efficient (or revenue-maximizing) mechanisms that are dominant-strategy incentive compatible and ex-post individually rational. In this section we also present and discuss the key Lemma 1 which characterizes the minimum mutual information between two ordered random variables with given marginal distributions. The complete proof of Lemma 1 is presented in Appendix A. Section 5 shows the usefulness of random pricing schemes when the participation constraint holds

only at the interim, and uniform price caps are imposed. Concluding remarks are given in Section 6.

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 does not specify 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.⁸

⁸Dworczak (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 problem is obtained *only* when we impose ex-post individual rationality, a restriction which is orthogonal to the condition identified in Dworczak (2020).

A parallel line of research examines auctions in the presence of signaling concerns. [Bos and Truys \(2021\)](#) study English and second-price auctions in which bidders’ payoffs include an additive term that depends on the inference an outsider forms about their value, given the winner’s identity and payment. They show that, in the unique equilibrium they identify, the second-price auction yields higher expected revenue than in the English auction. Using the same bidder preferences, [Bos and Pollrich \(2025\)](#) analyze auction and disclosure design and show that expected revenue is not monotonic in the amount of information revealed.

A related literature studies the effect of disclosure policies on particular post-auction interaction between the bidders and third parties. [Calzolari and Pavan \(2006b\)](#) study the optimal disclosure of information between an upstream and a downstream principal who contract sequentially with the same agent. [Calzolari and Pavan \(2006a\)](#) 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. [Haupt and Hitzig \(2024\)](#) study how a designer can implement a social choice rule while gradually eliciting agents’ private information in a way that minimizes the revelation of information irrelevant to the final decision.

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.

A second approach in computer science applies cryptographic tools to ensure that the communication between the seller and the bidders discloses only information that is necessary to run the mechanism. The early papers in this literature focused on guaranteeing the privacy of the bidder-bid relationship and the secrecy of the bids, while also ensuring the correctness and trustworthiness of the outcome (see [Naor, Pinkas and Sumner \(1999\)](#), [Parkes et al. \(2008\)](#) and the survey by [Alvarez and Nojournian \(2020\)](#)). However, this line of research is not applicable in our setting, where there is an *exogenous requirement* to publicly reveal the identity of the winner and the price paid.

More recently, [Canetti, Fiat and Genczarowski \(2023\)](#) introduced a novel approach to privacy that is complementary to ours. Using tools from cryptography, they show that a seller can credibly prove to bidders that he committed to a mechanism that is individually rational and incentive compatible without disclosing any information about the mechanism, and without relying on a third party for verification. This ensures that the only information that is disclosed is the outcome - which is the starting point of our paper.

3 Model

A seller owns one indivisible good, whose value to him is normalized to zero. There are n potential risk-neutral buyers. The number n is publicly known, but the identities of the bidders are not. The willingness to pay (i.e. “type”) of buyer $i \in N = \{1, \dots, n\}$, denoted θ_i , is privately and independently drawn from a distribution F over $[\underline{\theta}, \bar{\theta}]$ with $\bar{\theta} > \underline{\theta} \geq 0$. We restrict attention to distributions with strictly positive and continuously differentiable densities f over the interval $[\underline{\theta}, \bar{\theta}]$, which exhibit a monotone hazard rate – i.e., the ratio $(1 - F(\theta))/f(\theta)$ is decreasing in θ .⁹

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 ex-post:

⁹We use this condition in Step 2 of the proof of Theorem 1. We do not know if this condition is necessary for our result or if it can be further relaxed.

a buyer can opt out and get a payoff of zero after being informed of how much he has to pay (i.e., we impose the ex-post individual rationality constraint, EPIR). We assume that the seller can commit to the details of the mechanism.¹⁰

If one of the buyers wins the good, then the winner's identity and his payment are publicly disclosed. Until Section 4.4, 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_P^σ and G_W^σ denote the marginal distributions of P^σ and W^σ , respectively. $G_{W|P}^\sigma$ denotes the conditional distribution of W^σ given P^σ , while $G_{P|W}^\sigma$ denotes the conditional distribution of P^σ given W^σ . An outsider, who observes the winner's identity and payment (i.e. the realization of P^σ), updates her beliefs about the winner's type (the value of W^σ).

The privacy loss entailed by a mechanism M (along with its DSE σ) is quantified as the *mutual information* between the winner's willingness to pay (the random variable W^σ) and the paid price (the random variable P^σ). Mutual information measures the *reduction in uncertainty* about one variable given knowledge of the other. Rooted in information theory, it is calculated using the joint and marginal probability distributions of the variables.¹¹ Formally:

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):*

$$MI(W^\sigma, P^\sigma) = D_{KL}(G^\sigma || G_W^\sigma \otimes G_P^\sigma) \quad (\text{MI})$$

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

Equivalent representations of privacy loss. Mutual information can be com-

¹⁰E.g., the seller 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.

¹¹Unlike simpler measures such as covariance or correlation, which focus solely on linear dependencies, mutual information captures a broader range of statistical dependencies.

puted 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:

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

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

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

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 $H(\cdot)$ denotes differential entropy. Here, privacy loss is computed as the expected entropy reduction in the belief about winner type. Because the entropy $H(W^\sigma)$ is the same across all efficient mechanisms, this representation suggests that minimizing privacy loss is equivalent to maximizing expected entropy of the posterior belief (i.e., maximizing residual uncertainty) about the winner's 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, $G_W^\sigma(w) = w^2$ 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, $G_P^\sigma(p) = 1 - (1 - p)^2$, with a density of $2(1 - p)$. The joint distribution $G^\sigma(w, p)$ is uniform over the triangle $0 \leq p \leq w \leq 1$, with a density of 2. Plugging these into the KL-divergence formula, we obtain:

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

□

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 $\mathcal{M}$ 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 ex-post payoff is non-negative (i.e., ex-post 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:

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

Remark. There are at least two distinct approaches to studying privacy considerations in mechanism design. The first incorporates privacy concerns directly into agents’ preferences and studies their implications for incentive compatibility and behavior. Such concerns may stem from anticipated future interactions or from an intrinsic preference for privacy. The second approach, which we adopt in this paper, retains the standard rational benchmark for behavior and instead quantifies privacy loss using information-theoretic measures.¹²

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 $q : [\underline{\theta}, \bar{\theta}]^n \rightarrow \Delta(I)$ 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 $t_i : [\underline{\theta}, \bar{\theta}]^n \rightarrow \Delta(\mathbb{R}_+)$ 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 $T_i(\theta) = \mathbb{E}[t_i(\theta) | \theta]$ 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 mech-

¹²This approach is consistent with the observation that, in some contexts, individuals do not fully internalize their privacy concerns, a phenomenon often referred to as the “privacy paradox”; see, e.g., [Athey, Catalini and Tucker \(2017\)](#). The paradox captures the mismatch between stated privacy preferences and observed behavior. Put differently, individuals may narrowly bracket privacy decisions: when privacy is made salient, they express concern, yet when the possibility of privacy loss is only implicit, they fail to account for its consequences.

anisms 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.¹³

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

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

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

We then observe that $u_i(\underline{\theta}, \theta_{-i}) \leq 0$ because only buyers make payments to the seller, and because the good is never allocated to type $\underline{\theta}$ by Eq. (1). But ex-post individual rationality requires $u_i(\underline{\theta}, \theta_{-i}) \geq 0$, so $u_i(\underline{\theta}, \theta_{-i}) = 0$ for any θ_{-i} .

By plugging Eq. (1) into Eq. (DSIC), we obtain:¹⁴

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

Thus, the designer's problem reduces to the following: Among stochastic ex-post payment functions $t_1 \dots t_n$ that satisfy the incentive compatibility constraint Eq. (2), as well as the ex-post individual rationality constraint which requires every realized payment to satisfy $t_i(\theta) \leq \theta_i$ for the winner and $t_i(\theta) = 0$ for the losers

¹³I.e., the allocation is unique up to zero measure type profiles.

¹⁴It is well known that with deterministic payments, transfers in an efficient mechanism coincide with the second-price auction. See e.g. [Green and Laffont \(1977\)](#) and [Holmström \(1979\)](#).

(since we assume non-negative payments), find those that minimize the mutual information between winner's type and payment.

4 Second-Price Auction Minimizes Privacy Loss

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, 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 next.

A simple example is the following. Given a profile of bids where θ_1 is the highest bid and θ_2 is the second highest, the mechanism can determine the winner's price by randomizing between 0 and θ_1 with probabilities $1 - \theta_2/\theta_1$ and θ_2/θ_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 $\alpha = \theta_2$ and $\beta = \theta_1 - \theta_2$ that is supported on $[0, \theta_1]$, whose mean is precisely θ_2 .

Theorem 1 establishes that stochastic payments as in these examples will generate a higher privacy loss compared to the deterministic SPA.

Theorem 1 *The standard SPA with deterministic payments minimizes the privacy loss among all efficient, DSIC and ex-post individually rational mechanisms.*¹⁵

4.1 Minimum Mutual Information Coupling Under FOSD

In preparation for proving Theorem 1 we begin by deriving a key lemma. Given two distributions X and Y on $\mathbb{R}$, where X is non-atomic and $X(s) \leq Y(s) \forall s$, we derive a lower bound on the mutual information between any two jointly distributed

¹⁵It follows from the proof below that randomized payments strictly increase the privacy loss when the inverse hazard rate $\frac{1-F(\theta)}{f(\theta)}$ is strictly decreasing in θ .

random variables with marginal distributions X and Y .^{16,17} A key observation for this result is that the joint density that attains this lower bound has a multiplicative form.

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

$$\inf_{\lambda \in \mathcal{M}(X, Y)} D_{KL}(\lambda \parallel X \otimes Y) = -1 + \int_{\mathbb{R}} \log \frac{1}{Y(s) - X(s)} dX(s). \quad (3)$$

The infimum above is achieved as minimum whenever the RHS of Eq. (3) is finite, in which case the unique minimizer λ^* is the joint distribution defined by

$$\frac{d\lambda^*}{d(X \otimes Y)}(x, y) = \frac{1}{Y(x) - X(x)} \cdot e^{-\int_y^x \frac{1}{Y(s) - X(s)} dX(s)} \cdot \mathbf{1}_{Y(x) > X(x)} \cdot \mathbf{1}_{y \leq x}. \quad (4)$$

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. (3) is finite and if there exists $\hat{\lambda} \in \mathcal{M}(X, Y)$ such that $\frac{d\hat{\lambda}}{d(X \otimes Y)}(x, y) = h_1(x) \cdot h_2(y)$ for $(X \otimes Y)$ -almost every (x, y) with $y \leq x$ and for a pair of functions h_1, h_2 that are positive and bounded away from zero, then $\hat{\lambda} = \lambda^*$ as described above.

The proof of Lemma 1 appears in Appendix A. Applying it 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. (3) provides a lower bound on privacy loss, given by:

$$-1 + \int_{\mathbb{R}} \log \frac{1}{G_P(s) - G_W(s)} dG_W(s). \quad (5)$$

¹⁶To economize on notation we slightly abuse notation here by letting $X(s)$ and $Y(s)$ denote also the respective cumulative distribution functions.

¹⁷This result generalizes the bivariate case of Theorem 5.4 in Butucea et al. (2018) to environments where the marginal distributions may not admit densities; see also Arnold, Molchanov and Ziegel (2020). This generalization is important for our application, as the payment distribution is endogenously chosen and may not have a density (for example in Section 4.4 below). For this reason, our proof in Appendix A is also more involved than in the previous work.

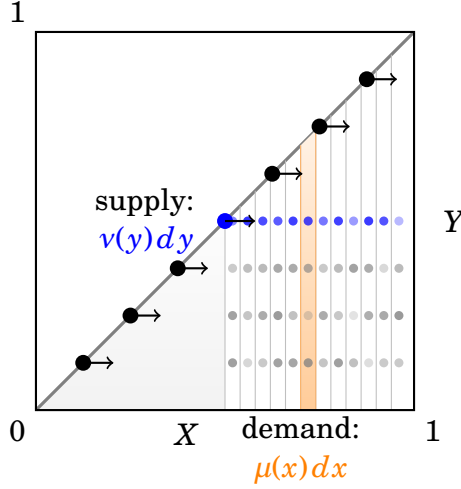


Figure 1: Supply and demand interpretation for Lemma 1

We emphasize that the condition $\mathcal{X} \geq \mathcal{Y}$ 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.

Optimal Transport Interpretation for Lemma 1

Lemma 1 characterizes the joint distribution λ^* in Equation (4) as the one that minimizes the mutual information between two ordered random variables with given marginal distributions. This minimization can be viewed as a variant of an *optimal transport* problem: given the marginals of the two random variables, the goal is to find a coupling that minimizes the expected value of a cost function $c(x, y)$, under the convention that any pair (x, y) with $x < y$ incurs infinite cost.¹⁸ A distinctive feature of the present problem, in contrast to classical optimal transport, is that the cost function itself depends on the chosen joint distribution. The objective therefore takes the form $\int c(x, y, \lambda) d\lambda(x, y)$. For instance, if the joint distribution λ has density $g(x, y)$, then $c(x, y, \lambda) = \log\left(\frac{g(x, y)}{g(x)g(y)}\right)$. Because the cost function depends on λ , the objective is *nonlinear* in the coupling λ and standard results in optimal transport do not apply.

Nonetheless, the connection to optimal transport can be used to give a supply-and-demand interpretation for Equation (4), providing intuition for the optimal joint distribution. To illustrate this, consider a simplified setting in which the

¹⁸We thank Yakov Babichenko for suggesting this convention.

cumulative distribution functions of X and Y are both continuous and supported on the unit interval $[0, 1]$. Let μ and ν denote their respective densities. Figure 1 illustrates this setup.

Suppose that at each point (y, y) along the main diagonal of the unit square $[0, 1]^2$ there is a supply point – a “warehouse” of infinitesimal size dy – containing mass $\nu(y)dy$ of some homogeneous good. In Figure 1, a few of these warehouses are illustrated as circles along the main diagonal. Each warehouse is permitted to transfer its contents only *to the right*, that is, from (y, y) to any destination (x, y) with $x \in [y, 1]$.

We seek a “transfer plan” that distributes mass from the warehouses to locations to their right, in a way that satisfies two key properties. First, each vertical slice of infinitesimal width dx at coordinate x must receive exactly $\mu(x)dx$ units of mass, combined from all warehouses. Note that such a transfer plan effectively induces a “joint distribution” over sources and destinations – specifying how much mass each destination receives from each source – which corresponds to the probability measure λ in our original formulation.

The second property we require is that, among all admissible transfer plans satisfying the marginal constraint (each vertical slice receives the required amount) and the directional constraint (mass can only be transferred to the right), we seek the one that introduces the *least structure* – that is, the plan that minimizes the mutual information between X and Y . Loosely speaking, we are looking for a plan under which knowing the origin warehouse reveals as little as possible, on average, about the distribution of delivery locations.

Equation (4) characterizes the optimal plan. It follows a simple rule. Fix a destination point x , and suppose all mass destined for locations less than x has already been allocated, and that now a decision has to be made regarding the mass that is allocated at x . Two quantities are now relevant:

1. the *local demand* at x , given by $\mu(x)dx$, and
2. the *overall remaining supply* in the system that can still be allocated to x , given by $Y(x) - X(x)$.

Here, $Y(x)$ is the total mass that could, in principle, be delivered to location x from all warehouses, and $X(x)$ is the total mass that has already been delivered to locations less than x . Their difference represents the undelivered mass still available to fulfill demand at x .

The optimal plan can be described as follows: from each warehouse y , it assigns to the vertical slice of width dx at position x a proportion $r(x)dx$ of the mass

from warehouse y that has not yet been allocated, where

$$r(x) = \frac{\mu(x)}{Y(x) - X(x)}. \quad (6)$$

The ratio $r(x)$ quantifies how large the demand at x is relative to the total supply still available to serve it. A higher value of the ratio indicates more urgent demand: the local need is high relative to what remains to be allocated, so warehouse y must contribute a greater share of its available supply to destination x . Equation (4), interpreted as the joint distribution of sources y and destinations x , follows from this rule by direct computation.¹⁹

An important property of this mass distribution rule is that it is *locally memoryless*: the proportion of mass directed from warehouse y to position x , out of the warehouse's yet-to-be-allocated supply, depends only on local quantities: the current demand, $\mu(x)$, and the total remaining unallocated supply from *all* warehouses, $Y(x) - X(x)$. It does not depend on the identity of the warehouse y , nor on how mass has been allocated to other destinations. In other words, all unallocated mass at x is treated identically, governed by the same local rule, regardless of its origin or delivery history.

This also explains why the marginal constraint is satisfied. Because every warehouse assigns to a slice of width dx at destination x the same proportion $r(x)dx$ of the mass it still has available when it reaches x , it follows that this slice receives the fraction $r(x)dx$ of the entire undelivered mass from all warehouses at x , namely $Y(x) - X(x)$. A direct computation shows that this mass equals $\mu(x)dx$, exactly matching the required demand.

To gain intuition for why the specified plan minimizes the mutual information between X and Y , it is helpful to look at the problem *backwards*: fix a destination x , pick a grain of mass located there, and ask – From which warehouse did this grain most likely come? For any two candidate sources $y_1 < y_2 \leq x$, the allocation

¹⁹To see this, let $M_y(x)$ denote the remaining mass from warehouse y that has not yet been allocated to destinations less than x . Then $M_y(x)$ evolves according to the differential equation

$$\frac{dM_y(x)}{dx} = -r(x) \cdot M_y(x).$$

with initial condition $M_y(y) = v(y)dy$. The solution is $M_y(x) = v(y)dy \cdot e^{-\int_y^x r(s)ds}$. The amount of mass delivered from y to x is therefore:

$$M_y(x) \cdot r(x)dx = \frac{1}{Y(x) - X(x)} \cdot e^{-\int_y^x \frac{\mu(s)}{Y(s) - X(s)} ds} \cdot \mu(x) \cdot v(y) \cdot dy \cdot dx,$$

which matches the expression in Equation (4), when written in terms of its density with respect to Lebesgue measure.

rule in Equation (4) implies

$$\frac{\lambda_{Y|X=x}^*(y_1)}{\lambda_{Y|X=x}^*(y_2)} = \frac{v(y_1)}{v(y_2)} e^{-\int_{y_1}^{y_2} r(s) ds}.$$

Crucially, this ratio is independent of the destination x . That is, while observing the destination x restricts the range of possible sources to those $y \leq x$, it does not change the relative likelihoods between any two sources within that range. In this sense, knowing the realization of the destination, X , reveals “relatively little” about the origin, Y . The proof of Lemma 1 formalizes this intuition, showing that the plan characterized by Equation (4) minimizes the mutual information $MI(X; Y)$ over all admissible allocation rules.

4.2 Proof of Theorem 1

In this subsection we prove Theorem 1. The proof proceeds in two steps. First, we show that the SPA induces a joint distribution of winner type and payment that achieves the mutual information lower bound established in Lemma 1, given its marginal distributions. Second, 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 weakly increases the mutual information lower bound compared to the one induced by the SPA.

Step 1. We show that the joint distribution of winner type and payment under the SPA achieves the mutual information lower bound in Eq. (3), given its marginal distributions. Note that $X = G_W = F^n$ is the CDF of 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, from the chain rule for Radon–Nikodym derivatives, we can calculate that

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

Define $h_1(w) = \frac{1}{nF(w)^{n-1}} \geq \frac{1}{n}$ and $h_2(p) = \frac{1}{1 - F(p)} \geq 1$. 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.²⁰

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.

Step 2. 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, $G_P(\bar{\theta}) = 1 = G_L(\bar{\theta})$. From this we will show that G_L minimizes the RHS of Eq. (5) among all possible G_P , which will prove Theorem 1.

We need to show that

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

Rearranging, this is equivalent to

$$\int_{\underline{\theta}}^{\bar{\theta}} \log \frac{G_L(s) - G_W(s)}{G_P(s) - G_W(s)} dG_W(s) \geq 0.$$

For any two real numbers $a > 0$ and $b \geq 0$, we have $\log \frac{b}{a} \leq \frac{b}{a} - 1$ and thus $\log \frac{a}{b} \geq \frac{a-b}{a}$. Thus,

$$\int_{\underline{\theta}}^{\bar{\theta}} \log \frac{G_L(s) - G_W(s)}{G_P(s) - G_W(s)} dG_W(s) \geq \int_{\underline{\theta}}^{\bar{\theta}} \frac{G_L(s) - G_P(s)}{G_L(s) - G_W(s)} dG_W(s).$$

We then observe that

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

²⁰To see this, write $\int_0^1 \log \frac{1}{n(1-x)x^{n-1}} d(x^n) = I_1 + I_2 + I_3$, where $I_1 = -\int_0^1 nx^{n-1} \log n dx$, $I_2 = -\int_0^1 nx^{n-1} \log(1-x) dx$ and $I_3 = -\int_0^1 nx^{n-1} \log x^{n-1} dx$. By straightforward computation, we have $I_1 = -\log n$. Using Eq. (4.293.8) in Gradshteyn and Ryzhik (2007), we have $I_2 = \sum_{k=1}^n \frac{1}{k}$. Integration by parts yields $I_3 = \frac{n-1}{n}$. Combining these results, we obtain $\int_0^1 \log \frac{1}{n(1-x)x^{n-1}} d(x^n) = \sum_{k=1}^n \frac{1}{k} - \log n + \frac{n-1}{n}$, which is finite for every $n \geq 2$.

Thus it suffices to show

$$\int_{\underline{\theta}}^{\bar{\theta}} (G_L(s) - G_P(s)) \cdot \frac{f(s)}{1 - F(s)} ds \geq 0. \quad (7)$$

From the mean-preserving spread property, we have $\int_{-\infty}^t (G_L(s) - G_P(s)) ds \leq 0$ for every $t \in \mathbb{R}$. Moreover, at $t = \bar{\theta}$ we have equality because $\int_{-\infty}^{\bar{\theta}} (G_L(s) - G_P(s)) ds$ evaluates to the difference between the mean of G_L and the mean of G_P , since they are both supported on $(-\infty, \bar{\theta}]$. Thus, from $\int_{-\infty}^{\bar{\theta}} (G_L(s) - G_P(s)) ds = 0 \geq \int_{-\infty}^t (G_L(s) - G_P(s)) ds$, we have

$$\int_t^{\bar{\theta}} (G_L(s) - G_P(s)) ds \geq 0 \quad \text{for every } t \leq \bar{\theta}. \quad (8)$$

Now note that $C(s) := \frac{f(s)}{1 - F(s)}$, which shows up in the desired integral inequality (7), is precisely the hazard rate that we assumed to be increasing in s . If we write $C(s) = C(\underline{\theta}) + \int_{\underline{\theta}}^s C'(t) dt$ with C' non-negative, then

$$\int_{\underline{\theta}}^{\bar{\theta}} (G_L(s) - G_P(s)) \cdot C(s) ds = C(\underline{\theta}) \cdot \int_{\underline{\theta}}^{\bar{\theta}} (G_L(s) - G_P(s)) ds + \int_{\underline{\theta}}^{\bar{\theta}} (G_L(s) - G_P(s)) \int_{\underline{\theta}}^s C'(t) dt ds.$$

The first term on the RHS above is non-negative by the inequality (8) at $t = \underline{\theta}$. The second term is a double integral that can be rearranged to $\int_{\underline{\theta}}^{\bar{\theta}} C'(t) \int_t^{\bar{\theta}} (G_L(s) - G_P(s)) ds dt$ after changing the order of integration, and it is also non-negative by (8). This proves the desired inequality (7), and thus among all possible payment distributions G_P , G_L minimizes the mutual information lower bound in Eq. (5).

4.3 On the Privacy Loss of the Second Price Auction

The proof of Theorem 1 establishes that the privacy loss associated with the standard second-price auction (with deterministic payments) is given by $\text{MI}_{\text{SPA}}(n) \equiv -1 + \int \log \frac{1}{G_L(s) - G_W(s)} dG_W(s)$, where G_W and G_L are the cumulative distribution functions of the winner type and payment, respectively (i.e. the first- and second-highest realizations out of n independent draws from F). Using the computation in Footnote 20, we can simplify this expression to the following:

$$\text{MI}_{\text{SPA}}(n) = -1 + \sum_{k=1}^n \frac{1}{k} - \log n + \frac{n-1}{n} = \sum_{k=1}^{n-1} \frac{1}{k} - \log n. \quad (9)$$

By Theorem 1, $\text{MI}_{\text{SPA}}(n)$ is the minimal privacy loss achievable by any effi-

cient, dominant-strategy incentive compatible, and ex-post individually rational auction with n players. Equation (9) leads to three key observations.

First, as indicated by the right-hand side of Eq. (9), the privacy loss of the second-price auction is *independent* of the underlying type distribution F .

Second, $MI_{SPA}(n)$ increases with the number of participating bidders, n ; this follows from the simple inequality that $\log(n+1) - \log n < \frac{1}{n}$. Intuitively, as n increases, both the highest type and the second highest type increase on average, but their expected difference becomes *smaller*. Thus, observing the second highest type imposes a tighter constraint on the possible values of the highest type, thereby increasing the privacy loss as the number of bidders grows.²¹

Finally, in light of this second observation, a natural question arises: as the number of players grows to infinity, does the privacy loss of the SPA diverge, or does it remain bounded? Eq. (9) shows that the latter is true. Specifically, by taking the limit of the right-hand side of Eq. (9) as n approaches infinity, we find that the privacy loss implied by the second-price auction is bounded from above and converges to a constant:

$$\lim_{n \rightarrow \infty} MI_{SPA}(n) = \lim_{n \rightarrow \infty} \left(\sum_{k=1}^{n-1} \frac{1}{k} - \log n \right) = \gamma. \quad (10)$$

where γ on the right-hand side of Eq. (10) is *Euler's constant*.

We record these three observations in the following proposition:

Proposition 1 *For the standard second-price auction with n players:*

1. *The privacy loss $MI_{SPA}(n)$ is independent of the underlying type distribution F ;²²*
2. *The privacy loss $MI_{SPA}(n)$ is strictly increasing in the number of bidders n ;*
3. *As the number of bidders n grows to infinity, the privacy loss $MI_{SPA}(n)$ converges to a finite limit, specifically the Euler constant γ .*

²¹An interesting direction for future work is to study the implications for the winner's privacy loss when the number of bidders is unknown.

²²Recall that our privacy loss notion measures the average reduction in uncertainty about the winner type, given observation of his payment. This result shows that the amount of reduction is independent of the prior belief F , even though residual uncertainty does depend on F .

4.4 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 $\theta_i = \max_{1 \leq j \leq n} \theta_j \geq r$. 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 $\mathcal{E}^\sigma$ that the good is allocated*. Similarly let P^σ denote the random variable of winner payment conditional on the same event $\mathcal{E}^\sigma$. 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:*

$$\mathbb{P}(\mathcal{E}^\sigma) \cdot MI(W^\sigma, P^\sigma)$$

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.*

The proof appears in Appendix B. We note, however, that Proposition 1 no longer holds because the optimal reserve price r depends on the distribution F .

5 Privacy with Uniform Price Caps

In the previous sections we required the individual rationality constraint to hold ex-post. One way to view this constraint is that buyers decide in the interim whether to participate or not, but at the ex-post stage they can still opt out if they are asked to pay above a cap that depends on their type.

In this section we show that these type-dependent price caps are what places a limit on privacy preservation and makes stochastic payments not helpful. To that end, we now analyze the seller's problem under the alternative assumption that participation is voluntary at the interim stage, and the highest price cannot exceed an exogenous price cap that is uniform across types.

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 realized payment. Namely, $t_i(\theta)$ is bounded above by K for all i and all θ .

For any $K \geq \bar{\theta}$, 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 $\theta \in [\underline{\theta}, \bar{\theta}]^n$ 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 2 *For any $K \geq \bar{\theta}$, 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, given P_M , the random variable $P_{\{0, K\}}$ is *conditionally independent* of the winner's type W . Therefore, by the Data Processing Inequality, we have:²³

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

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 with equality 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 $\bar{H}$ 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 $\bar{\mu}$ and $\underline{\mu}$ denote the expectation of $\bar{H}$ 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:

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

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

$$\mathbb{E}[P_M | W] = \alpha(W) \cdot \bar{\mu} + (1 - \alpha(W)) \cdot \underline{\mu} = \alpha(W) \cdot (\bar{\mu} - \underline{\mu}) + \underline{\mu}.$$

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 $\mathbb{E}[P_{\{0, K\}} | W]$ and $\mathbb{E}[P_M | W]$ must be equal for every value W of the winner type. Because $\alpha(W)$ is not constant in W ,²⁴ this equality implies $\underline{\mu} = 0$ and $\bar{\mu} = K$. Since the support of P_M is contained in the interval $[0, K]$, the distribution $\bar{H}$ must be

²³Given three random variables, X, Y, Z , that form a Markov chain $X \rightarrow Y \rightarrow Z$, the Data Processing Inequality states that $MI(X, Y) \geq MI(X, Z)$, with equality if and only if X and Y are conditionally independent given Z (see Theorem 2.8.1 in [Cover and Thomas \(2012\)](#)).

²⁴ $\alpha(W) \cdot K$ 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 point-mass at K while $\underline{H}$ is the point-mass at 0. This completes the proof that P_M is supported on $\{0, K\}$. $\blacksquare$

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

Proposition 3 *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 $\tau(\theta_i) = \mathbb{E}_{\theta_{-i}}[T_i(\theta_i, \theta_{-i}) \mid \theta_i \geq \max\{\theta_{-i}\}]$ denote the expected payment of a winner of 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 the remaining probability. Hence the privacy loss in the efficient $\{0, K\}$ -mechanism is given by:

$$\begin{aligned} D_{KL}(G \parallel G_W \otimes G_P) &= \mathbb{E}_W [D_{KL}(G_{P|W} \parallel G_P)] \\ &= \int_{\underline{\theta}}^{\bar{\theta}} \left(\frac{\tau(w)}{K} \log \frac{\tau(w)/K}{\mathbb{E}[\tau(W)]/K} + \left(1 - \frac{\tau(w)}{K} \right) \log \frac{1 - \tau(w)/K}{1 - \mathbb{E}[\tau(W)]/K} \right) dG_W(w) \end{aligned} \quad (11)$$

The integrand on the right-hand side of Eq. (11) is bounded above by

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

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

6 Conclusion

This paper takes a first step in exploring the implications of Bayesian privacy concerns for the design of efficient and optimal auctions. Since, in many settings, the identity of the auction winner and the price he paid are publicly disclosed, we focus on minimizing the privacy loss to the winner. We quantify this loss using the mutual information between the winner's type and his payment, and show that, when both dominant-strategy incentive compatibility and ex-post individual rationality are required, under mild conditions the second-price auction minimizes the winner's privacy loss among all auctions that satisfy the designer's primary objectives.

Our results highlight an unexplored property of the second-price auction: its attractive privacy characteristics. While we do not claim that this should be the primary reason to use the second-price auction in practical settings, our findings show that its determinism should not be viewed as a disadvantage, at least insofar as Bayesian privacy is concerned.

Our proof relies on a novel result that establishes a lower bound on the mutual information between two ordered random variables. We use this result, together with the property of dominant-strategy mechanisms that the winner’s payment is a mean-preserving spread of the second-highest bid, to establish our main theorems. A natural extension of this work would be to study privacy-preserving auctions under alternative solution concepts (such as Bayesian IC), while maintaining ex-post individual rationality.

References

- Alvarez, Ramiro, and Mehrdad Nojournian.** 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.
- Athey, Susan, Christian Catalini, and Catherine Tucker.** 2017. “The digital privacy paradox: Small money, small costs, small talk.” National Bureau of Economic Research.
- Bergemann, Dirk, and Johannes Hörner.** 2018. “Should First-Price Auctions Be Transparent?” *American Economic Journal: Microeconomics*, 10(3): 177–218.
- Bos, Olivier, and Martin Pollrich.** 2025. “Auctions with signaling bidders: Optimal design and information disclosure.” *Games and Economic Behavior*, 151: 95–107.
- Bos, Olivier, and Tom Truyts.** 2021. “Auctions with signaling concerns.” *Journal of Economics & Management Strategy*, 30(2): 420–448.
- 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.** 2006a. “Monopoly with Resale.” *The RAND Journal of Economics*, 37(2): 362–375.

- Calzolari, Giacomo, and Alessandro Pavan.** 2006b. “On the optimality of privacy in sequential contracting.” *Journal of Economic Theory*, 130(1): 168–204.
- Canetti, Ran, Amos Fiat, and Yannai A. Gonczarowski.** 2023. “Zero-Knowledge Mechanisms.”
- Cover, T.M., and J.A. Thomas.** 2012. *Elements of Information Theory*. Wiley.
- Cramton, Peter, and Jesse A. Schwartz.** 2000. “Collusive Bidding: Lessons from the FCC Spectrum Auctions.” *Journal of Regulatory Economics*, 17(3): 229–252.
- 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.
- Gradshteyn, I. S., and I. M. Ryzhik.** 2007. *Table of Integrals, Series, and Products.* . 7th ed., Academic Press.
- Green, Jerry, and Jean-Jacques Laffont.** 1977. “Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods.” *Econometrica*, 45(2): 427–438.
- Haupt, Andreas, and Zoë Hitzig.** 2024. “Contextually Private Mechanisms.” *Working paper*.
- Heffetz, Ori, and Katrina Ligett.** 2014. “Privacy and Data-Based Research.” *Journal of Economic Perspectives*, 28(2): 75–98.
- Holmström, Bengt.** 1979. “Groves’ Scheme on Restricted Domains.” *Econometrica*, 47(5): 1137–1144.
- 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.
- OECD.** 2025. “Implementing the OECD Recommendation on Public Procurement in OECD and Partner Countries: 2020–2024 Report.”
- Pai, Mallesh M., and Aaron Roth.** 2013. “Privacy and Mechanism Design.” *SIGecom Exch.*, 12(1): 829.
- Parkes, David C., Michael O. Rabin, Stuart M. Shieber, and Christopher Thorpe.** 2008. “Practical secrecy-preserving, verifiably correct and trustworthy auctions.” *Electronic Commerce Research and Applications*, 7(3): 294–312. Special Section: New Research from the 2006 International Conference on Electronic Commerce.

A Appendix: Proof of Lemma 1

This is a rather long proof, and we begin by 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

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

the integral of $v(x, y)$ with respect to the product measure $X \otimes Y$ over the region $y \leq x$. 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 $Y(s) \geq X(s)$ for every $s \in \mathbb{R}$.

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

A.1 Preliminary Results

As can be seen from the definition of λ^* in Eq. (4), 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 $Y_-(s) = \lim_{t \uparrow s} Y(t)$ 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. $Y(s) \geq Y_-(s) \geq X(s)$ holds for every s .

We then define the following sets:

$$\begin{aligned} A &= \{s \in \mathbb{R} : Y(s) = X(s)\}; \\ \bar{A} &= \{s \in \mathbb{R} : Y_-(s) = X(s)\}. \end{aligned}$$

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

Proof of Lemma 3. Since $Y(s) \geq Y_-(s) \geq X(s)$, any $s \in A$ necessarily also belongs to $\bar{A}$. Thus $\bar{A}$ contains A . To see that $\bar{A}$ is closed, consider any sequence $s_n \in \bar{A}$ 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 $s \in \bar{A}$. If instead s_n decreases in n , then $Y_-(s) \leq \lim_n Y_-(s_n) = \lim_n X(s_n) = X(s)$. But we discussed above that $Y_-(s) \geq X(s)$, so equality holds and s again belongs to $\bar{A}$. ■

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 pairwise disjoint open intervals $I_1, I_2, \dots$, which we fix in the sequel. Let us write $I_k = (a_k, b_k)$,²⁵ 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 $a_k \in \overline{A} \setminus A$ 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 $s = a_k \in \overline{A} \setminus A \subset A^c$. 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 $s \in \overline{A} \setminus A$, 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 .²⁶

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 $a_k \in \overline{A} \setminus A$. 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. ■

A.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

²⁵Here 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.

²⁶However, 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.

to $X \otimes Y$. In these cases the KL-divergence $D(\lambda \parallel X \otimes Y)$ 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

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

has λ -measure equal to $X(A) > 0$. But this set $T \setminus S$ is part of the 45-degree line, which has measure zero according to $X \otimes Y$.²⁷ Hence λ is not absolutely continuous with respect to $X \otimes Y$. ■

A.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

$$h^*(x, y) = \frac{1}{Y(x) - X(x)} \cdot e^{-\int_y^x \frac{1}{Y(s) - X(s)} dX(s)} \cdot 1_{Y(x) > X(x)} \cdot 1_{y \leq x}. \quad (12)$$

²⁷For 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.

While h^* is defined for any $y \leq x$, 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 $x \geq y$, $\int_y^x \frac{1}{Y(s)-X(s)} dX(s)$ is finite if and only if $x \in \hat{I}_k$. Consequently, $h^*(x, y)$ as defined in (12) is strictly positive if and only if $x \geq y$ 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 $b_k \in \bar{A}$, we have $Y_-(b_k) = X(b_k)$. Thus

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

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

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 $Y(s_n) - X(s_n) \rightarrow 0$. 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 $Y(s_n) - X(s_n) \rightarrow 0$ 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 $Y(s_n) - X(s_n) \rightarrow 0$ implies $s > y$ and $Y_-(s) = X(s)$. But this contradicts $s \in (y, x] \subset I_k \subset \bar{A}^c$. ■

A.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 $y \leq x$, 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

$$\int_{\mathbb{R}} h(x, y) dX(x) = 1 \quad \text{for } Y\text{-almost every } y; \quad (13)$$

$$\int_{\mathbb{R}} h(x, y) dY(y) = 1 \quad \text{for } X\text{-almost every } x. \quad (14)$$

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. (12) satisfies Eq. (13) for every $y \in A^c$ (i.e. $Y(y) > X(y)$).

Lemma 10 h^* defined in Eq. (12) satisfies Eq. (14) 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 $x \rightarrow b_k$. 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 $\lim_{x \rightarrow b_k^-} \alpha(x) = \alpha(b_k) = \infty$ 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 (14) 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 $y \leq x$. 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 $g(x) = 1$.²⁸ It remains to show that

$$\int_{[a_k, x]} g(y) dY(y) = Y(x) - X(x). \quad (15)$$

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 (15) 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 $[y_0, y_1] \cup (y_1, y_2] \cup \dots \cup (y_{n-1}, y_n]$ with $a_k = y_0 < y_1 < \dots < y_n = x$. 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 $\underline{g} \leq g \leq \overline{g}$ 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 (15), it suffices to show the following inequality for every partition:

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

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

²⁸In 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.

as finite sums. The above inequalities then become

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

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

For the first inequality, we prove by induction that

$$g(y_0) \cdot (Y(y_1) - Y_-(y_0)) + \sum_{i=1}^{m-1} g(y_i) \cdot (Y(y_{i+1}) - Y(y_i)) \leq g(y_m) \cdot (Y(y_m) - X(y_m)). \quad (16)$$

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

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

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

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

as we desire to show.²⁹ As for the induction step in (16) from m to $m + 1$, we need to verify that $g(y_m)(Y(y_m) - X(y_m)) + g(y_m)(Y(y_{m+1}) - Y(y_m)) \leq g(y_{m+1})(Y(y_{m+1}) - X(y_{m+1}))$. This reduces to

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

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 $\bar{g}$, we will similarly show by induction that

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

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

²⁹The final equality here follows by viewing the integral as a Riemann-Stieltjes integral, and making the substitution $z = X(s)$.

$g(y_{m+1})(Y(y_{m+1}) - X(y_{m+1}))$, 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 have³⁰

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

This proves the induction step and implies (17).

Therefore (15) holds and the lemma is proved. ■

A.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 $D_{KL}(\lambda \parallel X \otimes Y) \geq D_{KL}(\lambda^* \parallel X \otimes Y)$ 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 $\bar{A}$, 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.

³⁰Note 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

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

where $K(X, Y) = -1 + \int_{\mathbb{R}} \log \frac{1}{Y(s) - X(s)} dX(s)$. Consequently, $D_{KL}(\lambda^* \parallel X \otimes Y) = K(X, Y) \leq D_{KL}(\lambda \parallel X \otimes Y)$, 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 $D_{KL}(\lambda \parallel X \otimes Y)$ and $D_{KL}(\lambda \parallel \lambda^*)$ 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:

$$\begin{aligned} D_{KL}(\lambda \parallel X \otimes Y) &= \int_{\mathbb{R}^2} h(x, y) \log h(x, y) dX(x) dY(y) \\ &= \int_{\cup_k T_k} h(x, y) \log h(x, y) dX(x) dY(y) \\ &= \int_{\cup_k T_k} h(x, y) \log \frac{h(x, y)}{h^*(x, y)} dX(x) dY(y) + \int_{\cup_k T_k} h(x, y) \log h^*(x, y) dX(x) dY(y) \\ &= D_{KL}(\lambda \parallel \lambda^*) + \sum_k \int_{T_k} h(x, y) \log h^*(x, y) dX(x) dY(y). \end{aligned} \tag{18}$$

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

$$\begin{aligned} & \int_{T_k} h(x, y) \log h^*(x, y) dX(x) dY(y) \\ &= \int_{T_k} h(x, y) \log \left(\frac{1}{Y(x) - X(x)} \right) dX(x) dY(y) - \int_{T_k} h(x, y) \cdot \left(\int_y^x \frac{1}{Y(s) - X(s)} dX(s) \right) dX(x) dY(y). \end{aligned} \quad (19)$$

To simplify the first term on the RHS above, we recall that h is the density of $\lambda \in \mathcal{M}(X, Y)$, and thus satisfies the marginal requirements (13) and (14). In particular, (14) 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

$$\begin{aligned} \int_{T_k} h(x, y) \log \left(\frac{1}{Y(x) - X(x)} \right) dX(x) dY(y) &= \int_{\hat{I}_k} \log \frac{1}{Y(x) - X(x)} \cdot \left(\int_{\hat{I}_k} h(x, y) dY(y) \right) dX(x) \\ &= \int_{\hat{I}_k} \log \frac{1}{Y(x) - X(x)} dX(x). \end{aligned} \quad (20)$$

As for the second term on the RHS of (19), we have³¹

$$\begin{aligned} & \int_{T_k} h(x, y) \cdot \left(\int_{[y, x]} \frac{1}{Y(s) - X(s)} dX(s) \right) dX(x) dY(y) \\ &= \int_{y, s, x \in \hat{I}_k: y \leq s < x} h(x, y) \frac{1}{Y(s) - X(s)} dX(s) dX(x) dY(y) \\ &= \int_{s \in \hat{I}_k} \frac{1}{Y(s) - X(s)} \cdot \left(\int_{y, x \in \hat{I}_k: y \leq s, x > s} h(x, y) dX(x) dY(y) \right) dX(s). \end{aligned} \quad (21)$$

Now observe that the integral $\int_{y, x \in \hat{I}_k: y \leq s < x} h(x, y) dX(x) dY(y)$ is simply the measure that λ assigns to the region $\{(y, x) \in T_k : y \leq s < x\}$. 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 $\{(y, x) : y \leq s < x\}$, which is just $Y(s) - X(s)$. Hence, plugging in the RHS of (21), we obtain

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

³¹We 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 (20) and (22) into (19) and then back into (18), we arrive at

$$\begin{aligned}
D_{KL}(\lambda \parallel X \otimes Y) &= D_{KL}(\lambda \parallel \lambda^*) + \sum_k \left(\int_{\hat{I}_k} \log \frac{1}{Y(x) - X(x)} dX(x) - X(\hat{I}_k) \right) \\
&= D_{KL}(\lambda \parallel \lambda^*) + \int_{A^c} \log \frac{1}{Y(x) - X(x)} dX(x) - X(A^c) \\
&= D_{KL}(\lambda \parallel \lambda^*) + \int_{\mathbb{R}} \log \frac{1}{Y(x) - X(x)} dX(x) - 1 \\
&= D_{KL}(\lambda \parallel \lambda^*) + K(X, Y),
\end{aligned} \tag{23}$$

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

A.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*

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

Lemma 13 immediately implies that $\hat{\lambda}$ minimizes mutual information whenever the minimum is achieved. Thus $\hat{\lambda} = \lambda^*$ whenever the RHS of Eq. (3) 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

$$\begin{aligned}
D_{KL}(\lambda \parallel X \otimes Y) &= \int_{\mathbb{R}^2} h(x, y) \log h(x, y) dX(x) dY(y) \\
&= \int_{\mathbb{R}^2} h(x, y) \log \frac{h(x, y)}{h_1(x)h_2(y)} dX(x) dY(y) + \int_{\mathbb{R}^2} h(x, y) \log(h_1(x)h_2(y)) dX(x) dY(y) \\
&= D_{KL}(\lambda \parallel \hat{\lambda}) + \int_{\mathbb{R}^2} h(x, y) \log h_1(x) dX(x) dY(y) + \int_{\mathbb{R}^2} h(x, y) \log h_2(y) dX(x) dY(y).
\end{aligned} \tag{24}$$

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 $\int_{\mathbb{R}^2} h(x, y) \log h_2(y) dX(x) dY(y) = \int \log h_2(y) dY(y)$. Plugging these into Eq. (24), we obtain

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

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

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

as we desire to show. ■

B Proof of Theorem 2

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 $p \geq r$,

$$\mathbb{P}[W \leq w \mid P = p] = \frac{F(w) - F(p)}{1 - F(p)}. \quad (25)$$

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

From Eq. (25) 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:

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

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.³²

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

$$\int_r^{\bar{\theta}} \log \frac{1}{G_P(s) - G_W(s)} dG_W(s) \geq \int_r^{\bar{\theta}} \log \frac{1}{G_L(s) - G_W(s)} dG_W(s). \quad (26)$$

The proof is essentially the same as before, since we still have $\frac{g_W(s)}{G_L(s) - G_W(s)} = \frac{f(s)}{1 - F(s)}$ for any $s \geq r$.³³ Thus, the desired inequality (26) is implied by $\int_r^{\bar{\theta}} (G_L(s) - G_P(s)) \cdot \frac{f(s)}{1 - F(s)} ds \geq 0$, which is just the analogue of (8) with the range of integration changed to $[r, \bar{\theta}]$. The same proof that we had for (8) applies here.

³²As 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.

³³Intuitively, the reserve price r affects G_W and G_L by the same linear transformation.