

Privacy Preserving Signals*

Philipp Strack[†] Kai Hao Yang[‡]

August 19, 2024

Abstract

A signal is *privacy-preserving* with respect to a collection of *privacy sets* if the posterior probability assigned to every privacy set remains unchanged conditional on any signal realization. We characterize the privacy-preserving signals for arbitrary state space and arbitrary privacy sets. A signal is privacy-preserving if and only if it is a garbling of a *reordered quantile signal*. Furthermore, distributions of posterior means induced by privacy-preserving signals are exactly mean-preserving contractions of that induced by the *quantile signal*. We discuss the economic implications of our characterization for statistical discrimination, the revelation of sensitive information in auctions, and price discrimination.

Keywords: Privacy-preserving signal, privacy sets, independence, reordered quantile signal, protected characteristics.

JEL classification: C11, D63, D42, D83,

*We thank the coeditor, the anonymous referees, S. Nageeb Ali, Dirk Bergemann, Ben Brooks, Arjada Bardhi, Andy Choi, Joyee Deb, Laura Doval, Piotr Dworzak, Mira Frick, Joshua Gans, Paul Heidhues, Shota Ichihashi, Ryota Iijima, Emir Kamenica, Xiao Lin, Elliot Lipnowski, Ce Liu, Alessandro Lizzeri, Stephen Morris, Barry Nalebuff, Aniko Öry, Benjamin Polak, Doron Ravid, Aaron Roth, Fedor Sandomirskiy, Ludvig Sinander, Omer Tamuz, Nathan Yoder, Alexander Zentefis, and Jidong Zhou for their valuable comments and suggestions. We also thank Tsong-Hong Tenn and Jialun (Neil) He for their research assistance. All errors are our own.

[†]Department of Economics, Yale University, Email: philipp.strack@yale.edu

[‡]School of Management, Yale University, Email: kaihao.yang@yale.edu

1 Introduction

In many economic settings, there are constraints on what information can be used or revealed: Characteristics such as race, gender, and sexual orientation are protected in many contexts, and the information that can be revealed about them is limited due to legal, regulatory, or social norms. Motivated by this, we study the set of signals (Blackwell experiments) that are constrained to not reveal certain information.

For example, consider the case of a bank determining whether to grant a loan to an individual. In making this decision, the bank benefits from using an individual’s characteristics to predict whether they will default. However, the Equal Credit Opportunity Act prohibits discrimination against loan applicants on the basis of their “protected characteristics,” such as race, gender, or age. As a result, the bank is legally required to ensure that its loan decisions are not influenced by these protected characteristics. In other words, the information used by the bank in making loan decisions cannot be based on these characteristics.

This paper presents a framework for understanding information disclosure in situations where certain aspects of the state of the world must be kept private, or equivalently, where decisions must be taken independent of certain characteristics. Following [Blackwell \(1953\)](#), we model information as a signal about an abstract state of the world $\omega \in \Omega$. To capture a notion of protected information, we introduce a collection of events called *privacy sets*, which represent the information that cannot be disclosed. For instance, in the context of a bank loan, the privacy sets would include all the protected characteristics. We define a signal as *privacy-preserving* if, for any signal realization, the posterior probability of any privacy set remains unchanged and equals its prior probability. In other words, a privacy-preserving signal does not reveal any information about events that belong to the privacy sets.

We characterize all privacy-preserving signals. The privacy sets can always be represented as a random variable θ defined on the same state space Ω , so that a signal s is privacy-preserving if and only if it is independent of θ . In the bank loan example, the random variable θ would indicate an applicant’s protected characteristics. A privacy-preserving signal s can be informative about the applicant in many aspects, including their default probabilities, but must be independent of the applicant’s protected characteristics θ .

Our first main result characterizes all privacy-preserving signals: For any one-dimensional statistic $\phi : \Omega \rightarrow \mathbb{R}$ of the state ω , we define the *ϕ -quantile signal* as the signal that reveals the

empirical quantile of ϕ conditional on θ , plus potentially some noises when it has an atom. This signal is privacy-preserving, as it is uniformly distributed on $[0, 1]$ for every value of θ . A *reordered ϕ -quantile signal* is a signal obtained by further reordering (possibly randomly) the unit interval for each realization of θ while preserving its distribution.

[Theorem 1](#) establishes that for any statistic ϕ such that (ϕ, θ) fully reveals the state ω , a signal is privacy-preserving if and only if it is Blackwell-dominated by a reordered ϕ -quantile signal. As a result, the set of *all* privacy-preserving signals can essentially be generated by a *single* ϕ -quantile signal, via garbling and reordering. Moreover, these reordered ϕ -quantile signals are Blackwell-undominated among all privacy-preserving signals. Therefore, the reordered ϕ -quantile signals are exactly the frontier of all privacy-preserving signals in Blackwell’s sense.

When only the posterior *means* of a one-dimensional statistic ϕ are relevant, our second main result, [Theorem 2](#), shows that a CDF G is the distribution of posterior means of ϕ under some privacy-preserving signal if and only if G is a mean-preserving contraction of the distribution of posterior means induced by the ϕ -quantile signal. Consequently, in settings where the only economically relevant variables are the posterior means (e.g., when a decision-maker has a payoff that is affine in a statistic ϕ), *every* privacy-preserving signal leads to a lower expected payoff than the ϕ -quantile signal, and *any* privacy constraints can be completely summarized by a single majorization constraint.

Our results extend to the setting where privacy is defined *conditional* on another given random variable y . A signal is *conditionally privacy-preserving* if it is independent of θ *conditional on* y . This extension allows us to analyze settings where signals (or decisions) are allowed to reveal some information about the protected characteristics, as long as it is based on information contained in y . For example, banks cannot directly base discriminatory loan decisions on race, but can make predictions using applicants’ credit history, which is deemed “materially relevant,” even though credit histories may be correlated with race.

Having characterized the set of privacy-preserving signals, we then explore how to optimize over this set. Consider a decision-maker who chooses an action $a \in A$ to maximize their payoff $u(\omega, a)$, after observing a privacy-preserving signal. [Theorem 1](#) implies that it is without loss to restrict attention to reordered quantile signals. [Proposition 2](#) shows that the resulting optimization problem is equivalent to an optimal transport problem. When payoffs depend only on a single dimensional statistic ϕ and on θ , we fully characterize the optimal

privacy-preserving signals for a wide class of decision and persuasion problems, including when the payoff is single-crossing in (ϕ, a) or linear in ϕ , when the statistic ϕ takes binary values, and when actions are binary. This characterization covers a broad range of decision problems that appeared in the literature.

To illustrate the usefulness of our mathematical results, we apply them to various economic contexts. First, we consider statistical discrimination and fairness in algorithm design. In legal contexts, the notion of *disparate impact* describes the situation where different groups experience different outcomes even if the underlying policies are not explicitly based on group identities. Our results lead to a characterization of a decision-maker’s optimal signals that do not create disparate impacts, which generalizes existing results in the algorithmic fairness literature. Furthermore, we lay out an optimal and detail-free procedure for regulating statistical discrimination.

The second application considers disclosure in ad auctions without violating users’ privacy. A publisher runs a second price auction to sell targeted audiences to multiple advertisers. The publisher has rich information about the targeted audience, and can selectively disclose some information to the advertisers. However, some information about the audience, such as religious beliefs and sexual orientation, cannot or should not be disclosed. [Theorem 2](#), together with the results of [Bergemann, Heumann, Morris, Sorokin, and Winter \(2022\)](#), provides a characterization of optimal privacy-preserving information disclosure for the auctioneer.

The third application considers price discrimination and market segmentation in the spirit of [Bergemann, Brooks and Morris \(2015\)](#). In a setting where a monopolist is able to segment consumers based on their values, we consider a situation where consumers with different protected characteristics (e.g., gender and race), must face the same distribution of prices, even though the monopolist engages in third-degree price discrimination. Solving the optimal transport problem derived in [Proposition 2](#) allows us compute the market segmentation that maximizes the monopolist’s revenue while preventing price discrimination based on consumers’ protected characteristics.

Related Literature We follow the approach of [Blackwell \(1953\)](#) and model information as signals about an underlying state. We generalize Blackwell’s characterization of feasible signals by characterizing all feasible signals that do not reveal information about a given collection of events. While Blackwell shows that a signal is feasible if and only if it is

dominated by the signal which fully reveals the state, we show that a signal is feasible and privacy-preserving if and only if it is dominated by a “reordered quantile signal”.

In the literature on algorithmic fairness, one of the most common criteria for fairness requires the decisions to be statistically independent of protected characteristics (see, e.g., [Calders, Kamiran and Pechenizkiy 2009](#); [Hardt, Price and Srebro 2016](#); [Corbett-Davies, Pierson, Feller, Goel and Huq 2017](#)). These papers characterize optimal fair algorithms for decision problems with binary actions, binary states, or specific payoff structures.

[He, Sandomirskiy and Tamuz \(2023\)](#) introduce a Blackwell approach, and use tools from mathematical tomography, to characterize the optimal privacy-preserving signals for all decision problems with a binary state and general payoffs that do not depend on protected characteristics.¹ Our results unify and generalize their findings and results from the algorithmic fairness literature to more than two actions, more than two states, and general payoffs.

In the recent economic literature on discrimination and privacy, [Liang, Lu and Mu \(2023\)](#) and [Doval and Smolin \(2023\)](#) also consider fairness and algorithm design. They adopt different notions of fairness and characterize the Pareto frontier in terms of payoffs of different groups. [Bohren, Hull and Imas \(2024\)](#) develop a framework and a measurement of how early-stage disparate impacts can be exacerbated and lead to systemic discrimination. [Che, Kim and Zhong \(2021\)](#) show that in many markets, such discrimination could be fostered by endogenous ratings. [Eilat, Eliaz and Mu \(2021\)](#) consider mechanism design problems under privacy constraints, where a reduction in privacy is measured by the Kullback-Leibler divergence between the designer’s prior and posterior belief. [Ichihashi \(2020\)](#); [Galperti and Perego \(2023\)](#); [Rhodes and Zhou \(2024\)](#) further explore the implications and values of consumer privacy in a monopoly or oligopoly market.

Privacy-preserving signals also relate to the notion of *belief-invariant* Bayes correlated equilibria ([Forges 1993, 2006](#); [Liu 2015](#)) in games with incomplete information, which require the action recommendations received by each player to not affect the player’s belief about the unknown state and other players’ types. While both notions are stated in terms of statistical independence, we are interested in the informational content that can be provided subject to privacy constraints, rather than the strategic implications of correlating actions in games.

¹[He et al. \(2023\)](#) and our paper use different notions of Blackwell dominance, and we provide a detailed discussion in §6.1. Less directly related to our work, [He et al.](#) also introduce the multi-agent concept of “private private” signals, which are information structures such that no agent’s signal realization reveals information about the signal realization of other agents.

2 Model

States Throughout the paper, we fix the probability space $(\Omega \times [0, 1], \mathcal{F} \times \mathcal{B}, \mathbb{P})$ where (Ω, \mathcal{F}) is a standard Borel space, \mathcal{B} is the Borel sigma algebra on $[0, 1]$, and \mathbb{P} is the product measure induced by some probability measure over Ω and the Lebesgue measure over $[0, 1]$. We denote the outcome by

$$(\omega, r) \in \Omega \times [0, 1].$$

The first component of the outcome ω is the state, which captures all economically relevant information. The second part of the outcome r is the source of additional randomization that is independent of the state. We impose no further restrictions on the state ω . In particular, ω might be multi-dimensional and captures all relevant characteristics of an economic agent such as income, age, gender, race, address, etc.

Signals A signal is a random variable $s : \Omega \times [0, 1] \rightarrow S$.² This definition of a signal is equivalent to the definition of experiments given in Blackwell (1953).³ Intuitively, if the signal is random conditional on the state ω , this randomization is achieved by explicitly conditioning on the randomization device r . For any signal s , let $\mathbb{P}[\cdot | s]$ denote the conditional distribution given s .⁴

Definition 1 (Blackwell Dominance). A signal s is Blackwell-dominated by a signal \tilde{s} , denoted by $s \leq \tilde{s}$, if for every measurable function $u : \Omega \times A \rightarrow \mathbb{R}$ and action set A

$$\mathbb{E} \left[\sup_{a \in A} \mathbb{E} [u(\omega, a) | s] \right] \leq \mathbb{E} \left[\sup_{a \in A} \mathbb{E} [u(\omega, a) | \tilde{s}] \right].$$

²If one focuses on the case of countably many signal realizations, then this definition of a signal is equivalent to the one in Green and Stokey (2022). To see this, consider the partition of $\Omega \times [0, 1]$ induced by the signal realizations $\{(\omega, r) \in \Omega \times [0, 1] : s(\omega, r) = \hat{s}\}_{\hat{s} \in S}$. The posterior belief after observing signal realization \hat{s} is then the conditional distribution of \mathbb{P} over Ω given the event $\{(\omega, r) \in \Omega \times [0, 1] : s(\omega, r) = \hat{s}\}$. Conversely, for any countable partition of $\Omega \times [0, 1]$, one can define a random variable s as an indicator function of its partitional elements.

³To see this, note that for every signal we can define a (essentially) unique Blackwell experiment by $(S, \mathbb{P}[s \in \cdot | \omega]_{\omega \in \Omega})$. Conversely, as the probability space $(\Omega \times [0, 1], \mathcal{F} \times \mathcal{B}, \mathbb{P})$ is standard, for every Blackwell experiment we can construct a random variable s with the correct joint distribution (see von Neumann 1932; Rokhlin 1952).

⁴More precisely, since $(\Omega \times [0, 1], \mathcal{F} \times \mathcal{B})$ is a standard Borel space, a regular version of the conditional expectation given the σ -algebra generated by s exists, which we denote by $\mathbb{P}[\cdot | s]$ (Çinlar 2010, Theorem 2.7, pp. 151).

Privacy-Preserving Signals We are interested in signals about the state—or decisions that are made conditional on signal realizations—that do not reveal certain types of information. For example, in some economic contexts the signals might be required to preserve some notion of privacy; or actions taken by a firm might not be allowed to condition on protected characteristics, such as race and gender. For a collection of *privacy-sets*

$$\mathcal{P} \subseteq \mathcal{F}$$

that are closed under finite intersections,⁵ each $P \in \mathcal{P}$ defines an event that has to be “kept private,” and no information about it can be revealed. The number of privacy sets is allowed to be finite, countably infinite, or uncountable. Privacy sets encode what information (or characteristics) are protected in a given context, such as a person’s race, gender, or age.

Definition 2 (Privacy-Preserving Signals). A signal s is *privacy-preserving* (with respect to \mathcal{P}) if the prior and posterior probability of the state being in any privacy set coincide, i.e., for all $P \in \mathcal{P}$, almost surely,

$$\mathbb{P}[\omega \in P \mid s] = \mathbb{P}[\omega \in P]. \quad (1)$$

Our main result will establish a characterization of all Blackwell-undominated privacy-preserving signals.

Remark 1 (Privacy and Discrimination). One can think of the signal s as a decision taken that affects the outcomes of an economic agent (e.g., whether or not to grant a loan, or what interest rate to charge). The privacy sets encode protected characteristics based on which discrimination is illegal, such as race, age, or gender. The set of privacy-preserving signals then corresponds to the set of decision rules that are “non-discriminatory,” in the sense that the decision is independent of the agent’s protected characteristics. This connection between discrimination and privacy goes beyond the formal level: Often people want to keep certain information private, such as their sexual orientation, exactly because they fear discrimination.

⁵This requirement is equivalent to saying that, if two sets P_1 and P_2 are events that need to be kept private, then their intersection $P_1 \cap P_2$ has to be kept private too. For example, if a privacy-preserving signal is not allowed to reveal information about a person’s race or gender, then it is also not allowed to reveal information about whether a person is a white male or a non-white female.

3 Characterization of Privacy-Preserving Signals

In this section, we state our main results, [Theorem 1](#) and [Theorem 2](#). Formal proofs of the results can be found in the Appendix. To begin with, we first note that the privacy sets \mathcal{P} can always be encoded into a random variable θ , which takes values in some measurable space Θ . This will lead to a more succinct notation.

Lemma 1. *There exists a random variable $\theta : \Omega \rightarrow \Theta$ such that for any signal s , the following are equivalent: (i) s is privacy-preserving, (ii) s is independent of θ .*

For finite \mathcal{P} we can define $\theta = (\theta_P)_{P \in \mathcal{P}} \in \Theta = \{0, 1\}^{|\mathcal{P}|}$ to be a vector of indicators indicating which privacy sets the state belongs to, i.e., $\theta_P(\omega) = \mathbf{1}\{\omega \in P\}$. For example, if the privacy sets divide the population into female/male and white/non-white, then the realization of θ would correspond to non-white females; non-white males; white females; and white males. The proof for the infinite case can be found in the Appendix.

Given the random variable θ that encodes the privacy sets \mathcal{P} , we next introduce a class of signals that reveals the state to an outside observer who knows the characteristic θ .

Definition 3 (Conditionally Revealing Signals). A signal s is *conditionally revealing* if observing s and θ reveals ω .⁶

The definition of a conditionally revealing signal does not have direct implications for the informativeness of the signals s , as the characteristic θ is unknown. Moreover, a conditionally revealing signal may not be privacy-preserving.⁷ However, [Proposition 1](#) below establishes that every privacy-preserving signal is less informative than a conditionally revealing, privacy-preserving signal.

Proposition 1. *A signal is privacy-preserving if and only if it is Blackwell-dominated by some conditionally revealing privacy-preserving signal.*

According to [Proposition 1](#), a signal is privacy-preserving if and only if it is a garbling of some conditionally revealing privacy-preserving signal. In what follows, we construct explicitly a conditionally revealing privacy-preserving signal, and show that it can further generate *all* such signals, leading to a characterization of the set of privacy-preserving signals.

⁶Or more formally, for any $A \in \mathcal{F}$, $\mathbb{P}[\omega \in A \mid s, \theta] \in \{0, 1\}$ almost surely.

⁷For instance, the fully-revealing signal $s(\omega, r) = \omega$ is conditionally revealing, but not privacy-preserving.

Quantile Signals Consider any random variable $\phi : \Omega \rightarrow \mathbb{R}$ that summarizes the state into a single dimensional statistic. Let

$$F_\phi(z | \hat{\theta}) := \mathbb{P}[\phi(\omega) \leq z | \hat{\theta}]$$

be the distribution of ϕ conditional on $\hat{\theta}$, $F_\phi^-(\cdot | \hat{\theta})$ its left limit, and $F_\phi^{-1}(\cdot | \hat{\theta})$ its inverse.⁸

Definition 4 (Quantile Signal). The ϕ -quantile signal is defined as

$$q_\phi := rF_\phi(\phi | \theta) + (1 - r)F_\phi^-(\phi | \theta). \quad (2)$$

If $F_\phi(\cdot | \hat{\theta})$ is continuous for all $\hat{\theta} \in \Theta$, then q_ϕ is the empirical quantile of the realization of ϕ . For example, $q_\phi = 0.3$ reveals that ϕ takes a lower value than its current realization 30% percent of the time for each realization of θ . In the continuous case, $q_\phi = F_\phi(\phi | \theta)$ is uniformly distributed, and is thus independent of θ . The randomization in q_ϕ ensures that this property generalizes to discontinuous F_ϕ , as stated below.

Lemma 2. *The ϕ -quantile signal q_ϕ is privacy-preserving.*

The signal q_ϕ allows one to identify the realization of ϕ if θ is known. While q_ϕ may not be conditionally revealing in general, whenever ϕ is invertible, with its inverse ϕ^{-1} being measurable,⁹ the signal q_ϕ is conditionally revealing, since $\omega = \phi^{-1}(F_\phi^{-1}(q_\phi | \theta))$ almost surely.

Reordered Quantile Signals A measurable function $M : [0, 1] \rightarrow [0, 1]$ is a (Lebesgue) measure-preserving transformation if $\int_0^1 \mathbf{1}\{M(z) \leq x\} dz = x$ for all $x \in [0, 1]$. Two signals s, s' are said to be Blackwell-equivalent, denoted by $s \sim s'$, if $s \leq s'$ and $s' \leq s$.

Definition 5 (Reordered Quantile Signal). For any random variable $\phi : \Omega \rightarrow \mathbb{R}$, signal s is a reordered ϕ -quantile signal if there exists a family of measure-preserving transformations $(M_\theta)_{\theta \in \Theta}$ such that $M_\theta(s) \sim q_\phi$.

⁸Formally, for any CDF F , $F^-(z) := \lim_{\varepsilon \searrow 0} F(z - \varepsilon)$, for all $z \in \mathbb{R}$, while $F^{-1}(q) := \inf\{z \in \mathbb{R} : F(z) \geq q\}$, for all $q \in [0, 1]$.

⁹Such a function always exists, since (Ω, \mathcal{F}) is a standard Borel space and is thus isomorphic to a subset of \mathbb{R} with the Borel σ -algebra (see [von Neumann 1932](#); [Rokhlin 1952](#)).

For every ϕ -quantile signal q_ϕ and every family $(M_\theta)_{\theta \in \Theta}$ of measure-preserving transformations, there exists (up to Blackwell equivalence) a unique reordering of q_ϕ , which can be explicitly constructed.¹⁰ By definition, every reordered ϕ -quantile signal is privacy-preserving. Furthermore, as knowing θ and s reveals $q_\phi \sim M_\theta(s)$, it follows that s is conditionally revealing whenever q_ϕ is conditionally revealing (e.g., when ϕ is invertible).

We now present our first main result, which characterizes the set of privacy-preserving signals.

Theorem 1 (Characterization of Privacy-Preserving Signals). *Fix any conditionally revealing quantile signal q^* .*

- (i) *A signal is privacy-preserving if and only if it is Blackwell-dominated by some reordering of q^* .*
- (ii) *Every reordering of q^* is Blackwell-undominated among privacy-preserving signals.*

Part (i) of [Theorem 1](#) establishes that the set of *all* privacy-preserving signals can be generated from a *single* conditionally revealing quantile signal q^* , via reordering and garbling.¹¹ In particular, it is without loss to optimize only over reorderings of q^* instead of all privacy-preserving signals, as one can always ignore additional information. As we discuss in [§4](#), this is a drastic simplification, as it reduces the problem of optimizing over privacy-preserving signals to an optimization problem over measure-preserving transformations. Part (ii) establishes that every reordering of q^* is undominated. Thus, without imposing structures on the decision problem, no further restriction of the set of privacy-preserving signals is without loss.

Distributions of Posterior Means Our second main result—whose proof relies on the characterization given by [Theorem 1](#)—characterizes the set of distributions of posterior means of an arbitrary statistic ϕ induced by privacy-preserving signals. The expectation of a one-dimensional variable is a sufficient statistic in many economic contexts. For example, in the

¹⁰More details for the construction can be found in the Online Appendix.

¹¹Although some quantile signals are conditionally revealing, not all conditionally revealing privacy-preserving signals are equivalent to a quantile signal (e.g., the signal described by [Figure 2b](#) in [§5.3](#) is conditionally revealing, but is not equivalent to a ϕ -quantile signal for any ϕ). Nonetheless, part (i) of [Theorem 1](#) ensures that any conditionally revealing privacy-preserving signal is a *reordering* of some conditionally revealing quantile signal q^* . Intuitively, this is because a statistic ϕ does not allow for randomization, but a reordering does (see the Online Appendix for more details about this randomization).

context of a bank loan discussed in §1, this corresponds to the assumption that the bank’s preference depends only on whether the applicant will default, but not the race or gender of an applicant.

Consider any statistic $\phi : \Omega \rightarrow \mathbb{R}$. It is well-known that a CDF G is a distribution of posterior means $\mathbb{E}[\phi | s]$ under some signal s if and only if G is a mean-preserving contraction of the prior distribution $F_\phi(z) := \mathbb{P}[\phi(\omega) \leq z]$ (see e.g., [Hardy, Littlewood and Pólya 1929](#); [Strassen 1965](#)). However, not every mean-preserving contraction of F_ϕ can be the distribution of posterior means under a privacy-preserving signal. For example, if ϕ and θ are correlated, then F_ϕ could never be the distribution of posterior means.

Our second main result characterizes the distributions of posterior means induced by privacy-preserving signals. Let \bar{F}_ϕ be of posterior means $\mathbb{E}[\phi | q_\phi]$ induced by the ϕ -quantile signal q_ϕ . Namely,

$$\bar{F}_\phi(z) := \inf \{y \in [0, 1] : \mathbb{E}[F_\phi^{-1}(y | \theta)] \geq z\},$$

for all $z \in \mathbb{R}$, where the expectation is taken over θ .

Theorem 2 (Distributions of Posterior Means). *For any statistic $\phi : \Omega \rightarrow \mathbb{R}$, a CDF G is the distribution of posterior means $\mathbb{E}[\phi | s]$ induced by some privacy-preserving signal s if and only if G is a mean-preserving contraction of \bar{F}_ϕ .*

The formal proof of [Theorem 2](#) can be found in the Appendix. The “if” part follows from [Theorem 1](#) immediately, since any mean-preserving contraction G of \bar{F}_ϕ can be induced by a signal s that is Blackwell-dominated by—and hence is privacy-preserving according to [Theorem 1](#)—the quantile signal q_ϕ . The “only if” part also follows from [Theorem 1](#). Indeed, by [Theorem 1](#), it suffices to show that the posterior means induced by any reordering of q_ϕ must be less dispersed than $\mathbb{E}[\phi | q_\phi]$. To see this, let s be a reordering of q_ϕ defined by measure-preserving transformations $(M_\theta)_{\theta \in \Theta}$, so that $q_\phi \sim M_\theta(s)$. By definition,

$$\mathbb{E}[\phi | \hat{s}] = \mathbb{E}[F_\phi^{-1}(M_\theta(\hat{s}) | \theta)]$$

for all realization $\hat{s} \in [0, 1]$ of s . Let G be the distribution of $\mathbb{E}[\phi | s]$. Since s is a reordering of q_ϕ , s must also be uniformly distributed. Thus, $\mathbb{E}[\phi | s]$ and $G^{-1}(s)$ have the same distribution. As a result, G^{-1} can be rearranged in a way so that for (almost) all $\hat{s} \in [0, 1]$

(see Proposition 3 of [Ryff 1970](#)),

$$G^{-1}(\psi(\hat{s})) = \mathbb{E}[\phi \mid \hat{s}] = \mathbb{E}[F_\phi^{-1}(M_\theta(\hat{s}) \mid \theta)],$$

for some measure-preserving transformation $\psi : [0, 1] \rightarrow [0, 1]$. Together, for all $t \in [0, 1]$,

$$\begin{aligned} \int_t^1 G^{-1}(y) dy &= \int_0^1 \mathbf{1}\{\psi(z) \geq t\} G^{-1}(\psi(z)) dz = \int_0^1 \mathbf{1}\{\psi(z) \geq t\} \mathbb{E}[F_\phi^{-1}(M_\theta(z) \mid \theta)] dz \\ &\leq \int_0^1 \mathbf{1}\{z \geq t\} \mathbb{E}[F_\phi^{-1}(z \mid \theta)] dz \\ &= \int_t^1 \overline{F}_\phi^{-1}(z) dz, \end{aligned}$$

where the inequality follows because the covariance of two random variables with given marginals is maximized when they are comonotonic.¹² Therefore, G is a mean-preserving contraction of \overline{F}_ϕ (see, e.g., [Shaked and Shanthikumar \(2007\)](#), Section 3.A).

According to [Theorem 2](#), for any privacy-preserving signal s , $\mathbb{E}[\phi \mid s]$ must be less dispersed than $\mathbb{E}[\phi \mid q_\phi]$ under the convex order. Consequently, in decision problems where only posterior means of ϕ are relevant (e.g., when a decision-maker's payoff is affine in some statistic ϕ), the ϕ -quantile signal gives the highest expected payoff.

[Theorem 2](#) reduces the set of privacy-preserving signals to mean-preserving contractions of \overline{F}_ϕ in settings where only posterior means of a statistic ϕ are relevant. In particular, the privacy constraints, however complex they are, are entirely summarized by the mean-preserving-contraction upper bound \overline{F}_ϕ . As a result, economic insights developed in mean-measurable settings can be readily extended even if arbitrary privacy concerns are present, as we illustrate in [§5](#). Moreover, the mean-preserving contraction structure, together with recent results in [Kleiner, Moldovanu and Strack \(2021\)](#) and [Arieli, Babichenko, Smorodinsky and Yamashita \(2023\)](#), who characterize the extreme points of this set, allows one to focus on signals that either fully reveal the q_ϕ -quantile signal, or pool its realizations into at most two mass points on an interval.

Remark 2 (Conditionally Privacy-Preserving Signals). In many economic applications, a

¹²More precisely, this is a version of the Hardy-Littlewood-Polya inequality (see, e.g., [Hardy et al. 1929](#) and [Puccetti and Wang 2015](#), Theorem 2.1), as the function $(z_1, z_2) \mapsto \mathbf{1}\{z_1 \geq t\}z_2$ is supermodular for all $t \in [0, 1]$.

signal is only required to be privacy-preserving conditional on certain information. For example, if some signal $y : \Omega \times [0, 1] \rightarrow Y$ is already publicly available, then it would be natural to only restrict signals to not reveal *additional* information. This can be captured by defining a signal to be *conditionally privacy-preserving*, if for all privacy set $P \in \mathcal{P}$, a.s.,

$$\mathbb{P}[\omega \in P \mid s, y] = \mathbb{P}[\omega \in P \mid y].$$

Mathematically, [Theorem 1](#) and [Theorem 2](#) immediately extend to this case by simply applying them for each realization of y (see the Online Appendix).

We conclude this section with two examples that demonstrate how a conditionally revealing quantile signal can be constructed, as well as how reorderings are sometimes necessary.

Example 1 (The Single-Dimensional Case). Suppose that $\Omega = X \times \Theta$, where $x \in X \subseteq \mathbb{R}$ denotes an economic outcome (e.g., insurance risk, income, etc.), suppose that the privacy sets are defined by the component $\theta \in \Theta$ of the state,¹³ and suppose that the distribution $F(\cdot \mid \theta)$ of x conditional on θ is continuous. Then, the quantile signal $q^* := F(x \mid \theta)$ is conditionally revealing, as it reveals (x, θ) to an outside observer who knows θ . By [Theorem 1](#), every privacy-preserving signal is thus Blackwell-dominated by a reordering of q^* . By [Theorem 2](#), a CDF G is the distribution of posterior means $\mathbb{E}[x \mid s]$ induced by a privacy-preserving signal s if and only if G is a mean-preserving contraction of \bar{F} , where

$$\bar{F}(z) := \inf\{q \in [0, 1] : \mathbb{E}[F^{-1}(q \mid \theta)] \geq z\}.$$

Example 2 (Undominated Privacy-Preserving Signals). We next present an example illustrating that a privacy-preserving signal might seem to reveal little information but is in fact undominated. Consider the same setting as in [Example 1](#), and further suppose that $X = \{0, 1\}$. The quantile signal q^* can be represented by assigning to each point in the unit interval the value of x they reveal given θ (see [Figure 1a](#)). For example, a signal realization $q^* = 0.5$ reveals that $x = 1$ if $\theta = 0$, and that $x = 0$ if $\theta = 1$. This signal thus perfectly reveals that $x = 0$ for realizations $\hat{s} \leq 1/4$, and that $x = 1$ for realizations $\hat{s} \geq 3/4$, while the posterior over X remains uniform and assigns equal probability to $x = 1$ and $x = 0$ for realizations $\hat{s} \in (1/4, 3/4)$.

¹³That is, suppose that \mathcal{P} is the σ -algebra generated by the projection $(x, \theta) \mapsto \theta$.



Figure 1: Two examples of reordered quantile signals. The signal realizations are uniformly distributed over $[0, 1]$ and correspond to different points on the line. The color indicates which state is revealed by a given signal realization conditional on θ .

Meanwhile, the following reordering of q^* , defined by the measure-preserving transformations $M_0(s) = s$ and $M_1(s) = 1 - s$, can be represented by [Figure 1b](#). This signal induces a uniform posterior assigning equal probability to $x = 1$ and $x = 0$ for any realization of s .

Note that the distribution of posterior means under q^* is more dispersed than under its reordering, as stated by [Theorem 2](#). In fact, the distribution of posterior means under the reordered q^* is degenerate and assigns probability one to the prior mean. Nonetheless, according to [Theorem 1](#), these two signals are both Blackwell-undominated among all privacy-preserving signals. In particular, even though the reordered quantile signal given by [Figure 1b](#) does not reveal any information about x or θ separately, it is still Blackwell-undominated. This is because the signal given by [Figure 1b](#) perfectly reveals the *correlation* between x and θ , so that a decision-maker who cares only about whether $x = \theta$ can have their payoff maximized by observing this signal.

4 Optimizing over Privacy-Preserving Signals

4.1 Decision Problems

In this section, we apply [Theorem 1](#) and [Theorem 2](#) to characterize optimal privacy-preserving signals for a decision-maker who takes an action after observing the signal. Proofs for this section and the next section can be found in the Online Appendix. For the ease of exposition, we assume finitely many privacy sets so that θ takes finitely many values: $\Theta = \{1, \dots, J\}$. Consider a Bayesian decision problem (u, A) : After observing the signal s , the decision-maker

chooses an action $a \in A$ to maximize expected payoff

$$\mathbb{E}[u(\omega, a) \mid s]$$

where $u(\omega, a)$ denotes the decision-maker's ex-post payoff when the state is ω and the action is a . From [Theorem 1](#) and Blackwell's theorem, it follows that there always exists an optimal privacy-preserving signal that is a reordering of some conditionally revealing quantile signal. For any reordering of a conditionally revealing quantile signal s , denote by

$$\tilde{\omega}(s) := (\tilde{\omega}_1(s), \dots, \tilde{\omega}_J(s))$$

the vector of states revealed by the signal s , i.e., $\tilde{\omega}_j(\hat{s})$ is the state revealed by signal realization \hat{s} if $\theta = j$. The marginals of $\tilde{\omega}$ are fixed by $\tilde{\omega}_j \sim \mathbb{P}[\cdot \mid \theta = j]$. Let \mathcal{D} be the set of joint distributions ρ on Ω^J such that the marginal of the j -th coordinate is given by $\mathbb{P}[\cdot \mid \theta_j]$. Thus, any such signal corresponds to a coupling of states for different realizations of θ :

Lemma 3. *A distribution $\rho \in \Delta(\Omega^J)$ is the joint distribution of $\tilde{\omega}(s)$ for some reordering of a conditionally revealing quantile signal s if and only if $\rho \in \mathcal{D}$.*

With [Lemma 3](#), we can now characterize the optimal privacy-preserving signals for the decision-maker. Let $V : \Omega^J \rightarrow \mathbb{R}$ be defined as

$$V(\omega_1, \dots, \omega_J) := \sup_{a \in A} \left(\sum_{j=1}^J u(\omega_j, a) \mathbb{P}[\theta = j] \right), \quad (3)$$

for all $(\omega_j)_{j=1}^J \in \Omega^J$. The value of $V(\omega_1, \dots, \omega_J)$ can be interpreted as the decision-maker's indirect utility after learning that the state equals ω_j when $\theta = \theta_j$, without any further information about θ . We then have the following characterization:

Proposition 2 (Optimal Privacy-Preserving Signal). *The decision-maker's optimal value among all privacy-preserving signals is given by*

$$\sup_{s: s \perp \theta} \mathbb{E} \left[\sup_{a \in A} \mathbb{E}[u(\omega, a) \mid s] \right] = \sup_{\rho \in \mathcal{D}} \int_{\Omega^J} V(\omega_1, \dots, \omega_J) d\rho, \quad (4)$$

Moreover, fix any conditionally revealing quantile signal q^* , there exists an optimal privacy-preserving signal that is Blackwell-equivalent to a reordering s of q^* such that the distribution

of $\tilde{\omega}(s)$ is a solution of (4).

The optimization problem (4) is a multi-marginal optimal transport problem. The existence of solutions can be guaranteed if Ω is compact and if V is upper-semicontinuous, which we assume henceforth. While the optimal transport problem (4) may still be complex, the structure of this problem leads to explicit characterizations of optimal privacy-preserving signals for a wide range of economically relevant decision problems. Indeed, as we show below, a closed form solution can be obtained within a certain class of decision problems.

Supermodular Payoffs Many applications naturally admit supermodularity, which means that the objective depends on the state only through a single dimensional statistic ϕ , and higher (or lower) actions are optimal for higher values of the statistic. For example, if $\phi(\omega)$ measures the probability of a borrower repaying a loan and a is the interest rate a bank requires from a borrower, then it would be natural to assume that the bank wants to charge a lower interest rate to those borrowers who are more likely to repay.

Proposition 3 (Supermodular Payoffs). *Suppose that A is a totally ordered set, and that there exists a statistic $\phi : \Omega \rightarrow \mathbb{R}$ such that for all $\omega \in \Omega$ and for all $a \in A$,*

$$u(\omega, a) = h(\phi(\omega), \theta(\omega), a)$$

for some measurable $h : \mathbb{R} \times \Theta \times A \rightarrow \mathbb{R}$ that is supermodular in (ϕ, a) . Then the ϕ -quantile signal is optimal among all privacy-preserving signals.

According to [Proposition 3](#), for any decision-maker who chooses an action a to maximize a payoff $h(\phi, \theta, a)$ that is supermodular in (ϕ, a) , the ϕ -quantile signal is optimal. For example, for any prior, statistic ϕ , and collection of privacy sets \mathcal{P} , the loss $|\phi(\omega) - a|^p$ for $p > 1$ is minimized by revealing the ϕ -quantile signal.¹⁴

Binary Actions Another important special case is when the decision is only between two actions, e.g., a bank decides whether to extend a loan at an exogenously fixed interest rate.

Proposition 4 (Binary Actions). *Suppose that $A = \{0, 1\}$, then, among all privacy-preserving signals, the ϕ -quantile signal is optimal, where $\phi(\omega) = u(\omega, 1) - u(\omega, 0)$.*

¹⁴The optimal action $a : [0, 1] \rightarrow \mathbb{R}$ in this case is given as the solution to $\mathbb{E}[(\phi(\omega) - a(z))^{p-1} | q_\phi = z] = 0$.

[Proposition 4](#) follows immediately from [Proposition 3](#), since the ϕ -quantile signal is optimal as $u(\omega, a) = \phi(\omega) a + u(\omega, 0)$ is supermodular in (ϕ, a) . [Proposition 4](#) thus completely solves the decision problem for binary actions, arbitrary payoffs, and arbitrary privacy sets.

Separable Problems and Binary States In the meantime, [Theorem 2](#) also implies that for a specific class of “separable” decision problems, the quantile signal is optimal.

Proposition 5 (Separable Problems and Binary States). *Suppose there exists $\phi : \Omega \rightarrow \mathbb{R}$ and functions $f : \mathbb{R} \times A \rightarrow \mathbb{R}$, $g : A \rightarrow \mathbb{R}$, $h : A \times \Theta \rightarrow \mathbb{R}$ such that*

(i) $u(\omega, a) = \phi(\omega)g(a) + h(a, \theta(\omega))$ for all $\omega \in \Omega$, or

(ii) $u(\omega, a) = f(\phi(\omega), a)$ and $\phi(\omega) \in \{0, 1\}$ for all $\omega \in \Omega$.

Then the ϕ -quantile signal is optimal among all privacy-preserving signals.

Part (ii) of [Proposition 5](#) was previously obtained by [He et al. \(2023\)](#). Both part (i) and (ii) of [Proposition 5](#) follow immediately from [Theorem 2](#), which establishes that the quantile signal is the signal that induces the most dispersion in the posterior means. The assumption that $u(\omega, a) = \phi(\omega)g(a) + h(a, \theta)$ is satisfied in many applications. In particular, a payoff function u that is constant in θ and affine in ϕ satisfies this condition. For instance, in the bank loan example, this corresponds to assuming that the bank’s preference depends only on the expected amount repaid by the borrower, but not the race, gender, age, etc. of the borrower. Meanwhile, since posterior distributions over ϕ are one-dimensional when $\phi(\omega) \in \{0, 1\}$ for all $\omega \in \Omega$, the second part of the corollary follows.

4.2 Information Design

Consider the Bayesian persuasion setting where a sender discloses information about $\omega \in \Omega$ to a receiver who chooses an action $a \in A$, and suppose that the sender is restricted to choose only signals that are privacy-preserving.¹⁵ Let the sender’s payoff be $u_S : \Omega \times A \rightarrow \mathbb{R}$ and the receiver’s payoff be $u_R : \Omega \times A \rightarrow \mathbb{R}$. Let V_S^* be the sender’s value from choosing the optimal

¹⁵For example, the sender might be the prosecutor, as in [Kamenica and Gentzkow \(2011\)](#), trying to convince the judge that the defendant should not be released on bail, but is restricted to not using any information related to the race of the defendant even though such information might be predictive about the probability of reoffense.

privacy-preserving signal. For simplicity, suppose again that $|\Theta| = J < \infty$ and write Θ as $\{1, \dots, J\}$. Let $V_R : \Omega^J \times A \rightarrow \mathbb{R}$ be defined as

$$V_R(\omega_1, \dots, \omega_J, a) := \sum_{j=1}^J u_R(\omega_j, a) \mathbb{P}[\theta = j],$$

for any $(\omega_j)_{j=1}^J \in \Omega^J$. Moreover, for any $\rho \in \Delta(\Omega^J)$, let

$$V_S(\rho) := \mathbb{E}_\rho \left[\sum_{j=1}^J u_S(\omega_j, a^*(\rho)) \mathbb{P}[\theta = j] \right],$$

where $a^*(\rho)$ is the (sender-preferred) optimal action of the receiver that maximizes V_R when the posterior over $(\omega_j)_{j=1}^J$ is ρ . To ensure the existence of optimal signals, we assume that Ω is compact and that V_S is upper-semicontinuous. The next proposition characterizes the sender's value V_S^* .

Proposition 6 (Value of Persuasion). *Let \bar{V}_S be the concave closure of V_S . The sender's value V_S^* is given by*

$$V_S^* = \max_{\rho \in \mathcal{D}} \bar{V}_S(\rho),$$

Proposition 6 states that the sender's value can be found by a two-step procedure: First, fix a joint distribution ρ and find the optimal garbling of it by computing $\bar{V}_S(\rho)$. Then, optimize across $\rho \in \mathcal{D}$. Just as in standard persuasion problems, the characterization of **Proposition 6** requires computing the concave closure of the function V_S , which is typically computationally demanding. Nonetheless, when payoffs are such that the sender's indirect utility is measurable with respect to the posterior mean of some statistic ϕ , **Theorem 2** provides a much more tractable way to characterize optimal signals. Specifically, suppose that there exists a statistic $\phi : \Omega \rightarrow \mathbb{R}$ such that the sender's indirect utility is a function only of the posterior mean $\mathbb{E}[\phi | s]$, which we denote by $U_S : \mathbb{R} \rightarrow \mathbb{R}$. Then the sender's payoff given a signal can be written as $\int_{\mathbb{R}} U_S(x) dG$, where G is the CDF of posterior means induced by the signal. **Theorem 2** implies the following characterization:

Proposition 7 (Value of Mean-Measurable Persuasion). *Suppose that the sender's indirect utility is measurable with respect to posterior means and is denoted by $U_S : \mathbb{R} \rightarrow \mathbb{R}$. The*

sender's value V_S^* is given by

$$V_S^* = \sup_{G \leq_{\text{MPS}} \bar{F}_\phi} \int_{\mathbb{R}} U_S(x) dG, \quad (5)$$

As a result, since the objective of (5) is affine and since the feasible set is convex, one of the solutions must be an extreme point of the feasible set. If U_S is upper-semicontinuous and if \bar{F}_ϕ is continuous, then by Theorem 2 of [Kleiner et al. \(2021\)](#), there must be a solution G^* that takes the form of

$$G^*(x) = \begin{cases} \bar{F}_\phi(x), & \text{if } x \notin \cup_{i \in I} [\underline{x}_i, \bar{x}_i) \\ \bar{F}_\phi(\underline{x}_i), & \text{if } x \in [\underline{x}_i, \underline{y}_i) \\ x_i, & \text{if } x \in [\underline{y}_i, \bar{y}_i) \\ \bar{F}_\phi(\bar{x}_i), & \text{if } x \in [\bar{y}_i, \bar{x}_i) \end{cases},$$

for some collection of intervals $\{[\underline{x}_i, \bar{x}_i]\}_{i \in I}$ and $\{[\underline{y}_i, \bar{y}_i]\}_{i \in I}$ such that $[\underline{y}_i, \bar{y}_i] \subseteq [\underline{x}_i, \bar{x}_i]$, and for some $\{x_i\}_{i \in I} \subseteq \mathbb{R}$. Furthermore, G^* can be implemented by fully revealing the realizations q_ϕ of the ϕ -quantile signal whenever $q_\phi \notin \cup_{i \in I} [\underline{x}_i, \bar{x}_i)$, while pooling at most two signal realizations in each interval $[\underline{x}_i, \bar{x}_i)$ for all $i \in I$.

The characterization given by [Theorem 2](#) and [Proposition 7](#) is particularly convenient for introducing privacy concerns to mean-measurable information design problems (see, for instance, [Kolotilin, Li, Mylovanov and Zapechelnnyuk 2017](#); [Kolotilin 2018](#); [Dworczak and Martini 2019](#); [Ali, Haghpanah, Lin and Siegel 2022](#); [Bergemann et al. 2022](#); [Kolotilin and Wolitzky 2023](#); [Bergemann, Heumann and Morris 2023, 2024](#)) since the structure of the feasible privacy-preserving signals perfectly aligns with that of all feasible signals. In other words, the privacy constraints, however complex they are, are all summarized by the mean-preserving-spread upper bound \bar{F}_ϕ . We further demonstrate the value of this characterization in §5.

5 Economic Applications

To illustrate the relevance of privacy-preserving signals and to demonstrate the implications of our main results, we discuss several economic examples.

5.1 Statistical Discrimination and Algorithmic Fairness

Statistical discrimination refers to discriminatory outcomes that arise simply because some protected characteristics are statistically correlated with the payoff-relevant variables—and thus are used in decision-making. Legal studies and the literature on algorithmic fairness in computer science aim to explore optimal ways to discipline statistical discrimination that arise from (algorithmic-assisted) decision-making.

Informational Environment We first introduce a general environment that can be used to study discrimination. There is an underlying outcome $\gamma \in \Gamma$. A decision-maker observes covariates $(\theta, y, z) \in \Theta \times Y \times Z$ that are correlated with the outcome γ , and has to take an action $a \in A$. Among the observable covariates (θ, y, z) , θ denotes the “protected characteristics” (e.g., race and gender), y denotes the “materially relevant characteristics,” which are deemed acceptable to be used despite their correlations with θ (e.g., credit history or criminal history), and z denotes all other observable covariates (e.g., zipcode). The decision-maker’s payoff equals $\hat{u}(\gamma, a) \in \mathbb{R}$ when the outcome is γ and when the action is a .

A statistical model (or an algorithm) takes the inputs (θ, y, z) and outputs a prediction $s \in S$ for the outcome γ . The decision-maker then chooses an action a to maximize $\mathbb{E}[\hat{u}(\gamma, a) \mid s]$. As protected characteristics θ might be correlated with outcome γ , so might the prediction s . Therefore, the resulting outcome might exhibit statistical discrimination in the sense that a might be correlated with θ , even if \hat{u} does not depend on θ . A common regulatory approach is to discipline the degree of *disparate impact*—a legal term that refers to correlations between a and θ , regardless of whether the decision is made explicitly based on θ .¹⁶ This approach requires the action taken by the decision-maker to be independent of the protected characteristics θ , conditional on materially relevant characteristics y .¹⁷ This approach also coincides with a commonly adopted notion of fairness in computer science,

¹⁶For instance, in the case of *Griggs v. Duke Power Company*, the U.S. Supreme Court held that requiring a high school degree and an aptitude test for job transfers to certain departments creates disparate impacts for Black employees due to the history of segregation (i.e., transfer decision a is correlated with race θ). The court thus found such a policy in violation of Title VII of the Civil Rights Act, even though these requirements do not explicitly refer to race.

¹⁷In practice, an essential argument employers use in defense of disparate impact claims is the “business necessity” criterion. Namely, the inputs upon which the alleged discriminatory policies or rules are essential for the success of the defendant’s business. These necessities can be interpreted as the materially relevant characteristics y .

which is referred to as *demographic parity*.¹⁸

For example, suppose that a bank, which faces many loan applicants with observable characteristics (θ, y, z) , needs to make loan decisions a . The relevant outcome is whether an applicant will default in the future, denoted by $\gamma \in \{0, 1\}$. The Equal Credit Opportunity Act (15 U.S.C. 1691 et seq.) “*prohibits creditors from discriminating against credit applicants on the basis of race, color, religion, national origin, sex, marital status, age [..]*.” A concrete (although stringent) interpretation of this requirement taken in the literature is that the information about each individual’s default probability that the bank uses to make loan decisions must be independent of an individual’s protected characteristics (potentially conditional on materially relevant information, such as income). This requirement avoids the problem that even when restricting the bank to not condition on race directly, it might still do so indirectly through the use of covariates such as zip code, which is a well-known issue highlighted in the actuarial sciences and legal studies. For example, [Wiggins \(2020\)](#) states:¹⁹

“[...] race has become so highly correlated with other social statistics that actuarial science in general has developed a baked-in racial bias. Racial discrimination by proxy (e.g., zip code standing in for race) can be glimpsed in the disparate impact of data-driven decision-making in housing, healthcare, policing, sentencing, and more. Simply leaving out racial data in statistically aided decision-making distances institutions from claims of intentional discrimination, but a disparate, discriminatory impact lingers when other factors correlated with race power ac-

¹⁸See, e.g., [Darlington \(1971\)](#); [Calders et al. \(2009\)](#); [Dwork, Hardt, Pitassi, Reingold and Zemel \(2012\)](#); [Calders and Verwer \(2010\)](#); [Kamishima, Akaho and Sakuma \(2011\)](#); [Corbett-Davies et al. \(2017\)](#); [Gillis, McLaughlin and Spiess \(2021\)](#). Two other commonly adopted criteria are (i) *separation*, which requires balanced type-I and type-II errors (see, e.g., [Hardt et al. 2016](#)), or more generally, independence between a and θ conditional on the true outcome γ ; and (ii) *sufficiency*, which requires the action a to be a sufficient statistic for γ , so that the outcome γ is independent of θ conditional on a . It is well-known that none of any pairs of these three common fairness criteria can be satisfied at the same time, and hence the choice of a fairness criteria is necessary (see [Barocas, Hardt and Narayanan \(2019\)](#) and [Carey and Wu \(2023\)](#) for a comprehensive review of these criteria). With appropriate projections, our results can also be applied when the notion of separation, instead of independence, is adopted. See the Online Appendix for more details.

¹⁹Former Attorney Eric Holder also made a similar remark in the context of sentencing: “[...] basing sentencing decisions on static factors and immutable characteristics—like the defendant’s education level, socioeconomic background, or neighborhood—they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society.” See <https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th>.

tuarial analyses.”

Existing results in the fairness literature (see, e.g., [Calders et al. 2009](#); [Hardt et al. 2016](#); [Corbett-Davies et al. 2017](#)) take the “recommendation approach” (i.e., by taking $S = A$ so that the prediction s is a recommended action) and solve for the optimal recommendation subject to the constraint that a has to be statistically independent of θ . This approach is convenient, as it reduces the statistical problem of finding the optimal prediction s to a linear program, which is particularly tractable when the available actions are simple. The literature thus characterizes optimal algorithms in simple settings when the decision-maker’s choice is binary, e.g., when the bank only decides whether to grant a loan, and when the payoff is given by $\hat{u}(\gamma, a) = a \cdot (1 - \gamma - c)$ for some $c \in (0, 1)$.²⁰

However, in many applications, the decision-maker may need to make more than a binary choice. For example, a bank typically needs to decide—in addition to whether to grant the loan—how much to grant, what the interest rate should be, the amount of the down payment, and the form of the collateral. Moreover, in regulatory contexts, a regulator typically faces a wide-range of decision problems that are very different in their natures. For these problems, using the recommendation approach might not be as fruitful, since the approach relies on the specific payoff structure and benefits the most from the simplicity of the action space.²¹

A General Solution [Theorem 1](#) and [Theorem 2](#) lead to characterizations of optimal algorithms for arbitrary decision problems. To apply our results while incorporating the fact that only the covariates (θ, y, z) are observable, one may define the state as $\omega = (x, \theta) \in X \times \Theta = \Delta(\Gamma) \times \Theta$, where x corresponds to the distribution of outcomes γ conditional on (θ, y, z) .²² The expected payoff is then given by

$$u(x, \theta, a) := \int_{\Gamma} \hat{u}(\gamma, a) dx(\gamma).$$

By [Proposition 2](#), the optimal algorithms can be characterized by solving the optimal trans-

²⁰The optimal algorithms adopt different thresholds for different groups θ (conditional on materially relevant characteristics), and chooses action $a = 1$, e.g., grants the loan to an applicant, if and only if the conditional expectation $\mathbb{E}[\gamma \mid \theta, y, z]$, e.g., expected default probability, is below their group-specific thresholds.

²¹In fact, we are not aware of any paper that explicitly solves a model with more than two actions, or that solves for an optimal algorithm across many decision problems.

²²Formally, $x : \Theta \times Y \times Z \rightarrow \Delta(\Gamma)$ with $x(\theta, y, z) = \mathbb{P}[\gamma \in \cdot \mid \theta, y, z]$.

port problem (4). Moreover, [Proposition 3](#), [Proposition 4](#), and [Proposition 5](#) lead to explicit characterizations of an optimal fair algorithm under a wide class of economically relevant environments.

Perfectly Informative Covariates Suppose that $\Gamma \subseteq \mathbb{R}$ and that the covariates (θ, y, z) are perfectly informative, so that γ is a function of (θ, y, z) . In this case, X can be equivalently defined as Γ . If \hat{u} is supermodular in (γ, a) , then by [Proposition 3](#), the quantile signal is optimal. If \hat{u} is separable so that $\hat{u}(\gamma, a) = f(\gamma)g(a) + h(a)$, for some real-valued measurable functions f, g, h , then, according to [Proposition 5](#), the $\hat{\phi}$ -quantile signal is optimal, where $\hat{\phi}(\gamma, \theta) := f(\gamma)$. For instance, a large corporation decides whether to hire a worker, and if so, which position to assign the worker. Workers differ in their skill levels $\gamma \in [0, 1]$, and positions $a \in [0, 1]$ differ in their difficulties, with $a = 0$ being not hiring the worker. The employer’s payoff is given by $v(\gamma, a) - w(a)$, where v is a supermodular production function and w is the market wage schedule for different positions.

Binary Outcomes Suppose that $\Gamma = \{0, 1\}$. Then, $X = \Delta(\Gamma) = [0, 1]$, with x being the probability that $\gamma = 1$. Moreover,

$$u(x, \theta, a) = x\hat{u}(1, a) + (1 - x)\hat{u}(0, a)$$

is affine in x . [Proposition 5](#) (i) then implies that the quantile signal, which can simply be constructed by computing the quantile of the predicted score $x \in [0, 1]$ conditional on each protected characteristic, is optimal.²³ This set of assumptions is common in many relevant economic settings, such as loan decisions and bail decisions. For example, a lender makes loan decisions regarding an applicant. The applicant may either default ($\gamma = 0$) in the future or not ($\gamma = 1$). The actions of the bank could be highly complex, including deciding whether to give the loan, deciding interest rates, design payment schedules, and the choice of collaterals.

The Orthogonalization Procedure In the environments above, the quantile signals are optimal. In fact, the quantile signal can be generated by the following concrete and *detail-free*

²³This result generalizes the characterization of algorithmically fair optimal decisions obtained in Theorem 2 of [He et al. \(2023\)](#), which assumes the outcome γ to be binary, and the covariates to be *perfectly informative* about the outcome.

procedure that does not depend on the underlying decision-making problems:

1. Use the most efficient algorithm and all covariates to generate a (one-dimensional) prediction score ϕ for the relevant outcome γ .
2. Adopt a “post-processing step” after generating the raw prediction and *orthogonalize* these predictions by computing the quantile signal q_ϕ of the predicted scores for each group of protected characteristics, conditional on the materially relevant characteristics.
3. Decision-makers take an action based on the quantiles and the materially relevant characteristics.

If either the covariates are perfectly informative about a one-dimensional outcome $\gamma \in \mathbb{R}$, or the outcome γ is binary, as shown above, the quantile signal—and hence the output of the orthogonalization procedure—is always optimal. It is noteworthy that this procedure does not require *any* further details about the underlying decision problems. In particular, in the bank loan example, however complex a lender’s decision problem is, and regardless of how many different decision-makers a regulator faces, adopting the orthogonalization procedure always leads to an optimal outcome.

This is of practical importance, as it allows the reuse of existing econometric and statistical tools. The first step of the procedure can be carried out by an econometrician without any consideration for discrimination. The second step ensures that whatever procedure the econometrician uses, the final outcome will be non-discriminatory. While one might expect that such a procedure would lead to suboptimal decision rules, our result identifies conditions under which this two-step procedure will be optimal if the econometrician uses the best possible predictor of γ in the first step.²⁴

This procedure also suggests a concrete approach to ensuring fairness. A regulator could legally require every decision-maker to adopt a post-processing step and compute the quantiles after generating raw prediction scores, and enforce this requirement by audits or subpoena in legal proceedings.²⁵ This “output regulation” is distinct from the commonly adopted

²⁴Kamiran, Žilobaitė and Calders (2013) and Feldman, Friedler, Moeller, Scheidegger and Venkatasubramanian (2015) propose a similar procedure to “repair” unfair algorithms. See also: Calders and Verwer (2010) and Kamishima et al. (2011) for earlier work on repairing and regularizing unfair algorithms. This literature focuses on transforming any (potentially unfair) algorithm into a fair one. Our results imply that, not only does a similar procedure lead to a fair algorithm, it is in fact the *optimal* way to transform any algorithm into a fair one.

²⁵If the regulator requires the decision-maker to store the prediction scores, they can at any point in time

“input regulation” approach which limits the data available to the decision-maker (which is often inefficient due to covariates that are correlated with protected characteristics). Nevertheless, the implementation and enforcement can be fulfilled by similar regulatory tools.

5.2 Optimal Privacy-Preserving Disclosure in Auctions

A publisher runs a second-price auction to sell targeted audiences to $N \geq 1$ advertisers with private values $\{v_i\}_{i=1}^N \in \mathbb{R}_+^N$ that are independently drawn from F .²⁶ As in Bergemann et al. (2022), we assume that F is absolutely continuous. The publisher has control over how much information about the audience to give to the advertisers. Specifically, each advertiser observes a (conditionally independent) signal about their own value. Bergemann et al. (2022) study this problem and characterize the optimal symmetric signal to disclose to the advertisers. That is, they characterize the symmetric and independent signal $s : \mathbb{R}_+ \times [0, 1] \rightarrow \mathbb{R}$ that maximizes the publisher’s revenue, where each advertiser i observes a signal $s_i = s(v_i, r_i)$, with $\{r_i\}_{i=1}^n$ being the independent randomization device that is uniformly distributed on $[0, 1]$. They characterize the optimal signal as follows:

Proposition 8 (Bergemann et al. 2022, Theorem 1). *Any optimal symmetric information structure is equivalent, in terms of bidders’ expected values $\mathbb{E}[v_i | s_i]$, to a partitional signal $s : \mathbb{R}_+ \times [0, 1] \rightarrow \mathbb{R}$*

$$s(v_i, r_i) = \begin{cases} v_i & \text{if } F(v_i) < q_N^* \\ \mathbb{E}[v_i | F(v_i) \geq q_N^*] & \text{if } F(v_i) \geq q_N^* \end{cases},$$

where q_N^* is the unique root in $(0, 1)$ of a N -degree polynomial and does not depend on F .

In many economically relevant situations, disclosing some type of information about the audience is controversial. For example, Target was widely criticised for using pregnancy information in advertising.²⁷ Google and other advertising companies have been criticized for disclosing highly sensitive information to advertisers, including religious beliefs, ethnicities,

use these to check whether the procedure was applied correctly and apply appropriate penalties in case it was not.

²⁶We can interpret the values $\{v_i\}_{i=1}^N$ as the ex-post click-through rate by assuming that the value brought to each advertiser conditional on a click-through of an audience is a positive constant.

²⁷See <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=412314036668>.

diseases, disabilities, sexual orientation, and whether a user informed themselves online about incest and sexual abuse.²⁸

To model such a situation within our framework, suppose that there is some protected information that should not be disclosed to advertisers. Such information is correlated with each advertiser’s values v_i and is summarized by $\theta \in \Theta$. Assume that $\{v_i\}_{i=1}^N$ is independent and identically distributed conditional on θ .²⁹ Let $F(\cdot | \hat{\theta})$ be the distribution of v_i conditional on the realization $\hat{\theta}$ of θ . As before, we assume that $F(\cdot | \hat{\theta})$ is absolutely continuous for all $\hat{\theta} \in \Theta$.

The publisher is not allowed to disclose any information about θ and thus can only disclose a privacy-preserving signal to each advertiser. A symmetric privacy-preserving signal $s : \mathbb{R}_+ \times \Theta \times [0, 1] \rightarrow \mathbb{R}$ discloses $s(\hat{v}_i, \hat{\theta}, \hat{r}_i)$ to advertiser i when their value is $\hat{v}_i \in \mathbb{R}_+$, the protected characteristic is $\hat{\theta} \in \Theta$, and the randomization device is \hat{r}_i . A symmetric signal s is privacy-preserving if $s(v_i, \theta, r_i)$ is independent of θ for all i .

As the advertisers’ preferences only depend on their expected value $\mathbb{E}[v_i | s_i]$ after observing a signal s_i , we can apply [Theorem 2](#) to conclude that for any symmetric signal s , the distribution of $\mathbb{E}[v_i | s]$ must be a mean-preserving contraction of $\mathbb{E}[v_i | q_i]$, where q_i is the quantile signal

$$q_i := F(v_i | \theta),$$

whose distribution is given by $\bar{F}(z) := \inf\{y \in [0, 1] : \mathbb{E}[F^{-1}(y | \theta)] \geq z\}$. Moreover, for any symmetric privacy-preserving signal, advertisers’ conditional expected values are independently and identically distributed.³⁰ Together with the result from [Bergemann et al. \(2022\)](#), we obtain a complete characterization of optimal information disclosure in auctions subject to privacy constraints.

Proposition 9. *Any optimal symmetric information structure that does not reveal any information about θ is equivalent, in terms of bidder’s expected values $\mathbb{E}[v_i | s_i]$, to a partitional*

²⁸See <https://mashable.com/article/google-iab-gdpr-complaint>.

²⁹For instance, if θ denotes gender, then the values of each advertiser to target a female or male audience are assumed independently and identically distributed due to, say, heterogeneity among advertisers—just as they were assumed to be independent and identically without conditioning in [Bergemann et al. \(2022\)](#).

³⁰To see this, consider any symmetric privacy-preserving signal s . Let $s_i := s(v_i, \theta, r_i)$ for all i . Since $\{v_i\}_{i=1}^n$ is independent conditional on θ , $\mathbb{P}[s_1 \leq z_1, \dots, s_N \leq z_N] = \mathbb{E}[\mathbb{P}[s_1 \leq z_1, \dots, s_N \leq z_N | \theta]] = \mathbb{E}[\mathbb{P}[s_1 \leq z_1 | \theta] \cdots \mathbb{P}[s_N \leq z_N | \theta]]$. Moreover, since s is privacy-preserving, we have $\mathbb{E}[\mathbb{P}[s_1 \leq z_1 | \theta] \cdots \mathbb{P}[s_N \leq z_N | \theta]] = \mathbb{E}[\mathbb{P}[s_1 \leq z_1] \cdots \mathbb{P}[s_N \leq z_N]] = \mathbb{P}[s_1 \leq z_1] \cdots \mathbb{P}[s_N \leq z_N]$.

signal $s : \mathbb{R}_+ \times \Theta \times [0, 1] \rightarrow \mathbb{R}_+$

$$s(v_i, \theta, r_i) = \begin{cases} \mathbb{E}[v_i | F(v_i | \theta)] & \text{if } F(v_i | \theta) < q_N^* \\ \mathbb{E}[v_i | F(v_i | \theta) \geq q_N^*] & \text{if } F(v_i | \theta) \geq q_N^* \end{cases},$$

where q_N^* is the same as in [Proposition 8](#).

Example 3. Suppose that the willingness to pay for a good or service is higher among people with a certain sexual orientation, ethnic group, or among people with a specific disease. The auctioneer does not want to reveal any information about this sensitive information encoded in θ . For concreteness, suppose that the values are drawn from an exponential distribution conditional on protected information θ . Simple algebra shows that the optimal privacy-preserving signal is given as

$$s(v_i, \theta, r_i) = \begin{cases} v_i \times \frac{\mathbb{E}[v_i]}{\mathbb{E}[v_i | \theta]} & \text{if } F(v_i | \theta) < q_N^* \\ \mathbb{E}[v_i | F(v_i) \geq q_N^*] \times \frac{\mathbb{E}[v_i]}{\mathbb{E}[v_i | \theta]} & \text{if } F(v_i | \theta) \geq q_N^* \end{cases}.$$

This signal pools agents from different groups with different values in a way not present in the optimal signal without privacy concerns (given in [Proposition 8](#)): Values of agents from a group θ with a higher expected valuation $\mathbb{E}[v_i | \theta]$ are multiplicatively scaled down, while values of agents from a group with an ex-ante lower valuation are multiplicatively scaled up.

5.3 Price Discrimination

Our results can also be applied to settings of price discrimination in the spirit of [Bergemann et al. \(2015\)](#). Consider a monopolist who uses consumer data to price-discriminate consumers. The monopolist sells a single product to a unit mass of consumers. Each consumer demands a single unit and has value $x \in X := [\underline{x}, \bar{x}] \subset \mathbb{R}_+$ for the product. Moreover, each consumer belongs to one of the protected groups $\theta \in \Theta = \{\theta_1, \dots, \theta_J\}$. These characteristics are correlated with consumers' values. For each $j \in \{1, \dots, J\}$, let $F(\cdot | \theta_j)$ be the distribution of values of consumers with characteristic $\theta = \theta_j$. With different combinations of consumer data, the monopolist is able to charge different prices to different groups of consumers and engage in third-degree price discrimination.

While consumer data enables the monopolist to engage in price discrimination, it is often required by law or regulations that consumers cannot be price-discriminated based upon their protected characteristics. For example, a recent legislation (AB1287) in California specifically prohibits businesses from price-discriminating based on gender. Given such legal constraints, it is natural to ask: What market segmentations allow the monopolist to price-discriminate, but are not based on protected characteristics?

Clearly, the market segmentation that fully segments consumers by their values allows the monopolist to extract all the surplus. This, however, would typically lead to price discrimination based on protected characteristics, in the sense that consumers of different characteristics would face a different distribution of prices. Moreover, simply prohibiting the monopolist from using protected characteristics to segment consumers would not be privacy-preserving either, since the monopolist may have access to close proxies of these characteristics. For example, as noted by [The White House \(2015\)](#):

“Big data naturally raises concerns among groups that have historically been victims of discrimination. Given hundreds of variables to choose from, it is easy to imagine that statistical models could be used to hide more explicit forms of discrimination by generating customer segments that are closely correlated with race, gender, ethnicity, or religion [...], even if the profit motive is different from, and in many cases fundamentally inconsistent with, the sort of prejudice that our antidiscrimination laws seek to prohibit.”

Our results lead to a characterization of all market segmentations that prohibit the monopolist from price-discriminating consumers based on their protected characteristics, in the sense that consumers of different protected characteristics face the *same* distribution of prices.

Seller-Optimal Segmentations A natural question is what non-discriminatory market segmentation maximizes the seller’s profit. [Proposition 2](#) shows that this question reduces to a multi-marginal optimal transport problem. Specifically, let the state space be $\Omega := X \times \Theta$, and suppose that no information about θ can be revealed. For any vector of consumer values $(x_1, \dots, x_J) \in X^J$, let $V(x_1, \dots, x_J)$ be the maximal profit of the monopolist when

knowing that characteristic θ_j has value x_j , without knowing the characteristics:

$$V(x_1, \dots, x_n) = \max_{p \geq 0} \sum_{j=1}^J p \mathbf{1}\{x_j \geq p\} \mathbb{P}[\theta = \theta_j] = \max_{j \in \{1, \dots, J\}} \left\{ \sum_{i=j}^J \mathbb{P}[\theta = \theta_{(j)}] x_{(j)} \right\},$$

where $x_{(j)}$ is the j -th smallest element of (x_1, \dots, x_n) , and (j) is the index of the j -th smallest element $x_{(j)}$.

Proposition 2 implies that the profit-maximizing market segmentation can be identified by finding the joint distribution ρ of (x_1, \dots, x_J) that solves the optimal transport problem

$$\sup_{\rho \in \mathcal{D}} \int_{X^J} V(x_1, \dots, x_J) d\rho. \quad (6)$$

Suppose that $F(x | \theta_1) \geq \dots \geq F(x | \theta_J)$, so that $\{F(\cdot | \theta_j)\}_{j=1}^J$ is ranked by first-order stochastic dominance, and suppose that $(1 - \mathbb{P}[\theta = \theta_1]) \cdot \bar{x} \leq \underline{x}$ so that there is enough mass of the “lowest-type” consumers. Under this sufficient condition, we can obtain a closed-form solution of the transport problem (6), which in turn leads to a characterization of a seller-optimal segmentation.

Proposition 10. *Let q be the quantile signal. That is,*

$$q := rF(x | \theta) + (1 - r)F^-(x | \theta).$$

The market segmentation corresponding to quantile signal q maximizes the seller’s revenue. Moreover, under any optimal market segmentation:

- (i) The outcome is efficient and every consumer purchases the good.*
- (ii) Consumers with characteristic θ_1 always retain zero surplus, while consumers with characteristic $\theta_j \neq \theta_1$ retain positive surplus whenever $F(x | \theta_j) \neq F(x | \theta_1)$ for some $x \in X$.*
- (iii) The seller’s profit equals $\mathbb{E}[x | \theta_1]$. In particular, increases in consumers’ values with characteristics $\theta \neq \theta_1$ in the sense of FOSD do not affect the seller’s profit, while any increase in consumers’ values with characteristic θ_1 in the sense of FOSD is completely captured by the seller.*

Under the seller-optimal segmentation, consumers with different characteristics are pooled assortatively into the same segment. Since consumers’ values are FOSD-ranked based on



Figure 2: Panel (a) depicts the quantile signal that pools two groups of consumers assortatively. Panel (b) depicts the optimal signal, which pools as many consumers with the same values as possible, and then pools the remaining consumers negatively assortatively.

their characteristics, the lowest value in each segment must be of characteristic $\theta = \theta_1$. The assumption that $(1 - \mathbb{P}[\theta = \theta_1]) \cdot \bar{x} \leq \underline{x}$ then ensures that it is optimal for the seller to sell to every consumers in each segment by charging a price that equals the lowest value. Note that, however, the price distribution faced by each consumer characteristic θ is the same, which equals $F(\cdot | \theta_1)$.

This observation may serve as a cautionary tale, as in practice the legislation imposing privacy constraints typically mention explicitly that they are meant to protect groups that plausibly have a lower willingness to pay. However, according to [Proposition 10](#), this could mean that the group of consumers who statistically have a lower willingness to pay would have their surplus extracted, while groups of consumers with a higher willingness to pay would enjoy lower prices. In fact, when compared to uniform pricing (i.e., banning any forms of price discrimination), the group of consumers with the lowest willingness to pay must be worse-off compared to uniform pricing, while other groups of consumers might be better-off.³¹

When the assumptions in [Proposition 10](#) are violated, the quantile signal may not be optimal. In this case, non-trivial reorderings of the quantile signal would be necessary for optimality, as shown in the next example.

Example 4. Suppose that $X = \{1, 2, 3\}$, and $\Theta = \{\theta_1, \theta_2\}$, with $\mathbb{P}[\theta = \theta_1] = 1/2$. Suppose that the conditional distribution of x given $\theta = \theta_1$ is $(1/2, 1/3, 1/6)$; while the conditional distribution of x given $\theta = \theta_2$ equals $(1/6, 1/3, 1/2)$. One can show that the solution to the optimal transport problem (6) is given by the joint distribution ρ^* , where $\rho^*(1, 1) = \rho^*(3, 3) = 1/6$,

³¹See the Online Appendix for a concrete numerical example.

$\rho^*(2, 2) = \rho^*(1, 3) = 1/3$,³² corresponds to the reordered quantile signal generated by the measure-preserving transformations:

$$M_{\theta_1}(s) = \begin{cases} s, & \text{if } s \in [0, 1/6] \cup (5/6, 1] \\ 1/2 + (s - 1/6), & \text{if } s \in (1/6, 1/2] \\ 1/6 + (s - 1/2), & \text{if } s \in (1/2, 5/6] \end{cases} ; \quad M_{\theta_2}(s) = s.$$

The quantile signal and the optimal reordered quantile signal can be represented in the same way as in [Example 2](#) by [Figure 2](#).

6 Discussion

6.1 Relation to “Private Private Information”

[He et al. \(2023\)](#) study what information about a state taking finitely many values can be revealed to a group of n agents if their signals s_1, \dots, s_n are restricted to be independent, in which case the information structure is called *private private*. The concept of private private information structure relates to privacy-preserving signals, as an information structure $s = (s_1, \dots, s_n)$ is private private if and only if s_i is privacy-preserving with respect to the σ -algebra generated by s_{-i} for all i . An important difference between the concepts is that the information which needs to be kept private is exogenous in our model, while it is endogenous in the case of private private information.

Section 4 of [He et al.](#) discusses how the results they obtain for private private information structures can be used to gain insights into privacy-preserving signals. In our notation, consider a state space $\Omega = X \times \Theta$ an associated probability measure \mathbb{P} . By treating θ as the signal received by a dummy agent, a signal s is privacy-preserving if and only if (s, θ) is private private. They then establish the following result: For any (sufficiently nice) decision problem (u, A) , where $u : X \times A \rightarrow \mathbb{R}$ depends only on x ,

Theorem (Theorem 2 in [He et al. 2023](#)). *Whenever $X = \{0, 1\}$ is binary, there exists an optimal privacy-preserving signal s^* . The optimal signal is unique up to equivalence: the*

³²See the Online Appendix for detailed arguments.

distribution of $\mathbb{P}[x = 1 \mid s^*]$ is \overline{F}_x .³³ Furthermore, every privacy-preserving signal s reveals less information about x than s^* .

Our [Proposition 5](#) (ii) reproduces this insight. Our other results differ from [Theorem 2](#) in [He et al.](#) along two dimensions:

- (i) We do not restrict attention to binary states. In fact, our [Theorem 2](#) generalizes [Theorem 2](#) of [He et al.](#) from binary states to an arbitrary one-dimensional state space X for u affine in x .
- (ii) [Theorem 2](#) of [He et al.](#) considers decision problems in which the payoff does *not* depend on θ . Most of our results do not impose this restriction on the decision problem.³⁴ This restriction has meaningful implications, as (even when restricting to binary states) there does *not* exist a signal that is optimal for all decision problems (see [Example 2](#)), and thus [Theorem 2](#) in [He et al.](#) does not apply for the wider class of decision problems we study.³⁵ In other words, the notion of Blackwell dominance is different since the underlying state spaces ($X \times \Theta$ versus X) are different.

Finally, [He et al. \(2023\)](#) and the present paper use quite different mathematical methodologies; while [He et al.](#) use tools from tomography, our proofs are based upon majorization theory and optimal transport.

6.2 Relation to Differential Privacy

While we define privacy-preserving signals through an abstract collection of privacy sets, another notion of privacy is *differential privacy*, proposed by [Dwork, McSherry, Nissim and Smith \(2006\)](#).³⁶ Specifically, suppose that Ω is a finite product set $\Omega_1 \times \dots \times \Omega_n$, where each

³³Note that $\overline{F}_x^{-1}(q) := \mathbb{E}[F^{-1}(x \mid \theta)] = \mathbb{E}[\mathbf{1}\{\mathbb{P}[x = 1 \mid \theta] > 1 - q\}]$ for all $q \in [0, 1]$, and hence \overline{F}_x is exactly the conjugate of the distribution of the random variable $\bar{x}(\theta) = \mathbb{P}[x = 1 \mid \theta]$.

³⁴[Theorem 1](#), [Proposition 2](#), [Proposition 3](#), [Proposition 4](#), and [Proposition 5](#) (i) all allow for the utility to depend on θ . As the approach in [He et al.](#) is based on a characterization of feasible posterior beliefs about x , it is not clear how it can be adapted to this wider class of decision problems where these beliefs are no longer a sufficient statistic for payoffs.

³⁵In [Appendix C](#), [He et al.](#) demonstrate that even when the utility does not depend on θ , the uniqueness is a only feature for binary states, by showing that there is a continuum of privacy-preserving signals that can be optimal for some decision problem in an example with three states and binary θ .

³⁶Relatedly, [Schmutte and Yoder \(2022\)](#) characterize the distribution of posteriors that can be induced by signals satisfying ε -differential privacy and study an information design problem subject to differential privacy.

dimension Ω_i represents characteristics of a different agent. A signal s satisfies ε -*differential privacy* for $\varepsilon > 0$ if for any ω, ω' that differ only in the characteristic of a single agent i (i.e., $\omega_{-i} = \omega'_{-i}$), a.s.,

$$\left| \log \frac{\mathbb{P}[\omega | s]}{\mathbb{P}[\omega' | s]} - \log \frac{\mathbb{P}[\omega]}{\mathbb{P}[\omega']} \right| \leq \varepsilon.$$

Intuitively, the log-likelihood induced by the signal cannot be influenced by more than ε by each individual agent. Our notion of privacy considers signals that only depend on the characteristics of a single individual, and are restricted to not reveal certain information. In contrast, differential privacy considers signals that depend on a whole population of agents, but are only influenced—to a limited extent—by each individual agent. Mathematically, these notions are unrelated and aim to capture different aspects of privacy.

7 Conclusion

We provide a characterization of signals that do not reveal certain information, and among others presented applications to statistical discrimination, product information, auction, and price discrimination. An interesting avenue for future research is to use the mathematical characterization presented in this paper to understand the consequences of different notions of privacy and fairness.

Appendix

Lemma A.1. *A signal s is privacy-preserving with respect to $\mathcal{P} \subseteq \mathcal{F}$ if and only if it is privacy-preserving with respect to the σ -algebra generated by \mathcal{P} .*

Proof. Fix any nonempty collection $\mathcal{P} \subseteq \mathcal{F}$ that is closed under finite intersections. Consider any signal s that is privacy-preserving with respect to the σ -algebra generated by \mathcal{P} , denoted by $\sigma(\mathcal{P})$. Since $\mathcal{P} \subseteq \sigma(\mathcal{P})$, s is privacy-preserving with respect to \mathcal{P} . Conversely, consider any signal s that is privacy-preserving with respect to \mathcal{P} . Let $\mathcal{P}^s \subseteq \mathcal{F}$ be the collection of events for which (1) holds. Clearly, \mathcal{P}^s is nonempty since s is privacy-preserving with respect to \mathcal{P} . Moreover, from the fact that \mathbb{P} and $\mathbb{P}[\cdot | \hat{s}]$ are probability measures for all realization \hat{s} of s , it follows that \mathcal{P}^s is a λ -system. Therefore, by Dynkin's $\pi - \lambda$ theorem, since the π -system \mathcal{P}

is contained in the λ -system \mathcal{P}^s , the σ -algebra $\sigma(\mathcal{P})$ generated by \mathcal{P} must also be contained in \mathcal{P}^s . Therefore, s is privacy-preserving with respect to $\sigma(\mathcal{P})$. \square

Proof of Lemma 1. By Lemma A.1, it is without loss to assume that \mathcal{P} is a σ -algebra. Let $\Theta := \Omega$. Then (Θ, \mathcal{P}) is a measurable space. Moreover, let $\theta : (\Omega, \mathcal{F}) \rightarrow (\Theta, \mathcal{P})$ be defined as $\theta(\omega) := \omega$ for all $\omega \in \Omega$. Clearly, θ is measurable since for any $P \in \mathcal{P}$, $\theta^{-1}(P) := \{\omega \in \Omega : \theta(\omega) \in P\} = P \in \mathcal{F}$. Furthermore, for any signal $s : \Omega \times [0, 1] \rightarrow S$, for any $P \in \mathcal{P}$, and for any measurable $A \subseteq S$

$$\mathbb{P}[\theta \in P, s \in A] = \mathbb{P}[\omega \in \theta^{-1}(P), s \in A] = \mathbb{P}[\omega \in P, s \in A] = \mathbb{E}[\mathbf{1}\{s \in A\} \mathbb{E}[\mathbf{1}\{\omega \in P\} \mid s]],$$

where the last equality follows from the law of iterated expectation. Therefore, if s is privacy-preserving, then for any $P \in \mathcal{P}$ and for any measurable $A \subseteq S$,

$$\mathbb{P}[\theta \in P, s \in A] = \mathbb{E}[\mathbf{1}\{s \in A\} \mathbb{E}[\mathbf{1}\{\omega \in P\}]] = \mathbb{P}[\omega \in P] \cdot \mathbb{P}[s \in A] = \mathbb{P}[\theta \in P] \cdot \mathbb{P}[s \in A].$$

Conversely, if θ and s are independent, then for any $P \in \mathcal{P}$ and for any measurable $A \subseteq S$,

$$\begin{aligned} \mathbb{E}[\mathbf{1}\{s \in A\} \cdot \mathbf{1}\{\omega \in P\}] &= \mathbb{E}[\mathbf{1}\{s \in A\} \cdot \mathbf{1}\{\theta \in P\}] = \mathbb{P}[s \in A] \cdot \mathbb{P}[\theta \in P] \\ &= \mathbb{E}[\mathbf{1}\{s \in A\} \mathbb{P}[\theta \in P]] \\ &= \mathbb{E}[\mathbf{1}\{s \in A\} \mathbb{P}[\omega \in P]]. \end{aligned}$$

By the definition of conditional expectations, it then follows that

$$\mathbb{P}[\omega \in P \mid s] = \mathbb{P}[\omega \in P]$$

almost surely. \square

Proof of Lemma 2. Fix any $\hat{\theta} \in \Theta$, we show that $\mathbb{P}[q_\phi \leq z \mid \hat{\theta}] = z$ for all $z \in [0, 1]$. To see this, first note that since $F_\phi(\cdot \mid \hat{\theta})$ has at most countably many jumps, enumerated by $\{x_n\} \subseteq \mathbb{R}$, $F_\phi(\cdot \mid \hat{\theta})$ can be written as

$$F_\phi(x \mid \hat{\theta}) = H(x) + \sum_{n=1}^{\infty} \mathbf{1}\{x_n \leq x\} (F_\phi(x_n \mid \hat{\theta}) - F_\phi^-(x_n \mid \hat{\theta})),$$

for all $x \in \mathbb{R}$, where H is a nondecreasing and continuous function. For any $z \in [0, 1]$ and for any $x \in \mathbb{R}$, let

$$\Gamma_z(x) := \begin{cases} 0, & \text{if } z < F_\phi^-(x | \hat{\theta}) \\ \frac{z - F_\phi^-(x | \hat{\theta})}{F_\phi(x | \hat{\theta}) - F_\phi^-(x | \hat{\theta})}, & \text{if } z \in [F_\phi^-(x | \hat{\theta}), F_\phi(x | \hat{\theta})] \\ 1, & \text{if } z \geq F_\phi(x | \hat{\theta}) \end{cases},$$

and note that

$$\begin{aligned} \int_{\mathbb{R}} \Gamma_z(x) dF_\phi(x | \hat{\theta}) &= \int_{\mathbb{R}} \Gamma_z(x) dH(x) + \sum_{n=1}^{\infty} \Gamma_z(x_n) (F_\phi(x_n | \hat{\theta}) - F_\phi^-(x_n | \hat{\theta})) \\ &= \int_{\mathbb{R}} \mathbf{1}\{x \leq F_\phi^{-1}(z | \hat{\theta})\} dH(x) + \sum_{n=1}^{\infty} \Gamma_z(x_n) (F_\phi(x_n | \hat{\theta}) - F_\phi^-(x_n | \hat{\theta})) \\ &= H(F_\phi^{-1}(z | \hat{\theta})) + \sum_{n=1}^{\infty} \Gamma_z(x_n) (F_\phi(x_n | \hat{\theta}) - F_\phi^-(x_n | \hat{\theta})). \end{aligned}$$

Therefore, if $F_\phi(\cdot | \hat{\theta})$ is continuous at $F_\phi^{-1}(z | \hat{\theta})$, then $F_\phi(F_\phi^{-1}(z | \hat{\theta}) | \hat{\theta}) = z$ and $\Gamma_z(x_n) = \mathbf{1}\{z \geq F_\phi(x_n | \hat{\theta})\}$, and hence $\int_{\mathbb{R}} \Gamma_z(x) dF_\phi(x | \hat{\theta}) = z$. Meanwhile, if $F_\phi(\cdot | \hat{\theta})$ is discontinuous at $F_\phi^{-1}(z | \hat{\theta})$, then $z = F_\phi(x_k | \hat{\theta})$ for some $k \in \mathbb{N}$, $F_\phi(F_\phi^{-1}(z | \hat{\theta}) | \hat{\theta}) = F_\phi(x_k | \hat{\theta})$, and $\Gamma_z(x_k) = (z - F_\phi^-(x_k | \hat{\theta})) / (F_\phi(x_k | \hat{\theta}) - F_\phi^-(x_k | \hat{\theta}))$ while $\Gamma_z(x_n) = \mathbf{1}\{z \geq F_\phi(x_n | \hat{\theta})\}$ for $n \neq k$. Therefore,

$$\begin{aligned} \int_{\mathbb{R}} \Gamma_z(x) dF_\phi(x | \hat{\theta}) &= F_\phi(F_\phi^{-1}(z | \hat{\theta}) | \hat{\theta}) + z - F_\phi^-(x_k | \hat{\theta}) \\ &\quad + \sum_{n \neq k}^{\infty} \mathbf{1}\{x_n \leq F_\phi^{-1}(z | \hat{\theta})\} (F_\phi(x_n | \hat{\theta}) - F_\phi^-(x_n | \hat{\theta})) - \sum_{n=1}^{\infty} \mathbf{1}\{z \geq F_\phi(x_n | \hat{\theta})\} \\ &= F_\phi(x_k | \hat{\theta}) + z - F_\phi^-(x_k | \hat{\theta}) - (F_\phi(x_k | \hat{\theta}) - F_\phi^-(x_k | \hat{\theta})) = z. \end{aligned}$$

Consequently, for any $z \in [0, 1]$,

$$\mathbb{P}[q_\phi \leq z | \hat{\theta}] = \mathbb{P}[rF_\phi(\phi | \hat{\theta}) + (1-r)F_\phi^-(\phi | \hat{\theta}) | \hat{\theta}] = \int_{\mathbb{R}} \Gamma_z(x) dF_\phi(x | \hat{\theta}) = z,$$

as desired. Therefore, q_ϕ is independent of θ and thus, by [Lemma 1](#), is privacy-preserving. \square

Below, we introduce several auxiliary results before proving [Theorem 1](#) and [Theorem 2](#). To begin with, we first show that any signals Blackwell-dominated by a privacy-preserving signal is also privacy-preserving.

Lemma A.2. *Every signal Blackwell-dominated by a privacy-preserving signal is privacy-preserving.*

Proof. Consider any privacy-preserving signal s and consider a decision problem index by any arbitrary privacy set $P \in \mathcal{P}$, with $A = [0, 1]$ and

$$u(\omega, a) := -(\mathbf{1}\{\omega \in P\} - a)^2.$$

For each signal realization \hat{s} , the unique optimal action equals $\mathbb{P}[\omega \in P \mid \hat{s}]$, which by definition must equal $\mathbb{P}[\omega \in P]$. Thus, a signal is privacy-preserving if and only if it does not increase the value in any such problem. Since any Blackwell-dominated signal leads to a (weakly) lower payoff in any, and thus, these specific decision problems, it follows that any signal Blackwell-dominated by s must also be privacy-preserving. \square

Next, we introduce a result that leads to the proof of [Theorem 1](#) and [Theorem 2](#). Consider any one-dimensional statistic $\phi : \Omega \rightarrow \mathbb{R}$ of the state ω . Say that a signal s is a signal for (ϕ, θ) if s is measurable with respect to (ϕ, θ, r) .

Among all signals for (ϕ, θ) , we may analogously define the Blackwell order.

Definition A.1. *For any pair of signals s, \tilde{s} for (ϕ, θ) , s is said to be Blackwell-dominated by \tilde{s} in terms of (ϕ, θ) if for any action set A and any function $u : \Omega \times A \rightarrow \mathbb{R}$ that is measurable with respect to (ϕ, θ) ,*

$$\mathbb{E} \left[\sup_{a \in A} \mathbb{E}[u(\phi(\omega), \theta(\omega), a) \mid s] \right] \leq \mathbb{E} \left[\sup_{a \in A} \mathbb{E}[u(\phi(\omega), \theta(\omega), a) \mid \tilde{s}] \right].$$

In other words, the Blackwell order can be restricted to signals for (ϕ, θ) by comparing the information about (ϕ, θ) given by these signals.

Likewise, signals for (ϕ, θ) that reveal ϕ conditional on θ can be analogously defined.

Definition A.2. *For any statistic $\phi : \Omega \rightarrow \mathbb{R}$. A signal s for (ϕ, θ) conditionally reveals ϕ if for all (Borel) measurable $A \subseteq \mathbb{R}$, $\mathbb{P}[\phi \in A \mid s, \theta] \in \{0, 1\}$ almost surely.*

Proposition A.1. *Consider any random variable $\phi : \Omega \rightarrow \mathbb{R}$.*

- (i) *A signal for (ϕ, θ) is privacy-preserving if and only if it is Blackwell-dominated in terms of (ϕ, θ) by some reordered ϕ -quantile signal.*

(ii) Every reordered ϕ -quantile signal is Blackwell-undominated in terms of (ϕ, θ) among signals for (ϕ, θ) that are privacy-preserving.

The proof of [Proposition A.1](#) relies on the following lemmas.

Lemma A.3. *A signal for (ϕ, θ) is privacy-preserving if and only if it is Blackwell-dominated in terms of (ϕ, θ) by a privacy-preserving signal that conditionally reveals ϕ .*

Proof. Sufficiency follows from the same arguments that prove [Lemma A.2](#), with ω replaced by θ . For necessity, consider any privacy-preserving signal \tilde{s} for (ϕ, θ) . Let $\xi, \zeta : [0, 1] \rightarrow [0, 1]$ be two independently and uniformly distributed random variables. Then there exists another privacy-preserving signal $s \sim \tilde{s}$ that is measurable with respect to (ω, ξ) and is independent of ζ and θ . Let $F_\phi(\cdot | \theta, s)$ be the conditional distribution of ϕ given θ and s , and let $t := \zeta F_\phi(\phi | \theta, s) + (1 - \zeta)F_\phi^-(\phi | \theta, s)$. Then, since s and θ are independent of ζ , by the same arguments as the proof of [Lemma 2](#), t is uniformly distributed conditional on θ and s . Therefore, the signal (t, s) is privacy-preserving. Moreover, by construction, (t, s) is measurable with respect to (ϕ, θ, r) and thus is a signal for (ϕ, θ) . Meanwhile, since (t, s) reveals more information than s , $\tilde{s} \sim s \leq (t, s)$. Lastly, by construction $F_\phi^{-1}(t | \theta, s) = \phi$ almost surely. Thus, (t, s) conditionally reveals ϕ . This completes the proof. \square

Lemma A.4. *Any privacy-preserving signal for (ϕ, θ) that conditionally reveals ϕ is Blackwell-undominated in terms of (ϕ, θ) among all privacy-preserving signals for (ϕ, θ) .*

Proof. Let Γ be the set of probability measures on $\mathbb{R} \times \Theta$, whose marginal over Θ equals $\nu(\cdot) := \mathbb{P}[\theta \in \cdot]$ (i.e., for all $\mu \in \Gamma$, $\mu(\mathbb{R} \times B) = \mathbb{P}[\theta \in B] =: \nu(B)$). Then for any $\mu \in \Gamma$, since \mathbb{R} is standard-Borel, there exists a transition probability $T[\mu] : \Theta \rightarrow \Delta(\mathbb{R})$ (see, e.g., [Çinlar 2010](#), Theorem 2.18, pp. 154),³⁷ such that

$$\mu((-\infty, x] \times B) = \int_B T[\mu](x | \theta) d\nu(\theta),$$

for all $x \in \mathbb{R}$ and for all measurable $B \subseteq \Theta$.

³⁷By the proof of [Lemma 1](#), the measurable Θ can be taken as a subset of \mathbb{R} . Therefore, $\Delta(\mathbb{R} \times \Theta)$ can be endowed with the weak-* topology and the associated Borel σ -algebra, which makes $\Delta(\mathbb{R} \times \Theta)$ a Polish space and hence is standard Borel. Therefore, $T[\cdot]$ can be regarded as a measurable function from $\Delta(\mathbb{R} \times \Theta)$ to \mathcal{G}^Θ , where \mathcal{G} is the space of CDFs on \mathbb{R} , with the Borel σ -algebra generated by the weak-* topology, and the σ -algebra over \mathcal{G}^Θ is the product σ -algebra. Moreover, since $T[\mu]$ is a transition probability, for all $\mu, \tilde{\mu} \in \Gamma$, $\mu \neq \tilde{\mu}$ implies $T[\mu] \neq T[\tilde{\mu}]$.

Note that for any privacy-preserving signal s for (ϕ, θ) , let λ^s be the induced distribution of posterior beliefs. Then $\lambda^s(\Gamma) = 1$. If, furthermore, s conditionally reveals ϕ , then for λ^s -almost all $\mu \in \Delta(\mathbb{R} \times \Theta)$, $T[\mu](\cdot | \theta)$ is a step function with two steps for ν -almost all $\theta \in \Theta$.

Suppose now that a privacy-preserving signal s for (ϕ, θ) , which conditionally reveals ϕ , is Blackwell-dominated by a privacy-preserving signal \tilde{s} for (ϕ, θ) . Suppose also that s is not Blackwell-equivalent to \tilde{s} . Then, by Theorem 2 of [Strassen \(1965\)](#), there exists a dilation $K : \Delta(\mathbb{R} \times \Theta) \rightarrow \Delta\Delta(\mathbb{R} \times \Theta)$ such that $\lambda^{\tilde{s}}(E) = \int_{\Delta(\mathbb{R} \times \Theta)} K(E | \mu) d\lambda^s(\mu)$, for all measurable $E \subseteq \Delta(\mathbb{R} \times \Theta)$. Since s is not Blackwell-equivalent to \tilde{s} , there exists a measurable set $E_0 \subseteq \Delta(\mathbb{R} \times \Theta)$ with positive λ^s -measure, such that for all $\mu \in E_0$, $K(\cdot | \mu)$ is not degenerate. Moreover, since \tilde{s} is privacy-preserving, $\lambda^{\tilde{s}}(\Gamma) = 1$, and hence $K(\Gamma | \mu) = 1$ for λ^s -almost all $\mu \in \Delta\Delta(\mathbb{R} \times \Theta)$. Lastly, since K is a dilation, for any $\mu \in E_0$, $\int_{\Delta(\mathbb{R} \times \Theta)} \tilde{\mu}(A) dK(\tilde{\mu} | \mu) = \mu(A)$, for all $A \subseteq \mathbb{R} \times \Theta$.

As a result, for any $\mu \in E_0$, for all $x \in \mathbb{R}$ and for all measurable $B \subseteq \Theta$,

$$\begin{aligned} \int_B T[\mu](x | \theta) d\nu(\theta) &= \mu((-\infty, x] \times B) = \int_{\Delta(\mathbb{R} \times \Theta)} \tilde{\mu}((-\infty, B) dK(\tilde{\mu} | \mu) \\ &= \int_{\Delta(\mathbb{R} \times \Theta)} \left(\int_B T[\tilde{\mu}](x | \theta) d\nu(\theta) \right) dK(\tilde{\mu} | \mu) \\ &= \int_B \left(\int_{\Delta(\mathbb{R} \times \Theta)} T[\tilde{\mu}](x | \theta) dK(\tilde{\mu} | \mu) \right) d\nu(\theta), \end{aligned}$$

and hence for ν -almost all $\theta \in \Theta$,

$$T[\mu](x | \theta) = \int_{\Delta(\mathbb{R} \times \Theta)} T[\tilde{\mu}](x | \theta) dK(\tilde{\mu} | \mu),$$

a contradiction, since for ν -almost all θ , $T[\mu](\cdot | \theta)$ is a step function with two steps and hence cannot be written as a mixture of distinct CDFs. Therefore, s cannot be Blackwell-dominated by \tilde{s} . \square

Lemma A.5. *For any signal \tilde{s} , there exists a signal $s : \Omega \times [0, 1] \rightarrow [0, 1]$ such that $s \sim \tilde{s}$ and that $\mathbb{P}[s \leq x] = x$ for all $x \in [0, 1]$.*

Proof. Consider any signal \tilde{s} . Let $\xi, \zeta : [0, 1] \rightarrow [0, 1]$ be two independently and uniformly distributed random variables. Then there exists another signal $\bar{s} \sim \tilde{s}$ that is measurable with respect to (ω, ξ) and is independent of ζ . Define another signal $s' : \Omega \times [0, 1] \rightarrow \Delta(\Omega)$ as

$s'(\omega, r) := \mathbb{P}[\omega \in \cdot \mid \bar{s}(\omega, r)]$. Since $\Delta(\Omega)$, with the Borel- σ algebra induced by the weak- $*$ topology, is standard Borel, there exists an isomorphism $\psi : \Delta(\Omega) \rightarrow [0, 1]$. Since ψ is invertible, $\psi(s') \sim \bar{s}$. Let F_ψ be the distribution of $\psi(s')$, so that $F_\psi(z) := \mathbb{P}[\psi(s') \leq z]$ for all $z \in [0, 1]$. Then, let $s := \zeta F_\psi(\psi(s')) + (1 - \zeta)F_\psi^-(\psi(s'))$. By the same arguments as the proof of [Lemma 2](#), s is uniformly distributed. Moreover, since $F_\psi^{-1}(s) = \psi(s')$ a.e., $s \sim \psi(s') \sim \bar{s} \sim \tilde{s}$. This completes the proof. \square

Lemma A.6. *Every privacy-preserving signal for (ϕ, θ) that conditionally reveals ϕ is Blackwell-equivalent to some reordered ϕ -quantile signal.*

Proof. Consider any privacy-preserving signal $s : \Omega \times [0, 1] \rightarrow S$ for ϕ that conditionally reveals ϕ . By [Lemma A.5](#), it is without loss to assume that $S = [0, 1]$ and that the marginal distribution of s is uniform. Since s conditionally reveals ϕ , there exists a measurable function $\eta : [0, 1] \times \Theta \rightarrow \mathbb{R}$ such that $\phi = \eta(s, \theta)$ almost surely. Therefore, for any $\hat{\theta} \in \Theta$,

$$\mathbb{P}[\eta(s, \theta) \leq x \mid \hat{\theta}] = \mathbb{P}[\phi \leq x \mid \hat{\theta}] = F_\phi(x \mid \hat{\theta}),$$

and hence, for any $\hat{\theta} \in \Theta$, the nondecreasing rearrangement of $\eta(\cdot, \hat{\theta})$ is $F_\phi^{-1}(\cdot \mid \hat{\theta})$. Thus, by Proposition 3 of [Ryff \(1970\)](#), there exists a family $(M_{\hat{\theta}})_{\hat{\theta} \in \Theta}$ of measure-preserving transformations such that

$$\eta(s, \hat{\theta}) = F_\phi^{-1}(M_{\hat{\theta}}(s) \mid \hat{\theta}),$$

for all $\hat{\theta} \in \Theta$. Meanwhile, let t be the M -reordered ϕ -quantile signal, the posterior belief conditional on realization \hat{t} and on $\hat{\theta}$ is $F_\phi^{-1}(M_{\hat{\theta}}(\hat{t}) \mid \hat{\theta})$. Together, $s \sim t$, as desired. \square

Proof of Proposition A.1. (i) follows from [Lemma A.2](#), [Lemma A.3](#), and [Lemma A.6](#). (ii) follows from [Lemma A.4](#) and [Lemma A.6](#). \square

Proof of Proposition 1. Take $\phi : \Omega \rightarrow \mathbb{R}$ to be an invertible statistic with ϕ^{-1} being measurable. Then any signal is a signal for (ϕ, θ) . Moreover, a signal \tilde{s} Blackwell-dominates s if and only if \tilde{s} Blackwell-dominates s in terms of (ϕ, θ) . [Proposition 1](#) then follows from [Lemma A.2](#) and [Lemma A.3](#). \square

Proof of Theorem 1. For any statistic ϕ such that the ϕ -quantile signal q_ϕ is conditionally revealing, the σ -algebra generated by (ϕ, θ) must be \mathcal{F} . Therefore, every signal s is a signal

for (ϕ, θ) . Since Ω is a standard Borel space, there exists an invertible $\phi : \Omega \rightarrow \mathbb{R}$ such that ϕ^{-1} is measurable, and hence there exists a conditionally revealing ϕ -quantile signal. As a result, [Theorem 1](#) follows from [Proposition A.1](#). \square

Proof of [Theorem 2](#). For sufficiency, since \overline{F}_ϕ is induced by the ϕ -quantile signal q_ϕ , every mean-preserving contraction G of \overline{F}_ϕ can be induced by a signal for (ϕ, θ) that is Blackwell-dominated by q_ϕ by Strassen's theorem ([Strassen 1965](#)). Together with [Proposition A.1](#), G can be induced by a privacy-preserving signal. To prove necessity, by [Proposition A.1](#), it suffices to show that the distribution G of posterior means induced by any reordered ϕ -quantile signal is a mean-preserving contraction of \overline{F}_ϕ . To see this, observe that for any family $(M_{\hat{\theta}})_{\hat{\theta} \in \Theta}$, the posterior mean conditional on a realization \hat{s} of the M -reordered ϕ -quantile signal s is given by

$$\mathbb{E}[F_\phi^{-1}(M_\theta(\hat{s}) \mid \theta)]. \quad (\text{A.7})$$

Let G be the distribution of the random variable $\mathbb{E}[F_\phi^{-1}(M_\theta(s) \mid \theta)]$, where the expectation is taken over θ . Since s is uniformly distributed, the quantile function G^{-1} is the nondecreasing arrangement of the function $\hat{s} \mapsto \mathbb{E}[F_\phi^{-1}(M_\theta(\hat{s}) \mid \theta)]$. Thus, by [Proposition 3 of \[Ryff \\(1970\\)\]\(#\)](#), there exists a Lebesgue measure-preserving transformation $\psi : [0, 1] \rightarrow [0, 1]$ such that for almost all $\hat{s} \in [0, 1]$,

$$G^{-1}(\psi(\hat{s})) = \mathbb{E}[F_\phi^{-1}(M_\theta(\hat{s}) \mid \theta)].$$

As ψ and $M_{\hat{\theta}}$ are Lebesgue measure-preserving transformations for all $\hat{\theta} \in \Theta$, we have that for all $t \in [0, 1]$,

$$\begin{aligned} \int_t^1 G^{-1}(y) dy &= \int_0^1 \mathbf{1}\{\psi(z) \geq t\} G^{-1}(\psi(z)) dz = \int_0^1 \mathbf{1}\{\psi(z) \geq t\} \mathbb{E}[F_\phi^{-1}(M_\theta(z) \mid \theta)] dz \\ &\leq \int_0^1 \mathbf{1}\{z \geq t\} \mathbb{E}[F_\phi^{-1}(z \mid \theta)] dz \\ &= \int_t^1 \overline{F}_\phi^{-1}(z) dz, \end{aligned}$$

where the first equality follows since ψ is Lebesgue measure-preserving, and the inequality follows from Fubini's theorem and the Hardy-Littlewood-Polya inequality (see, e.g., [Hardy et al. 1929](#) and [Puccetti and Wang 2015](#), [Theorem 2.1](#)). Thus, G^{-1} is majorized by \overline{F}_ϕ^{-1} , which implies that G is a mean-preserving contraction of \overline{F}_ϕ (see, e.g., [Shaked and Shanthikumar](#)

(2007), Section 3.A). □

References

- ALI, S. N., N. HAGHPANAH, X. LIN, AND R. SIEGEL (2022) “How to Sell Hard Information,” *Quarterly Journal of Economics*, 137 (1), 619–678.
- ARIELI, I., Y. BABICHENKO, R. SMORODINSKY, AND T. YAMASHITA (2023) “Optimal Persuasion via Bi-Pooling,” *Theoretical Economics*, 18 (1), 15–36.
- BAROCAS, S., M. HARDT, AND A. NARAYANAN (2019) *Fairness and Machine Learning: Limitations and Oppurtunities*: MIT Press.
- BERGEMANN, D., B. BROOKS, AND S. MORRIS (2015) “The Limits of Price Discrimination,” *American Economic Review*, 105 (3), 921–957.
- BERGEMANN, D., T. HEUMANN, AND S. MORRIS (2023) “Screening with Persuasion,” Working Paper.
- (2024) “Bidder-Optimal Information Structures in Auctions,” Working Paper.
- BERGEMANN, D., T. HEUMANN, S. MORRIS, C. SOROKIN, , AND E. WINTER (2022) “Optimal Information Disclosure in Classic Auctions,” *American Economic Review: Insights*, 4 (3), 371–388.
- BLACKWELL, D. (1953) “Equivalent Comparisons of Experiments,” *Annals of Mathematical Statistics*, 24 (2), 265–272.
- BOHREN, J. A., P. HULL, AND A. IMAS (2024) “Systemic Discrimination: Theory and Measurement,” Working Paper.
- CALDERS, T., F. KAMIRAN, AND M. PECHENIZKIY (2009) “Building Classifiers with Independency Constraints,” in *IEEE International Conference on Data Mining Workshops*, 13–18, [10.1109/ICDMW.2009.83](https://doi.org/10.1109/ICDMW.2009.83).
- CALDERS, T. AND S. VERWER (2010) “Three Naive Bayes Approaches for Discrimination-Free Classification,” *Data Mining and Knowledge Discovery*, 21, 277–292.
- CAREY, A. N. AND X. WU (2023) “The Statistical Fairness Field Guide: Perspectives from Social and Formal Sciences,” *AI and Ethics*, 3, 1–23.
- CHE, Y.-K., K. KIM, AND W. ZHONG (2021) “Statistical Discrimination in Ratings-Guided Markets,” Working Paper.
- ÇINLAR, E. (2010) *Probability and Stochastics*: Springer.

- CORBETT-DAVIES, S., E. PIERSON, A. FELLER, S. GOEL, AND A. HUQ (2017) “Algorithmic Decision Making and the Cost of Fairness,” in *23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797–806, [10.1145/3097983.3098095](https://doi.org/10.1145/3097983.3098095).
- DARLINGTON, R. B. (1971) “Another Look at Cultural Fairness,” *Journal of Educational Measurement*, 8 (2), 71–82.
- DOVAL, L. AND A. SMOLIN (2023) “Persuasion and Welfare,” Technical report, CEPR Discussion Papers.
- DWORCZAK, P. AND G. MARTINI (2019) “The Simple Economics of Optimal Persuasion,” *Journal of Political Economy*, 127, 1993–2048.
- DWORK, C., M. HARDT, T. PITASSI, O. REINGOLD, AND R. ZEMEL (2012) “Fairness through Awareness,” *ACM ITCS Proceedings*, 214–226.
- DWORK, C., F. MCSHERRY, K. NISSIM, AND A. SMITH (2006) “Calibrating noise to sensitivity in private data analysis,” in *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, 265–284, Springer.
- EILAT, R., K. ELIAZ, AND X. MU (2021) “Bayesian privacy,” *Theoretical Economics*, 16 (4), 1557–1603.
- FELDMAN, M., S. A. FRIEDLER, J. MOELLER, C. SCHEIDEGGER, AND S. VENKATASUBRAMANIAN (2015) “Certifying and Removing Disparate Impact,” in *21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 259–268, [10.1145/2783258.2783311](https://doi.org/10.1145/2783258.2783311).
- FORGES, F. (1993) “Five Legitimate Definitions of Correlated Equilibrium in Games with Incomplete Information,” *Theory and Decision*, 35, 277–310.
- (2006) “Correlated Equilibrium in Games with Incomplete Information Revisited,” *Theory and Decision*, 61, 329–344.
- GALPERTI, S. AND J. PEREGO (2023) “Privacy and the Value of Data,” *American Economic Association: Paper and Proceedings*, 113, 197–203.
- GILLIS, T., B. MCLAUGHLIN, AND J. SPIESS (2021) “On the Fairness of Machine-Assisted Human Decisions,” Working Paper.
- GREEN, J. R. AND N. L. STOKEY (2022) “Two Representations of Information Structures and their Comparisons,” *Decisions in Economics and Finance*, 45 (2), 541–547.
- HARDT, M., E. PRICE, AND N. SREBRO (2016) “Equality of Opportunity in Supervised

- Learning,” arXiv preprint arXiv:1610.02413.
- HARDY, G., J. E. LITTLEWOOD, AND G. PÓLYA (1929) “Some Simple Inequalities Satisfied by Convex Functions,” *Messenger Math*, 58 (4), 145–152.
- HE, K., F. SANDOMIRSKIY, AND O. TAMUZ (2023) “Private Private Information,” arXiv preprint arXiv:2112.14356.
- ICHIHASHI, S. (2020) “Online Privacy and Information Disclosure by Consumers,” *American Economic Review*, 110 (2), 569–595.
- KAMENICA, E. AND M. GENTZKOW (2011) “Bayesian Persuasion,” *American Economic Review*, 101 (6), 2560–2615.
- KAMIRAN, F., I. ŽILOBAITĚ, AND T. CALDERS (2013) “Quantifying Explianable Discrimination and Removing Illegal Discrimination in Automated Decision Making,” *Knowledge and Information Systems*, 35 (3), 613–644.
- KAMISHIMA, T., S. AKAHO, AND J. SAKUMA (2011) “Fairness-aware Learning through Regularization Approach,” in *IEEE International Conference on Data Mining Workshops*, 643–650, [10.1109/ICDMW.2011.83](https://doi.org/10.1109/ICDMW.2011.83).
- KLEINER, A., B. MOLDOVANU, AND P. STRACK (2021) “Extreme Points and Majorization: Economic Applications,” *Econometrica*, 89 (4), 1557–1593.
- KOLOTILIN, A. (2018) “Optimal information disclosure: A linear programming approach,” *Theoretical Economics*, 13, 607–635.
- KOLOTILIN, A., M. LI, T. MYLOVANOV, AND A. ZAPECHELNYUK (2017) “Persuasion of a Privately Informed Receiver,” *Econometrica*, 85 (6), 1949–1964.
- KOLOTILIN, A. AND A. WOLITZKY (2023) “The Economics of Partisan Gerrymandering,” Working Paper.
- LIANG, A., J. LU, AND X. MU (2023) “Algorithm Design: A Fairness-Accuracy Frontier,” Working Paper.
- LIU, Q. (2015) “Correlation and Common Priors in Games with Incomplete Information,” *Journal of Economic Theory*, 157, 49–75.
- VON NEUMANN, J. (1932) “Einige sätze über messbare abbildungen,” *Ann. of Math.(2)*, 33 (3), 574–586.
- PUCETTI, G. AND R. WANG (2015) “Extremal Dependence Concepts,” *Statistical Science*, 30 (4), 485–571.
- RHODES, A. AND J. ZHOU (2024) “Personalization and Privacy Choice,” Working Paper.

- ROKHLIN, V. (1952) *On the Fundamental Ideas of Measure Theory: (Matematicheskii Sbornik (ns) 25 (67) 107-150 (1949))* (71): American Mathematical Society.
- RYFF, J. V. (1970) “Measure Preserving Transformations and Rearrangements,” *Journal of Mathematical Analysis and Applications*, 31, 449–458.
- SCHMUTTE, I. M. AND N. YODER (2022) “Information Design for Differential Privacy,” Working Paper.
- SHAKED, M. AND J. G. SHANTHIKUMAR (2007) *Stochastic Orders*: Springer.
- STRASSEN, V. (1965) “The Existence of Probability Measures with Given Marginals,” *Annals of Mathematical Statistics*, 36, 423–439.
- TCHEN, A. H. (1980) “Inequalities for Distributions with Given Marginals,” *Annals of Probability*, 8 (4), 814–827.
- THE WHITE HOUSE (2015) “Big Data and Differential Pricing.”
- WIGGINS, B. (2020) *Calculating Race: Racial Discrimination in Risk Assessment*: Oxford University Press.