

Stochastic Optimization and Coupling^{*}

Frank Yang[†]

Kai Hao Yang[‡]

March 12, 2026

Abstract

We study optimization problems in which a linear functional is maximized over probability measures that are dominated by a given measure according to an integral stochastic order in an arbitrary dimension. We show that the following four properties are equivalent for any such order: (i) the test function cone is closed under pointwise minimum, (ii) the value function is affine, (iii) the solution correspondence has a convex graph with decomposable extreme points, and (iv) every ordered pair of measures admits an order-preserving coupling. As corollaries, we derive the extreme and exposed point properties involving integral stochastic orders such as multidimensional mean-preserving spreads and stochastic dominance. Applying these results, we generalize Blackwell's theorem by completely characterizing the comparisons of experiments that admit two equivalent descriptions—through instrumental values and through information technologies. We also show that these results immediately yield new insights into information design, mechanism design, and decision theory.

Keywords: Stochastic orders, extreme points, Blackwell order, Strassen's theorem, comparisons of experiments, Bayes' rule, information design, mechanism design.

^{*}We thank Ryota Iijima, Ravi Jagadeesan, Jesse Shapiro, Ludvig Sinander, Andrzej Skrzypacz, and Tomasz Strzalecki for helpful comments and suggestions.

[†]Department of Economics, Harvard University. Email: fyang@fas.harvard.edu.

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

1 Introduction

Many economic problems involve comparing or choosing probability measures. A celebrated example is the Blackwell order, which compares the distributions of posterior beliefs to understand which posterior distributions result from a more informative experiment that uniformly improves the payoff of a decision maker for all decision problems. The Blackwell order is an example of *integral stochastic orders* where the comparisons are defined with respect to a set of test functions, which in Blackwell's case is the set of convex functions representing the indirect utility functions from decision problems.

We study abstract optimization problems involving probability measures that are dominated by a given measure according to an integral stochastic order. Formally, we aim to understand the properties of

$$\max_{\nu \leq_{\mathcal{C}} \mu} \int f(x) \nu(dx),$$

where μ is a given measure on some abstract set X , $\mathcal{C}$ is a cone of test functions that defines an integral stochastic order, and ν is the endogenous measure being chosen. We obtain new structural properties of this optimization problem. As corollaries, we derive the extreme and exposed point properties involving integral stochastic orders such as multidimensional mean-preserving spreads and stochastic dominance. Applying these results, we generalize Blackwell's theorem by completely characterizing the comparisons of experiments that admit two equivalent descriptions—through instrumental values and through information technologies. We also show that these results immediately yield new insights into information design, mechanism design, and decision theory.

Abstract Results. Before discussing the economic applications, we start by describing our abstract results. Let $V_f^*(\mu)$ denote the *value function* of the optimization problem and $X_f^*(\mu)$ denote its *solution correspondence*. It is easy to see that $V_f^*(\mu)$ is a concave functional and $X_f^*(\mu)$ is convex-valued, but in general both objects can be very complex. Our results pertain to necessary and sufficient conditions on the stochastic order under which the optimization problem drastically simplifies and the stochastic comparison itself yields a tractable structure.

We say that the cone of test functions $\mathcal{C}$ is *min-closed* if it is closed under pointwise minimum, i.e., $\min\{g_1, g_2\} \in \mathcal{C}$ for $g_1, g_2 \in \mathcal{C}$. We say that the stochastic order $\leq_{\mathcal{C}}$ admits *order-preserving couplings* if for any $\nu \leq_{\mathcal{C}} \mu$ there exists a Markov kernel P such that

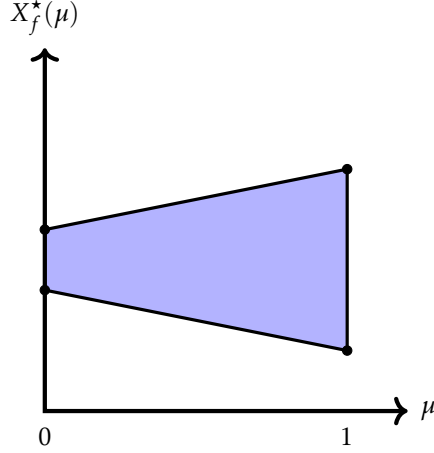


Figure 1: Trapezoid Graph. $X = \{0, 1\}$.

The colored area represents the graph of a solution correspondence X_f^* . The extreme points of this set are pairs (μ, v) , where μ is an extreme point of $\Delta(X)$ and v is an extreme point of $X_f^*(\mu)$.

$\nu = P * \mu$ and $P * \delta_x \leq_{\mathcal{C}} \delta_x$ for all x , i.e., there exists a Strassen-type coupling.¹

Our key structural result is an equivalence theorem ([Theorem 1](#)) that establishes the following four-way equivalence:

- (a) $\mathcal{C}$ is min-closed.
- (b) $V_f^*(\mu)$ is affine for all f .
- (c) $\leq_{\mathcal{C}}$ admits order-preserving couplings.
- (d) $X_f^*(\mu)$ has a convex graph whose extreme points are “decomposable” for all f .

By “decomposable” extreme points, we mean that every extreme point (μ, v) of the graph of X_f^* is one where μ is an extreme point of its convex domain, and v is an extreme point of $X_f^*(\mu)$ given μ —we refer to such a graph as a **trapezoid graph**, as illustrated by [Figure 1](#).

The equivalence theorem immediately yields two sets of applications. The first set of applications derives properties of stochastic orders defined by min-closed test functions, and then derives consequences about the value function and the solution correspondence—in particular, the result shows that (i) we can solve such problems in a pointwise fashion using an appropriate $\mathcal{C}$ -envelope of f ([Corollary 1](#)), (ii) the extreme points of a chain of ordered measures decompose ([Corollary 2](#)), (iii) every nested optimization problem where the second mover chooses a dominated measure according to the order admits a

¹As we explicitly define later, we use the notation $P * \mu$ to denote the measure obtained by applying the transition kernel P to μ . That is, $P * \mu(dy) := \int P(dy | x) \mu(dx)$.

decomposable extreme-point characterization, and when the first mover’s objective is linear, it reduces to a linear optimization problem (Corollary 3), and (iv) every extreme and exposed point of the set of measures dominated by a given measure (“orbits”) can be characterized in a pointwise fashion by a unique order-preserving coupling (Proposition 1).

To illustrate these results, as examples, we consider stochastic orders defined by multidimensional mean-preserving spreads and stochastic dominance. In particular, since concave functions are closed under pointwise minimum, we immediately derive the exposed point structure of the $\text{MPS}(\mu)$ orbit given a fixed measure μ (Proposition 2), which nests the one-dimensional case studied in Kleiner, Moldovanu, and Strack (2021) as a special case. Moreover, since convex functions are *not* closed under pointwise minimum, by Theorem 1, we also show that the $\text{MPC}(\mu)$ orbit (mean-preserving contraction) cannot admit a structurally similar characterization, which provides a precise explanation about the difference between these two orbits observed by Kleiner, Moldovanu, and Strack (2021). In contrast, the stochastic orbit defined by stochastic dominance upper bound $\text{LSD}(\mu)$ and that defined by stochastic dominance lower bound $\text{HSD}(\mu)$ behave completely symmetrically because both nondecreasing functions and nonincreasing functions are cones that satisfy the min-closure property—in particular, just like $\text{MPS}(\mu)$, our structural results immediately yield a characterization of the exposed points for these two orbits (Proposition 3), which nests the one-dimensional case studied in Yang and Zentefis (2024).

Since Theorem 1 is an equivalence theorem, besides applying it for optimization problems, powerful applications of the result are axiomatic in nature—our second set of applications takes a structural property of stochastic orders as given and characterizes the set of stochastic orders satisfying the structural property. In particular, the $(a) \implies (c)$ direction of Theorem 1 is an immediate consequence of the well-known Strassen’s theorem (Strassen 1965). Theorem 1 shows that Strassen’s theorem has a converse—the only way to admit order-preserving couplings in the Strassen sense is to have the min-closure property for the test functions, which is also equivalent to the value function being affine in the stochastic optimization problem. Given the wide application of Strassen’s theorem in various settings, these equivalence results are extremely helpful for delineating the limits of possible stochastic comparisons that satisfy desirable properties such as the Blackwell order. To further illustrate, we now discuss in detail how we apply the structural result to fully characterize comparisons of experiments that generalize Blackwell’s theorem by admitting two equivalent descriptions in the Blackwell sense.

Consistent Comparisons of Experiments. We study the comparisons of posterior belief distributions. Let $X = \Delta(\Omega)$ denote the simplex of beliefs where Ω is some state space. The Blackwell order $\leq_B$ is an order over the distributions of posterior beliefs $\mu \in \Delta(X)$. The celebrated Blackwell theorem (Blackwell 1951) shows that the Blackwell order has two equivalent descriptions:

- (i) A posterior distribution higher in the Blackwell order leads to a higher expected value for any convex indirect utility function. **(Value Description)**
- (ii) A posterior distribution higher in the Blackwell order can be obtained from the other distribution by a martingale transition kernel. **(Information Description)**

Thus, the Blackwell order $\leq_B$ on $\Delta(X)$ can be thought of as either comparing belief distributions according to the instrumental value they generate for solving a class of problems, or comparing belief distributions according to whether one can be obtained from the other via additional information.

We characterize the set of all orders $\leq$ on $\Delta(X)$ that admit such a dual representation in the Blackwell sense. Consider any cone of indirect utility functions $\mathcal{C}$, which may or may not be a subset of convex functions, representing the payoffs obtained given posterior beliefs.² Any such $\mathcal{C}$ via the **(Value Description)** defines a comparison $\leq_{\mathcal{C}}$ on the belief distributions. Now, consider any set of belief transition kernels $\mathcal{P}$, which may or may not be a subset of martingales, representing the feasible belief transitions upon observing new information. Any such $\mathcal{P}$ via the **(Information Description)** defines a comparison $\leq^{\mathcal{P}}$ on the belief distributions as well.

We say that an order $\leq$ comparing distributions of beliefs is **Blackwell-consistent** if there exist some $\mathcal{C}$ and some $\mathcal{P}$ such that $\leq = \leq_{\mathcal{C}} = \leq^{\mathcal{P}}$, in which case we call the dual representation $(\mathcal{C}, \mathcal{P})$ a **consistent pair**. A Blackwell-consistent order has the desirable property that any description of information (through $\mathcal{P}$) has a dual representation in terms of its instrumental values (through $\mathcal{C}$). Applying Theorem 1, our second main result (Theorem 2 and Theorem 3) characterizes all Blackwell-consistent orders—it turns out that the following are equivalent:

- (a) $\leq$ is Blackwell-consistent.
- (b) $\leq = \leq_{\mathcal{C}}$ for some $\mathcal{C}$ that is closed under pointwise maximum.
- (c) $\leq = \leq^{\mathcal{P}}$ for some $\mathcal{P}$ that is closed under kernel composition.

²For example, a subset of convex functions can be interpreted as a subset of decision problems, and a superset of convex functions can be interpreted as including additional *persuasion* problems.

Moreover, we also characterize the set of consistent pairs $(\mathcal{C}, \mathcal{P})$. We say that a pair $(\mathcal{C}, \mathcal{P})$ is **Blackwell-invariant** if (i) $\mathcal{C}$ is the set of indirect utility functions for which the payoff weakly increases under any transition kernel from $\mathcal{P}$ starting from any prior, and (ii) $\mathcal{P}$ is the set of transition kernels under which the payoff weakly increases for any indirect utility function from $\mathcal{C}$ under any prior. [Theorem 3](#) shows that $(\mathcal{C}, \mathcal{P})$ is a consistent pair if and only if $(\mathcal{C}, \mathcal{P})$ is Blackwell-invariant, i.e., they form a **fixed point** when iterating on the **(Value Description)** and **(Information Description)**. As a consequence, we show that every Blackwell-consistent order admits a **unique** pair of descriptions $(\mathcal{C}, \mathcal{P})$, exactly pinned down by Blackwell invariance.

If the agent updates according to Bayes' rule, then the test functions must at least include all the affine functions (action-independent payoffs). Then, for a consistent comparison in our sense, the closure under pointwise maximum implies that the test functions must include all convex functions, and hence the order can only be a **strengthening** of the Blackwell order—the Blackwell order is the weakest Bayes-plausible Blackwell-consistent order. Any strict weakening of the Blackwell order such as the Lehmann order ([Lehmann 1988](#)) cannot be Blackwell-consistent while maintaining Bayes plausibility. This shows that the Blackwell order is, in fact, necessary for consistent comparisons under Bayes' rule. A natural question is whether Bayes' rule itself is necessary to admit a consistent order if one wants to compare all experiments that are comparable in the Blackwell-garbling sense. We formulate this question among all prior-dependent systematic distortion updating rules and show that an updating rule can yield a consistent comparison in the Blackwell sense if and only if it is a **divisible updating rule** ([Cripps 2018](#)) which is exactly Bayes' rule under a homeomorphic transformation. Together, these results show that both Bayes' rule and the Blackwell order itself are almost necessary to admit two equivalent descriptions if we want to compare all Blackwell-comparable experiments.

Whenever an order is consistent, by [Theorem 1](#), it also implies the stochastic optimization problem yields a tractable structure. Exploiting this implication, we also show how our characterizations also deliver insights into constrained and non-Bayesian information design problems. Indeed, under Bayes' rule, a **constrained information design** problem defines a subset of martingales $\mathcal{P}$ achievable with the constrained technology—if the subset $\mathcal{P}$ is composition-closed, then by [Theorem 1](#) and [Theorem 2](#), we immediately obtain an envelope characterization. The constrained envelope is defined by the set of test functions $\mathcal{C}$ that is consistent with $\mathcal{P}$ and is weakly below the concave envelope in the unconstrained case. Using the envelope characterization, we show that if constrained information design is beneficial whenever unconstrained information design is beneficial (i.e., the constraints are not too limiting), then the optimal constrained signal *cannot* be less

informative (in the Blackwell sense) than the optimal unconstrained signal, and is always *more* informative (in the Blackwell sense) with two states. An example of a composition-closed technology is privacy-preserving signals (Strack and Yang 2024), for which we immediately obtain a new optimality characterization. In the case of *non-Bayesian information design*, we show that for any prior-dependent systematic distortion updating rule (De Clippel and Zhang 2022), dynamic information design in multiple stages, unlike the Bayesian case, always yields a strictly higher payoff under *some* objective of the designer, unless the updating rule is a divisible updating rule in which case a one-shot signal is sufficient.

Nested Optimization and Stackelberg Principals. For any order $\leq_{\mathcal{C}}$ where $\mathcal{C}$ satisfies our min-closure property, as a consequence of Theorem 1, even *nested optimization* problems admit a simple structure due to the trapezoid graph property. In our final set of applications (Section 5), we consider a class of problems that we call *Stackelberg-principal* problems where (i) the leading principal first selects a measure μ from a compact convex set M , and (ii) the second principal then selects a measure ν dominated by μ according to some order $\leq_{\mathcal{C}}$. Whenever $\leq_{\mathcal{C}}$ satisfies min-closure, every leader-optimal equilibrium of this game can be succinctly characterized, and there always exists one where μ^* is an extreme point of M and ν^* is an extreme point of the orbit given μ^* .

We then illustrate the consequences of this general characterization via a series of examples arising from information design, decision theory, and mechanism design. First, we consider the *sequential persuasion* problem (Li and Norman 2021) and show that there always exists a sequential extreme equilibrium where every sender chooses a belief distribution that is an extreme point of the mean-preserving spread of the belief distribution chosen by the previous sender. In such an equilibrium, unlike the silent equilibria in Li and Norman (2021) and Wu (2023), every sender sends a signal, and along any history, the signals sent have the same structure *as if* there were no following senders. Second, we consider the *robust persuasion* problem (Dworczak and Pavan 2022)—there the first principal is the sender and the second principal is the adversarial nature. We illustrate how to apply our results there and show connections to their separation theorem which characterizes the robust solutions that are best-case optimal among worst-case optimal ones. Third, we consider the model of *objective ambiguity aversion* (Olszewski 2007; Ahn 2008) where ambiguity aversion is modeled via preferences over menus/ambiguity sets of lotteries. We ask when an outside observer by observing choices over lotteries μ can distinguish between (i) an objectively ambiguity-averse DM who has some ambiguity set $A(\mu)$ and (ii) an expected utility DM who faces no ambiguity. We show that in the

case where $A(\mu)$ is described by a stochastic orbit (i.e., $\{\nu : \nu \leq_{\mathcal{C}} \mu\}$), the outside observer can distinguish between these two models essentially if and only if $\mathcal{C}$ is not min-closed. In particular, for ambiguity sets of the form where the actual lottery could be a mean-preserving contraction of the reference lottery μ , we can distinguish these two models; however, for ambiguity sets taking the form of mean-preserving spreads or stochastic dominance shifts, we cannot distinguish between these two models. Lastly, we consider the model of *optimal property right* design (Dworczak and Muir 2024). Here, the leading principal is a designer who posts a menu of lotteries over a good and associated prices, which become a buyer’s outside options. The second principal is a seller who contracts with the buyer as in a standard mechanism design framework but must satisfy the type-dependent IR constraints resulting from the designer’s initial menu. The designer and the seller have different preferences over the final allocations. However, the main result of Dworczak and Muir (2024) shows that the optimal property design by the designer is a simple option-to-own menu that consists of one posted price, just like in standard monopoly pricing. Dworczak and Muir (2024) show this by developing a generalized ironing procedure and observing that the ironing procedure is actually a linear operator of the outside option function, which then leads to a surprising extreme point structure. We show that this argument can be more broadly viewed as the trapezoid property of the nested optimization problem where the designer first chooses a distribution of posted prices, and then the seller chooses another distribution of posted prices subject to giving every type a weakly higher payoff, which defines an integral stochastic order.

Related Literature. This paper is related to several streams of literature. Our main abstract result is closely related to Strassen’s theorem and convex duality. Strassen’s theorem (Strassen 1965) shows that the concave order admits an order-preserving coupling. It is known that the concave order of Strassen’s theorem can be further generalized to integral stochastic orders defined by min-closed test functions (Meyer 1966). Our abstract result expands this to a four-way equivalence. In particular, we show that the existence of order-preserving couplings also implies that the stochastic order must be defined by min-closed test functions, which is the converse to Strassen’s theorem. Moreover, we show that these are also equivalent to having affine value functions and trapezoid solution graphs. In that regard, our results are also related to optimization problems with mean-preserving spreads constraints, where the values are characterized by concave envelopes (e.g., Klemperer 1968; Myerson 1981; Aumann and Maschler 1995; Kamenica and Gentzkow 2011; Beiglböck and Nutz 2014). Our four-way equivalence explains precisely why some stochastic optimization problems are more tractable than others, due to

the equivalence relation between the affine value property/trapezoid graph property and the min-closure property.

The main characterization leads to an immediate characterization of extreme points of chains of measures ordered by a single integral stochastic order, which generalizes the characterization of [Ciosmak \(2023\)](#) for the convex order. Moreover, we characterize extreme points of multidimensional mean-preserving spread and first-order stochastic dominance orbits, which extends the characterization of [Kleiner, Moldovanu, and Strack \(2021\)](#) and [Yang and Zentefis \(2024\)](#) respectively.³ Our multidimensional MPS orbits characterization also complements [Kleiner et al. \(2024\)](#), who characterize the exposed points of MPC orbits. A one-dimensional probability measure can be equivalently viewed as a monotone function via its CDF; however, this equivalence breaks down for multidimensional probability measures.⁴ [Yang and Yang \(2025\)](#) characterize the extreme points of multidimensional monotone functions and show applications to mechanism and information design problems.

Our second main result contributes to the literature on comparisons of experiments pioneered by [Blackwell \(1951\)](#). We generalize Blackwell’s theorem by completely characterizing all possible comparisons of experiments that admit two equivalent descriptions in the Blackwell sense. Our characterization establishes a five-way equivalence and identifies if-and-only-if conditions under which an integral stochastic order on the marginal distributions admits an equivalent coupling order. As we show, this simultaneously connects Strassen’s theorem on the existence of order-preserving couplings, the Skorokhod-type embedding theorems of general Markov processes ([Rost 1971](#)), and the characterizations of optimal values of stopping problems via excessive functions ([Dynkin 1963](#)). Since we identify all orders of experiments that can be equivalently represented by the instrumental values of the experiments, our results are related to other orderings of experiments (e.g., [Lehmann 1988](#); [Chen 2025](#)), as well as attempts to compare experiments and persuasion problems with non-Bayesian updating rules and heterogeneous priors (e.g., [Alonso and Câmara 2016](#); [De Clippel and Zhang 2022](#); [Kobayashi 2025](#); [Azrieli and Das 2025](#)). We show that the only non-Bayesian updating rules consistent with Blackwell-style comparisons are exactly the divisible updating rules defined by [Cripps \(2018\)](#)—which implies that dynamic information design yields a strict benefit under some objec-

³See also [Nikzad \(2023\)](#) and [Candogan and Strack \(2021\)](#) for extreme points of mean-preserving contraction orbits with linear constraints; and [Augias and Uhe \(2025\)](#) for extreme points of mean-preserving spread intervals.

⁴Indeed, recall that a bivariate CDF F must also satisfy $F(x'_1, x'_2) - F(x_1, x'_2) - F(x'_1, x_2) + F(x_1, x_2) \geq 0$ for all $x_1 < x'_1$ and $x_2 < x'_2$.

tive of the designer if and only if the updating rule is not divisible.⁵ As we demonstrate, our characterization result also helps to identify constrained information design problems that are particularly tractable and admit a similar belief-based characterization as in the unconstrained problems. These problems include privacy-preserving persuasion (Strack and Yang 2024; He, Sandomirskiy, and Tamuz 2024), as well as sequential sampling and persuasion problems (e.g., Wald 1945; Henry and Ottaviani 2019; Morris and Strack 2019; Ball and Espín-Sánchez 2021; Ni, Shen, and Tang 2023).

In terms of other economic applications, as we demonstrate, many mechanism and information design problems have a nested structure—which we term Stackelberg-principal problems—in which the trapezoid-graph property delivers immediate insights. This includes sequential persuasion with multiple senders (Li and Norman 2021); robust persuasion with a worst-case Nature who supplies additional information (Dworczak and Pavan 2022); a particular form of objective ambiguity aversion where the ambiguity sets are stochastic orbits (Olszewski 2007; Ahn 2008); and optimal property right design (Dworczak and Muir 2024).

The remainder of the paper proceeds as follows. Section 2 introduces the notation and the primitives. Section 3 presents our main abstract result. Section 4 presents our main application to the comparisons of experiments and its implications for constrained and non-Bayesian information design. Section 5 presents our additional economic applications with nested optimization. Section 6 concludes.

2 Primitives

Notation. Let X be a compact Polish space endowed with the Borel σ -algebra $\mathcal{B}(X)$. Let $C(X)$ denote the set of continuous functions on X , endowed with the sup-norm. Let $\Delta(X)$ be the set of probability measures on X , endowed with the weak- $*$ topology.

For any $\mu \in \Delta(X)$ and any measurable Markov kernel $P : X \rightarrow \Delta(X)$, denote by $P * \mu$ the induced measure on X . That is,

$$[P * \mu](A) := \int_X P(A | y) \mu(dy) \quad (\text{Transition Kernel})$$

for all $A \in \mathcal{B}(X)$. For any measurable function $g : X \rightarrow \mathbb{R}$, denote by $g * P$ the expected

⁵Divisible updating rules are homeomorphic to Bayes' rule but they are *not* Blackwell-monotone (Whitmeier forthcoming). As we discuss, combining Whitmeier (forthcoming) and our result, it follows that the garbling-based comparison admits a dual representation if and only if the updating rule is divisible, and the class of dualizing functions is the class of decision problems if and only if it is Bayes' rule.

value of g under P . That is,

$$[g * P](x) := \int_X g(y)P(dy | x), \quad (\text{Conditional Expectation})$$

for all $x \in X$. For any two transition kernels P_1, P_2 , denote by $P_1 \circ P_2$ the composition of P_1 and P_2 . That is,

$$[P_1 \circ P_2](A | x) := \int_X P_1(A | y)P_2(dy | x), \quad (\text{Composition})$$

for all measurable $A \subseteq X$ and for all $x \in X$. For any $x \in X$, let δ_x be the Dirac measure on x . For any convex and compact subset M of $\Delta(X)$, $\text{ext}(M)$ denotes the set of *extreme points* of M , and $\text{exp}(M)$ denotes the set of *exposed points* of M .

Integral Stochastic Orders. Let $\mathcal{C}$ be a convex cone of bounded upper semicontinuous functions such that (i) $\mathcal{C} \cap C(X)$ is sup-norm closed, (ii) $\mathcal{C}$ contains all constants, (iii) $\mathcal{C}$ admits a subset of continuous functions that can approximate any function in $\mathcal{C}$ from above,⁶ and (iv) $\mathcal{C}$ is closed under bounded decreasing pointwise limits.⁷

We say that $\mathcal{C}$ is *min-closed* if the functions in $\mathcal{C}$ are closed under pointwise minimum:

$$\min\{g_1, g_2\} \in \mathcal{C} \quad \text{for all } g_1, g_2 \in \mathcal{C}. \quad (\text{Min-Closure})$$

Examples of min-closed $\mathcal{C}$ include nondecreasing, nonincreasing, concave, nondecreasing concave, nonincreasing concave functions (when X carries the relevant order/convex structure). Non-examples include convex functions (they are max-closed).

For any bounded upper semicontinuous function f , define the $\mathcal{C}$ -*envelope* of f as

$$\bar{f}(x) := \inf \{g(x) : g \in \mathcal{C}, g \geq f\} \quad (\mathcal{C}\text{-Envelope})$$

for all $x \in X$. It can be readily shown that $\bar{f} \in \mathcal{C}$ if $\mathcal{C}$ is min-closed (see [Lemma 2](#) in the appendix).

For any $\mu, \nu \in \Delta(X)$, we say that μ $\mathcal{C}$ -*dominates* ν , denoted by

$$\nu \leq_{\mathcal{C}} \mu,$$

if for any $g \in \mathcal{C}$,

$$\int_X g(x)\nu(dx) \leq \int_X g(x)\mu(dx).$$

⁶That is, for all $g \in \mathcal{C}$, there exists a decreasing sequence $\{g_n\} \subseteq \mathcal{C} \cap C(X)$ such that $\{g_n\} \downarrow g$ pointwise.

⁷That is, if $\{g_n\} \downarrow g$ pointwise and $\{g_n\} \subseteq \mathcal{C}$, then $g \in \mathcal{C}$.

Note that any integral stochastic order $\leq_C$ is reflexive and transitive, but need not be antisymmetric. Thus, formally, $\leq_C$ is a preorder. Throughout the paper, we use the term order to mean a preorder.⁸ We say that $\leq_C$ admits *order-preserving couplings* if for every $\nu \leq_C \mu$, there exists some P such that $\nu = P * \mu$ and $P * \delta_x \leq_C \delta_x$ for all x .

Stochastic Optimization. Fix any bounded upper semicontinuous function $f : X \rightarrow \mathbb{R}$. Our main results study the solutions to the following stochastic optimization problem

$$\sup_{\nu \in \Delta(X) : \nu \leq_C \mu} \int_X f \, d\nu. \quad (\text{Optimization})$$

Since $\Delta(X)$ is compact and the orbit set $\{\nu : \nu \leq_C \mu\}$ is closed and hence compact (see Lemma 3 in the appendix), (Optimization) has a solution.

For any $\mu \in \Delta(X)$, let

$$V_f^*(\mu) := \max_{\nu \in \Delta(X) : \nu \leq_C \mu} \int_X f \, d\nu \quad (\text{Value Function})$$

be the *value function* of (Optimization) and let

$$X_f^*(\mu) := \operatorname{argmax}_{\nu \in \Delta(X) : \nu \leq_C \mu} \int_X f \, d\nu \quad (\text{Solution Correspondence})$$

be the *solution correspondence* of (Optimization). By standard arguments and Berge's maximum theorem, it is clear that V_f^* is concave and upper semicontinuous on $\Delta(X)$, and X_f^* is compact-valued and upper-hemicontinuous on $\Delta(X)$.

For any closed and convex subset M of $\Delta(X)$, let

$$G_f^M := \{(\mu, \nu) : \mu \in M, \nu \in X_f^*(\mu)\} \subseteq \Delta(X) \times \Delta(X) \quad (\text{Solution Graph})$$

be the *graph* of the solution correspondence X_f^* given domain M . Note that since X_f^* is upper-hemicontinuous, G_f^M is compact.

We say that X_f^* has the *trapezoid graph* property if for any closed convex M , (i) its graph G_f^M is convex, and (ii) the extreme points of the graph are decomposable:

$$\operatorname{ext}(G_f^M) = \{(\mu, \nu) : \mu \in \operatorname{ext}(M), \nu \in \operatorname{ext}(X_f^*(\mu))\}. \quad (\text{Decomposability})$$

⁸One can also impose antisymmetry without affecting any of our results, in which case the term order means a partial order.

Nested Stochastic Optimization. In the above problem, we take the reference measure μ as given, but our results also pertain to a nested optimization problem where μ is chosen by a first mover. In particular, for any continuous quasi-convex functional $W : G_f^M \rightarrow \mathbb{R}$, consider the problem of selecting a pair of measures from the graph G_f^M to maximize W ,

$$\max_{(\mu, \nu) \in G_f^M} W(\mu, \nu). \quad (\text{Nested Optimization})$$

That is, the first mover has preferences over (μ, ν) given by W ; the second mover has a linear objective given by f , with indifferences broken in favor of the first mover.

3 Abstract Results

Theorem 1 (Equivalence Theorem). *The following are equivalent:*

- (a) $\mathcal{C}$ is min-closed.
- (b) V_f^* is affine for all f .
- (c) $\leq_{\mathcal{C}}$ admits order-preserving couplings.
- (d) X_f^* has the trapezoid graph property for all f .

The proof of [Theorem 1](#) can be found in the appendix. Below, we briefly sketch the main steps of the proof. We prove the equivalence through the order:

$$(a) \implies (b) \implies (c) \implies (d) \implies (a).$$

To prove $(a) \implies (b)$, we use standard arguments in convex duality. In particular, we apply the Fenchel-Rockafellar duality theorem. However, if we were to assume that f is continuous, that every $g \in \mathcal{C}$ is continuous, and that $\mathcal{C}$ is closed under the sup norm, then there is a more direct proof in the form of the minimax theorem. To see this, note that the dual cone of $\mathcal{C}$ is given by

$$\mathcal{C}^* = \left\{ \eta \in M(X) : \int g(x) \eta(dx) \geq 0 \text{ for all } g \in \mathcal{C} \right\},$$

where $M(X)$ is the space of finite signed Borel measures (equipped with the weak-* topology). The primal problem ([Optimization](#)) can then be written as:

$$\sup_{\nu \in \Delta(X)} \int f(x) \nu(dx) \text{ subject to } \mu - \nu \in \mathcal{C}^*. \quad (\text{Primal})$$

By the bipolar theorem, the dual cone of $\mathcal{C}^*$ is $\mathcal{C}$ itself, and hence we can identify the “multiplier” of the above constraint as $g \in \mathcal{C}$. Thus, the Lagrangian is

$$\int f(x) \nu(dx) + \int g(x)(\mu - \nu)(dx).$$

In particular, the value of the problem is equal to

$$V_f^*(\mu) = \sup_{\nu \in \Delta(X)} \inf_{g \in \mathcal{C}} \int f(x) \nu(dx) + \int g(x)(\mu - \nu)(dx),$$

which, by Sion’s minimax theorem, is equal to

$$\inf_{g \in \mathcal{C}} \sup_{\nu \in \Delta(X)} \int f(x) \nu(dx) + \int g(x)(\mu - \nu)(dx),$$

which is equivalent to

$$\inf_{g \in \mathcal{C}} \left\{ \int g(x) \mu(dx) + \max_{x \in X} [f(x) - g(x)] \right\}.$$

Now, note that since $\mathcal{C}$ includes all constants, and the above value is invariant to shifting g by a constant, the above problem can be written as

$$\inf_{g \in \mathcal{C}} \int g(x) \mu(dx) \text{ subject to } g(x) - f(x) \geq 0 \text{ for all } x. \quad (\text{Dual})$$

This proves that there is no duality gap. Now, importantly, note that **(Min-Closure)** of $\mathcal{C}$ implies that an optimal solution to the above dual problem is simply $g = \bar{f} \in \mathcal{C}$: Indeed, for any $g \in \mathcal{C}$ and $g \geq f$, by the definition of $\bar{f}$, we have that for every $x \in X$,

$$\bar{f}(x) \leq g(x),$$

and hence $\int \bar{f}(x) \mu(dx) \leq \int g(x) \mu(dx)$. This proves dual attainment, and hence strong duality. Importantly, the dual solution $\bar{f}$ is independent of μ . It follows immediately that

$$V_f^* = \int \bar{f}(x) \mu(dx),$$

which is affine in μ .

The proof of (b) $\implies$ (c) relies on the Krein-Milman theorem. Specifically, we first

show that the orbit

$$\mathcal{K}_\mu := \{\nu : \nu \leq_C \mu\}$$

is convex and compact. Then, we show that (b) implies that every exposed point ν of $\mathcal{K}_\mu$ must be such that

$$\nu = P * \mu$$

for a transition kernel P with $P * \delta_x \leq_C \delta_x$ for all x . Indeed, fix an exposed point ν of $\mathcal{K}_\mu$ and a continuous function f for which ν is the unique solution to [\(Optimization\)](#). Since V_f^* is affine, we have

$$\int f \, d\nu = V_f^*(\mu) = \int V_f^*(\delta_x) \mu(dx) = \int \left(\max_{\eta: \eta \leq_C \delta_x} \int f \, d\eta \right) \mu(dx).$$

That is, the value of the problem [\(Optimization\)](#) is the same as the average value of a family of pointwise optimization problems

$$\max_{\eta: \eta \leq_C \delta_x} \int f \, d\eta. \quad \textbf{(Pointwise Optimization)}$$

Using standard measurable selection arguments, we select a transition kernel P where $P(\cdot \mid x)$ is a solution of the pointwise optimization problem for all $x \in X$. Then, $P * \mu$ must also solve [\(Optimization\)](#). But since ν is the unique solution by construction, we have $\nu = P * \mu$. With (c) established for exposed points, we can then apply the Krein-Milman theorem for exposed points ([Klee jr. 1958](#)) to represent every element $\nu \in \mathcal{K}_\mu$ as the limit of a sequence $\{\nu_n\}$ of convex combinations of exposed points. For each element ν_n along the sequence, by linearity, there also exists a transition kernel P_n with $P_n * \delta_x \leq_C \delta_x$ for all x such that $\nu_n = P_n * \mu$. By appropriate approximation arguments, (c) then follows after taking limits.

The proof of (c) $\implies$ (d) is based on first principles. It is clear that (c) implies that V_f^* is affine, which in turn implies that G_f^M has a convex graph. To see that G_f^M has decomposable extreme points, note that for any extreme point (μ, ν) of G_f^M , clearly ν is an extreme point of $X_f^*(\mu)$. To see that μ must be an extreme point of M , suppose the contrary. Then, there exist distinct $\mu_1, \mu_2 \in M$ such that $\frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 = \mu$. Since ν is in $X_f^*(\mu)$, (c) implies that there exists a transition kernel P such that $\nu = P * \mu$, and that P solves [\(Pointwise Optimization\)](#) for μ -almost every $x \in X$. Now, let $\nu_i := P * \mu_i$ for $i \in \{1, 2\}$. It follows that $\nu_i \leq_C \mu_i$ for all $i \in \{1, 2\}$, and that—since $P(\cdot \mid x)$ is a solution of the pointwise

maximization problem μ -almost surely and hence μ_i -almost surely as $\mu_i \ll \mu$ —we have

$$\int f \, d\nu_i = V_f^*(\mu_i)$$

for $i \in \{1, 2\}$. This implies that $(\mu_i, \nu_i) \in G_f^M$. However, $(\mu, \nu) = \frac{1}{2}(\mu_1, \nu_1) + \frac{1}{2}(\mu_2, \nu_2)$ is a mixture of two distinct elements in G_f^M , contradicting the extremality of (μ, ν) .

Lastly, the proof of $(d) \implies (a)$ uses a separating hyperplane argument. First, we show that (d) immediately implies that V_f^* is affine for all upper semicontinuous f . Then, suppose that $\mathcal{C}$ is not min-closed. Then there exist $g_1, g_2 \in \mathcal{C}$ such that $k := \min\{g_1, g_2\}$ is not in $\mathcal{C}$. Therefore, there exists a separating hyperplane that separates k and $\mathcal{C}$. Together with the properties of $\mathcal{C}$, we then show that this separating hyperplane corresponds to two probability measures $\mu, \nu \in \Delta(X)$ such that $\nu \leq_{\mathcal{C}} \mu$, and

$$\int k \, d\nu > \int k \, d\mu.$$

Then, for any $x \in X$ and for any $\eta \leq_{\mathcal{C}} \delta_x$, since $g_1, g_2 \in \mathcal{C}$,

$$\int k \, d\eta \leq \min \left\{ \int g_1 \, d\eta, \int g_2 \, d\eta \right\} \leq \min\{g_1(x), g_2(x)\} = k(x).$$

Therefore, $V_k^*(\delta_x) = k(x)$. But then, since $\nu \leq_{\mathcal{C}} \mu$ and V_k^* is affine, we have

$$\int k \, d\nu \leq V_k^*(\mu) = \int V_k^*(\delta_x) \mu(dx) = \int k \, d\mu < \int k \, d\nu,$$

a contradiction.

Remark 1. It is noteworthy that throughout the proof, we do not use any existing results besides the duality theorem, the Krein-Milman theorem, and the separating hyperplane theorem. In particular, $(a) \implies (c)$ is in fact the celebrated Strassen theorem ([Strassen 1965](#); [Meyer 1966](#)). Our approach leads to an alternative proof using duality and the Krein-Milman theorem. More importantly, we also prove that Strassen's theorem has a converse—the existence of order-preserving couplings implies that the stochastic order must be defined by a min-closed test function cone.

Remark 2. We state [Theorem 1](#) for the optimization problem where the constraint set is $\{\nu : \nu \leq_{\mathcal{C}} \mu\}$. Of course, we can also flip the order by considering $\{\nu : \mu \leq_{\mathcal{C}} \nu\}$, in which case all results in [Theorem 1](#) continue to hold with the corresponding directional changes: (i) min-closure replaced by max-closure, (ii) the definition of order-preserving coupling replaced by the opposite direction, and (iii) the $\mathcal{C}$ -envelope of f that characterizes the

value function V_f^* replaced by the $[-\mathcal{C}]$ -envelope of f .

3.1 Consequences under Min-Closure

Theorem 1 immediately implies a series of structural properties about optimization involving integral stochastic orders satisfying the min-closure property.

Corollary 1 (Pointwise Optimization and Envelope). *For any $\leq_{\mathcal{C}}$ where $\mathcal{C}$ is min-closed,*

$$V_f^*(\mu) = \int_X V_f^*(\delta_x) \mu(dx) = \int_X \bar{f}(x) \mu(dx).$$

Corollary 1 identifies the affine value function by the $\mathcal{C}$ -envelope of f and observes that the original optimization admits a pointwise solution where the solution can be obtained by treating each x separately and maximizing over distributions $\eta_x \leq_{\mathcal{C}} \delta_x$.

Corollary 2 (Extreme Points of Chains). *Consider any $\leq_{\mathcal{C}}$ where $\mathcal{C}$ is min-closed and any compact convex subset M of $\Delta(X)$ that is closed under $\leq_{\mathcal{C}}$. The following are equivalent:*

- (a) $(\mu_i)_{i=1}^n$ is an extreme point of the set $\{(\tilde{\mu}_i)_{i=1}^n \in M^n : \tilde{\mu}_{i+1} \leq_{\mathcal{C}} \tilde{\mu}_i\}$.
- (b) μ_1 is an extreme point of M and μ_i is an extreme point of the set $\{\tilde{\mu} \in M : \tilde{\mu} \leq_{\mathcal{C}} \mu_{i-1}\}$ for all $i \in \{2, \dots, n\}$.

Corollary 2 shows that the extreme points of a chain of ordered measures $(\mu_i)_{i=1, \dots, n}$ admit a simple decomposable structure whenever the order $\leq_{\mathcal{C}}$ satisfies the min-closure property. An example of **Corollary 2** is when $\mathcal{C}$ consists of the set of concave functions on $\mathbb{R}^m$, and $M = \Delta(X)$, in which case we recover the extreme point structure of a pair of measures ordered by the convex order as characterized in Ciosmak (2023).

Corollary 3 (Reduction of Nested Optimization). *For any $\leq_{\mathcal{C}}$ where $\mathcal{C}$ is min-closed and any **(Nested Optimization)**, there exists (μ, ν) solving **(Nested Optimization)** such that $\mu \in \text{ext}(M)$ and $\nu \in \text{ext}(X_f^*(\mu))$.*

Corollary 3 shows that **(Nested Optimization)** admits simple solutions where the first mover simply chooses an extreme point from the feasible space M even after taking into account the response by the second mover. In fact, as **Theorem 1** reveals, when the first mover's objective $W(\mu, \nu)$ is a linear functional and hence can be written as $\int w_A(x) \mu(dx) + \int w_B(x) \nu(dx)$, there exists a modified linear objective $\widetilde{W}$ where w_B is modified to be $\widetilde{w}_B := w_B * P_f^*$ and $P_f^*(\cdot | x)$ is the pointwise solution for the second mover, such that the first mover can simply optimize the modified linear objective as if the second

mover did not exist. We describe our results and applications of nested optimization in detail in [Section 5](#).

For any (μ, ν) with $\nu \leq_C \mu$, we say that they are *uniquely rationalized* if they admit a unique order-preserving kernel P , i.e., there exists a μ -a.e. unique P such that $\nu = P * \mu$ and $P * \delta_x \leq_C \delta_x$ for μ -a.e. x . The following result characterizes structural properties of extreme and exposed points of stochastic orbits for any integral stochastic order satisfying the min-closure property:

Proposition 1 (Extreme and Exposed Points of Stochastic Orbits). *For any $\leq_C$ and any $\mu \in \Delta(X)$, let*

$$\mathcal{K}_\mu := \left\{ \nu : \nu \leq_C \mu \right\}.$$

If $\mathcal{C}$ is min-closed, then:

- (i) $\nu \in \text{ext}(\mathcal{K}_\mu)$ if and only if (μ, ν) is uniquely rationalized by an order-preserving kernel P where $P * \delta_x$ is an extreme point of $\mathcal{K}_{\delta_x}$ μ -a.e.
- (ii) $\nu \in \text{exp}(\mathcal{K}_\mu)$ if and only if (μ, ν) is uniquely rationalized by an order-preserving kernel P where $P * \delta_x$ is an exposed point of $\mathcal{K}_{\delta_x}$ μ -a.e. by the same continuous function f .

[Proposition 1](#) shows that for any convex orbit set $\mathcal{K}_\mu$ where $\leq_C$ satisfies the min-closure property, the extreme and exposed points admit a simple structure: (i) ν is an extreme point if and only if it can be uniquely constructed via an order-preserving kernel from μ that is pointwise extremal; (ii) ν is an exposed point if and only if it can be uniquely constructed via an order-preserving kernel from μ that is pointwise exposed by the same function f . [Proposition 1](#) nests the extreme-point characterization of [Roy \(1976\)](#) for the convex order as a special case. As we will demonstrate, these properties are extremely helpful for explicitly identifying the extreme and exposed points of the orbits defined by multidimensional mean-preserving spreads and multidimensional stochastic dominance.

3.2 Exposed Points of Multidimensional MPS and FOSD

In what follows, we demonstrate the application of [Proposition 1](#) by characterizing exposed points of multidimensional MPS and FOSD orbits.⁹

⁹For simplicity, we focus on the characterization of exposed points, as they are already dense in the set of extreme points by Straszewicz's theorem ([Klee jr. 1958](#)). Alternatively, we could directly apply part (i) of [Proposition 1](#) to characterize the extreme points. Unlike in the one-dimensional case, not all extreme points are exposed, but the characterizations are largely similar.

Multidimensional MPS. Let

$$\text{MPS}(\mu) := \{ \nu : \nu \leq_{\text{concave}} \mu \},$$

where $\leq_{\text{concave}}$ is the concave order, i.e., $\mathcal{C}$ is the set of concave functions on $\mathbb{R}^n$. Since concave functions are closed under pointwise minimum, by [Theorem 1](#), optimization problems involving $\text{MPS}(\mu)$ admit simple structures—in particular, [Corollaries 1 to 3](#) and [Proposition 1](#) immediately apply. Now, to illustrate, we apply [Proposition 1](#) to further characterize the exposed points of $\text{MPS}(\mu)$.

For any f , the $\mathcal{C}$ -envelope $\bar{f}$ is simply the concavification of f , which is a.e. differentiable. Let $\mathcal{A}$ be the set of differentiability points of $\bar{f}$. For any $x \in \mathcal{A}$, let h^x be the unique supporting hyperplane of $\bar{f}$. Let $\mathcal{H} := \{h^x\}_{x \in \mathcal{A}}$. Define

$$S := \{x : \bar{f}(x) = f(x)\}.$$

For any $h \in \mathcal{H}$, let

$$R_h := \{x : \bar{f}(x) = h(x)\},$$

and let $S_h^c := R_h \cap S^c$. Then, note that $\{S_h^c\}_{h \in \mathcal{H}}$ is a partition of $S^c \cap \mathcal{A}$. We refer to $\{S_h^c\}_{h \in \mathcal{H}}$ as the **affine envelope regions**. For any $h \in \mathcal{H}$, let

$$T_h := \{x : f(x) = h(x)\}.$$

We say that f has a **strict contact** if $x \notin \text{conv}(T_{h^x} \setminus \{x\})$ for a.e. $x \in S \cap \mathcal{A}$. We say an affine envelope region S_h^c is **simplicial-touching** if

$$R_h \text{ is a simplex and } \text{ext}(R_h) = T_h.$$

If S_h^c is simplicial-touching, then for any $x \in S_h^c$, there exists a unique mixture of extreme points of R_h that represents x —call it the unique **barycentric splitting** of x .

Proposition 2 (Exposed Points of Multidimensional MPS). *For any full-dimensional compact convex $X \subseteq \mathbb{R}^n$ and any $\mu \in \Delta(X)$ that is mutually absolutely continuous with respect to the Lebesgue measure, the following statements are equivalent:*

- (a) ν is an exposed point of $\text{MPS}(\mu)$.
- (b) There exists a continuous f with a strict contact and simplicial-touching affine envelope regions $S_{h^x}^c$ for a full-measure set of $x \in S^c$, and $\nu = P^f * \mu$, where $P^f(\cdot | x) = \delta_x$ for $x \in S$ and $P^f(\cdot | x)$ is the unique barycentric splitting for the full-measure set of $x \in S^c$.

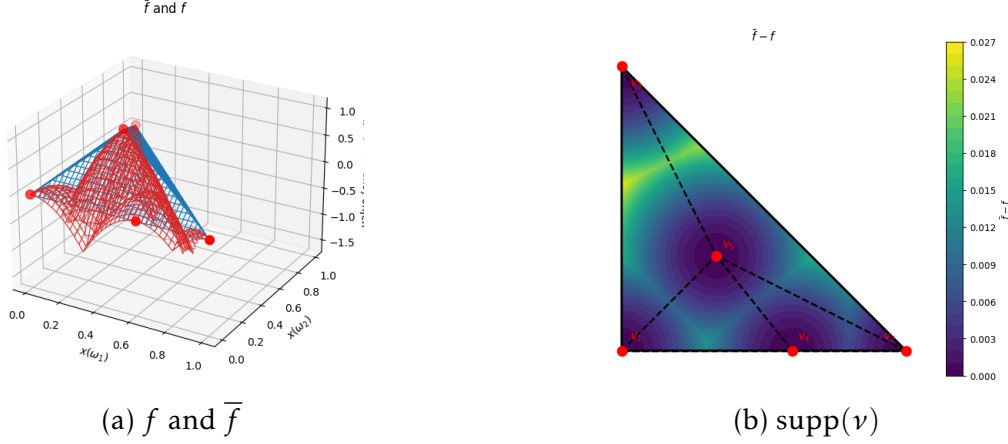


Figure 2: An exposed point ν of $\text{MPS}(\mu)$

Figure 2 illustrates an exposed point ν of the set $\text{MPS}(\mu)$. Figure 2a plots the associated exposing function f (in red) and its concave envelope $\bar{f}$ (in blue), and marks the points where $\bar{f} = f$. It can be readily seen from Figure 2a that f has a strict contact and simplicial-touching affine envelope regions. Figure 2b plots the heat map of $\bar{f} - f$, and the dots correspond to the points where $\bar{f} = f$, which indicate the support of ν .

An immediate consequence of Proposition 2 is that any exposed point of $\text{MPS}(\mu)$ takes the form of $P * \mu$, where the kernel P has the property that either (i) it does not move the mass; or (ii) on a collection of non-overlapping simplices (e.g., triangles in $\mathbb{R}^2$), it splits each point in a simplex into the vertices of the simplex. This immediately generalizes the exposed point characterization of Kleiner, Moldovanu, and Strack (2021) in the one-dimensional case, where the convex regions are simply intervals.

Multidimensional MPC. Now, consider the orbit

$$\text{MPC}(\mu) := \{ \nu : \mu \leq_{\text{concave}} \nu \} = \{ \nu : \nu \leq_{\text{convex}} \mu \}.$$

Since the set of convex functions is not closed under pointwise minimum—it is closed under pointwise maximum—it follows that optimization involving $\text{MPC}(\mu)$ generally admits complex structures. For example, by Theorem 1, the value function V_f^* must be strictly concave on $\Delta(X)$ for some f . See Kleiner et al. (2024) for the characterization of the structure of Lipschitz exposed points of $\text{MPC}(\mu)$ for finitely-supported μ .

Multidimensional Stochastic Dominance. Let

$$\text{LSD}(\mu) := \{ \nu : \nu \leq_{\text{nondecreasing}} \mu \},$$

where $\leq_{\text{nondecreasing}}$ is the stochastic dominance order, i.e., $\mathcal{C}$ is the set of nondecreasing functions on $\mathbb{R}^n$. Similarly, let

$$\text{HSD}(\mu) := \{ \nu : \mu \leq_{\text{nondecreasing}} \nu \} = \{ \nu : \nu \leq_{\text{nonincreasing}} \mu \}.$$

Since both nondecreasing functions and nonincreasing functions are closed under pointwise minimum, by [Theorem 1](#), optimization problems involving $\text{LSD}(\mu)$ and $\text{HSD}(\mu)$ admit simple structures—in particular, [Corollaries 1 to 3](#) and [Proposition 1](#) immediately apply. Now, to illustrate, we apply [Proposition 1](#) to further characterize the exposed points of $\text{LSD}(\mu)$ and $\text{HSD}(\mu)$.

We start with $\text{LSD}(\mu)$. For any f , the $\mathcal{C}$ -envelope $\bar{f}$ is simply the monotone envelope of f , i.e. $\bar{f}(x) = \max_{y \leq x} f(y)$. Define

$$S := \{ x : \bar{f}(x) = f(x) \}.$$

For any $z \in \text{Ran}(\bar{f})$, let

$$R_z := \{ x : \bar{f}(x) = z \},$$

and

$$T_z := S \cap R_z = \{ x : \bar{f}(x) = f(x) = z \}.$$

We say that f has a **strict monotone contact** if for a.e. $x \in S$, we have $f(x) > f(y)$ for all $y \leq x$ and $y \neq x$ (in the coordinatewise order). For any $z \in \text{Ran}(\bar{f})$, let

$$S_z^c := R_z \cap S^c.$$

Clearly, $\{S_z^c\}_{z \in \text{Ran}(\bar{f})}$ is a partition of S^c . We write $z^x = \bar{f}(x)$. We refer to $\{S_z^c\}$ as **monotone envelope regions**. We say a monotone envelope region S_z^c is **downward staircase-like** if there exists some upward-closed set $A_z \subset [0, 1]^n$ such that

$$R_z \setminus A_z = \bigsqcup_{\underline{x} \in T_z} \left(\{ x : x \geq \underline{x} \} \setminus A_z \right).$$

If S_z^c is downward staircase-like, then for any $x \in R_z \setminus A_z$, there exists a unique **downward transport** that deterministically transports x to $\underline{x}$ for some $\underline{x} \in T_z$.

Proposition 3 (Exposed Points of Multidimensional FOSD). *For any $\mu \in \Delta([0, 1]^n)$ that is mutually absolutely continuous with respect to the Lebesgue measure, the following statements are equivalent:*

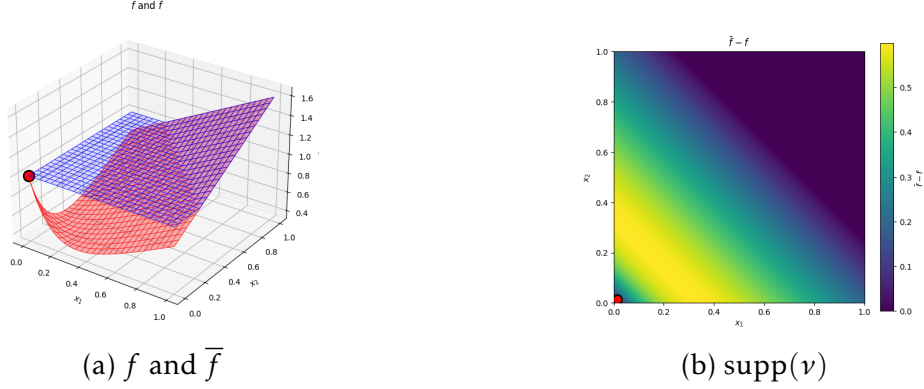


Figure 3: An exposed point ν of $\text{LSD}(\mu)$

- (a) ν is an exposed point of $\text{LSD}(\mu)$.
- (b) There exists a continuous f with a strict monotone contact and downward staircase-like monotone envelope regions $S_{z^x}^c$ for a full-measure set of $x \in S^c$ with $x \in R_{z^x} \setminus A_{z^x}$, and $\nu = P^f * \mu$, where $P^f(\cdot | x) = \delta_x$ for $x \in S$ and $P^f(\cdot | x)$ is the unique downward transport for the full-measure set of $x \in S^c$.

Figure 3 illustrates an exposed point ν of $\text{LSD}(\mu)$. Figure 3a plots the associated exposing function f and its monotone envelope $\bar{f}$. It can be seen from Figure 3a that f has a strict monotone contact. Figure 3b plots the heat map of $\bar{f} - f$. The region where $\bar{f} = f$ consists of the point $(0,0)$ and the top-right corner. Every point not in this region has a unique downward transport to $(0,0)$.

An immediate consequence of Proposition 3 is that any exposed point of $\text{LSD}(\mu)$ takes the form of $P * \mu$, where the kernel P has the property that either (i) it does not move the mass; or (ii) on a collection of non-overlapping differences of two nested upsets that are downward staircase-like, it transports each point in any such region down to a corresponding unique minimal element. This immediately generalizes the exposed point characterization of Yang and Zentefis (2024) in the one-dimensional case, where the differences of two nested upsets are simply intervals.

The case of $\text{HSD}(\mu)$ is exactly symmetric to $\text{LSD}(\mu)$, in contrast to the asymmetry that arises between $\text{MPS}(\mu)$ and $\text{MPC}(\mu)$. In particular, it can be characterized symmetrically by using *decreasing envelope regions*, the differences of nested downward-closed sets that form *upward staircase-like* shapes, and the *upward transport* maps. We omit the details here.

4 Consistent Comparisons of Experiments

In this section, we apply our main characterization to generalize Blackwell's theorem. In a pioneering paper, [Blackwell \(1951\)](#) identifies an order on experiments that has a dual representation. On the one hand, the Blackwell order has an instrumental-value-based representation: an experiment Blackwell dominates another if and only if one yields a higher value under *all* statistical decision problems than the other. On the other hand, the Blackwell order has an information-technology-based interpretation: an experiment Blackwell dominates another if and only if one can be obtained by adding noise to (i.e., garbling) the other. In the space of distributions of posterior beliefs induced by an experiment under a given prior, this dual representation essentially identifies an order $\leq$ over $\Delta(X)$, a convex cone of test functions (i.e., a class of problems) $\mathcal{C}$, and a set of information technologies $\mathcal{P}$, such that the order $\leq$ can be simultaneously represented by comparing values of problems $g \in \mathcal{C}$, as well as being coupled by some information technologies $P \in \mathcal{P}$. In this regard, Blackwell's theorem identifies such an order where $\mathcal{C}$ is the set of convex functions and $\mathcal{P}$ is the set of martingales.

Given Blackwell's theorem, one natural question is: in general, which orders $\leq$ over $\Delta(X)$ admit such dual representations? When is it the case that an abstract order $\leq$ over $\Delta(X)$ can be simultaneously interpreted as an order that reflects instrumental values, as well as an order that reflects the information content of distributions over posterior beliefs? In this section, we characterize all such orders. To begin with, we first introduce some notation and terminology.

Throughout this section, fix a compact Polish state space Ω , and let $X := \Delta(\Omega)$ be the space of beliefs over Ω . Let $\mathcal{C}_0$ be the set of bounded lower semicontinuous functions on X and $\mathcal{P}_0$ be the set of measurable Markov kernels. For any $\mathcal{P} \subseteq \mathcal{P}_0$ and for any $x \in X$, let

$$\mathcal{P}_x := \left\{ P(\cdot \mid x) : P \in \mathcal{P} \right\}.$$

We say that $\mathcal{P} \subseteq \mathcal{P}_0$ is **rectangular** if $P \in \mathcal{P}$ whenever $P(\cdot \mid x) \in \mathcal{P}_x$ for all $x \in X$. We say that a pair $(\mathcal{C}, \mathcal{P})$ is **regular** if $\mathcal{C} \subseteq \mathcal{C}_0$ is a convex cone such that $\mathcal{C} \cap C(X)$ is sup-norm closed, contains all constant functions, admits continuous approximations from below, and is closed under bounded increasing pointwise limits; while $\mathcal{P} \subseteq \mathcal{P}_0$ has a closed graph,¹⁰ contains the identity kernel, and is convex and rectangular. We impose regularity on $(\mathcal{C}, \mathcal{P})$ throughout this section unless stated otherwise.

¹⁰That is, the graph $\{(x, \eta) : x \in X, \eta \in \mathcal{P}_x\}$ is closed in $X \times \Delta(X)$.

For any $\mathcal{C}$, define the following comparison:

$$\mu \leq_{\mathcal{C}} \nu \text{ if } \int g(x) \nu(dx) \geq \int g(x) \mu(dx), \forall g \in \mathcal{C}. \quad (\text{Value Description})$$

For any $\mathcal{P}$, define the following comparison:

$$\mu \leq^{\mathcal{P}} \nu \text{ if } \exists P \in \mathcal{P} \text{ s.t. } \nu = P * \mu. \quad (\text{Information Description})$$

For any order $\leq$ on $\Delta(X)$, we say that $\leq$ is **Blackwell-consistent** if there exists some (regular) $(\mathcal{C}, \mathcal{P})$ such that

$$\leq_{\mathcal{C}} = \leq = \leq^{\mathcal{P}}, \quad (\text{Blackwell Consistency})$$

in which case we call $(\mathcal{C}, \mathcal{P})$ a **Blackwell-consistent pair**.

We say that $\mathcal{C}$ is **max-closed** if for any $g_1, g_2 \in \mathcal{C}$, $\max\{g_1, g_2\} \in \mathcal{C}$. As an immediate consequence of [Theorem 1](#), our next result characterizes all Blackwell-consistent orders:

Theorem 2. *An order $\leq$ is Blackwell-consistent if and only if $\leq = \leq_{\mathcal{C}}$ for a max-closed $\mathcal{C}$.*

Proof of Theorem 2. The “if” direction: Suppose $\leq = \leq_{\mathcal{C}}$ for a max-closed $\mathcal{C}$. Let

$$\mathcal{P} := \left\{ P \in \mathcal{P}_0 : \delta_x \leq_{\mathcal{C}} P * \delta_x \text{ for all } x \right\}.$$

Note that $\mu \leq \nu \iff \nu \leq_{[-\mathcal{C}]} \mu$ where $[-\mathcal{C}]$ is a convex cone that is min-closed. Thus, for any $\mu \leq \nu$, by [Theorem 1](#),¹¹ there exists some P such that $\nu = P * \mu$ where $P * \delta_x \leq_{[-\mathcal{C}]} \delta_x$, i.e., $\delta_x \leq_{\mathcal{C}} P * \delta_x$ for all x , and hence $P \in \mathcal{P}$. Therefore, $\leq_{\mathcal{C}} \implies \leq^{\mathcal{P}}$. Moreover, for any $\nu = P * \mu$ for some $P \in \mathcal{P}$, since $\leq_{\mathcal{C}}$ is an integral stochastic order and $\delta_x \leq_{\mathcal{C}} P * \delta_x$ for all x , it follows immediately that $\mu \leq_{\mathcal{C}} \nu$. Thus, $\leq^{\mathcal{P}} \implies \leq_{\mathcal{C}}$ and hence these two are equivalent. Thus, $\leq$ is Blackwell-consistent.¹²

The “only if” direction: Suppose $\leq$ is Blackwell-consistent. Then there exists some regular $(\mathcal{C}, \mathcal{P})$ such that $\leq_{\mathcal{C}} = \leq = \leq^{\mathcal{P}}$. We claim that $\leq_{[-\mathcal{C}]}$ admits order-preserving couplings. Indeed, fix any $\nu \leq_{[-\mathcal{C}]} \mu$. Then, $\mu \leq_{\mathcal{C}} \nu$ and hence $\nu = P * \mu$ for some $P \in \mathcal{P}$. For any x , note that by definition $\delta_x \leq^{\mathcal{P}} P * \delta_x$ and hence $\delta_x \leq_{\mathcal{C}} P * \delta_x$ by Blackwell consistency. Thus, for all $x \in X$ and all $g \in \mathcal{C}$, we have

$$\int g(y) P(dy | x) \geq g(x)$$

¹¹It can be verified that the technical assumptions required by [Theorem 1](#) hold for $[-\mathcal{C}]$ if $\mathcal{C}$ is regular.

¹²It can be verified that the constructed $\mathcal{P}$ is regular; see [Lemma 6](#) in the appendix.

and hence

$$\int [-g](y)P(dy \mid x) \leq [-g](x).$$

Thus, $P * \delta_x \leq_{[-\mathcal{C}]} \delta_x$ for all x . Therefore, $\leq_{[-\mathcal{C}]}$ admits order-preserving couplings. By [Theorem 1](#), $[-\mathcal{C}]$ is min-closed and hence $\mathcal{C}$ is max-closed, as desired. $\square$

[Theorem 2](#) asserts that $\leq$ is Blackwell-consistent—admitting simultaneous value-based and information-based representations—if and only if the class of test functions $\mathcal{C}$ in its value description is closed under pointwise maximum. Intuitively, it means that whenever $g_1, g_2 \in \mathcal{C}$ are two problems, the problem whose value at each posterior belief x is $\max\{g_1(x), g_2(x)\}$ —corresponding to choosing the better of the two problems at each belief—must also belong to $\mathcal{C}$. This “choose the better problem” closure property turns out to be the *exact* condition that bridges the value-based and information-based descriptions of an order.

Now, a few natural questions arise. What is the key structure of $\mathcal{P}$ in the dual space that mirrors the max-closure property of $\mathcal{C}$? Which pairs $(\mathcal{C}, \mathcal{P})$ are consistent? Given a Blackwell-consistent order, are there multiple pairs $(\mathcal{C}, \mathcal{P})$ that represent the same order? We now answer all of them in our next result.

For any $P_1, P_2 \in \mathcal{P}_0$, recall from [\(Composition\)](#) that $P_1 \circ P_2$ denotes the composition of the two kernels. We say that $\mathcal{P} \subseteq \mathcal{P}_0$ is **composition-closed** if for any $P_1, P_2 \in \mathcal{P}$, $P_1 \circ P_2 \in \mathcal{P}$. As we show, it turns out that composition-closure property of $\mathcal{P}$ is the *exact* condition that mirrors the max-closure property of $\mathcal{C}$, each of which is equivalent to Blackwell consistency.

To characterize the consistent pairs, we build on the construction given in the proof of [Theorem 2](#). To start, recall the notation from [\(Conditional Expectation\)](#), $[g * P](x) := \int g(y)P(dy \mid x)$. Define two operators Φ and Ψ as follows: For any $\mathcal{C} \subseteq \mathcal{C}_0$, let

$$\Psi(\mathcal{C}) := \left\{ P \in \mathcal{P}_0 : g * P \geq g, \forall g \in \mathcal{C} \right\} \quad \textbf{(Improving Kernels)}$$

denote the set of belief transitions that are uniformly improving across all test functions in $\mathcal{C}$. For any $\mathcal{P} \subseteq \mathcal{P}_0$, let

$$\Phi(\mathcal{P}) := \left\{ g \in \mathcal{C}_0 : g * P \geq g, \forall P \in \mathcal{P} \right\} \quad \textbf{(Improving Utilities)}$$

denote the set of test functions that are uniformly improving across all belief transitions

in $\mathcal{P}$. We say that $(\mathcal{C}, \mathcal{P})$ is **Blackwell-invariant** if

$$\Psi(\mathcal{C}) = \mathcal{P} \text{ and } \Phi(\mathcal{P}) = \mathcal{C}. \quad (\text{Blackwell Invariance})$$

A Blackwell-invariant pair $(\mathcal{C}, \mathcal{P})$ has a straightforward interpretation. It means that the available information technologies $\mathcal{P}$ are exactly the ones that are valuable for problems in $\mathcal{C}$ (i.e., $\mathcal{P} = \Psi(\mathcal{C})$). In the meantime, the problems $\mathcal{C}$ are exactly the ones for which the available information technologies $\mathcal{P}$ are valuable (i.e., $\mathcal{C} = \Phi(\mathcal{P})$).

Finally, we say that $\mathcal{P}$ is **one-shot sufficient** if for all f and all x , we have

$$\sup_{P \in \mathcal{P}} \int f(y) P(dy | x) = \sup_{P \in \mathcal{P} \otimes \mathcal{P}} \int f(y) P(dy | x), \quad (\text{One Shot})$$

where

$$\mathcal{P} \otimes \mathcal{P} := \{P_1 \circ P_2 : P_1, P_2 \in \mathcal{P}\}$$

denotes the set of all (twice) compositions of elements in $\mathcal{P}$. It is clear that if $\mathcal{P}$ is composition-closed, then it is one-shot sufficient.

In the case of the Blackwell order, $\mathcal{C}$ is the cone of convex functions and $\mathcal{P}$ is the set of martingale transition kernels. Clearly, $\mathcal{C}$ is max-closed and $\mathcal{P}$ is composition-closed. Moreover, $\Psi(\mathcal{C}) = \mathcal{P}$ and $\Phi(\mathcal{P}) = \mathcal{C}$. That is, the test functions $\mathcal{C}$ and transition kernels $\mathcal{P}$ associated with the Blackwell order are in fact Blackwell-invariant. This turns out not to be a coincidence. Our next result shows that (i) the composition-closure property of $\mathcal{P}$ mirrors the max-closure property of $\mathcal{C}$, each of which characterizes Blackwell consistency, and (ii) Blackwell consistency is equivalent to Blackwell invariance, which turns out to be also equivalent to the one-shot sufficiency property of the set of transition kernels for optimization problems.

In general, different convex cones $\mathcal{C}$ and $\mathcal{C}'$ or different sets of kernels $\mathcal{P}$ and $\mathcal{P}'$ might induce the same comparisons, i.e., $\leq_{\mathcal{C}} = \leq_{\mathcal{C}'}$ or $\leq^{\mathcal{P}} = \leq^{\mathcal{P}'}$. Since Blackwell consistency is only a property about the induced order, one might naturally expect that there might be multiplicity involved in determining the consistent pairs. However, perhaps surprisingly, every Blackwell-consistent order admits a **unique** consistent pair that characterizes it.

Theorem 3 (Blackwell-Consistent Pairs). *The following are equivalent:*

- (a) $(\mathcal{C}, \mathcal{P})$ is Blackwell-consistent.
- (b) $(\mathcal{C}, \mathcal{P})$ is Blackwell-invariant.
- (c) $\mathcal{P}$ is one-shot sufficient, and $\mathcal{C} = \Phi(\mathcal{P})$.

(d) $\mathcal{P}$ is composition-closed, and $\mathcal{C} = \Phi(\mathcal{P})$.

(e) $\mathcal{C}$ is max-closed, and $\mathcal{P} = \Psi(\mathcal{C})$.

Moreover, every Blackwell-consistent order $\leq$ admits a unique consistent pair $(\mathcal{C}, \mathcal{P})$.

4.1 Which Orders are Consistent under Bayesian Updating?

With [Theorem 2](#) and [Theorem 3](#), we now further specialize to Blackwell's environment, and characterize which orders of Blackwell experiments are Blackwell-consistent. Clearly, if beliefs are updated via Bayes' rule, then by Blackwell's theorem, the Blackwell order is Blackwell-consistent. We characterize all orders of Blackwell experiments that are Blackwell-consistent when beliefs are updated via Bayes' rule in this section, and characterize all updating rules that admit Blackwell-consistent orders in the next section.

4.1.1 Bayes-Plausible and Blackwell-Consistent Orders

To begin with, we say that an order $\leq$ on $\Delta(X)$ is *Bayes-plausible* if

$$\mu \leq \nu \implies \int x \mu(dx) = \int x \nu(dx), \quad (\text{Bayes Plausibility})$$

where the integral is in the sense of the Pettis integral, i.e., the two belief distributions must have the same barycenter.

An immediate consequence of [Theorem 2](#) is the following:

Corollary 4. *For any pair $(\mathcal{C}, \mathcal{P})$, the following are equivalent:*

(a) $\leq_{\mathcal{C}} = \leq^{\mathcal{P}}$ is Bayes-plausible.

(b) $\mathcal{C}$ is a max-closed superset of all convex functions, and $\mathcal{P} = \Psi(\mathcal{C})$.

(c) $\mathcal{P}$ is a composition-closed subset of all martingale kernels, and $\mathcal{C} = \Phi(\mathcal{P})$.

In particular, any Bayes-plausible and Blackwell-consistent order strengthens the Blackwell order, and the Blackwell order is the weakest such order.

Indeed, [Corollary 4](#) follows because Bayes plausibility implies that all affine functions must be in the cone of test functions, which combined with [Theorem 2](#) and [Theorem 3](#) implies that all convex functions must be in the test cone since the test cone must be max-closed and any convex function is a pointwise supremum of a collection of affine functions. Therefore, any Bayes-plausible weakening of the Blackwell order through a

subset of convex functions, such as the Lehmann order (Lehmann 1988) or the linear convex order (Chen 2025), cannot be consistent in our sense.¹³ Corollary 4 shows that any Bayes-plausible and Blackwell-consistent order $\leq$ must imply the Blackwell order—therefore, the Blackwell order is necessary for a consistent comparison, while maintaining Bayes plausibility. In other words, any Bayes-plausible, Blackwell-consistent order $\leq$ must rely on a set $\mathcal{C}$ of problems that includes all decision problems, and on a set $\mathcal{P}$ of information technologies that is a subset of martingales.

4.1.2 Constrained Information Design

An alternative interpretation of Corollary 4 is a characterization of what kinds of constrained information technologies admit a consistent representation based on the instrumental value of information. In particular, Corollary 4 identifies a class of **constrained information design** problems—where the feasible information technology is constrained to a subset $\mathcal{P}$ of martingales—that are particularly tractable and preserve many of the desirable properties of the unconstrained problem.

A rectangular set of transition kernels $\mathcal{P}$ equivalently defines a set of feasible experiments $\mathcal{P}_x$ at each x . Thus, $\mathcal{P}$ defines a constrained information design problem:

$$V_f^{\mathcal{P}}(\delta_x) := \max_{\delta_x \leq^{\mathcal{P}} \nu} \int f(y) \nu(dy),$$

where ν is the posterior distribution given the prior x and an experiment chosen from $\mathcal{P}_x$. In persuasion problems, the function f is the indirect utility function of the sender. In flexible learning problems with a uniformly posterior separable cost function, the function $f(x) = v(x) - h(x)$ where the convex function v is the indirect utility function from a decision problem, and the convex function h is the measure of uncertainty.

By Theorem 3, if $\mathcal{P}$ is composition-closed, then $\leq^{\mathcal{P}} = \leq_{\mathcal{C}}$ for a max-closed $\mathcal{C}$, which by Theorem 1 then implies the following envelope characterization:

$$V_f^{\mathcal{P}}(\mu) = \int \left(\max_{\delta_x \leq_{\mathcal{C}} \nu} \int f(y) \nu(dy) \right) \mu(dx) = \int \hat{f}(x) \mu(dx),$$

where $\hat{f}$ is the $[-C]$ -envelope of f . Importantly, the envelope $\hat{f}$ here is **prior-independent**, just like in the unconstrained case. This contrasts with the case where the constraints are moment constraints on the posterior distributions ν —such as the ones studied by

¹³It is immediate to verify that neither of these orders is defined by a set of test functions that is max-closed.

Doval and Skreta (2024), who show that the value of the constrained problem can be found by concavifying the Lagrangian given the optimal multipliers which depend on the prior. Indeed, such moment constraints on the posterior distributions are generally *not* composition-closed in terms of the implied set $\mathcal{P}$, which as our results show are also necessary for the existence of prior-independent envelope characterization. An immediate consequence of this prior-independent envelope is the following characterization generalizing the unconstrained case:

Corollary 5. *For any composition-closed $\mathcal{P}$, a feasible ν is constrained-optimal if and only if*

$$\text{supp}(\nu) \subseteq \{y \in X : f(y) = \hat{f}(y)\}$$

and

$$\int f(y) \nu(dy) = \hat{f}(x),$$

where x is the prior and $\hat{f}$ is the $[-\Phi(\mathcal{P})]$ -envelope of f . The value of a constrained optimal signal is $\hat{f}(x)$, and the designer benefits from constrained signals if and only if $\hat{f}(x) > f(x)$.

Corollary 5 shows that many structural properties derived in Kamenica and Gentzkow (2011) can be immediately generalized to the constrained case where signals need not be fully flexible as long as $\mathcal{P}$ is composition-closed. The key geometric structure can be obtained by replacing the concave envelope with the $[-\Phi(\mathcal{P})]$ -envelope.

How does the constrained optimal signal compare to the KG signal? We say that the constrained problem is *value-eliminating* if there exists some prior $x \in X$ under which the constrained designer cannot benefit from information design while the unconstrained designer strictly benefits. A consequence of Corollary 5 is the following comparison:

Proposition 4 (Constrained vs. KG Signals). *For finite Ω , composition-closed $\mathcal{P}$, and any f ,*

- (i) *either the constrained problem is value-eliminating;*
- (ii) *or at any prior x where the unconstrained optimal signal is unique, any constrained optimal signal cannot be strictly Blackwell-dominated by the unconstrained optimal signal, and always weakly Blackwell-dominates the unconstrained optimum when $|\Omega| = 2$.*

Indeed, to see Proposition 4, note that if the constrained problem is not value-eliminating, then by Corollary 5

$$\hat{f}^{\text{cav}}(x) > f(x) \implies \hat{f}^{\mathcal{C}}(x) > f(x),$$

where we use the superscripts to distinguish the concave envelope versus $\mathcal{C}$ -envelope where $\mathcal{C} = -\Phi(\mathcal{P})$. By Corollary 5 again, this implies that any constrained optimal ν^*

must satisfy

$$\nu^* \left(\text{conv}(\text{supp}(\nu_{\text{KG}})) \setminus \text{supp}(\nu_{\text{KG}}) \right) = 0,$$

since otherwise there exists some y in the above set such that $\hat{f}^{\text{cav}}(y) = \hat{f}^{\mathcal{C}}(y) = f(y)$ which would then contradict that ν_{KG} is the unique unconstrained optimum. This then implies that ν^* cannot be strictly Blackwell-dominated by ν_{KG} with any finite Ω , and must weakly Blackwell-dominate ν_{KG} with two states.

Next, we show two examples of constrained signals where $\mathcal{P}$ is composition-closed.

Privacy-Preserving Signals. Consider the setting adopted from [Strack and Yang \(2024\)](#). Let Ω be a compact Polish space and let $\mathcal{E}$ be a sub- σ -algebra on Ω . An event $E \in \mathcal{E}$ describes a *privacy event* that needs to be protected. Given a prior x_0 , a signal is said to be *privacy-preserving* if under any signal realization, the posterior belief about any event $E \in \mathcal{E}$ remains the same as the prior $x_0(E)$. For any $x \in X = \Delta(\Omega)$, let

$$K(x) := \{y \in X : y(E) = x(E), \forall E \in \mathcal{E}\}$$

be the set of posterior beliefs that preserves privacy when the prior is x . The set of privacy-preserving signals can be equivalently represented by a set $\mathcal{P}$ of martingales such that for all $P \in \mathcal{P}$ and for all $x \in X$,

$$P(K(x) \mid x) = 1.$$

Note that $\mathcal{K} := \{K(x) : x \in X\}$ forms a partition of X . As a result, $\mathcal{P}$ is composition-closed.

The order $\leq^{\mathcal{P}}$ has a natural interpretation. Two distributions $\mu, \nu \in \Delta(X)$ are ordered by $\leq^{\mathcal{P}}$ if and only if ν can be attained from μ via privacy-preserving signals $P \in \mathcal{P}$ that do not further reveal any information about events in $\mathcal{E}$ than what μ had already revealed.¹⁴ Since $\mathcal{P}$ is composition-closed, [Theorem 2](#) ensures that $\leq^{\mathcal{P}}$ is a Blackwell-consistent order. As a result, any distribution of posterior beliefs ν that can be obtained from μ , via disclosing privacy-preserving information, gives a higher instrumental value in all problems $g \in \mathcal{C}$, where

$$\mathcal{C} := \Phi(\mathcal{P}).$$

The set $\mathcal{C}$ of problems also has an intuitive description: every $g \in \mathcal{C}$ must be such that g is convex when restricted to each partitional element $K \in \mathcal{K}$. In other words, information disclosed by privacy-preserving signals can be ordered by their instrumental values for

¹⁴In the language of [Strack and Yang \(2024\)](#), this means that ν is the distribution of posterior beliefs induced by a conditionally privacy-preserving signal given the information revealed by μ .

decision problems, restricted to the domain of beliefs that preserve privacy. In particular, for a fixed prior x_0 , the order over privacy-preserving signals defined by $\mathcal{P}_{x_0}$ coincides with the Blackwell order. In general, however, the order $\leq^{\mathcal{P}} = \leq_{\mathcal{C}}$ is a strengthening of the Blackwell order, and there are signals that are Blackwell-ordered but are not ordered by the privacy-preserving order. These are exactly signals that disclose more information in ways that violate privacy.

An immediate consequence of [Theorem 3](#) here is an envelope characterization of the optimal value of privacy-preserving persuasion. Given a public signal $\mu \in \Delta(X)$, a sender chooses a (conditionally) privacy-preserving signal that does not reveal more information regarding $\mathcal{E}$. The sender's problem, by [Theorem 3](#), can be written as:

$$\max_{\nu: \mu \leq_{\mathcal{C}} \nu} \int f \, d\nu,$$

where $f : X \rightarrow \mathbb{R}$ is the indirect utility of the sender. Since $\mathcal{C}$ is max-closed, $-\mathcal{C}$ is min-closed. [Theorem 1](#) then implies that the value of the problem is characterized by

$$V_f^*(\mu) = \int \hat{f}^{\mathcal{C}} \, d\mu, \quad (\text{Privacy-Preserving Persuasion})$$

where

$$\hat{f}^{\mathcal{C}}(x) := \inf\{g(x) : g \in -\mathcal{C}, g \geq f\}$$

is the $[-\mathcal{C}]$ -envelope of f . The expression of [\(Privacy-Preserving Persuasion\)](#) further simplifies and generalizes Proposition 6 of [Strack and Yang \(2024\)](#), which characterizes the value of privacy-preserving persuasion for a fixed prior via concavification and optimal transport. In contrast, [\(Privacy-Preserving Persuasion\)](#) does not involve solving an optimal transport problem, and allows for arbitrary public information μ . Operationally, the $[-\mathcal{C}]$ -envelope is also often easier to compute relative to the concave closure. Indeed, instead of taking the concave closure of the entire function f , computing the $[-\mathcal{C}]$ -envelope only requires taking the lower-dimensional concave envelope of f on each partitional element $K \in \mathcal{K}$.

For instance, consider a simple privacy-preserving persuasion problem where the sender is a doctor and the receiver is an insurance company. The doctor can provide information about a patient to the insurance company. The state of the world could be one of three states: ω_1 , where the patient is genetically prone to cancer; ω_2 , where the patient is not genetically prone but often exposed to chemicals; and ω_3 , where the patient is unlikely to have cancer. The receiver can choose whether to insure the patient. The

doctor prefers the patient to be insured: they get a payoff of 1 if the patient is insured, and a payoff of zero otherwise. The insurance company prefers to insure a healthy patient: it gets payoff 1 if $\omega = \omega_3$, 0 if $\omega = \omega_2$, and -1 if $\omega = \omega_1$. The privacy events $\mathcal{E}$ are $\{\{\omega_1\}, \{\omega_2, \omega_3\}\}$, since HIPAA prohibits the doctor from disclosing genetic information about a patient. Suppose that there is already some public information about the patient known to the insurance company. Denote by $\mu \in \Delta(X)$ the receiver's initial distribution of posterior beliefs. The sender's privacy-preserving persuasion problem can be written as

$$\max_{v: \mu \leq_{\mathcal{C}} v} \int f \, dv,$$

where $f(x) := \mathbb{1}\{x(\omega_3) \geq x(\omega_1)\}$ is the sender's indirect utility as a function of posterior beliefs. It can be readily shown that the $[-\mathcal{C}]$ -envelope of f is given by

$$\hat{f}^{\mathcal{C}}(x) := \mathbb{1}\left\{x(\omega_1) \leq \frac{1}{2}\right\} \cdot \min\left\{1, \frac{x(\omega_3)}{x(\omega_1)}\right\},$$

and therefore the sender's value is

$$\int_X \mathbb{1}\left\{x(\omega_1) \leq \frac{1}{2}\right\} \cdot \min\left\{1, \frac{x(\omega_3)}{x(\omega_1)}\right\} \mu(dx).$$

In contrast, the concave envelope of f is given by

$$\hat{f}^{\text{cav}}(x) := \min\left\{1, 1 - x(\omega_1) + x(\omega_3)\right\},$$

which is greater than $\hat{f}^{\mathcal{C}}$. Thus, the sender's value, had there been no privacy constraints, would be

$$\int_X \min\left\{1, 1 - x(\omega_1) + x(\omega_3)\right\} \mu(dx) \geq \int_X \mathbb{1}\left\{x(\omega_1) \leq \frac{1}{2}\right\} \cdot \min\left\{1, \frac{x(\omega_3)}{x(\omega_1)}\right\} \mu(dx).$$

Figure 4 plots these two envelopes, where the functions are plotted on the $(x(\omega_3), x(\omega_1))$ -plane, using the fact that $x(\omega_2) = 1 - x(\omega_1) - x(\omega_3)$.

Sequential Sampling. Another class of technologies that is automatically composition-closed is *dynamic sampling/stopping* technologies. For example, [Brocas and Carrillo \(2007\)](#), [Henry and Ottaviani \(2019\)](#), [Morris and Strack \(2019\)](#), and [Ball and Espín-Sánchez \(2021\)](#) consider the classic Wald sampling model ([Wald 1945](#)) adopted in information design settings. We describe a general setting in discrete time. Fix a finite state space Ω . The

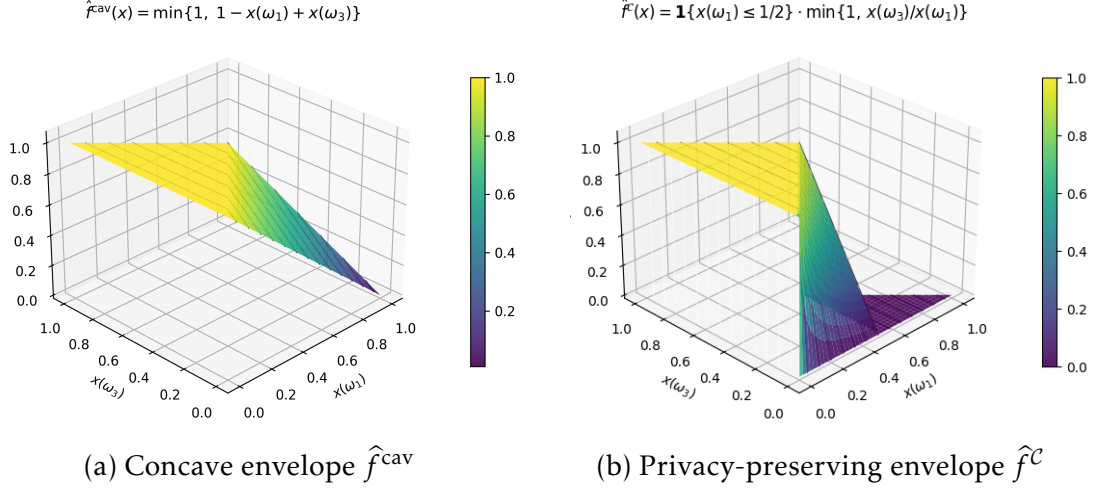


Figure 4: Comparison of envelopes

belief process $(x_n)_{n \geq 0}$ taking values in $X = \Delta(\Omega)$ under any monitoring technology in the style of [Wald \(1945\)](#) would evolve according to a time-homogeneous Markov martingale with some transition kernel Q .

A designer chooses a *randomized stopping time* τ with respect to the natural filtration of $(x_n)_{n \geq 0}$. The induced posterior distribution is the law of the stopped belief x_τ . Every such τ generates a transition kernel P^τ on X defined by

$$P^\tau(\cdot \mid x) = \mathbb{P}[x_\tau \in \cdot \mid x_0 = x].$$

Let

$$\mathcal{P} := \left\{ P^\tau : \tau \text{ is a randomized stopping time for } (x_n)_{n \geq 0} \right\}.$$

Then, it is evident that $\mathcal{P}$ is composition-closed and regular. Therefore, by [Theorem 3](#), there exists an order $\leq_{\mathcal{C}}$ stronger than the Blackwell order on $\Delta(X)$ such that

$$\mu \leq_{\mathcal{C}} \nu \iff \text{there exists a randomized stopping time } \tau \text{ s.t. } x_\tau \sim \nu \text{ with } x_0 \sim \mu.$$

Therefore, any dynamic sampling/stopping problem is equivalent to a *reduced-form constrained information design* for which [Corollary 5](#) and [Proposition 4](#) hold immediately—in particular, they admit envelope characterizations and satisfy structural properties just like in the unconstrained case studied in [Kamenica and Gentzkow \(2011\)](#). Moreover, as [Proposition 4](#) shows, unless the constrained problem is value-eliminating, the induced posterior distributions for any such problem must weakly Blackwell dominate the KG signal with a binary state, and cannot be strictly Blackwell dominated by the KG signal

with any finite state.

What is the appropriate envelope here? By [Corollary 5](#), the appropriate envelope $\hat{f}$ is defined by $[-\Phi(\mathcal{P})]$ -envelope of f . By definition, $g \in -\Phi(\mathcal{P})$ if and only if

$$\int g(y)P(dy | x) \leq g(x) \text{ for all } x \text{ and all } P \in \mathcal{P},$$

which is equivalent to

$$\int g(y)P^\tau(dy | x) \leq g(x) \text{ for all } x \text{ and all randomized stopping times } \tau.$$

Since the Markov process is time-homogeneous, note that the above is equivalent to

$$\int g(y)Q(dy | x) \leq g(x) \text{ for all } x,$$

where Q is the transition kernel of the Markov process. In the language of optimal stopping theory, such a function is called a *superharmonic/excessive function*. This provides a point of connection to the well-known *Dynkin's theorem* ([Dynkin 1963](#)) in optimal stopping. It is further generalized in [Dayanik and Karatzas \(2003\)](#) and states that the value function

$$V = \inf \left\{ g : g \text{ is excessive and } g \geq f \right\}, \quad (\text{Dynkin})$$

which our results immediately recover.

Combining [Theorem 3](#) and the above observation about time-homogeneous Markov chains, we also immediately obtain:

$$\int g d\mu \geq \int g d\nu \text{ for all excessive } g \iff \exists \tau \text{ s.t. } x_\tau \sim \nu \text{ with } x_0 \sim \mu, \quad (\text{Rost})$$

which is the well-known *Rost's theorem* ([Rost 1971](#)) in the theory of Skorokhod embeddings for Markov processes. Rost's theorem is extremely useful for identifying the set of inducible posterior distributions from stopping a Markov chain, and has been applied in, e.g., [Morris and Strack \(2019\)](#).

More generally, our results apply to the cases where (i) the Markov chain need not be time-homogeneous and (ii) the choice need not be an optimal stopping problem but a *control problem*. Indeed, suppose in each period, the designer can choose from a feasible set of experiments (e.g., as studied in [Ni, Shen, and Tang 2023](#)), which corresponds to a base set of Markov transition kernels. By the same logic as above, it follows immediately that the problem admits an envelope characterization, and the inducible posterior belief

distributions can be characterized by an integral stochastic order.

4.2 Which Updating Rules Admit Consistent Orders?

In [Section 4.1](#), we show that any Blackwell-consistent Bayes-plausible order on $\Delta(X)$ must be a strengthening of the Blackwell order. In particular, if the updating rule is fixed to be Bayes' rule, then the garbling order is necessary for there to be a Blackwell-consistent value-based order of information. Given this result, a natural question arises: are there other Blackwell-consistent value-based orders of information if the updating rule is allowed to be non-Bayesian? Without Bayesian updating, orders on $\Delta(X)$ need not be Bayes-plausible. We now apply [Theorem 2](#) and [Theorem 3](#) to characterize all possible updating rules that can admit a consistent order $\leq$ on the set of posterior belief distributions (which need not be Bayes-plausible). It turns out that if we want to compare all experiments that are comparable by Blackwell's garbling criterion, then essentially the only possible updating rules are homeomorphic to Bayes' rule. To begin with, we first introduce a general framework for non-Bayesian updating.

4.2.1 A General Framework for Non-Bayesian Updating

Let Ω be a finite set of states, and let $X := \Delta(\Omega)$ be the set of beliefs over Ω . [Cripps \(2018\)](#) develops a general theory for *distorted updating rules* which distort Bayesian posterior beliefs. A (*prior-dependent*) *systematic distortion* of belief updates is a continuous function $d : X \times X \rightarrow X$, where $d(x, y) \in X$ denotes the distorted posterior belief when the prior is x and the Bayesian posterior is y . For example, $d^B(x, y) := y$ corresponds to the Bayesian updating rule without any distortion. We impose several regularity conditions on d : (i) d is continuous; (ii) for any $x \in X^0 := \text{int } \Delta(\Omega)$, $z \mapsto d(x, z)$ is injective on X^0 ; (iii) $d(x, x) = x$ for all $x \in X$.¹⁵ Let $\mathcal{D}$ be the set of such updating rules.

For any $d \in \mathcal{D}$, it is convenient to define

$$\hat{d}(x_B, x_D, y) := d\left(x_D, \frac{x_D \oslash x_B \odot y}{\langle x_D \oslash x_B, y \rangle}\right),$$

for all $x_B, x_D, y \in X^0$, where $\oslash$ denotes the component-wise division of two vectors, and $\odot$ denotes the component-wise product of two vectors.¹⁶ The function $\hat{d}(x_B, x_D, y)$ corresponds to the distorted belief where the distorted prior is x_D , and the Bayesian posterior

¹⁵These conditions, together with the convention that distorted updating rules are defined in terms of beliefs, correspond to Axioms 1-3 of [Cripps \(2018\)](#).

¹⁶Since $d(x, y)$ is continuous, $\hat{d}$ can be extended continuously on $X \times X \times X$.

is y , which is updated from a (potentially different) Bayesian prior $x_B \in X$. When the Bayesian prior is different from the distorted prior, the Bayesian posterior updated from the distorted prior is distinct from, but is determined by (up to a likelihood ratio), the Bayesian posterior updated from the Bayesian prior x_B . The second argument of the right-hand side makes this correction. It follows immediately from the definition that

$$\hat{d}(x_B, x_B, y) = d(x_B, y), \quad \text{and} \quad \hat{d}(x_B, x_D, x_B) = x_D.$$

That is, when the Bayesian prior x is the same as the distorted prior, the distorted posterior given the Bayesian posterior is exactly given by $d(x, y)$. Moreover, when the Bayesian posterior equals the Bayesian prior x_B (i.e., there is no Bayesian updating), the distorted posterior must be the same as the distorted prior x_D (i.e., there is no distorted updating either).

While an updating rule $d \in \mathcal{D}$ describes only how beliefs will be updated (relative to Bayesian updating) from a single signal, it implies a natural description of how beliefs will be updated from multiple signals that arrive sequentially. To see this, for any updating rule $d \in \mathcal{D}$, let $d_{\text{II}} : X \times X \times X \rightarrow X$ be defined by

$$d_{\text{II}}(x, z; y) := \hat{d}(y, d(x, y), z) = d\left(d(x, y), \frac{d(x, y) \odot y \odot z}{\langle d(x, y) \odot y, z \rangle}\right),$$

for all $x, y, z \in X^0$.¹⁷ Under this definition, $d_{\text{II}}(x, z; y)$ describes the terminal distorted posterior belief under a two-stage updating process, where one signal is drawn from a Blackwell experiment first, and then a second signal is drawn from another Blackwell experiment (which could potentially depend on the realization of the first signal). In this process, the prior is x , the first-stage Bayesian posterior belief is y (and hence the first-stage interim distorted posterior belief is $d(x, y)$), and the second-stage Bayesian posterior belief is z . Clearly, if the updating rule is Bayesian, then sequentially updating must lead to the same outcome as if updating is done at once, and the terminal posterior belief would not depend on the realization of the interim belief y , conditional on z itself. That is,

$$d_{\text{II}}^{\text{B}}(x, z; y) = d^{\text{B}}(x, z) = z,$$

for all $x, y, z \in X^0$. We refer to updating rules that satisfy the same property as being **divisible**. That is, we say that $d \in \mathcal{D}$ is divisible if $d(x, \cdot)$ is bijective for all $x \in X$, and

$$d_{\text{II}}(x, z; y) = d(x, z), \quad \text{(Divisibility)}$$

¹⁷Since $d(x, y)$ is continuous, d_{II} can be extended continuously on $X \times X \times X$.

for all $x, y, z \in X^0$. Divisible updating rules have a compact characterization, due to [Cripps \(2018\)](#). Namely, for $d \in \mathcal{D}$, d is divisible if and only if there exists a homeomorphism $F : X \rightarrow X$ such that

$$d(x, y) = F^{-1} \left(\frac{F(x) \otimes x \odot y}{\langle F(x) \otimes x, y \rangle} \right)$$

for all $x, y \in X^0$. When F is the identity function, $d(x, y)$ is the Bayesian updating rule. In a sense, a divisible updating rule d is *essentially* Bayesian: it distorts the prior via a homeomorphism F , updates via Bayes' rule, and distorts the posterior back again via F^{-1} .

4.2.2 Consistent Comparisons under Non-Bayesian Updating

Let $\mathcal{P}_M$ be the set of all martingale kernels. For any $P \in \mathcal{P}_M$ and for any $x \in X$, we may interpret $P(\cdot | x)$ as the distribution over (Bayesian) posterior beliefs induced by a Blackwell experiment when the (Bayesian) prior is x . As noted before, the garbling order of Blackwell experiments can be represented by the order $\leq^{\mathcal{P}_M}$ on $\Delta(X)$. Therefore, for there to be a Blackwell-consistent, value-based order on $\Delta(X)$, there has to be a class of problems under which the expected value under any $P(\cdot | x)$ is higher than the value at x , for all priors x . When $d \in \mathcal{D}$ is the Bayesian updating rule, by Blackwell's theorem, the class of problems are exactly decision problems (i.e., convex functions of posterior beliefs). When the updating rule $d \in \mathcal{D}$ is non-Bayesian, however, this is generally not the case.

To evaluate the value of information under a potentially non-Bayesian updating rule $d \in \mathcal{D}$, a natural perspective is to evaluate the expected value of a problem *paternalistically*, as in [Whitmeyer \(forthcoming\)](#). From the paternalistic perspective, a decision is made under the distorted belief, whereas the expected value is computed using the Bayesian belief. Therefore, a problem would be described by a function $g : X \times X \rightarrow \mathbb{R}$, where $g(x_B, x_D)$ denotes the indirect value of a problem when the Bayesian belief, which is used to evaluate the expected value, is x_B , and when the distorted belief, which is used to determine the action, is x_D . For example, if the problem is a decision problem, with action set A and payoff function u , then

$$g(x_B, x_D) = \mathbb{E}_{\omega \sim x_B} [u(\omega, a^*(x_D))]$$

is the indirect utility, where $a^*(x_D)$ is the optimal action under belief x_D . However, the problem need not be a decision problem; for example, it can be a non-Bayesian persuasion problem where $a^*(x_D)$ is the optimal action taken under the distorted belief x_D and an objective $\tilde{u}$ different from u , with the payoff evaluated under the Bayesian posterior x_B

and the sender's objective u .

For the garbling order to have a Blackwell-consistent, value-based order, the relevant class of problems must be such that

$$\int g(y, \hat{d}(x_B, x_D, y)) P(dy | x_B) \geq g(x_B, x_D).$$

That is, given any pair of Bayesian and distorted priors (x_B, x_D) , the expected paternalistic value of the problem with more information—given by $P(\cdot | x_B)$ —must be higher than the value without any information. Let $\mathcal{C}^d \subseteq B(X \times X)$ be such functions.

Naturally, one might conjecture that the class of problems $\mathcal{C}^d$ dualizes the garbling-based comparison under the updating rule d . However, our next result shows that this is in general not the case. In fact, we show the garbling-based comparison admits a Blackwell-consistent value-based order if and only if d is divisible. As a result, there is generally no instrumental-value-based representation of the garbling order once the updating rule is non-Bayesian, with the only exception being the case when the updating rule is divisible (i.e., being “essentially” Bayesian).

For any (x_B, x_D) , and for any $P \in \mathcal{P}_M$, let Y be the random variable that follows $P(\cdot | x_B)$ and consider the distribution Q of

$$(Y, \hat{d}(x_B, x_D, Y)).$$

If we interpret x_B as the Bayesian prior, x_D as a distorted prior, and $P(\cdot | x_B)$ as a Blackwell experiment (identified by the distribution of *Bayesian* posterior beliefs), then $Q(\cdot | x_B, x_D)$ is the joint distribution of the Bayesian and d -distorted posterior beliefs under the same experiment. Let $\mathcal{Q}_0$ be the set of transition kernels on $X \times X$. Let

$$\mathcal{Q}_{(x_B, x_D)}^d := \left\{ \text{Law}(Y, \hat{d}(x_B, x_D, Y)) : Y \text{ is a random variable with } \mathbb{E}[Y] = x_B \right\}.$$

Let

$$\mathcal{Q}^d := \left\{ Q \in \mathcal{Q}_0 : Q(\cdot | x_B, x_D) \in \mathcal{Q}_{(x_B, x_D)}^d \text{ for all } x_B, x_D \right\}$$

denote the set of all distributions of posterior beliefs under the distorted updating rule d . Note that $\mathcal{Q}^d$ is regular. Moreover, $\mathcal{Q}^d$ defines a binary relation on $\Delta(X \times X)$, where $\mu \leq^{\mathcal{Q}^d} \nu$ if and only if $\nu = Q * \mu$ for some $Q \in \mathcal{Q}^d$. By the definition of $\mathcal{Q}^d$, the comparison given by $\leq^{\mathcal{Q}^d}$ encodes the garbling order of Blackwell experiments: for any (x_B, x_D) , by construction, $\delta_{(x_B, x_D)} \leq^{\mathcal{Q}^d} \nu$ if and only if ν is the joint distribution of the Bayesian posterior and the distorted posterior under the updating rule d for *some* experiment P .

Our next result characterizes updating rules that admit consistent comparisons in the Blackwell sense by applying [Theorem 2](#) to the space $X \times X$.

Proposition 5 (Consistent Comparisons under Non-Bayesian Updating). *The garbling-based comparison $\leq^{\mathcal{Q}^d}$ is a Blackwell-consistent order if and only if d is divisible.*

Naturally, one might conjecture that the garbling comparison always has a dual representation in terms of its instrumental value just like in Blackwell’s theorem under Bayes’ rule. However, [Proposition 5](#) shows that this is *false*. According to [Proposition 5](#), when the updating rule d is not divisible, there cannot be a value-based order that represents the garbling-based comparison $\leq^{\mathcal{Q}^d}$. In particular, there exist distributions of posterior beliefs $\mu, \nu \in \Delta(X \times X)$ such that $\nu = Q * \mu$ and the instrumental value for any problem $g \in \mathcal{C}^d$ is increased:

$$\int g(x_B, x_D) \mu(dx_B, dx_D) \leq \int g(x_B, x_D) \nu(dx_B, dx_D),$$

and yet $Q \notin \mathcal{Q}^d$, i.e., Q cannot be represented via a garbling of an experiment under the updating rule d . Moreover, [Proposition 5](#) shows that when d is divisible, $\leq^{\mathcal{Q}^d}$ is a Blackwell-consistent order—equivalent to the value-based order $\leq_{\mathcal{C}^d}$ —and thus Blackwell’s theorem is preserved in this regard.

Given that the garbling-based comparison only has a dual representation under divisible updating rules, which are homeomorphic to Bayes’ rule, one might naturally further conjecture that the class of problems $\mathcal{C}^d$, which dualizes the garbling order under a divisible updating rule d , is the class of decision problems just like in Blackwell’s theorem. However, this is again *false*. Indeed, by the characterization of [Whitmeyer \(forthcoming\)](#), any non-Bayesian divisible updating rule is *not* Blackwell-monotone, and hence there exists a decision problem under which a more informative experiment according to the garbling order leads to a lower paternalistic payoff. Nonetheless, by [Proposition 5](#), a divisible updating rule is Blackwell-consistent, and therefore such a decision problem cannot appear in the class of problems that dualizes the garbling order under the divisible updating rule. In other words, combining [Whitmeyer \(forthcoming\)](#) and [Proposition 5](#), we then know that the class of problems that dualizes the garbling order under a divisible, non-Bayesian updating rule is *not* the class of all decision problems. It necessarily excludes *some* decision problem and may include *some* persuasion problem or costly information acquisition problem—and yet, it dualizes the garbling order for a divisible updating rule according to [Proposition 5](#).

4.2.3 Dynamic Values and Optimization under Non-Bayesian Updating

Similar to [Section 4.1](#), by virtue of [Theorem 1](#), our results have immediate implications for optimization problems. In particular, we now show how [Proposition 5](#) yields new insights into information design problems with non-Bayesian updating (see, e.g., [Alonso and Câmara 2016](#); [De Clippel and Zhang 2022](#); [Azrieli and Das 2025](#); [Kobayashi 2025](#)).

In information design problems with a non-Bayesian receiver, the sender's problem is equivalent to choosing a martingale $P \in \mathcal{P}_M$ to maximize

$$\int f(y, \hat{d}(x_B, x_D, y)) P(dy \mid x_B)$$

for a fixed (potentially heterogeneous) prior (x_B, x_D) , where $f : X \times X \rightarrow \mathbb{R}$ is the (Bayesian) sender's indirect utility that depends on the sender's Bayesian belief and the receiver's d -distorted belief. The optimal value of such a persuasion problem, as shown by [Alonso and Câmara \(2016\)](#) and [De Clippel and Zhang \(2022\)](#), can be geometrically characterized by the concave envelope of the function $y \mapsto f(y, \hat{d}(x_B, x_D, y))$ evaluated at the prior (x_B, x_D) . However, a crucial implication of [Proposition 5](#) is that, when the updating rule is not divisible, such envelopes do not correspond to the functions that dualize the garbling order. In particular, the envelope is prior-dependent. Indeed, by [Theorem 3](#), the sender may improve their payoff by sending signals *sequentially*, and static/one-shot persuasion would not be optimal in general.

To formalize this, for any $d \in \mathcal{D}$ and $x_B, x_D \in X$, consider the problem

$$\sup_{Q \in \mathcal{Q}^d \otimes \mathcal{Q}^d} \int f(y_B, y_D) Q(dy_B, dy_D \mid x_B, x_D). \quad (\text{Dynamic})$$

In particular, [\(Dynamic\)](#) could be interpreted as a dynamic information design problem with a non-Bayesian receiver, potentially with a different prior than the Bayesian sender, in which case $f(y_B, y_D) = v(y_B, a^*(y_D))$ is the sender's payoff as a function of the sender's own Bayesian prior and the receiver's optimal action taken in the second period given the distorted posterior belief. When information is costless and the sender only cares about the receiver's action, $v(y_B, a) = \mathbb{E}_{\omega \sim x_B}[u(\omega, a)]$ would be affine in y_B ; however, when information is costly to the sender as in [Gentzkow and Kamenica \(2014\)](#), or when the sender cares about the information intrinsically for their own action, $v(y_B, a)$ would not be affine in y_B . More generally, our formulation captures all information design problems in which the sender's payoff depends on the Bayesian and distorted posterior beliefs.

When d is the Bayesian updating rule, in our setting, it is well-known that dynamic

information design does not have any additional value compared to one-shot information design. Without Bayesian updating, there could be a benefit for dynamic information design. We say that $d \in \mathcal{D}$ admits *no dynamic value* if for all $x_B, x_D \in X$, and for any upper semicontinuous $f : X \times X \rightarrow \mathbb{R}$,

$$\sup_{Q \in \mathcal{Q}^d \otimes \mathcal{Q}^d} \int f(y_B, y_D) Q(dy_B, dy_D \mid x_B, x_D) = \sup_{Q \in \mathcal{Q}^d} \int f(y_B, y_D) Q(dy_B, dy_D \mid x_B, x_D).$$

Corollary 6. *An updating rule $d \in \mathcal{D}$ admits no dynamic value if and only if d is divisible.*

The sufficiency part of [Corollary 6](#) immediately reproduces the result of [Kobayashi \(2025\)](#), and extends it to *all* information design problems, including persuasion problems where the sender's payoff is state-dependent. Moreover, [Corollary 6](#) also answers the question raised in [Azrieli and Das \(2025\)](#) and [Kobayashi \(2025\)](#) by proving the converse—whenever the updating rule is not divisible, there exists *some* information design problem where the sender can strictly benefit from sending signals sequentially.

5 Nested Optimization and Stackelberg Principals

We now describe in more detail the nested optimization problem and give a few examples that arise in information design, decision theory, and mechanism design.

5.1 General Framework

There are two principals A and B . Principal A is the *Stackelberg leader* who chooses a measure $\mu \in M \subseteq \Delta(X)$, where M is a compact convex set. Principal B is the *Stackelberg follower* who chooses a measure $\nu \leq_C \mu$ after principal A has moved, where $\leq_C$ is an integral stochastic order. Principal B solves a stochastic optimization problem as in [Section 3](#):

$$\max_{\nu \leq_C \mu} \int f(x) \nu(dx).$$

Recall that $X_f^*(\mu)$ is the solution correspondence of this problem.

Principal A solves a nested optimization problem anticipating the effect on the follower. We assume that principal B breaks ties in favor of principal A so that principal A 's problem always has a solution. Formally, principal A cares about the pair of measures (μ, ν) selected at the end of the game, and thus solves the nested optimization as in

Section 3:

$$\max_{(\mu, \nu) \in G_f^M} W(\mu, \nu),$$

where W is some continuous quasi-convex functional describing the principal's preferences and G_f^M is the graph of the solution correspondence of principal B , i.e.,

$$G_f^M = \{(\mu, \nu) : \mu \in M, \nu \in X_f^*(\mu)\} \subseteq \Delta(X) \times \Delta(X).$$

Formally, our solution concept is **leader-optimal subgame perfect equilibria** (also known as strong Stackelberg equilibria).

By [Corollary 3](#), which is an immediate consequence of [Theorem 1](#), we know that whenever $\mathcal{C}$ satisfies the min-closure property, there must exist an equilibrium in which (i) principal A selects $\mu \in \text{ext}(M)$ and then (ii) principal B selects $\nu \in \text{ext}(X_f^*(\mu))$.

Moreover, in the case where $W(\mu, \nu)$ is a linear functional and hence can be written as $\int w_A(x)\mu(dx) + \int w_B(x)\nu(dx)$, we can explicitly construct all equilibria by invoking [Theorem 1](#). To see this, let

$$w_B^*(x) := \max_{\eta \in X_f^*(\delta_x)} \int w_B(y)\eta(dy) \quad \text{and} \quad \mathcal{P}_B^*(x) := \text{argmax}_{\eta \in X_f^*(\delta_x)} \int w_B(y)\eta(dy).$$

Then, we immediately have:

Corollary 7. *Suppose $\mathcal{C}$ is min-closed. The leader value of any leader-optimal equilibrium is*

$$\max_{\mu \in M} \int (w_A(x) + w_B^*(x))\mu(dx). \quad \textbf{(Modified Objective)}$$

Moreover, (μ, ν) is a leader-optimal equilibrium if and only if

- (i) μ maximizes **(Modified Objective)** ;
- (ii) $\nu = P * \mu$ for some P such that $P(\cdot | x) \in \mathcal{P}_B^*(x)$ for μ -a.e. x .

Moreover, if the convex set M for principal A is also described by a stochastic orbit of the form $\{\mu : \mu \leq_{\mathcal{C}_A} \gamma\}$ where $\mathcal{C}_A$ is min-closed and γ is an exogenous measure, then by [Theorem 1](#) again, the leader value is simply given by

$$\int \overline{w^*}(x)\gamma(dx)$$

where $w^* = w_A + w_B^*$ and the envelope above is taken with respect to the set $\mathcal{C}_A$ and any leader-optimal strategy μ can only be supported on $\{x : \overline{w^*}(x) = w^*(x)\}$. While the above

setup is described via the leader-optimal selection, any selection rule is compatible with our approach with the only difference being that the definition of w_B^* and $\mathcal{P}_B^*$ will be modified accordingly.

5.2 Economic Applications

We now describe a series of examples where our general results apply.

Sequential Persuasion. Consider the following setting adapted from [Li and Norman \(2021\)](#). There is a finite state space Ω . There are two senders. Sender A designs a Blackwell experiment first, and then after observing A 's choice, sender B designs an additional Blackwell experiment that can depend on the signal realization of A 's experiment. The receiver observes both experiments and both realized signals and takes an action from a finite set. This model corresponds to the case where $X = \Delta(\Omega)$, $\leq_C$ is the concave order, and $M = \{\mu : \mu \leq_C \delta_{x_0}\}$ where x_0 is the common prior about the state. As an immediate consequence of our results ([Corollary 7](#)), there always exists a *sequential extremal equilibrium* in which every sender i selects an extreme point μ_i of $\text{MPS}(\mu_{i-1})$ where μ_{i-1} is the belief distribution selected by the previous sender. In particular, in this equilibrium, along any history, *every* sender sends a signal with support size at most $|\Omega|$ as if the subsequent senders do not exist, unlike in the *one-shot equilibria* characterized in [Li and Norman \(2021\)](#) in which the first sender splits beliefs in the set of “stable beliefs” and the subsequent senders all stay silent (also known as *silent equilibria* in [Wu 2023](#)).

Robust Persuasion. Consider the following setting adapted from [Dworczak and Pavan \(2022\)](#). Again, there is a finite state space Ω and two senders who move sequentially just like in sequential persuasion. However, the second sender, who chooses an additional experiment after observing the realized signal of the first sender, is the *adversarial nature* whose objective is to minimize the payoff of the first sender. Because the second mover has completely opposite preferences to the first mover, the leader-optimal selection rule here has no bite: The set of all SPEs—which correspond to the *worst-case optimal signals* in [Dworczak and Pavan \(2022\)](#)—is exactly the leader-optimal SPEs, characterized by [Corollary 7](#). In particular, the modified objective in [Corollary 7](#) is given by the *lower convex envelope* of the initial sender's indirect utility function V (assuming the receiver has a unique best response at any posterior belief). Since the initial sender's problem has the constraint set $M = \{\mu : \mu \leq_C \delta_{x_0}\}$ also given by a stochastic orbit, the leader value of this game must be given by the concavification of the lower convex envelope V_{LCX} , evaluated

at the prior x_0 , which is the *full information payoff* where the initial sender just releases all information. Moreover, a posterior distribution $\mu \in \Delta(X)$ is induced by a worst-case optimal signal if and only if

$$\text{supp}(\mu) \subseteq \left\{ x : V_{\text{lcx}}(x) = \widehat{V_{\text{lcx}}}^{\text{cav}}(x) = V_{\text{full}}(x) \right\} =: Y,$$

where $V_{\text{full}}(x)$ is the affine function describing the payoff of the initial sender given prior x and release of full information. Dworczak and Pavan (2022) consider a refinement of worst-case optimal signals, the *robust solutions*, which maximize the initial sender's indirect utility function among the worst-case optimal signals and therefore corresponds to solving a standard persuasion problem with a support restriction that $\text{supp}(\mu) \subseteq Y$. For any non-affine V_{lcx} , note that any interior belief $x \in \text{int}(\Delta(\Omega))$ cannot be in Y because then by construction it implies that V_{lcx} is an affine function. Therefore, as observed in Dworczak and Pavan (2022), any worst-case optimal signal must induce *separation* of states where the support of any worst-case optimal μ must rule out some states. In fact, for any subset $B \subset \Omega$, the same argument implies that for any $x \in \text{int}(\Delta(B))$ to be in Y , $V_{\text{lcx}}|_{\Delta(B)}$ must be an affine function coinciding with $V_{\text{full}}|_{\Delta(B)}$ which happens if and only if $V|_{\Delta(B)} \geq V_{\text{full}}|_{\Delta(B)}$; conversely, if $V_{\text{lcx}}|_{\Delta(B)}$ coincides with $V_{\text{full}}|_{\Delta(B)}$, then any $x \in \Delta(B)$ is in Y . Together, these characterize a set $\mathcal{F}$ of subsets $B \subset \Omega$ such that $x \in Y$ if and only if $\text{supp}(x) \in \mathcal{F}$, which is the main separation theorem (Theorem 1) of Dworczak and Pavan (2022).

Objective Ambiguity Aversion. Consider the following setting adapted from Olszewski (2007) and Ahn (2008) who study *objective ambiguity aversion* of a decision maker by interpreting ambiguity as a menu of lotteries. The set of outcomes X is a compact set in $\mathbb{R}^n$. A *lottery* $\mu \in \Delta(X)$ specifies a distribution over outcomes. A *menu* A is a set of lotteries. Olszewski (2007) and Ahn (2008) characterize preferences over menus of lotteries and formalize the notion of objective ambiguity aversion. Specifically, under a standard set of axioms, Olszewski (2007) shows that the representation of objective ambiguity-averse preference takes the following form:

$$\mathcal{V}(A) := \alpha \cdot \max_{\mu \in A} \int u(x) \mu(dx) + (1 - \alpha) \cdot \min_{\mu \in A} \int u(x) \mu(dx),$$

for some function u , and some constant $\alpha \in [0, 1]$. In other words, the objective ambiguity-aversion representation takes the Hurwicz max-min form: the decision maker evaluates menus by computing the “best-case” payoff and the “worst-case” expected utility (evalu-

ated by u) from lotteries in that menu, and takes a convex combination of the best and the worst case payoff. For any u and any $\alpha \in [0, 1]$, the representation $\mathcal{V}$ defines a preference $\leq$ by

$$A \leq B \iff \mathcal{V}(A) \leq \mathcal{V}(B),$$

for all menus A and B .

Suppose now that the menus are parametrized by a single lottery via a stochastic order. That is, each feasible menu A takes the form of $A = \{v : v \leq_{\mathcal{C}} \mu\} =: A_{\mathcal{C}}(\mu)$, for some convex cone $\mathcal{C}$ of test functions that contains $-1, 1$ and can be approximated by continuous functions from above. As menus are parametrized by single lotteries, both the ambiguity-aversion preference $\leq$ and its representation $\mathcal{V}$ immediately imply a preference $\tilde{\leq}$ and a representation $\tilde{\mathcal{V}}$ over lotteries, via

$$\mu_1 \tilde{\leq} \mu_2 \iff A_{\mathcal{C}}(\mu_1) \leq A_{\mathcal{C}}(\mu_2)$$

and

$$\tilde{\mathcal{V}}(\mu) := \mathcal{V}(A_{\mathcal{C}}(\mu)).$$

Two natural questions then arise: (i) when can the representation $\tilde{\mathcal{V}}$ be distinguished from a vNM *expected utility representation* over lotteries (i.e., when is it the case that $\tilde{\mathcal{V}}(\mu) = \int \hat{u} d\mu$ for some $\hat{u} : X \rightarrow \mathbb{R}$)? (ii) When can the preference $\tilde{\leq}$ be distinguished from an *expected utility preference* (i.e., when does the preference $\tilde{\leq}$ violate the vNM expected utility axioms)?

Applying our results, we can answer these questions. Indeed, under this parameterization of menus, note that the representation of an ambiguity-averse decision maker is exactly a Stackelberg leader who chooses a lottery μ anticipating the nature as a follower to choose $v \leq_{\mathcal{C}} \mu$ —with probability α , the nature wants to maximize the DM's payoff (an optimistic conjecture) and with probability $1 - \alpha$, the nature wants to minimize the DM's payoff (a pessimistic conjecture).

To state our result, recall that $\mathcal{C}$ is not min-closed if there exist $g_1, g_2 \in \mathcal{C}$ such that $\min\{g_1, g_2\} \notin \mathcal{C}$. By [Theorem 1](#), for any continuous $f : X \rightarrow \mathbb{R}$, and for any $\mu \in \Delta(X)$, the dual minimization problem

$$\min_{g \in \mathcal{C}, g \geq f} \int g d\mu \tag{Dual}$$

has a unique solution (up to a μ -measure zero set) whenever $\mathcal{C}$ is min-closed (and the solution equals the $\mathcal{C}$ -envelope of f). Motivated by this, we say that $\mathcal{C}$ is **strongly non-min-closed** if there exist a continuous function $f : X \rightarrow \mathbb{R}$ and a measure $\mu_0 \in \Delta(X)$ that is

absolutely continuous with respect to the Lebesgue measure and has a density bounded away from zero, such that the dual minimization problem **(Dual)** has two distinct continuous solutions g_1, g_2 that are not affinely related. Note that $\mathcal{C}$ is not min-closed whenever it is strongly non-min-closed.¹⁸ As an example, the set of convex functions (which contains all affine functions, and is not min-closed) is in fact strongly non-min-closed.¹⁹

Proposition 6. *Consider stochastic-orbit ambiguity sets defined by $\leq_{\mathcal{C}}$. Then:*

- (i) *There exists an ambiguity-aversion representation $\tilde{V}$ that is not an expected utility representation if and only if $\mathcal{C}$ is not min-closed.*
- (ii) *There exists an ambiguity-aversion preference $\tilde{\succsim}$ that is not an expected utility preference if $\mathcal{C}$ is strongly non-min-closed.*

Consequently, whenever $\mathcal{C}$ is min-closed, any ambiguity-averse preference, when translated to the space of lotteries, can in fact be represented by an expected utility *as if* it were an expected utility preference. For example, when the ambiguity menus are given by mean-preserving spread orbits, every ambiguity-aversion preference over these menus has an expected utility representation. Conversely, an outside observer who observes all binary choices over lotteries can distinguish ambiguity-aversion preferences from expected utility preferences whenever $\mathcal{C}$ is strongly non-min-closed. For instance, when the ambiguity menus are given by the mean-preserving contraction orbits, an outside observer who observes all binary choices over lotteries can distinguish an ambiguity-averse DM from an expected-utility DM who faces no ambiguity.

Property Right Design. Consider the following setting adapted from Dworczak and Muir (2024). There is a *seller* and a *buyer* who contract on the sale of a single good. Let $x \in [0, 1]$ denote the allocation probabilities and $t \in \mathbb{R}_+$ denote the transfers. The buyer has private value $\theta \in [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}_+$ for the good. The seller maximizes some linear objective $\mathbb{E}[V_S(\theta)x(\theta) + \lambda_S t(\theta)]$, where $\lambda_S \geq 0$, by posting a menu $M_S = \{(x, t)\}$ of lotteries and prices for the buyer to choose. However, before the seller moves, a *designer* first designs the property right by posting an initial menu $M_D = \{(x, t)\}$ that the buyer

¹⁸Indeed, suppose that $\mathcal{C}$ is min-closed. Then $\min\{g_1, g_2\} \in \mathcal{C}$ must also solve the dual problem, which implies that $\int g_1 d\mu_0 = \int g_2 d\mu_0 = \int \min\{g_1, g_2\} d\mu_0$, and hence $g_1 = g_2$ μ_0 -a.e., a contradiction.

¹⁹For example, consider any convex X with affine dimension at least 2. Let x_0 be the barycenter of the Lebesgue measure on X . Take two non-affinely-related affine functions l_1, l_2 that cross at x_0 : $l_1(x_0) = l_2(x_0)$. Let $f := \min\{l_1, l_2\}$, and let μ_0 be the Lebesgue measure. By Jensen's inequality, every $g \in \mathcal{C}$ with $g \geq f$ yields $\int g d\mu_0 \geq g(x_0) \geq f(x_0)$, and $\int l_1 d\mu_0 = \int l_2 d\mu_0 = l_1(x_0) = l_2(x_0) = f(x_0)$. Thus, both l_1 and l_2 are solutions of the dual problem. One can also verify that the set of convex functions is strongly non-min-closed even if X has affine dimension 1.

treats as *outside options* and hence can select if declining to contract with the seller. The designer maximizes a different linear objective $\mathbb{E}[V_D(\theta)x(\theta) + \lambda_D t(\theta)]$, where $\lambda_D \geq 0$.

Without loss of generality, we may assume the seller always induces the buyer to contract with them. Thus, the seller's problem is equivalent to a mechanism design problem with a *type-dependent participation constraint* $U(\theta) \geq R(\theta)$, where $U(\theta)$ is the indirect utility function of the buyer under the seller's menu and $R(\theta)$ is the indirect utility function of the buyer under the designer's menu. Equivalently, as in Dworczak and Muir (2024), the seller maximizes

$$\max_{x, \underline{u} \geq 0} \int_{\underline{\theta}}^{\bar{\theta}} W_S(\theta) x(\theta) d\theta - \alpha \underline{u}$$

subject to x being non-decreasing, and

$$\underline{u} + \int_{\underline{\theta}}^{\theta} x(s) ds \geq \underline{u}_0 + \int_{\underline{\theta}}^{\theta} x_0(s) ds \text{ for all } \theta,$$

where $W_S(\theta)$ is an appropriate virtual value function for the seller, $\underline{u}_0 = R(\underline{\theta})$ and $x_0(\theta) = R'(\theta)$. Anticipating the seller solving this problem, the designer chooses $(\underline{u}_0, x_0)$ in the first stage where x_0 is non-decreasing and $\underline{u}_0 \geq 0$. Theorem 1 of Dworczak and Muir (2024) shows that despite the designer not directly controlling the allocation rule and having a different objective from the seller, the designer's optimal property right menu is very simple: $M_D = \{(1, p)\}$ interpreted as an *option-to-own* property right where the buyer can always own the good at a price of p . Dworczak and Muir (2024) proves this result by first developing a *generalized ironing procedure* that solves the seller's problem subject to the type-dependent IR constraints, and then uncovering a surprising feature that the seller's optimal mechanism depends on the outside-option function $R(\theta)$ only through a *linear transformation* via the ironing procedure. Thus, there exists a designer's solution that is an extreme point of the feasible $R(\theta)$ functions, corresponding to a menu $\{(1, p)\}$.

We now describe how to understand this result via our framework and the trapezoid property. Since we assume the transfers are non-negative, as is standard in this environment, any IC mechanism can be equivalently represented as a *randomized posted price* (with the price potentially lower than $\underline{\theta}$).²⁰ Thus, consider the lotteries of the posted prices $\mu, \nu \in \Delta([0, \bar{\theta}])$ chosen by the designer and the seller, respectively. In this language,

²⁰Indeed, any such mechanism can be extended to be an IC mechanism on $[0, \bar{\theta}]$ where type 0 gets 0 payoff given non-negative transfers. Thus, there exists a randomized posted price to replicate the original mechanism.

the type-dependent IR constraint is exactly that

$$\int \max\{\theta - p, 0\} \nu(dp) \geq \int \max\{\theta - p, 0\} \mu(dp) \text{ for all } \theta \in [0, \bar{\theta}] ,$$

which is exactly that

$$\nu \succeq_{\text{decreasing convex}} \mu$$

or equivalently

$$\nu \preceq_{\text{increasing concave}} \mu.$$

Let $\mathcal{C}$ be the set of **increasing concave** functions on $[0, \bar{\theta}]$. Both the seller and the designer's objectives are linear functionals of ν . Therefore, the seller solves the problem

$$\max_{\nu: \nu \preceq_{\mathcal{C}} \mu} \int f(p) \nu(dp),$$

for some function f . The feasible set of μ here is unrestricted, i.e., $M = \Delta([0, \bar{\theta}])$. The designer solves the problem

$$\max_{(\mu, \nu) \in G_f^M} \int w(p) \nu(dp),$$

for some function w . Since the set of increasing concave functions is min-closed, by [Corollary 3](#), we immediately obtain that there exists (μ^*, ν^*) solving the nested optimization such that $\mu^* \in \text{ext}(M)$, i.e., $\mu^* = \delta_p$ for some p , and $\nu^* \in \text{ext}(\{\nu : \nu \preceq_{\mathcal{C}} \mu^*\})$ which takes the form of $\nu^* = \alpha \delta_{p_1} + (1 - \alpha) \delta_{p_2}$ where $p_1 \leq p_2$. Therefore, the designer's optimal menu takes the form $\{(1, p)\}$, i.e., an option-to-own design, and then subsequently the seller's optimal menu takes the form $\{(\alpha, \alpha p_1), (1, \alpha p_1 + (1 - \alpha)p_2)\}$.

6 Conclusion

We study optimization problems in which a linear functional is maximized over probability measures that are dominated by a given measure according to an integral stochastic order in an arbitrary dimension. We show that the following four properties are equivalent for any such order: (i) the test function cone is closed under pointwise minimum, (ii) the value function is affine, (iii) the solution correspondence has a convex graph with decomposable extreme points, and (iv) every ordered pair of measures admits an order-preserving coupling. As corollaries, we derive the extreme and exposed point properties involving integral stochastic orders such as multidimensional mean-preserving spreads and stochastic dominance. Applying these results, we generalize Blackwell's theorem

by completely characterizing the comparisons of experiments that admit two equivalent descriptions—through instrumental values and through information technologies. We also show that these results immediately yield new insights into information design, mechanism design, and decision theory.

A Omitted Proofs

A.1 Technical Lemmas

A.1.1 Technical Lemmas for [Section 3](#)

Lemma 1. *There exists a countable set of continuous functions $\hat{\mathcal{C}} \subseteq \mathcal{C} \cap C(X)$ such that for all $g \in \mathcal{C}$, there exists a sequence $\{g_n\} \subseteq \hat{\mathcal{C}}$ such that $g_n \geq g$ for all n , and $g_n \rightarrow g$ pointwise.*

Proof. Since X is a compact metric space, $C(X) \cap \mathcal{C}$ is separable under the sup-norm. Take a countable dense subset $\tilde{\mathcal{C}} = \{h_m\}_{m=1}^\infty$ of $\mathcal{C} \cap C(X)$. Let

$$\hat{\mathcal{C}} := \{h_m + q : m \in \mathbb{N}, q \in \mathbb{Q}_+\}.$$

Since $\mathcal{C}$ is a convex cone that contains all constants, and since the sum of a continuous function and a constant is continuous, it follows that $\hat{\mathcal{C}} \subseteq \mathcal{C} \cap C(X)$. Clearly, $\hat{\mathcal{C}}$ is also countable.

We first claim that for any $h \in \mathcal{C} \cap C(X)$ and for any $\varepsilon > 0$, there exists $\hat{h} \in \hat{\mathcal{C}}$ such that $h < \hat{h} < h + \varepsilon$. Indeed, for any such $h \in \mathcal{C} \cap C(X)$ and ε , take $m \in \mathbb{N}$ so that $\|h_m - h\| < \varepsilon/3$ and $q \in \mathbb{Q}_+ \cap (\varepsilon/3, 2\varepsilon/3)$. Then, for all $x \in X$,

$$h_m(x) + q > h(x) - \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = h(x),$$

and

$$h_m(x) + q < h(x) + \frac{\varepsilon}{3} + \frac{2\varepsilon}{3} = h(x) + \varepsilon,$$

proving the claim.

Now, consider any $g \in \mathcal{C}$. Since g is approximated by continuous functions in $\mathcal{C}$ from above, there exists $\{g_n\} \subseteq \mathcal{C} \cap C(X)$ such that $\{g_n\} \downarrow g$. For each $n \in \mathbb{N}$, take $\hat{h}_n \in \hat{\mathcal{C}}$ such that $g_n < \hat{h}_n < g_n + 2^{-n}$. Together, we have

$$g \leq g_n < \hat{h}_n < g_n + 2^{-n},$$

for all $n \in \mathbb{N}$. Therefore, for all $x \in X$,

$$\limsup_{n \rightarrow \infty} |\hat{h}_n(x) - g(x)| \leq \limsup_{n \rightarrow \infty} (|g_n(x) - g(x)| + 2^{-n}) = 0.$$

Therefore, the sequence $\{\hat{h}_n\} \subseteq \hat{\mathcal{C}}$ satisfies that $\hat{h}_n \geq g$ for all n and $\hat{h}_n \rightarrow g$ pointwise, as desired. $\square$

Lemma 2. Suppose $\mathcal{C}$ is min-closed. Fix any upper semicontinuous function $f : X \rightarrow \mathbb{R}$. Let

$$\bar{f}(x) := \inf \{g(x) : g \in \mathcal{C}, g \geq f\},$$

for all $x \in X$. Then $\bar{f} \in \mathcal{C}$.

Proof. Fix a countable approximating class $\hat{\mathcal{C}} \subseteq \mathcal{C} \cap C(X)$ from [Lemma 1](#). Let

$$\hat{\mathcal{C}}_f := \{h \in \hat{\mathcal{C}} : h \geq f\}$$

which is non-empty by [Lemma 1](#) since f is bounded from above and $\mathcal{C}$ contains all constants and thus there exists some $h \in \hat{\mathcal{C}}$ with $h \geq f$. Enumerate $\hat{\mathcal{C}}_f$ as $\{h_n\}_{n=1}^\infty$ and write

$$k_n := \min\{h_1, \dots, h_n\}.$$

Since $\mathcal{C}$ is min-closed, note that $k_n \in \mathcal{C} \cap C(X)$. Moreover, the sequence $\{k_n\}$ is decreasing with $k_n \geq f$. Fix any $x \in X$. We claim that

$$\lim_{n \rightarrow \infty} k_n(x) = \inf_{h \in \hat{\mathcal{C}}_f} h(x) = \inf_{g \in \mathcal{C}, g \geq f} g(x) = \bar{f}(x).$$

Indeed, the first equality holds by definition, and $\inf_{h \in \hat{\mathcal{C}}_f} h(x) \geq \bar{f}(x)$ follows by that $\hat{\mathcal{C}}_f \subseteq \{g \in \mathcal{C} : g \geq f\}$. Now, to see the reverse inequality, fix any $g \in \mathcal{C}$ with $g \geq f$, and any $\varepsilon > 0$. By [Lemma 1](#), there exists $h \in \hat{\mathcal{C}}$ such that $g(x) \leq h(x) \leq g(x) + \varepsilon$. Since $g \geq f$, we have $h \geq f$ and hence $h \in \hat{\mathcal{C}}_f$. Therefore,

$$\inf_{h \in \hat{\mathcal{C}}_f} h(x) \leq g(x) + \varepsilon.$$

Taking $\varepsilon \downarrow 0$ gives

$$\inf_{h \in \hat{\mathcal{C}}_f} h(x) \leq g(x),$$

for all $g \in \mathcal{C}$ with $g \geq f$. Therefore, we have

$$\inf_{h \in \hat{\mathcal{C}}_f} h(x) \leq \inf_{g \in \mathcal{C}, g \geq f} g(x) = \bar{f}(x),$$

as claimed. Thus, $\{k_n\} \subseteq \mathcal{C} \cap C(X)$ is a bounded decreasing sequence that converges pointwise to $\bar{f}$. Since $\mathcal{C}$ is closed under pointwise convergence from above, $\bar{f} \in \mathcal{C}$. $\square$

Lemma 3. *For any $\mu \in \Delta(X)$, the set*

$$\mathcal{K}_\mu := \left\{ \nu \in \Delta(X) : \nu \leq_{\mathcal{C}} \mu \right\}$$

is closed and hence compact.

Proof. For any $\{\nu_n\} \subseteq \mathcal{K}_\mu$ such that $\{\nu_n\} \rightarrow \nu$ for some $\nu \in \Delta(X)$, since $\nu_n \leq_{\mathcal{C}} \mu$,

$$\int g \, d\nu_n \leq \int g \, d\mu$$

for all $g \in \mathcal{C}$. Therefore, for all $g \in \mathcal{C}$ that is continuous,

$$\int g \, d\nu = \lim_{n \rightarrow \infty} \int g \, d\nu_n \leq \int g \, d\mu.$$

Since for any $g \in \mathcal{C}$, there exists a sequence of continuous functions $\{g_m\} \subseteq \mathcal{C}$ such that $\{g_m\} \downarrow g$. As a result, by the monotone convergence theorem,

$$\int g \, d\nu = \lim_{m \rightarrow \infty} \int g_m \, d\nu \leq \lim_{m \rightarrow \infty} \int g_m \, d\mu = \int g \, d\mu.$$

Therefore, $\nu \in \mathcal{K}_\mu$. That is, $\mathcal{K}_\mu \subseteq \Delta(X)$ is closed, and hence, compact. $\square$

Lemma 4. *Let $X \subset \mathbb{R}^n$ be a full-dimensional compact convex set. Consider any upper semi-continuous concave function $f : X \rightarrow \mathbb{R}$. For each differentiability point x of f , let*

$$h^x(y) := f(x) + \langle \nabla f(x), y - x \rangle$$

be the supporting hyperplane of f at x , and let

$$R_{h^x} := \left\{ y \in X : f(y) = h^x(y) \right\}.$$

Then, for Lebesgue-almost every $x \in X$, we have

$$x \in \text{ri}(R_{h^x}).$$

Proof. Define $g : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ by

$$g(y) := \begin{cases} -f(y), & \text{if } y \in X, \\ +\infty, & \text{otherwise.} \end{cases}$$

Since f is upper semicontinuous and concave on the compact convex set X , the function g is proper, lower semicontinuous, and convex. Its conjugate is therefore

$$g^*(p) = \sup_{y \in \mathbb{R}^n} [\langle p, y \rangle - g(y)] = \sup_{y \in X} [\langle p, y \rangle + f(y)],$$

which is convex, as it is the pointwise supremum of a family of affine functions. It is also finite for every $p \in \mathbb{R}^n$ as X is compact and f is upper semicontinuous.

Let D denote the set of differentiability points of f in the interior of the full-dimensional X . Since f is concave and ∂X has Lebesgue measure zero, D is a full-Lebesgue-measure subset of X . Thus, it suffices to show that

$$x \in \text{ri}(R_{h^x})$$

for a full-Lebesgue-measure subset of D .

To this end, fix any $x \in D$ and let

$$p_x := -\nabla f(x).$$

Since $\nabla f(x)$ is a supergradient of f at x ,

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle$$

for all $y \in X$. Equivalently, we have

$$\langle p_x, y \rangle - g(y) \leq \langle p_x, x \rangle - g(x)$$

for all $y \in \mathbb{R}^n$. Note that the equality holds if and only if $y \in R_{h^x}$ —indeed, if the equality

holds, then we must have $y \in X$, and

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle = h^x(y),$$

and thus $y \in R_{h^x}$; conversely, if $y \in R_{h^x}$, then $y \in X$ and the above also hold. Therefore, we have

$$R_{h^x} = \operatorname{argmax}_{y \in \mathbb{R}^n} [\langle p_x, y \rangle - g(y)].$$

We claim that, for every $p \in \mathbb{R}^n$, we have

$$\operatorname{argmax}_{y \in \mathbb{R}^n} [\langle p, y \rangle - g(y)] = \partial g^*(p),$$

where $\partial g^*(p)$ is the subdifferential of g^* at p . Indeed, for any

$$y \in \operatorname{argmax}_{y' \in \mathbb{R}^n} [\langle p, y' \rangle - g(y')],$$

we have

$$g^*(p) = \langle p, y \rangle - g(y),$$

and therefore for every $q \in \mathbb{R}^n$, by definition of g^* , we have

$$g^*(q) \geq \langle q, y \rangle - g(y) = g^*(p) + \langle y, q - p \rangle,$$

and hence $y \in \partial g^*(p)$. Conversely, if $y \in \partial g^*(p)$, then we have

$$g^*(q) \geq g^*(p) + \langle y, q - p \rangle,$$

for all $q \in \mathbb{R}^n$. Rearranging, we have

$$\langle q, y \rangle - g^*(q) \leq \langle p, y \rangle - g^*(p)$$

for all $q \in \mathbb{R}^n$. Since g is proper, lower semicontinuous, and convex, we have $g = g^{**}$. Now, taking the supremum of the above over q and using $g = g^{**}$, we obtain

$$g(y) = g^{**}(y) \leq \langle p, y \rangle - g^*(p).$$

By Fenchel's inequality, we also have

$$g(y) \geq \langle p, y \rangle - g^*(p).$$

Therefore, we conclude that

$$g^*(p) = \langle p, y \rangle - g(y),$$

and thus y attains the maximum of $\langle p, y' \rangle - g(y')$ over all $y' \in \mathbb{R}^n$, proving the claim.

Applying the previous claim with $p = p_x$ yields

$$\partial g^*(p_x) = \operatorname{argmax}_{y \in \mathbb{R}^n} [\langle p_x, y \rangle - g(y)] = R_{h^x}.$$

Therefore, we have

$$B := \left\{ x \in D : x \notin \operatorname{ri}(R_{h^x}) \right\} = \left\{ x \in D : x \in \operatorname{rb}(R_{h^x}) \right\} = \left\{ x \in D : x \in \operatorname{rb}(\partial g^*(p_x)) \right\},$$

where $\operatorname{rb}(A)$ denotes the relative boundary of a convex set A . Hence, we conclude that

$$B \subseteq \bigcup_{p \in \mathbb{R}^n} \operatorname{rb}(\partial g^*(p)).$$

Since $g^* : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, by Theorem 3.3 of [Drusvyatskiy and Lewis \(2011\)](#), we conclude that

$$\bigcup_{p \in \mathbb{R}^n} \operatorname{rb}(\partial g^*(p))$$

has Lebesgue measure zero. Thus, B has Lebesgue measure zero, as desired. $\square$

A.1.2 Technical Lemmas for [Section 4](#)

Lemma 5. *For any rectangular $\mathcal{P}, \mathcal{P}'$, $\mathcal{P} = \mathcal{P}'$ if and only if $\leq^{\mathcal{P}} = \leq^{\mathcal{P}'}$.*

Proof. The “only if” direction is clear. To prove the “if” direction, fix any $x \in X$. Note that for any $P \in \mathcal{P}_0$, $\nu = P * \delta_x$ if and only if $\nu = P(\cdot \mid x)$. Since $\leq^{\mathcal{P}} = \leq^{\mathcal{P}'}$, we have

$$\{\nu : \nu = P * \delta_x \text{ for some } P \in \mathcal{P}\} = \{\nu : \delta_x \leq^{\mathcal{P}} \nu\} = \{\nu : \delta_x \leq^{\mathcal{P}'} \nu\} = \{\nu : \nu = P * \delta_x \text{ for some } P \in \mathcal{P}'\}.$$

Thus,

$$\mathcal{P}_x = \left\{ P(\cdot \mid x) : P \in \mathcal{P} \right\} = \left\{ P(\cdot \mid x) : P \in \mathcal{P}' \right\} = \mathcal{P}'_x.$$

Since this holds for all x , and $\mathcal{P}, \mathcal{P}'$ are rectangular, it follows that $\mathcal{P} = \mathcal{P}'$, as desired. $\square$

Lemma 6. *The operators Φ and Ψ have the following properties:*

- (a) Φ and Ψ are decreasing in the set-inclusion order.
- (b) For any $\mathcal{C} \subseteq \mathcal{C}_0$ that admits continuous approximations from below, $\Psi(\mathcal{C}) \subseteq \mathcal{P}_0$ is convex, rectangular, composition-closed, and has a closed graph.
- (c) For any $\mathcal{P} \subseteq \mathcal{P}_0$, $\Phi(\mathcal{P})$ is a max-closed convex cone that contains $1, -1$ and is closed under bounded increasing pointwise limits, with $\Phi(\mathcal{P}) \cap C(X)$ being closed under the sup-norm.
- (d) For any $\mathcal{C} \subseteq \mathcal{C}_0$, $\mathcal{C} \subseteq \Phi(\Psi(\mathcal{C}))$
- (e) For any $\mathcal{P} \subseteq \mathcal{P}_0$, $\mathcal{P} \subseteq \Psi(\Phi(\mathcal{P}))$.

Proof. (a): Consider any $\mathcal{C}, \mathcal{C}'$ such that $\mathcal{C} \subseteq \mathcal{C}' \subseteq \mathcal{C}_0$. For any $P \in \Psi(\mathcal{C}')$, we have $g * P \geq g$ for all $g \in \mathcal{C}'$ by definition. Since $\mathcal{C} \subseteq \mathcal{C}'$, we have $g * P \geq g$ for all $g \in \mathcal{C}$ as well. Thus, $P \in \Psi(\mathcal{C})$. Likewise, consider any $\mathcal{P}, \mathcal{P}'$ such that $\mathcal{P} \subseteq \mathcal{P}' \subseteq \mathcal{P}_0$. For any $g \in \Phi(\mathcal{P}')$, we have $g * P \geq g$ for all $P \in \mathcal{P}'$ by definition. Since $\mathcal{P} \subseteq \mathcal{P}'$, we have $g * P \geq g$ for all $P \in \mathcal{P}$ as well. Thus, $g \in \Phi(\mathcal{P})$.

(b): It is easy to see that $\Psi(\mathcal{C})$ is convex and rectangular by definition. To see that the graph of $x \mapsto \Psi(\mathcal{C})_x$ is closed, note that the graph can be written as:

$$\text{Gr} = \bigcap_{g \in \mathcal{C} \cap C(X)} \left\{ (x, \eta) : \int g \, d\eta \geq g(x) \right\}$$

by the assumption that any $g \in \mathcal{C}$ can be approximated from below via $\{g_n\} \subseteq \mathcal{C} \cap C(X)$ and the monotone convergence theorem. Since the graph is an intersection of a collection of closed sets, it is closed. To see that $\Psi(\mathcal{C})$ is composition-closed, take any $P_1, P_2 \in \Psi(\mathcal{C})$ and note that for any $g \in \mathcal{C}$ and for any $x \in X$

$$[g * (P_1 \circ P_2)](x) = \int \left(\int g(z) P_1(dz \mid y) \right) P_2(dy \mid x) \geq \int g(y) P_2(dy \mid x) \geq g(x).$$

Thus, $P_1 \circ P_2 \in \Psi(\mathcal{C})$.

(c): It is easy to see that $\Phi(\mathcal{P})$ is a convex cone and that $1, -1 \in \Phi(\mathcal{P})$ since $P(\cdot \mid x)$ is a probability measure for all $x \in X$. To see that $\Phi(\mathcal{P})$ is max-closed, consider any

$g_1, g_2 \in \Phi(\mathcal{P})$. For any $P \in \mathcal{P}$ and for any $x \in X$,

$$\begin{aligned} [\max\{g_1, g_2\} * P](x) &= \int \max\{g_1(y), g_2(y)\} P(dy | x) \geq \max \left\{ \int g_1(y) P(dy | x), \int g_2(y) P(dy | x) \right\} \\ &\geq \max\{g_1(x), g_2(x)\}. \end{aligned}$$

Thus, $\max\{g_1, g_2\} \in \Phi(\mathcal{P})$.

To see that $\Phi(\mathcal{P}) \cap C(X)$ is closed under the sup-norm, consider any sequence $\{g_n\} \subseteq \Phi(\mathcal{P}) \cap C(X)$ and suppose that $g_n \rightarrow g$ under the sup-norm. Since $C(X)$ is closed under the sup-norm, $g \in C(X)$. Moreover, since $g_n \in \Phi(\mathcal{P})$ for all n , for any $P \in \mathcal{P}$ and for any $x \in X$, we have

$$\begin{aligned} 0 &\geq \limsup_{n \rightarrow \infty} \left[g_n(x) - \int g_n(y) P(dy | x) \right] \\ &= g(x) - \int g(y) P(dy | x), \end{aligned}$$

where the equality follows from $\{g_n\} \rightarrow g$ in the sup-norm, and hence pointwise, as well as the dominated convergence theorem.

Lastly, to see that $\Phi(\mathcal{P})$ is closed under bounded increasing pointwise limits, consider any $\{g_n\} \subseteq \Phi(\mathcal{P})$ such that $\{g_n\} \uparrow g$ for some g . Then, since g is the limit of an increasing sequence of lower semicontinuous functions, g is lower semicontinuous. Combining the previous argument and the monotone convergence theorem, we have $g \in \Phi(\mathcal{P})$.

(d): Consider any $\mathcal{C} \subseteq \mathcal{C}_0$ and any $g \in \mathcal{C}$. For any $P \in \Psi(\mathcal{C})$, by definition, $g * P \geq g$. Therefore, $g \in \Phi(\Psi(\mathcal{C}))$.

(e): Consider any $\mathcal{P} \subseteq \mathcal{P}_0$ and any $P \in \mathcal{P}$. For any $g \in \Phi(\mathcal{P})$, by definition $g * P \geq g$. Therefore, $P \in \Psi(\Phi(\mathcal{P}))$. $\square$

Lemma 7. For any regular $\mathcal{P}$ and for any $\mu \in \Delta(X)$, let

$$M(\mu) := \{P * \mu : P \in \mathcal{P}\}.$$

Then, $M(\mu)$ is a compact and convex subset of $\Delta(X)$.

Proof. Clearly, $M(\mu)$ is convex since $\mathcal{P}$ is regular, and in particular, convex. To see that $M(\mu)$ is compact, it suffices to show that $M(\mu)$ is closed, since $M(\mu) \subseteq \Delta(X)$ and $\Delta(X)$ is compact. To this end, consider any sequence $\{\nu_n\} \subseteq M(\mu)$ such that $\{\nu_n\} \rightarrow \nu$ for some

$\nu \in \Delta(X)$. We show that $\nu = P * \mu$ for some $P \in \mathcal{P}$. To begin with, for each $n \in \mathbb{N}$, since $\nu_n \in M(\mu)$, there exists $P_n \in \mathcal{P}$ such that $\nu_n = P_n * \mu$.

Since $C(X)$ is separable, there exists a countable set $\{h_m\} \subseteq C(X)$ with $\|h_m\| \leq 1$ that is dense in the unit ball $\{h \in C(X) : \|h\| \leq 1\}$. For each $m \in \mathbb{N}$ and $n \in \mathbb{N}$, let

$$g_{n,m}(x) := \int_X h_m(y) P_n(dy | x),$$

for all $x \in X$. Since $\|h_m\| \leq 1$ and since $P_n(\cdot | x) \in \Delta(X)$, we have that $|g_{n,m}(x)| \leq 1$ for all $n, m \in \mathbb{N}$ and for all $x \in X$. Note that, by definition,

$$\int_X g_{n,m} d\mu = \int_X \left(\int_X h_m(y) P_n(dy | x) \right) \mu(dx) = \int_X h_m dP_n * \mu = \int_X h_m d\nu_n,$$

for all $n, m \in \mathbb{N}$, and hence, since $\{\nu_n\} \rightarrow \nu$,

$$\lim_{n \rightarrow \infty} \int_X g_{n,m} d\mu = \lim_{n \rightarrow \infty} \int_X h_m d\nu_n = \int_X h_m d\nu, \quad (1)$$

for all $m \in \mathbb{N}$.

Since $g_{n,m}$ has bounded $L^1(\mu)$ -norm for all $n, m \in \mathbb{N}$, by Komlós' theorem, there exists a subsequence $\{g_{n_j^{(1)},1}\} \subseteq \{g_{n,1}\}$ such that the Cesàro sum

$$\frac{1}{k} \sum_{j=1}^k g_{n_j^{(1)},1}$$

converges as $k \rightarrow \infty$, μ -almost surely. Then, for this subsequence, by Komlós' theorem again, there exists a further subsequence $\{g_{n_j^{(2)},2}\} \subseteq \{g_{n_j^{(1)},2}\} \subseteq \{g_{n,2}\}$ such that the Cesàro sum converges. Recursively, for each m , we obtained a subsequence $\{g_{n_j^{(m)},m}\} \subseteq \{g_{n_j^{(m-1)},m}\} \subseteq \{g_{n,m}\}$ such that the Cesàro sum converges μ -almost surely. Now, using the Cantor diagonalization argument, and the fact that countable union of μ -measure zero sets is still μ -measure zero, there exists a subsequence $\{n_j\}$ and a measurable set $X_\infty \subseteq X$ such that $\mu(X_\infty) = 1$ and that the Cesàro sum

$$\bar{g}_{k,m}(x) := \frac{1}{k} \sum_{j=1}^k g_{n_j,m}(x)$$

converges as $k \rightarrow \infty$ for all m and all $x \in X_\infty$. Let g_m be (in the sense of μ -a.s. equivalence class) the pointwise limit of $\{\bar{g}_{k,m}\}$ for all $m \in \mathbb{N}$. As a μ -a.s. pointwise limit, g_m is

measurable. Moreover, since $|g_{n,m}| \leq 1$ for all $n, m \in \mathbb{N}$, $|g_m| \leq 1$ for all $m \in \mathbb{N}$ as well.

In the meantime, for each $k \in \mathbb{N}$, and for the same subsequence $\{n_j\}$, let

$$\bar{P}_k := \frac{1}{k} \sum_{j=1}^k P_{n_j}.$$

Since $\mathcal{P}$ is regular, and thus is convex, $\bar{P}_k \in \mathcal{P}$ for all $k \in \mathbb{N}$. In particular, $\{\bar{P}_k(\cdot | x)\}_{k=1}^\infty \subseteq \mathcal{P}_x$ for all $x \in X$. Then, by definition,

$$\bar{g}_{k,m}(x) = \frac{1}{k} \sum_{j=1}^k g_{n_j,m}(x) = \frac{1}{k} \sum_{j=1}^k \int_X h_m(y) P_{n_j}(dy | x) = \int_X h_m(y) \bar{P}_k(dy | x), \quad (2)$$

for all $x \in X$, and for all $k, m \in \mathbb{N}$. Since $\mathcal{P}$ has a closed graph, $\mathcal{P}_x$ is closed and hence compact. Fix any $x \in X_\infty$. Since $\mathcal{P}_x$ is compact, there exists a subsequence $\{\bar{P}_{k_j}(\cdot | x)\}$ that converges to some η_x . We claim that every convergent subsequence of $\{\bar{P}_k(\cdot | x)\}$ converges to the same η_x , and thus—since $\mathcal{P}_x$ is a compact metric space—the entire sequence $\{\bar{P}_k(\cdot | x)\}$ converges to η_x . Indeed, for any $m \in \mathbb{N}$, for any two convergent subsequences $\{P_{k_j^{(1)}}(\cdot | x)\}$ and $\{P_{k_j^{(2)}}(\cdot | x)\}$ of $\{\bar{P}_k(\cdot | x)\}$, let $\eta_x^{(1)}$ and $\eta_x^{(2)}$ be their limits, respectively. By definition, we have

$$\lim_{j \rightarrow \infty} \int_X h_m(y) \bar{P}_{k_j^{(i)}}(dy | x) = \int_X h_m d\eta_x^{(i)},$$

for $i \in \{1, 2\}$ and for all $m \in \mathbb{N}$. In the meantime, since $x \in X_\infty$, $\{\bar{g}_{k,m}(x)\} \rightarrow g_m(x)$ as $k \rightarrow \infty$. Together with (2), we have

$$\int_X h_m d\eta_x^{(i)} = \lim_{j \rightarrow \infty} \int_X h_m(y) \bar{P}_{k_j^{(i)}}(dy | x) = \lim_{j \rightarrow \infty} \bar{g}_{k_j^{(i)},m}(x) = g_m(x), \quad (3)$$

for $i \in \{1, 2\}$, and for all $m \in \mathbb{N}$. In particular, for all $m \in \mathbb{N}$,

$$\int_X h_m d\eta_x^{(1)} = \int_X h_m d\eta_x^{(2)}.$$

Moreover, since $\{h_m\}$ is dense in $\{h \in C(X) : \|h\| \leq 1\}$, for any $h \in C(X)$ with $\|h\| \leq 1$, there exists $\{h_{m_l}\} \subseteq \{h_m\}$ such that $\|h_{m_l} - h\| \rightarrow 0$ as $l \rightarrow \infty$. Therefore, for any continuous function $h \in C(X)$ with $\|h\| \leq 1$, by the dominated convergence theorem,

$$\int_X h d\eta_x^{(1)} = \lim_{l \rightarrow \infty} \int_X h_{m_l} d\eta_x^{(1)} = \lim_{l \rightarrow \infty} \int_X h_{m_l} d\eta_x^{(2)} = \int_X h d\eta_x^{(2)}.$$

Thus, for any $h \in C(X)$ with $h \neq 0$,

$$\int_X h d\eta_x^{(1)} = \|h\| \cdot \int_X \frac{h}{\|h\|} d\eta_x^{(1)} = \|h\| \cdot \int_X \frac{h}{\|h\|} d\eta_x^{(2)} = \int_X h d\eta_x^{(2)}.$$

As a result, by Riesz's representation theorem, $\eta_x^{(1)} = \eta_x^{(2)} = \eta_x$. Furthermore, as $\{\bar{P}_k(\cdot | x)\} \rightarrow \eta_x$ and $\mathcal{P}_x$ is closed, we have that $\eta_x \in \mathcal{P}_x$ as well. Thus, for all $x \in X_\infty$, $\{\bar{P}_k(\cdot | x)\} \rightarrow \eta_x$ for some $\eta_x \in \mathcal{P}_x$. Since $\mathcal{P}$ is rectangular and $x \mapsto \mathcal{P}_x$ has a closed graph and compact values, by the Kuratowski–Ryll–Nardzewski measurable selection theorem, there exists a measurable selection $s : X \rightarrow \Delta(X)$ so that $s(x) \in \mathcal{P}_x$ for all $x \in X$. Let

$$P(\cdot | x) := \begin{cases} \eta_x, & \text{if } x \in X_\infty; \\ s(x), & \text{otherwise.} \end{cases}$$

Since $x \mapsto \eta_x$ is the pointwise limit of the transition kernels $\{\bar{P}_k\}$ on X_∞ and the pointwise limit of measurable functions into a metrizable space is measurable, we have that $x \mapsto \eta_x$ is measurable on X_∞ . Moreover, since s is measurable, P by construction is a valid Markov kernel. Furthermore, since $\mathcal{P}$ is rectangular, $\eta_x \in \mathcal{P}_x$ for all $x \in X_\infty$, and $s(x) \in \mathcal{P}_x$ for all $x \in X$, we have that $P \in \mathcal{P}$.

Now, for each $k \in \mathbb{N}$, let

$$\bar{\nu}_k := \bar{P}_k * \mu = \frac{1}{k} \sum_{j=1}^k P_{n_j} * \mu = \frac{1}{k} \sum_{j=1}^k \nu_{n_j}.$$

Since $\{\nu_n\} \rightarrow \nu$, $\{\bar{\nu}_k\} \rightarrow \nu$ as well. For any $m, k \in \mathbb{N}$, by (2), we have

$$\int_X h_m d\bar{\nu}_k = \int_X \left(\int_X h_m(y) \bar{P}_k(dy | x) \right) \mu(dx) = \int_X \bar{g}_{k,m} d\mu.$$

In the meantime, since $\{\bar{g}_{k,m}\} \rightarrow g_m$ μ -almost surely for all m , and since $\|\bar{g}_{k,m}\| \leq 1$, by the dominated convergence theorem,

$$\lim_{k \rightarrow \infty} \int_X \bar{g}_{k,m} d\mu = \int_X g_m d\mu.$$

Together, since $\{\bar{\nu}_k\} \rightarrow \nu$, we have that for all $m \in \mathbb{N}$,

$$\int_X h_m d\nu = \lim_{k \rightarrow \infty} \int_X h_m d\bar{\nu}_k = \lim_{k \rightarrow \infty} \int_X \bar{g}_{k,m} d\mu = \int_X g_m d\mu$$

$$\begin{aligned}
&= \int_X \left(\int_X h_m(y) P(dy | x) \right) \mu(dx) \\
&= \int_X h_m dP * \mu,
\end{aligned}$$

where the fourth equality follows from (3) with the conclusion that $\eta_x^{(1)} = \eta_x^{(2)} = \eta_x$ for all $x \in X_\infty$ and from the definition of P —in particular, from $P(\cdot | x) = \eta_x$ for all $x \in X_\infty$. Again, since $\{h_m\}$ is dense in $\{h \in C(X) : \|h\| \leq 1\}$, for any $h \in C(X)$ with $\|h\| \leq 1$, there exists $\{h_{m_l}\} \subseteq \{h_m\}$ such that $\|h_{m_l} - h\| \rightarrow 0$ as $l \rightarrow \infty$, and hence, by the dominated convergence theorem,

$$\int_X h d\nu = \lim_{l \rightarrow \infty} \int_X h_{m_l} d\nu = \lim_{l \rightarrow \infty} \int_X h_{m_l} dP * \mu = \int_X h dP * \mu.$$

As a result, for any $h \in C(X)$ with $h \neq 0$,

$$\int_X h d\nu = \|h\| \cdot \int_X \frac{h}{\|h\|} d\nu = \|h\| \cdot \int_X \frac{h}{\|h\|} dP * \mu = \int_X h dP * \mu.$$

Thus, by Riesz's representation theorem again, we have $\nu = P * \mu$. Since $P \in \mathcal{P}$, it follows that $\nu \in M(\mu)$, as desired. $\square$

Lemma 8. *Let X be a compact convex metrizable subset of a locally convex Hausdorff topological vector space and let Y be a metric space. Let $f : X \times Y \rightarrow \mathbb{R}$ be a bounded lower semi-continuous function where $x \mapsto f(x, y)$ is convex for all $y \in Y$. Then, there exists a sequence of continuous functions $f_n : X \times Y \rightarrow \mathbb{R}$ such that:*

- (i) $f_n \leq f_{n+1} \leq f$ for all n ,
- (ii) $f_n \uparrow f$ pointwise on $X \times Y$,
- (iii) for every $y \in Y$, the function $x \mapsto f_n(x, y)$ is convex.

Proof. Let $\mathcal{A}(X)$ denote the space of continuous affine functions on X . Since X is compact and metrizable, $C(X)$ is separable in the sup-norm. Since $\mathcal{A}(X)$ is a closed subspace of $C(X)$, $\mathcal{A}(X)$ is also separable. Fix a countable sup-norm dense subset

$$\{a_k\} \subset \mathcal{A}(X).$$

For each $k \in \mathbb{N}$, define

$$m_k(y) := \min_{x \in X} (f(x, y) - a_k(x)),$$

for all $y \in Y$. Since $(x, y) \mapsto f(x, y) - a_k(x)$ is lower semicontinuous and X is compact, the minimum is attained, and the map $m_k : Y \rightarrow \mathbb{R}$ is lower semicontinuous by Berge's theorem of maximum.

Since Y is a metric space and m_k is real-valued and lower semicontinuous, there exists an increasing sequence of continuous functions $\phi_{k,l} : Y \rightarrow \mathbb{R}$ such that $\{\phi_{k,l}\}_l \uparrow m_k$ pointwise on Y .

Now, for each k and each l , let

$$h_{k,l}(x, y) := a_k(x) + \phi_{k,l}(y).$$

Then, by construction, $h_{k,l}$ is continuous on $X \times Y$, and for each fixed y , $x \mapsto h_{k,l}(x, y)$ is affine. Moreover, since $\phi_{k,l} \leq m_k$ for all l and for all k , we have

$$h_{k,l}(x, y) = a_k(x) + \phi_{k,l}(y) \leq a_k(x) + m_k(y) \leq f(x, y)$$

for all $(x, y) \in X \times Y$, where the last inequality follows from the definition of m_k . Thus, for all k and all l , we have that $h_{k,l} \leq f$ is a continuous function where $x \mapsto h_{k,l}(x, y)$ is affine for all y .

Enumerate the countable family $\{h_{k,l}\}$ as $\{h_j\}_{j \in \mathbb{N}}$, and define

$$f_n(x, y) := \max \{h_1(x, y), \dots, h_n(x, y)\},$$

for all $n \in \mathbb{N}$ and for all $(x, y) \in X \times Y$. By construction, for each $n \in \mathbb{N}$, $f_n \leq f$ and f_n is continuous on $X \times Y$. Moreover, for every fixed y , the map $x \mapsto f_n(x, y)$ is convex, since it is the pointwise maximum of finitely many affine functions. Also, $f_n \leq f_{n+1}$ for all $n \in \mathbb{N}$.

It remains to prove that $f_n \uparrow f$ pointwise. Since $f_n \uparrow \sup_j h_j$ pointwise, it is enough to show that

$$\sup_{j \in \mathbb{N}} h_j(x, y) = f(x, y)$$

for all $(x, y) \in X \times Y$. To see this, fix $(x_0, y_0) \in X \times Y$ and $\varepsilon > 0$. Let

$$g(x) := f(x, y_0),$$

for all $x \in X$. Then g is a lower semicontinuous, convex function on X . We claim that there exists $a \in \mathcal{A}(X)$ such that $a \leq g$ and $a(x_0) \geq g(x_0) - \varepsilon/4$. Indeed, let $E \supset X$ be the ambient locally convex topological vector space, and extend g to a function $\tilde{g} : E \rightarrow (-\infty, +\infty]$ on

E by

$$\tilde{g}(x) := \begin{cases} g(x), & \text{if } x \in X, \\ +\infty, & \text{otherwise.} \end{cases}$$

Since X is compact and hence closed, $\tilde{g}$ is a proper lower semicontinuous convex function. By the Fenchel–Moreau theorem (see, e.g., [Rockafellar 1970](#)), a proper lower semicontinuous convex function is the supremum of its continuous affine minorants. Thus, we have

$$\tilde{g}(x_0) = \sup \left\{ b(x_0) : b \text{ is a continuous affine function on } E, b \leq \tilde{g} \right\}.$$

Then, since for any continuous affine b on E with $b \leq \tilde{g}$, $b|_X$ is a continuous affine function on X with $b|_X \leq g$, and since $x_0 \in X$, it follows that

$$g(x_0) = \tilde{g}(x_0) \leq \sup \left\{ a(x_0) : a \in \mathcal{A}(X), a \leq g \right\} \leq g(x_0).$$

Thus, we have

$$g(x_0) = \sup \left\{ a(x_0) : a \in \mathcal{A}(X), a \leq g \right\}.$$

Therefore, there exists $a \in \mathcal{A}(X)$ such that $a \leq g$ on X and $a(x_0) \geq g(x_0) - \varepsilon/4$, as desired.

Now, since $\{a_k\}$ is dense in $\mathcal{A}(X)$, there exists $k \in \mathbb{N}$ such that $\|a_k - a\|_\infty < \varepsilon/4$. Then

$$a_k(x_0) \geq a(x_0) - \varepsilon/4 \geq g(x_0) - \varepsilon/2 = f(x_0, y_0) - \varepsilon/2.$$

Meanwhile, for every $x \in X$, we have

$$a_k(x) \leq a(x) + \varepsilon/4 \leq g(x) + \varepsilon/4 = f(x, y_0) + \varepsilon/4,$$

which implies that

$$m_k(y_0) = \min_{x \in X} (f(x, y_0) - a_k(x)) \geq -\varepsilon/4.$$

Since $\{\phi_{k,l}(y_0)\}_l \uparrow m_k(y_0)$, there exists some $l \in \mathbb{N}$ such that

$$\phi_{k,l}(y_0) \geq -\varepsilon/2.$$

Then, we have

$$h_{k,l}(x_0, y_0) = a_k(x_0) + \phi_{k,l}(y_0) \geq f(x_0, y_0) - \varepsilon.$$

Since $h_{k,l} \leq f$, it follows that

$$f(x_0, y_0) - \varepsilon \leq \sup_{j \in \mathbb{N}} h_j(x_0, y_0) \leq f(x_0, y_0).$$

Taking $\varepsilon \downarrow 0$ gives

$$\sup_{j \in \mathbb{N}} h_j(x_0, y_0) = f(x_0, y_0).$$

Since (x_0, y_0) is chosen arbitrarily, it follows that $f_n \uparrow f$ pointwise, as desired. $\square$

A.2 Proof of Theorem 1

(a) $\implies$ (b): Suppose $\mathcal{C}$ is min-closed. First, observe that for any $\nu \leq_{\mathcal{C}} \mu$,

$$\int_X f \, d\nu \leq \int_X \bar{f} \, d\nu \leq \int_X \bar{f} \, d\mu,$$

where the second inequality follows from $\bar{f} \in \mathcal{C}$ by Lemma 2. Therefore,

$$\max_{\nu \in \Delta(X): \nu \leq_{\mathcal{C}} \mu} \int_X f \, d\nu \leq \int_X \bar{f} \, d\mu.$$

We now show that

$$\int_X \bar{f} \, d\mu \leq \max_{\nu \in \Delta(X): \nu \leq_{\mathcal{C}} \mu} \int_X f \, d\nu.$$

To this end, let $C(X)$ be the normed linear space of continuous functions on X under the sup-norm. By Riesz's representation theorem, the dual $[C(X)]^*$ of $C(X)$ is isomorphic to the set of regular Radon measures on X . Define $\Theta : C(X) \rightarrow \mathbb{R} \cup \{\infty\}$ and $\Xi : C(X) \rightarrow \mathbb{R} \cup \{\infty\}$ as:

$$\Theta(g) := \begin{cases} 0, & \text{if } g \geq f \\ +\infty, & \text{otherwise} \end{cases}, \quad \text{and} \quad \Xi(g) := \begin{cases} \int_X g \, d\mu, & \text{if } g \in \tilde{\mathcal{C}} \\ +\infty, & \text{otherwise} \end{cases},$$

where

$$\tilde{\mathcal{C}} = \mathcal{C} \cap C(X).$$

Note that $\tilde{\mathcal{C}}$ is a convex cone. Therefore, Θ and Ξ are convex functions with values in $\mathbb{R} \cup \{+\infty\}$.

Next, define the dual functions $\Theta^* : [C(X)]^* \rightarrow \mathbb{R} \cup \{\infty\}$ and $\Xi^* : [C(X)]^* \rightarrow \mathbb{R} \cup \{\infty\}$

$$\Theta^*(\nu) := \sup_{g \in C(X)} \left[\int_X g \, d\nu - \Theta(g) \right] = \sup_{g \geq f} \int_X g \, d\nu$$

and

$$\Xi^*(\nu) := \sup_{g \in C(X)} \left[\int_X g \, d\nu - \Xi(g) \right] = \sup_{g \in \tilde{\mathcal{C}}} \left(\int_X g \, d\nu - \int_X g \, d\mu \right).$$

Note that for any $\nu \in [C(X)]^*$ that is not a positive measure, there exists $\varphi \geq 0$ such that

$$\int_X \varphi \, d\nu < 0$$

and hence by considering the sequence of functions $g_t := \varphi \cdot t + M$, where M is a continuous function above f , we have

$$\inf_{g \geq f} \int_X g \, d\nu = -\infty.$$

Meanwhile, note that for any $\nu \in [C(X)]^*$ such that $\nu \not\leq_{\mathcal{C}} \mu$, there exists $g \in \mathcal{C}$ such that

$$\int_X g \, d\nu > \int_X g \, d\mu,$$

which implies that the same holds for some $\tilde{g} \in \tilde{\mathcal{C}}$ as well since any such g can be approximated by a decreasing sequence $\{g_n\} \subseteq \tilde{\mathcal{C}}$ that converges pointwise to g and hence $\int g_n \, d(\nu - \mu) \rightarrow \int g \, d(\nu - \mu) > 0$ by the dominated convergence theorem. Thus, since $\tilde{\mathcal{C}}$ is a cone, we have

$$\inf_{g \in \tilde{\mathcal{C}}} \left(\int_X g \, d\mu - \int_X g \, d\nu \right) = -\infty.$$

Together, since $1, -1 \in \tilde{\mathcal{C}}$, it then follows that

$$\inf_{g \geq f} \int_X g \, d\nu + \inf_{g \in \tilde{\mathcal{C}}} \left(\int_X g \, d\mu - \int_X g \, d\nu \right) > -\infty$$

if and only if $\nu \in \Delta(X)$ and $\nu \leq_{\mathcal{C}} \mu$. Lastly, note that for any $\nu \in \Delta(X)$ such that $\nu \leq_{\mathcal{C}} \mu$,

$$\inf_{g \in \tilde{\mathcal{C}}} \left(\int_X g \, d\mu - \int_X g \, d\nu \right) = 0.$$

Consequently, we have

$$\begin{aligned}
\sup_{\nu \in [C(X)]^*} [-\Theta^*(-\nu) - \Xi^*(\nu)] &= \sup_{\nu \in [C(X)]^*} \left[\inf_{g \geq f} \int_X g \, d\nu + \inf_{g \in \tilde{\mathcal{C}}} \left(\int_X g \, d\mu - \int_X g \, d\nu \right) \right] \\
&= \sup_{\nu \in \Delta(X): \nu \leq_C \mu} \left[\inf_{g \geq f} \int_X g \, d\nu + \inf_{g \in \tilde{\mathcal{C}}} \left(\int_X g \, d\mu - \int_X g \, d\nu \right) \right] \\
&= \sup_{\nu \in \Delta(X): \nu \leq_C \mu} \inf_{g \geq f} \int_X g \, d\nu \\
&= \max_{\nu \in \Delta(X): \nu \leq_C \mu} \int_X f \, d\nu,
\end{aligned}$$

where the last equality also uses that any bounded upper semicontinuous function on a compact metric space admits a decreasing sequence of continuous functions approximating it pointwise from above.

By [Lemma 2](#), $\bar{f} \in \mathcal{C}$ and hence $\bar{f} + 1 \in \mathcal{C}$. Moreover, by assumption, there exists a sequence $\{h_n\}$ where $h_n \in \tilde{\mathcal{C}}$ such that h_n converges to $\bar{f} + 1$ pointwise from the above. Let $g_0 = h_1$. Then $g_0 \in \tilde{\mathcal{C}}$ and $g_0 \geq f + 1$. Note that at this g_0 , we have that (i) Θ and Ξ are finite and (ii) Θ is continuous at g_0 in the sup norm. Therefore, by the Fenchel-Rockafellar duality theorem ([Villani 2003](#), Theorem 1.9), we have

$$\begin{aligned}
&\max_{\nu \in \Delta(X): \nu \leq_C \mu} \int_X f \, d\nu \\
&= \sup_{\nu \in [C(X)]^*} \left[\inf_{g \geq f} \int_X g \, d\nu + \inf_{g \in \tilde{\mathcal{C}}} \left(\int_X g \, d\mu - \int_X g \, d\nu \right) \right] \\
&= \sup_{\nu \in [C(X)]^*} [-\Theta^*(-\nu) - \Xi^*(\nu)] \\
&= \inf_{g \in C(X)} [\Theta(g) + \Xi(g)] \\
&= \inf_{g \in \tilde{\mathcal{C}}, g \geq f} \int_X g \, d\mu \\
&\geq \int_X \bar{f} \, d\mu,
\end{aligned}$$

where the last inequality follows from that for any $g \in \tilde{\mathcal{C}}$ and $g \geq f$, by the definition of $\bar{f}$, we have that for all x

$$g(x) \geq \bar{f}(x).$$

This completes the proof.

(b) $\implies$ (c): The proof is by the Krein-Milman theorem. Consider the orbit set

$$\mathcal{K}_\mu = \left\{ \nu \in \Delta(X) : \nu \leq_{\mathcal{C}} \mu \right\},$$

which is clearly convex. Moreover, by [Lemma 3](#), $\mathcal{K}_\mu$ is also compact.

We first claim that for any exposed point $\nu \in \exp(\mathcal{K}_\mu)$, there exists some P such that $\nu = P * \mu$ and $P * \delta_x \leq_{\mathcal{C}} \delta_x$ for all x . To see this, consider any exposed point $\nu \in \exp(\mathcal{K}_\mu)$. By definition, there exists some continuous f such that ν is the unique solution to [\(Optimization\)](#). Thus,

$$\int f \, d\nu = V_f^*(\mu) = \int V_f^*(\delta_x) \mu(dx) = \int \left(\max_{\eta: \eta \leq_{\mathcal{C}} \delta_x} \int f \, d\eta \right) \mu(dx),$$

where the second equality follows from (b) since $V_f^*(\mu)$ is an upper semicontinuous, and hence Baire-one, affine function and thus satisfies the barycentric formula (see Corollary 4.22 of [Lukeš et al. 2009](#)). By the Kuratowski–Ryll–Nardzewski measurable selection theorem, the correspondence

$$x \mapsto \operatorname{argmax}_{\eta \leq_{\mathcal{C}} \delta_x} \int f \, d\eta$$

admits a measurable selection. Fix any such measurable selection $x \mapsto \eta_x$, and define P by $P(\cdot | x) = \eta_x$ for all $x \in X$. Then, P is a transition kernel with $P * \delta_x \leq_{\mathcal{C}} \delta_x$ for all $x \in X$. Moreover, by construction,

$$\int f \, dP * \mu = \int \left(\int f(y) P(dy | x) \right) \mu(dx) = \int V_f^*(\delta_x) \mu(dx) = V_f^*(\mu) = \max_{\nu: \nu \leq_{\mathcal{C}} \mu} \int f \, d\nu,$$

and hence $P * \mu$, which satisfies $P * \mu \leq_{\mathcal{C}} \mu$, must also be an optimal solution. But since ν is the unique solution, it follows that $\nu = P * \mu$, proving the claim.

Next, we show that the same property holds for any $\nu \in \operatorname{co}(\exp(\mathcal{K}_\mu))$. Indeed, for any $\nu \in \operatorname{co}(\exp(\mathcal{K}_\mu))$, $\nu = \sum_{i=1}^n \lambda_i \nu_i$ for some $\{\nu_i\}_{i=1}^n \subseteq \exp(\mathcal{K}_\mu)$, and some $\{\lambda_i\} \subseteq \mathbb{R}_+$ with $\sum_{i=1}^n \lambda_i = 1$. For each $i \in \{1, \dots, n\}$, as shown above, since $\nu_i \in \exp(\mathcal{K}_\mu)$, there exists a transition kernel P_i such that $\nu_i = P_i * \mu$ and $P_i * \delta_x \leq_{\mathcal{C}} \delta_x$ for all $x \in X$. Let

$$P := \sum_{i=1}^n \lambda_i P_i.$$

Then,

$$\nu = \sum_{i=1}^n \lambda_i \nu_i = \sum_{i=1}^n \lambda_i P_i * \mu = P * \mu.$$

Moreover, for any $g \in \mathcal{C}$, and for any $x \in X$

$$\int g(y)P(dy | x) = \sum_{i=1}^n \lambda_i \int g(y)P_i(dy | x) \leq \sum_{i=1}^n \lambda_i g(x) = g(x).$$

Thus, $P * \delta_x \leq_{\mathcal{C}} \delta_x$ for all $x \in X$.

Now, consider any $\nu \in \mathcal{K}_\mu$. Since $\mathcal{K}_\mu$ is a compact convex subset of a compact metric space, by a theorem of [Klee jr. \(1958\)](#) (in particular, see Theorem 2.51 of [Lukeš et al. 2009](#)), there exists a sequence $\{\nu_n\} \subseteq \text{co}(\exp(\mathcal{K}_\mu))$ such that $\{\nu_n\} \rightarrow \nu$. For each $n \in \mathbb{N}$, as shown above, there exists a transition kernel P_n such that $\nu_n = P_n * \mu$ and $P_n * \delta_x \leq_{\mathcal{C}} \delta_x$ for all $x \in X$. For each $n \in \mathbb{N}$, define $\gamma_n \in \Delta(X \times X)$ as

$$\gamma_n(B \times A) := \int_B P_n(A | x) \mu(dx),$$

for all measurable $A, B \subseteq X$. Since X is a compact Polish space, $\Delta(X \times X)$ is also a compact Polish space. Therefore, by possibly taking a subsequence, $\{\gamma_n\}$ converges to some $\gamma \in \Delta(X \times X)$. By the disintegration theorem (see, e.g., [Çinlar, 2010](#), pp. 154, Theorem 2.18), there exists a transition kernel P such that

$$\gamma(B \times A) = \int_B P(A | x) \mu(dx),$$

for all measurable $A, B \subseteq X$. As a result, for any continuous function $h : X \rightarrow \mathbb{R}$.

$$\begin{aligned} \int_X h d\nu &= \lim_{n \rightarrow \infty} \int_X h d\nu_n = \lim_{n \rightarrow \infty} \int_X \left(\int_X h(y) P_n(dy | x) \right) \mu(dx) \\ &= \lim_{n \rightarrow \infty} \int_{X \times X} h(y) \gamma_n(dx, dy) \\ &= \int_{X \times X} h(y) \gamma(dx, dy) \\ &= \int_X \left(\int_X h(y) P(dy | x) \right) \mu(dx) \\ &= \int_X h d(P * \mu). \end{aligned}$$

Therefore, $\nu = P * \mu$. From this P , we now construct a transition kernel $\tilde{P}$ such that $\tilde{P} * \delta_x \leq_{\mathcal{C}} \delta_x$ for all $x \in X$ and that $\nu = \tilde{P} * \mu$. To begin with, we first show that for any

continuous $g \in \mathcal{C}$, there exists a measurable set $X_g \subseteq X$ with $\mu(X_g) = 1$, such that

$$\Phi^g(x) := \int g(y)P(dy | x) \leq g(x)$$

for all $x \in X_g$. Indeed, fix any $g \in \mathcal{C}$. For each $n \in \mathbb{N}$, let

$$\Phi_n^g(x) := \int g(y)P_n(dy | x),$$

for all $x \in X$. By construction, since $P_n * \delta_x \leq_{\mathcal{C}} \delta_x$ for all x and since $g \in \mathcal{C}$, we have $\Phi_n^g(x) \leq g(x)$ for all x . Therefore, for any nonnegative continuous function $\psi : X \rightarrow \mathbb{R}$,

$$\begin{aligned} \int_X \psi(x)\Phi^g(x)\mu(dx) &= \int_X \psi(x) \left(\int_X g(y)P(dy | x) \right) \mu(dx) \\ &= \int_{X \times X} \psi(x)g(y)\gamma(dx, dy) \\ &= \lim_{n \rightarrow \infty} \int_{X \times X} \psi(x)g(y)\gamma_n(dx, dy) \\ &= \lim_{n \rightarrow \infty} \int_X \psi(x) \left(\int g(y)P_n(dy | x) \right) \mu(dx) \\ &= \lim_{n \rightarrow \infty} \int_X \psi(x)\Phi_n^g(x)\mu(dx) \\ &\leq \int_X \psi(x)g(x)\mu(dx), \end{aligned}$$

where the third equality follows from $\psi \cdot g$ being continuous on $X \times X$ and $\{\gamma_n\} \rightarrow \gamma$. Thus, $\psi \mapsto \int \psi[g - \Phi^g]d\mu$ is a positive linear functional and hence by the Riesz-Markov-Kakutani representation theorem, $[g(x) - \Phi^g(x)]\mu(dx)$ is identified by a unique positive Radon measure, which immediately implies that $g(x) - \Phi^g(x) \geq 0$ for μ -a.e. x . Therefore, there exists a measurable set $X_g \subseteq X$ with $\mu(X_g) = 1$ such that $\Phi^g(x) \leq g(x)$ for all $x \in X_g$.

Next, we show that there exists a set X_∞ where this property holds uniformly across all $g \in \mathcal{C}$. By [Lemma 1](#), there exists a countable set $\{g_m\}_{m=1}^\infty \subseteq \mathcal{C} \cap C(X)$ that can approximate any $g \in \mathcal{C}$ from above. As shown above, for each $m \in \mathbb{N}$, since $g_m \in \mathcal{C}$ is continuous, there exists a measurable set X_m such that $\mu(X_m) = 1$ and $\Phi^{g_m}(x) \leq g_m(x)$ for all $x \in X_m$. Let $X_\infty := \cap_{m=1}^\infty X_m$. Clearly, $\mu(X_\infty) = 1$ as well. Now, fix any $g \in \mathcal{C}$. Let $\{g_{m_k}\} \subseteq \{g_m\}$ be a sequence such that $g_{m_k} \geq g$ for all k and $g_{m_k} \rightarrow g$ as $k \rightarrow \infty$. Then, for all $x \in X_\infty$, $\Phi^{g_{m_k}}(x) \leq g_{m_k}(x)$ for all $k \in \mathbb{N}$. Therefore, for all $x \in X_\infty$ and all $k \in \mathbb{N}$, we have

$$\int g(y)P(dy | x) \leq \int g_{m_k}(y)P(dy | x) = \Phi^{g_{m_k}}(x) \leq g_{m_k}(x),$$

which, by taking limit over k , implies that

$$\Phi^g(x) = \int g(y)P(dy | x) \leq \lim_{k \rightarrow \infty} g_{m_k}(x) = g(x).$$

Since g is an arbitrary element of $\mathcal{C}$, we have for all $g \in \mathcal{C}$, $\Phi^g \leq g$ on X_∞ where $\mu(X_\infty) = 1$.

Now, let

$$\tilde{P}(\cdot | x) := \begin{cases} P(\cdot | x), & \text{if } x \in X_\infty; \\ \delta_x, & \text{otherwise.} \end{cases}$$

By construction, since P is a measurable Markov kernel, $\tilde{P}$ is also a measurable Markov kernel. Moreover, for all $x \in X$ and all $g \in \mathcal{C}$,

$$\int g(y)\tilde{P}(dy | x) = \begin{cases} \int g(y)P(dy | x) \leq g(x), & \text{if } x \in X_\infty \\ g(x), & \text{otherwise.} \end{cases}$$

Therefore, $\tilde{P} * \delta_x \leq_{\mathcal{C}} \delta_x$ for all $x \in X$. Meanwhile, since $P(\cdot | x) = \tilde{P}(\cdot | x)$ for μ -almost all $x \in X$, and since $\nu = P * \mu$, we have $\nu = \tilde{P} * \mu$ as well. Together, we have that for any $\nu \in \mathcal{K}_\mu$, there exists a transition kernel $\tilde{P}$ such that $\nu = \tilde{P} * \mu$ and $\tilde{P} * \delta_x \leq_{\mathcal{C}} \delta_x$ for all $x \in X$, as desired.

(c) $\implies$ (d): Fix any closed and convex subset M of $\Delta(X)$. Fix any upper semicontinuous $f : X \rightarrow \mathbb{R}$. Let

$$h(x) := \max_{P: P * \delta_x \leq_{\mathcal{C}} \delta_x} \int f(y)P(dy | x),$$

for any $x \in X$. From (c), note that for any $\mu \in \Delta(X)$, we have

$$V_f^*(\mu) = \max_{\nu \leq_{\mathcal{C}} \mu} \int f d\nu = \int \left(\max_{P: P * \delta_x \leq_{\mathcal{C}} \delta_x} \int f(y)P(dy | x) \right) \mu(dx) = \int h d\mu,$$

where the second equality also uses the Kuratowski–Ryll–Nardzewski measurable selection theorem as before. Therefore, V_f^* is affine and satisfies the barycentric formula. This immediately implies that G_f^M has a convex graph. Indeed, for any $(\mu_1, \nu_1), (\mu_2, \nu_2) \in G_f^M$ and any $\alpha \in (0, 1)$, we have

$$\nu := \alpha \nu_1 + (1 - \alpha) \nu_2 \leq_{\mathcal{C}} \alpha \mu_1 + (1 - \alpha) \mu_2 =: \mu$$

and thus

$$\int f(x)\nu(dx) = \alpha V_f^*(\mu_1) + (1 - \alpha) V_f^*(\mu_2) = V_f^*(\mu),$$

and therefore $\nu \in X_f^*(\mu)$. To see that G_f^M has decomposable extreme points, observe that the above also shows that $\nu \in X_f^*(\mu)$ if and only if $\nu = P * \mu$ for some P such that $P(\cdot | x)$ solves

$$\max_{\eta: \eta \leq_C \delta_x} \int f d\eta,$$

for μ -almost all $x \in X$. It is easy to see that if $\mu \in \text{ext}(M)$ and $\nu \in \text{ext}(X_f^*(\mu))$, then $(\mu, \nu) \in \text{ext}(G_f^M)$. Indeed, if $\mu \in \text{ext}(M)$, $\nu \in \text{ext}(X_f^*(\mu))$, and $(\mu, \nu) = \frac{1}{2}(\mu_1, \nu_1) + \frac{1}{2}(\mu_2, \nu_2)$ where $(\mu_1, \nu_1), (\mu_2, \nu_2) \in G_f^M$, then we must have $\mu_1 = \mu_2 = \mu$ since $\mu \in \text{ext}(M)$, which implies that $\nu_1, \nu_2 \in X_f^*(\mu)$ and hence $\nu_1 = \nu_2 = \nu$ since $\nu \in \text{ext}(X_f^*(\mu))$.

Now, we prove the other direction. Fix any $(\mu, \nu) \in \text{ext}(G_f^M)$. Clearly, we have that $\nu \in \text{ext}(X_f^*(\mu))$. Now, suppose for contradiction that $\mu \notin \text{ext}(M)$. Then $\mu = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2$ for some $\mu_1 \neq \mu_2 \in M$. Note that μ_1 and μ_2 are both absolutely continuous with respect to μ . By the above observation, there exists some P such that $P * \mu = \nu$ and $P(\cdot | x)$ solves the above pointwise maximization problem μ -a.e. Let $\nu_1 = P * \mu_1$ and $\nu_2 = P * \mu_2$. Since μ_i are absolutely continuous with respect to μ , it follows immediately that $\nu_i \leq_C \mu_i$, and

$$\int f(x) \nu_i(dx) = \int f(y) P(dy | x) \mu_i(dx) = \int h(x) \mu_i(dx) = V_f^*(\mu_i),$$

where the second equality uses the μ -a.e., and hence μ_i -a.e., pointwise optimality of P , and the third equality uses the barycentric formula for V_f^* . Therefore, $\nu_i \in X_f^*(\mu_i)$. It follows that $(\mu_i, \nu_i) \in G_f^M$ for $i \in \{1, 2\}$. However, note that, by construction, $\frac{1}{2}(\mu_1, \nu_1) + \frac{1}{2}(\mu_2, \nu_2) = (\mu, \nu)$ and $\mu_1 \neq \mu_2$ —contradicting (μ, ν) is an extreme point of G_f^M .

(d) $\implies$ (a): Note that since G_f^M is convex for $M = \Delta(X)$, we must have that V_f^* is affine. Indeed, for any $\mu_1, \mu_2 \in \Delta(X)$. Fix any $\nu_1 \in X_f^*(\mu_1), \nu_2 \in X_f^*(\mu_2)$. By convexity of the graph, we have that for any $\alpha \in (0, 1)$, $\alpha \nu_1 + (1 - \alpha) \nu_2 \in X_f^*(\alpha \mu_1 + (1 - \alpha) \mu_2)$ and thus

$$V_f^*(\alpha \mu_1 + (1 - \alpha) \mu_2) = \alpha \int f(x) \nu_1(dx) + (1 - \alpha) \int f(x) \nu_2(dx) = \alpha V_f^*(\mu_1) + (1 - \alpha) V_f^*(\mu_2).$$

Therefore, V_f^* is an upper semicontinuous affine function on $\Delta(X)$. Now, suppose for contradiction that $\mathcal{C}$ is not min-closed. Then there exist $g_1, g_2 \in \mathcal{C}$ and

$$k := \min\{g_1, g_2\} \notin \mathcal{C}.$$

By assumption, there exists a decreasing sequence $\{g_i^n\} \subseteq \mathcal{C} \cap C(X)$ such that $g_i^n \downarrow g_i$ for each $i \in \{1, 2\}$. Let $k^n := \min\{g_1^n, g_2^n\}$. Note that each k^n is continuous and the decreasing

sequence $\{k^n\} \downarrow k$ by construction. Therefore, there must exist some n such that $k^n \notin \mathcal{C}$, because otherwise by assumption we must have $k \in \mathcal{C}$. Fix such n . Then, $k^n \in C(X)$ but $k^n \notin \mathcal{C} \cap C(X)$.

Let $\tilde{\mathcal{C}} := \mathcal{C} \cap C(X)$. Since $\tilde{\mathcal{C}}$ is a closed convex set in the space of $C(X)$ under the sup-norm, which is a locally convex topological vector space, by the strict Hahn–Banach separation theorem (see, e.g., [Lukeš et al. 2009](#), Theorem A.1), there exists a continuous linear functional that separates $\{k^n\}$ and $\tilde{\mathcal{C}}$. In particular, there exists a signed finite Borel measure σ and some $\alpha \in \mathbb{R}$ such that

$$\int k^n(x) \sigma(dx) > \alpha \geq \sup_{g \in \tilde{\mathcal{C}}} \int g(x) \sigma(dx).$$

Since $\tilde{\mathcal{C}}$ is a cone, it must be that

$$\sup_{g \in \tilde{\mathcal{C}}} \int g(x) \sigma(dx) \leq 0.$$

Since $\tilde{\mathcal{C}}$ contains -1 and 1 , we must also have $\sigma(X) = 0$, and $\int k^n(x) \sigma(dx) > \alpha \geq 0$. By the Jordan decomposition theorem, there exist Borel measures σ^+ and σ^- on X such that $\sigma = \sigma^+ - \sigma^-$. Since $\sigma(X) = 0$ and $\int k^n(x) \sigma(dx) > 0$, $\sigma^+(X) = \sigma^-(X) > 0$. Let $m := \sigma^+(X)$, and let $\nu := \sigma^+/m$, $\mu := \sigma^-/m$. Then, $\mu, \nu \in \Delta(X)$, and for any $g \in \tilde{\mathcal{C}}$, we have

$$\int_X g d\nu - \int_X g d\mu = \frac{1}{m} \left(\int_X g d\sigma^+ - \int_X g d\sigma^- \right) \leq 0,$$

and hence for any $g \in \mathcal{C}$, we have

$$\int_X g d\nu \leq \int_X g d\mu$$

by the assumption that any $g \in \mathcal{C}$ admits a sequence of continuous approximations from above and the monotone convergence theorem. Therefore, $\nu \leq_{\mathcal{C}} \mu$. Moreover, by construction, we also have

$$\int k^n(x) \nu(dx) > \int k^n(x) \mu(dx).$$

Now, note that for any $x \in X$ and for any $\eta \leq_{\mathcal{C}} \delta_x$, since $g_1^n, g_2^n \in \mathcal{C}$, we have

$$\int k^n d\eta = \int \min\{g_1^n, g_2^n\} d\eta \leq \min \left\{ \int g_1^n d\eta, \int g_2^n d\eta \right\} \leq \min\{g_1^n(x), g_2^n(x)\} = k^n(x).$$

It then follows that, for all $x \in X$,

$$V_{k^n}^*(\delta_x) = k^n(x).$$

Together, since $\nu \leq_C \mu$ and since $V_{k^n}^*$ is affine, we have

$$\int k^n d\nu \leq V_{k^n}^*(\mu) = \int V_{k^n}^*(\delta_x) \mu(dx) = \int k^n d\mu < \int k^n d\nu,$$

where the first equality uses that $V_{k^n}^*$ is an upper semicontinuous, and hence Baire-one, affine function and thus satisfies the barycentric formula (see Corollary 4.22 of [Lukáš et al. 2009](#))—a contradiction. $\square$

A.3 Proof of [Proposition 1](#)

(i): Consider any $\nu \in \mathcal{K}_\mu$ such that ν is uniquely rationalized by some order-preserving P with $P(\cdot | x)$ being an extreme point of $\mathcal{K}_{\delta_x}$ μ -a.e.. Suppose for contradiction that there exist $\nu_1, \nu_2 \in \mathcal{K}_\mu$ and $\lambda \in (0, 1)$, with $\nu_1 \neq \nu_2$, such that $\nu = \lambda \nu_1 + (1 - \lambda) \nu_2$. Since $\nu_i \in \mathcal{K}_\mu$, by [Theorem 1](#), there exists an order-preserving kernel P_i such that $P_i * \delta_x \leq_C \delta_x$ for all x and $i \in \{1, 2\}$. Thus,

$$P * \mu = \nu = \lambda \nu_1 + (1 - \lambda) \nu_2 = (\lambda P_1 + (1 - \lambda) P_2) * \mu,$$

which is also order-preserving. Since ν is uniquely rationalized, we have that

$$P(\cdot | x) = \lambda P_1(\cdot | x) + (1 - \lambda) P_2(\cdot | x),$$

for μ -almost all $x \in X$. However, since $\nu_1 \neq \nu_2$, there exists a subset $A \subseteq X$ with $\mu(A) > 0$ such that $P_1(\cdot | x) \neq P_2(\cdot | x)$ for all $x \in A$. Then, for this μ -positive measure of $x \in A$, we have that $P(\cdot | x)$ is not an extreme point of $\mathcal{K}_{\delta_x}$, a contradiction.

For the converse, consider any extreme point ν of $\mathcal{K}_\mu$. By [Theorem 1](#), $\nu = P * \mu$ for some order-preserving P such that $P * \delta_x \leq_C \delta_x$ for all $x \in X$. We first show that ν must be uniquely rationalized. Indeed, suppose for contradiction that there exist order-preserving kernels P_1, P_2 where P_1, P_2 differ on a μ -positive measure set such that $P_i * \delta_x \leq_C \delta_x$ for μ -a.e. x and

$$\nu = P_i * \mu,$$

for all $i \in \{1, 2\}$. Since X is Polish, $\mathcal{B}(X)$ is generated by a countable algebra $\mathcal{A}$. Therefore, for any x , if $P_1(\cdot | x) \neq P_2(\cdot | x)$, then there exists $B \in \mathcal{A}$ such that $P_1(B | x) \neq P_2(B | x)$. By

countability of $\mathcal{A}$, this implies that there exists a single measurable set $B \subseteq X$ such that $\mu(\{x : P_1(B | x) \neq P_2(B | x)\}) > 0$. Let

$$A := \{x \in X : P_1(B | x) > P_2(B | x)\}.$$

By possibly relabeling, it is without loss of generality to assume that $\mu(A) > 0$. Then, let

$$Q_1(\cdot | x) := \begin{cases} P_1(\cdot | x), & \text{if } x \in A \\ P_2(\cdot | x), & \text{otherwise} \end{cases}, \quad \text{and} \quad Q_2(\cdot | x) := \begin{cases} P_2(\cdot | x), & \text{if } x \in A \\ P_1(\cdot | x), & \text{otherwise} \end{cases}.$$

Clearly, Q_1, Q_2 are transition kernels since A is a measurable set. Moreover, since $P_i * \delta_x \leq_{\mathcal{C}} \delta_x$ for μ -a.e. x and for $i \in \{1, 2\}$, we have that $Q_i * \delta_x \leq_{\mathcal{C}} \delta_x$ for μ -a.e. $x \in X$ and $i \in \{1, 2\}$.

Let $\nu_i := Q_i * \mu$ for all $i \in \{1, 2\}$. It then follows that $\nu_1, \nu_2 \in \mathcal{K}_\mu$. Furthermore, by construction, we have

$$\frac{1}{2}Q_1 + \frac{1}{2}Q_2 = \frac{1}{2}P_1 + \frac{1}{2}P_2,$$

and thus,

$$\nu = \left(\frac{1}{2}P_1 + \frac{1}{2}P_2\right) * \mu = \frac{1}{2}Q_1 * \mu + \frac{1}{2}Q_2 * \mu = \frac{1}{2}\nu_1 + \frac{1}{2}\nu_2.$$

In the meantime, by construction, we have

$$\begin{aligned} \nu_1(B) - \nu_2(B) &= Q_1 * \mu(B) - Q_2 * \mu(B) \\ &= \int_A (P_1(B | x) - P_2(B | x)) \mu(dx) + \int_{A^c} (P_2(B | x) - P_1(B | x)) \mu(dx) \\ &\geq \int_A (P_1(B | x) - P_2(B | x)) \mu(dx) \\ &> 0, \end{aligned}$$

where the first inequality follows from $P_1(B | x) \leq P_2(B | x)$ for all $x \in A^c$ by construction; and the second inequality follows from $P_1(B | x) > P_2(B | x)$ for all $x \in A$ and $\mu(A) > 0$. Therefore, $\nu_1 \neq \nu_2$, and hence ν is not an extreme point, a contradiction.

Next, we show that for the (essentially) unique P such that $\nu = P * \mu$, it must be that $P(\cdot | x) \in \text{ext}(\mathcal{K}_{\delta_x})$ for μ -almost all $x \in X$. To prove this, we identify any μ -measurable Markov kernel P as a joint distribution $\gamma \in \Delta(X \times X)$ whose first marginal is μ . We claim that the set of order-preserving μ -measurable Markov kernels can be identified by

$$\Gamma_\mu := \left\{ \gamma : (\pi_1)_* \gamma = \mu, \int h(x)g(y)\gamma(dx, dy) \leq \int h(x)g(x)\mu(dx) \text{ for all } h \in C_+(X), g \in \mathcal{C} \cap C(X) \right\},$$

where $C_+(X)$ is the set of nonnegative continuous functions on X and $(\pi_1)_*\gamma$ denotes the first marginal of γ . By inspection, any order-preserving P can be identified by an element in Γ_μ . To see the converse, fix any $\gamma \in \Gamma_\mu$. Note that by the monotone convergence theorem and the assumption that any element $g \in \mathcal{C}$ can be approximated pointwise from above by continuous functions in $\mathcal{C}$, we have

$$\int h(x)g(y)\gamma(\mathrm{d}x, \mathrm{d}y) \leq \int h(x)g(x)\mu(\mathrm{d}x) \text{ for all } h \in C_+(X), g \in \mathcal{C}.$$

By the disintegration theorem, there exists a Markov kernel Q such that $\gamma(\mathrm{d}x, \mathrm{d}y) = Q(\mathrm{d}y \mid x)\mu(\mathrm{d}x)$, and hence

$$\int h(x) \left[\int g(y)Q(\mathrm{d}y \mid x) \right] \mu(\mathrm{d}x) \leq \int h(x)g(x)\mu(\mathrm{d}x) \text{ for all } h \in C_+(X), g \in \mathcal{C}.$$

Thus, for any $g \in \mathcal{C}$, we have

$$\int h(x) \left[g(x) - \int g(y)Q(\mathrm{d}y \mid x) \right] \mu(\mathrm{d}x) \geq 0 \text{ for all } h \in C_+(X),$$

which implies that

$$q_g(x) := g(x) - \int g(y)Q(\mathrm{d}y \mid x) \geq 0 \text{ for } \mu\text{-a.e. } x.$$

Indeed, to see the above, note that $h \mapsto \int h q_g \mathrm{d}\mu$ is a positive linear functional and hence by the Riesz-Markov-Kakutani representation theorem, $q_g(x)\mu(\mathrm{d}x)$ is identified by a unique positive Radon measure, which immediately implies that $q_g(x) \geq 0$ for μ -a.e. x . Let $X_g := \{x : q_g(x) \geq 0\}$. Then, $\mu(X_g) = 1$ for all $g \in \mathcal{C}$. Let $\{g_m\}_{m \in \mathbb{N}} \subseteq \mathcal{C} \cap C(X)$ be the countable collection given by [Lemma 1](#). Let

$$X_\infty := \bigcap_{m=1}^{\infty} X_{g_m}.$$

Note that X_∞ is measurable and satisfies $\mu(X_\infty) = 1$. Now, fix any $x \in X_\infty$ and any $g \in \mathcal{C}$. By [Lemma 1](#), there exists a sequence $\{g_{m_k}\} \subseteq \{g_m\}_{m \in \mathbb{N}}$ such that $g_{m_k} \geq g$ for all k , and $g_{m_k}(x) \rightarrow g(x)$ as $k \rightarrow \infty$. Then, by construction, for any $k \in \mathbb{N}$, we have

$$\int g(y)Q(\mathrm{d}y \mid x) \leq \int g_{m_k}(y)Q(\mathrm{d}y \mid x) \leq g_{m_k}(x),$$

where the last inequality uses that $x \in X_\infty$. Taking $k \rightarrow \infty$ then gives that

$$\int g(y)Q(dy \mid x) \leq g(x).$$

Since this holds for all $g \in \mathcal{C}$, it follows immediately that

$$Q * \delta_x \leq_{\mathcal{C}} \delta_x$$

for all $x \in X_\infty$. Since $\mu(X_\infty) = 1$, modifying Q on a μ -null set if necessary, we conclude that Q is an order-preserving kernel.

We equip $\Delta(X \times X)$ with the weak-* topology. Note that $\Delta(X \times X)$ is compact. Moreover, Γ_μ is a closed subset of $\Delta(X \times X)$ and hence compact. Now consider the map $T : \Delta(X \times X) \rightarrow \Delta(X)$ defined by: for all $\gamma \in \Delta(X \times X)$,

$$T(\gamma) := (\pi_2)_* \gamma,$$

i.e., the projection to the second marginal of γ . Clearly, T is a continuous linear map given the weak-* topology. Moreover, note that $T(\Gamma_\mu) = \mathcal{K}_\mu$ by construction and our previous argument. Since T is continuous and Γ_μ is compact, $\mathcal{K}_\mu$ is compact. By Proposition 8.26 of [Simon \(2011\)](#), we conclude that

$$\text{ext}(\mathcal{K}_\mu) \subseteq T(\text{ext}(\Gamma_\mu)).$$

For any $\nu \in \text{ext}(\mathcal{K}_\mu)$, by our previous argument, there exists a unique $\gamma \in \Gamma_\mu$ such that $(\pi_2)_* \gamma = \nu$. Thus, the unique $\gamma \in \text{ext}(\Gamma_\mu)$.

It remains to argue that for the unique $\gamma \in \text{ext}(\Gamma_\mu)$, its μ -a.e. unique disintegration satisfies $Q(\cdot \mid x) \in \text{ext}(\mathcal{K}_{\delta_x})$ for μ -a.e. x . Note that, by the disintegration theorem, without loss of generality, we can take Q as a Borel-measurable kernel. Now, suppose for contradiction that the set

$$S := \left\{ x \in X : Q(\cdot \mid x) \notin \text{ext}(\mathcal{K}_{\delta_x}) \right\}$$

satisfies $\mu(S) > 0$. Define

$$\Sigma := \left\{ (x, \eta_1, \eta_2) \in X \times \Delta(X) \times \Delta(X) : \eta_1 \in \mathcal{K}_{\delta_x}, \eta_2 \in \mathcal{K}_{\delta_x}, Q(\cdot \mid x) = \frac{1}{2}\eta_1 + \frac{1}{2}\eta_2, \eta_1 \neq \eta_2 \right\}.$$

Since the set

$$\{(x, \eta) \in X \times \Delta(X) : \eta \in \mathcal{K}_{\delta_x}\} = \bigcap_{g \in \mathcal{C} \cap C(X)} \left\{ (x, \eta) : \int g d\eta \leq g(x) \right\}$$

is an intersection of closed sets and hence closed; the set

$$\left\{ (x, \eta_1, \eta_2) : Q(\cdot | x) = \frac{1}{2}\eta_1 + \frac{1}{2}\eta_2 \right\}$$

is Borel since the Markov kernel Q is Borel-measurable; and the set $\{(\eta_1, \eta_2) : \eta_1 \neq \eta_2\}$ is open. Therefore, Σ is a Borel set. Moreover, Note that

$$S = \pi_X(\Sigma),$$

i.e., the set S is exactly the projection of Σ , and thus S is an analytic set. By the Jankov-von Neumann uniformization theorem ([Kechris 2012](#), Theorem 18.1), it follows that there exists a universally measurable selection $x \mapsto (\eta_1(x), \eta_2(x))$ mapping from S to $\Delta(X) \times \Delta(X)$. Now, for $i = 1, 2$, let

$$Q_i(\cdot | x) := \begin{cases} \eta_i(x) & \text{if } x \in S, \\ Q(\cdot | x) & \text{otherwise.} \end{cases}$$

By construction, $x \mapsto Q_i(\cdot | x)$ is μ -measurable, and hence Q_i 's are μ -measurable Markov kernels. Moreover, $Q_i(\cdot | x) \in \mathcal{K}_{\delta_x}$ for μ -a.e. x by construction. Thus, each Q_i is identified by some $\gamma_i \in \Gamma_\mu$. By construction, note that $\frac{1}{2}\gamma_1 + \frac{1}{2}\gamma_2 = \gamma$. Finally, to verify $\gamma_1 \neq \gamma_2$, fix a countable collection of determining functions $\{\psi_n\}$ where $\psi_n \in C(X)$. For each n , define

$$S_n^+ := \left\{ x \in S : \int \psi_n d\eta_1(x) > \int \psi_n d\eta_2(x) \right\} \text{ and } S_n^- := \left\{ x \in S : \int \psi_n d\eta_1(x) < \int \psi_n d\eta_2(x) \right\}.$$

Since $\{\psi_n\}$ is a determining class, every $x \in S$ belongs to $S_n^+ \cup S_n^-$ for some n . Hence, $S \subseteq \bigcup_n S_n^+ \cup S_n^-$. Since $\mu(S) > 0$, there exists some n such that $\mu(S_n^+) > 0$ or $\mu(S_n^-) > 0$. Without loss of generality, say $\mu(S_n^+) > 0$. Then

$$\begin{aligned} & \int \int \mathbb{1}_{S_n^+}(x) \psi_n(y) \gamma_1(dx, dy) - \int \int \mathbb{1}_{S_n^+}(x) \psi_n(y) \gamma_2(dx, dy) \\ &= \int_{S_n^+} \left[\int \psi_n d\eta_1(x) - \int \psi_n d\eta_2(x) \right] \mu(dx) \\ &> 0, \end{aligned}$$

and thus $\gamma_1 \neq \gamma_2$. Therefore, $\gamma \notin \text{ext}(\Gamma_\mu)$ —a contradiction.

(ii): Suppose that there exist some continuous f and some order-preserving kernel P such that $\nu = P * \mu$, and μ -a.e. we have $\int f(y)\eta(dy) = \bar{f}(x)$ and $\eta \leq_{\mathcal{C}} \delta_x \iff \eta = P_x$. Let A be the set under which the second claim holds. So $\mu(A) = 1$. We claim that $\nu = P * \mu$ must be an exposed point. Fix the objective given by f . Suppose for contradiction that $\nu' \neq \nu$ is also optimal under f . Then by [Theorem 1](#), there exists an order-preserving kernel P' such that μ -a.e. P'_x is optimal for the pointwise optimization problem, i.e., there exists a set B with $\mu(B) = 1$ such that for all $x \in B$, we have P'_x is a solution to:

$$\max_{\eta \leq_{\mathcal{C}} \delta_x} \int f(y)\eta(dy).$$

Clearly, $\mu(A \cap B) = 1$ and for all $x \in A \cap B$, we have that

$$\max_{\eta \leq_{\mathcal{C}} \delta_x} \int f(y)\eta(dy) = \bar{f}(x),$$

and $P'_x = P_x$. But then immediately we have $\nu' = P' * \mu = P * \mu = \nu$, a contradiction.

For the converse, suppose ν is an exposed point. Let f be any continuous function such that $\int f(x)\nu(dx)$ exposes ν . Let

$$S(x) := \operatorname{argmax}_{\eta \leq_{\mathcal{C}} \delta_x} \int f(y)\eta(dy).$$

Note that by [Theorem 1](#), ν is optimal for f if and only if

$$\nu \in \mathcal{S} := \left\{ \nu = P * \mu \text{ and } \mu\left(P_x \in S(x)\right) = 1 \right\}.$$

Suppose for contradiction that

$$\mu(|S(x)| > 1) > 0.$$

Let E be the above set with $\mu(E) > 0$. Fix a countable set $\{\varphi_n\}_n \subset C(X)$ that separates measures on $\Delta(X)$. Let

$$R_n(x) := \left\{ \int \varphi_n(y)\eta(dy) : \eta \in S(x) \right\} \subset \mathbb{R}.$$

Since $S(x)$ is compact and $\eta \mapsto \int \varphi_n(y)\eta(dy)$ is continuous, $R_n(x)$ is compact. Moreover, $R_n(x)$ is also convex, since $S(x)$ is convex. Thus, $R_n(x)$ is a compact interval.

For any $S(x)$ with $|S(x)| > 1$, there exist two distinct $\eta_1, \eta_2 \in S(x)$. Since $\{\varphi_n\}$ separates measures, there exists some n such that

$$\int \varphi_n(y) \eta_1(dy) \neq \int \varphi_n(y) \eta_2(dy),$$

and hence $|R_n(x)| > 1$. Let

$$E_n := \{x : |R_n(x)| > 1\}.$$

Then, we have

$$E \subseteq \bigcup_{n \geq 1} E_n,$$

and hence for some n , $\mu(E_n) > 0$. Fix such n , and let $\varphi := \varphi_n$.

Let

$$S^+(x) := \operatorname{argmax}_{\eta \in S(x)} \int \varphi(y) \eta(dy) \quad S^-(x) := \operatorname{argmin}_{\eta \in S(x)} \int \varphi(y) \eta(dy).$$

Note that each is nonempty compact-valued. By the Kuratowski–Ryll–Nardzewski theorem, there exist measurable selections

$$\eta_x^+ \in S^+(x) \quad \eta_x^- \in S^-(x).$$

Note that by construction, for any $x \in E_n$, we have

$$\int \varphi(y) \eta_x^+(dy) - \int \varphi(y) \eta_x^-(dy) = \int \varphi_n(y) \eta_x^+(dy) - \int \varphi_n(y) \eta_x^-(dy) > 0.$$

Fix any P such that $\nu = P * \mu$ and $P_x \in S(x)$ μ -a.e. Construct P^+ and P^- by

$$P_x^+ = \begin{cases} \eta_x^+ & \text{if } x \in E_n \\ P_x & \text{otherwise.} \end{cases} \quad P_x^- = \begin{cases} \eta_x^- & \text{if } x \in E_n \\ P_x & \text{otherwise.} \end{cases}$$

It follows immediately that $\nu^+ := P^+ * \mu \leq_c \mu$ and $\nu^- := P^- * \mu \leq_c \mu$. Moreover, by construction, we also have that $P_x^+ \in S(x)$ μ -a.e. and similarly $P_x^- \in S(x)$ μ -a.e. Therefore, ν^+ and ν^- must be optimal for f . Now, note that by construction

$$\int \varphi(x) \nu^+(dx) - \int \varphi(x) \nu^-(dx) = \int_{E_n} \left(\int \varphi(y) \eta_x^+(dy) - \int \varphi(y) \eta_x^-(dy) \right) \mu(dx) > 0,$$

where the strict inequality follows since $\mu(E_n) > 0$ and on E_n , the difference is strictly

positive pointwise. Therefore, $\nu^+ \neq \nu^-$ —a contradiction. $\square$

A.4 Proof of Proposition 2

(a) $\implies$ (b): Suppose ν is an exposed point and let f be the continuous function that exposes ν . Let $\mathcal{A}$ be the set of differentiability points of $\bar{f}$, which is a Lebesgue full-measure set in X . By Proposition 1, since μ is mutually absolutely continuous with the Lebesgue measure, for Lebesgue-almost all x , we have that the pointwise problem

$$\max_{\eta \preceq_{\text{concave}} \delta_x} \int f(y) \eta(dy) = \bar{f}(x)$$

admits a unique solution. We claim that f must have a strict contact—indeed, if there exists a positive measure of $x \in S \cap \mathcal{A}$ such that $x \in \text{conv}(T_{h^x} \setminus \{x\})$, then for such a positive measure of x , which by definition satisfies $f(x) = \bar{f}(x)$, the pointwise maximization problem admits multiple solutions, contradicting the above.

Now, we claim that for a full-measure set of $x \in S^c \cap \mathcal{A}$, the affine envelope region $S_{h^x}^c$ must be simplicial-touching. We first note that, by Lemma 4, there exists a full-measure set $\mathcal{B}$ in X such that x is in the relative interior of R_{h^x} for all $x \in \mathcal{B}$.

Now, consider the set

$$\mathcal{Z} := \left\{ x \in S^c \cap \mathcal{A} \cap \mathcal{B} : \text{pointwise problem admits a unique solution at } x \right\}.$$

By Proposition 1 and that $\mathcal{B}$ has full measure, $\mathcal{Z}$ is a full-measure subset of $S^c \cap \mathcal{A}$. We claim that for all $x \in \mathcal{Z}$, the affine envelope region $S_{h^x}^c$ must be simplicial-touching. To see this, fix any such x . Let $h := h^x$ to simplify the notation.

Note that by construction, $h \geq \bar{f} \geq f$, and hence

$$T_h \subseteq R_h.$$

Note that R_h is a convex set. Thus, $\text{conv}(T_h) \subseteq R_h$. Now, we show that $R_h \subseteq \text{conv}(T_h)$. Fix any $z \in R_h$. By definition, we have $\bar{f}(z) = h(z)$. Moreover, note that

$$V_f^*(\delta_z) = \bar{f}(z) = h(z).$$

By compactness, as argued before, the pointwise problem always admits a solution. In particular, there exists a solution η_z to the pointwise problem at z . Then, $\eta \preceq_{\text{concave}} \delta_z$ and $\int f d\eta_z = \bar{f}(z) = h(z) = \int h d\eta_z$. Hence, η_z must be supported on T_h and have barycenter z .

Thus, $z \in \text{conv}(T_h)$. Therefore, $R_h \subseteq \text{conv}(T_h)$, proving that $\text{conv}(T_h) = R_h$.

Note that by construction x is in the relative interior of $R_h = \text{conv}(T_h)$. This implies that T_h must be affinely independent, because otherwise at the fixed $x \in \text{conv}(T_h)$, the optimization problem would admit multiple pointwise solutions that are supported on T_h . Indeed, if T_h is not affinely independent, by Radon's Theorem, there exists some point $w \in \text{conv}(T_h)$ that has two distinct barycentric representations. Since $x \in \text{ri}(\text{conv}(T_h))$, there exists some $\lambda \in [0, 1)$ and some $y \in \text{conv}(T_h)$ such that

$$x = \lambda y + (1 - \lambda)w.$$

Therefore, the point x must have two distinct barycentric representations supported on T_h . Note that each of which would be optimal for the pointwise problem, contradicting the uniqueness property of the pointwise problem at x . Therefore, T_h is affinely independent, and hence we have

$$\text{ext}(R_h) = \text{ext}(\text{conv}(T_h)) = T_h,$$

and hence R_h is a simplex.

By [Proposition 1](#), there exists an order-preserving coupling P^f such that $\nu = P^f * \mu$ and that $P^f * \delta_x$ is exposed by f for μ -almost all, and hence Lebesgue-almost all, $x \in X$. This implies that $P^f = \delta_x$ for a full-measure set of $x \in S$, and P^f is the unique barycentric splitting for Lebesgue-almost all $x \in \mathcal{Z}$, which is a full-measure set in S^c , since, as we have shown, the unique pointwise solution for all $x \in \mathcal{Z}$ is given by the barycentric splitting. Thus, if necessary, modifying the kernel $P^f(\cdot | x)$ to be δ_x on a μ -null set gives the desired construction, concluding the proof.

(b) $\implies$ (a): Fix any such f, P^f , and $\nu = P^f * \mu$. We verify the conditions in [Proposition 1](#). Consider the set of differentiability points $\mathcal{A}$ of $\bar{f}$. We claim that for Lebesgue-almost all $x \in \mathcal{A}$, which is a Lebesgue full-measure set in X by concavity of $\bar{f}$, we have

$$\int f(y)\eta(dy) = \bar{f}(x) \text{ and } \eta \leq_{\text{concave}} \delta_x \iff \eta = P_x^f.$$

Indeed, consider the full-measure set of $x \in S \cap \mathcal{A}$ for which $x \notin \text{conv}(T_{h^x} \setminus \{x\})$. Fix any such x . Note that $\bar{f}(x) = f(x)$ and hence $\eta = P_x^f = \delta_x$ clearly satisfies the left side; moreover, if there exists some η satisfying the left side, then η must be supported on T_{h^x} and have barycenter x , but then it must be that $\eta = \delta_x$.

Now, consider the full-measure set of $x \in S^c \cap \mathcal{A}$ such that $S_{h^x}^c$ is simplicial-touching. Fix any such x . Let $h := h^x$. Since S_h^c is simplicial-touching, if $\eta = P_x^f$, then by construction

we have the left side; conversely, if the left side is satisfied by η , then η must be supported on T_h , but since $T_h = \text{ext}(R_h)$, where R_h is a simplex, is affinely independent, there exists a unique mean-preserving splitting of x supported on T_h given by P_x^f , and hence $\eta = P_x^f$.

Since the equivalence relation holds almost everywhere on $\mathcal{A}$ and $\mathcal{A}$ is a full-measure set in X , by [Proposition 1](#), ν must be an exposed point. $\square$

A.5 Proof of [Proposition 3](#)

(a) $\implies$ (b): Suppose ν is an exposed point of $\text{LSD}(\mu)$. Let f be a continuous function such that $\int f d\nu$ exposes ν . By [Proposition 1](#), there exists an order-preserving kernel P^f such that $\nu = P^f * \mu$, and for μ -a.e. $x \in X$, and hence Lebesgue a.e. $x \in X$ (since μ is mutually absolutely continuous with Lebesgue measure), we have

$$\int f(y)\eta(dy) = \bar{f}(x) \text{ and } \eta \leq_{\text{nondecreasing}} \delta_x \iff \eta = P_x^f. \quad (4)$$

We claim that f must have a strict monotone contact—indeed, if there exists a positive-Lebesgue-measure set of $x \in S$ such that there exists $y \leq x$ and $y \neq x$ with $f(y) = f(x) = \bar{f}(x)$, then for such a positive measure of x , the pointwise maximization problem admits multiple solutions, contradicting the above.

Now, fix $z \in \text{Ran}(\bar{f})$. Let

$$A_z := \left\{ x : \text{there exist two distinct } t, t' \in T_z \text{ with } t \leq x, t' \leq x \right\} \cup \left\{ x : \bar{f}(x) > z \right\}.$$

Note that A_z is upward-closed. We claim that

$$R_z \setminus A_z = \bigcup_{\underline{x} \in T_z} \left(\{ x : x \geq \underline{x} \} \setminus A_z \right).$$

Indeed, if $x \in R_z \setminus A_z$, then we have $\bar{f}(x) = z = \max_{y \leq x} f(y)$, and hence there exists $t \leq x$ such that $f(t) = z$. Since $t \leq x$, we have

$$\bar{f}(t) \leq \bar{f}(x) = z,$$

but with $\bar{f}(t) \geq f(t) = z$, we have $t \in T_z$. Moreover, since $x \notin A_z$, there exists a unique $t \leq x$ such that $t \in T_z$. Thus, x belongs to exactly one piece on the right-hand side $\{ x : x \geq t \} \setminus A_z$. Conversely, if $x \geq t$ for some $t \in T_z$ and $x \notin A_z$, then

$$\bar{f}(x) \geq \bar{f}(t) = f(t) = z,$$

and moreover,

$$\bar{f}(x) \leq z$$

since $x \notin A_z$, and hence

$$\bar{f}(x) = z,$$

and so $x \in R_z \setminus A_z$. This then proves that

$$R_z \setminus A_z = \bigsqcup_{\underline{x} \in T_z} \left(\{x : x \geq \underline{x}\} \setminus A_z \right),$$

with the disjointness follows from that, as we have shown, any element in the LHS belongs to a unique piece of the RHS.

Now, fix any $x \in S^c$ such that (4) holds, which is a full-measure set in S^c by [Proposition 1](#). Let $z^x := \bar{f}(x)$. By definition, $x \in R_{z^x}$. We show that $x \notin A_{z^x}$. Indeed, suppose for contradiction that $x \in A_{z^x}$. Then, since $\bar{f}(x) = z^x$, we must have that there exist two distinct $t, t' \in T_{z^x}$ with $t, t' \leq x$ —however, then there exist two distinct solutions to the pointwise problem at x —a contradiction. Therefore, $x \in R_{z^x} \setminus A_{z^x}$. Then, by our previous argument, it follows that there exists a unique $\underline{x} \in T_{z^x}$ with $\underline{x} \leq x$. Since P_x^f is the unique pointwise solution at x , it must be that $P_x^f = \delta_{\underline{x}}$, which is exactly the unique downward transport of x .

Thus, for a full-measure set of $x \in S^c$, the region $S_{z^x}^c$ is downward staircase-like, $x \in R_{z^x} \setminus A_{z^x}$, and $P^f(\cdot \mid x)$ is the unique downward transport. Moreover, since f has a strict monotone contact, $P^f(\cdot \mid x) = \delta_x$ for a full-measure set of $x \in S$. Thus, without loss of generality, we may take $P^f(\cdot \mid x) = \delta_x$ for all $x \in S$, yielding the desired construction.

(b) $\implies$ (a): Now suppose (b) holds. Fix any such f, P^f , and $\nu = P^f * \mu$. We verify the pointwise uniqueness condition in [Proposition 1](#).

First, fix any x in the full-measure subset of S where strict monotone contact holds. Since $\bar{f}(x) = f(x)$, δ_x is feasible and optimal. Moreover, any optimal η solving the pointwise problem at x must be supported on points $y \leq x$ with $f(y) = f(x)$, which, by strict monotone contact, implies that $\eta = \delta_x$.

Now, fix any x in the full-measure subset of S^c from statement (b). Let $z^x := \bar{f}(x)$. By assumption,

$$x \in R_{z^x} \setminus A_{z^x} = \bigsqcup_{\underline{x} \in T_{z^x}} \left(\{x' : x' \geq \underline{x}\} \setminus A_{z^x} \right),$$

and hence there exists a unique $\underline{x} \in T_{z^x}$ with $\underline{x} \leq x$. By construction, $P^f(\cdot \mid x) = \delta_{\underline{x}}$, which

is a feasible solution and in fact optimal since $f(\underline{x}) = z^x = \bar{f}(x)$. Moreover, if

$$\eta \leq_{\text{nondecreasing}} \delta_x \text{ and } \int f(y) \eta(dy) = \bar{f}(x) = z^x,$$

then η must be supported on points $y \leq x$ with $f(y) = z^x$. For any such y , in fact, we have

$$\bar{f}(y) \leq \bar{f}(x) = z^x \text{ and } \bar{f}(y) \geq f(y) = z^x,$$

we must have

$$\bar{f}(y) = f(y) = z^x,$$

and hence $y \in T_{z^x}$. Hence, η must be supported on points of T_{z^x} that are weakly below x . However, as we have argued, $\underline{x}$ is the unique such point, and hence $\eta = \delta_{\underline{x}} = P^f(\cdot | x)$.

Therefore, we conclude that for a full-measure set of $x \in X$, $P^f(\cdot | x)$ is the unique solution to the pointwise problem at x with the objective given by f . By [Proposition 1](#), ν is an exposed point of $\text{LSD}(\mu)$, completing the proof. $\square$

A.6 Proof [Theorem 3](#)

We prove this in the order of

$$(a) \implies (b) \implies (c) \implies (d) \implies (e) \implies (a).$$

$(a) \implies (b)$: We first show that $\mathcal{P} = \Psi(\mathcal{C})$. First, note that $\mathcal{P} \subseteq \Psi(\mathcal{C})$. Indeed, for any $P \in \mathcal{P}$ and for any $x \in X$, since $P(\cdot | x) = P * \delta_x$ and since $(\mathcal{C}, \mathcal{P})$ is Blackwell-consistent, $P(\cdot | x) \geq_{\mathcal{C}} \delta_x$. Thus, for all $g \in \mathcal{C}$

$$\int g(y) P(dy | x) \geq g(x)$$

for all $x \in X$, which implies that $P \in \Psi(\mathcal{C})$. As a result,

$$\nu = P * \mu \text{ for some } P \in \mathcal{P} \implies \nu = P * \mu \text{ for some } P \in \Psi(\mathcal{C}).$$

Next, consider any $\mu, \nu \in \Delta(X)$ such that $\nu = P * \mu$ for some $P \in \Psi(\mathcal{C})$. Since $P \in \Psi(\mathcal{C})$, we have that $g * P \geq g$ for all $g \in \mathcal{C}$. Therefore, for any $g \in \mathcal{C}$,

$$\int g d\nu = \int g d(P * \mu) = \int g * P d\mu \geq \int g d\mu.$$

That is, $\mu \leq_{\mathcal{C}} \nu$. As a result, since $(\mathcal{C}, \mathcal{P})$ is Blackwell-consistent, we have

$$\nu = P * \mu \text{ for some } P \in \Psi(\mathcal{C}) \implies \mu \leq_{\mathcal{C}} \nu \iff \nu = P * \mu \text{ for some } P \in \mathcal{P}.$$

Together, these imply that

$$\nu = P * \mu \text{ for some } P \in \mathcal{P} \iff \nu = P * \mu \text{ for some } P \in \Psi(\mathcal{C}).$$

Furthermore, by [Lemma 6](#), $\Psi(\mathcal{C})$ is rectangular. Also, $\mathcal{P}$ is rectangular as it is regular. Therefore, by [Lemma 5](#), we conclude that $\mathcal{P} = \Psi(\mathcal{C})$.

Next, we show that $\mathcal{C} = \Phi(\mathcal{P})$. Since $\mathcal{P} = \Psi(\mathcal{C})$, [Lemma 6](#) implies that $\mathcal{C} \subseteq \Phi(\Psi(\mathcal{C})) = \Phi(\mathcal{P})$. We show that $\Phi(\mathcal{P}) \subseteq \mathcal{C}$. By the proof of the “only if” part of [Theorem 2](#), since the order $\leq_{\mathcal{C}}$ is Blackwell-consistent, $\mathcal{C}$ must be max-closed. Now, fix any $g \in \Phi(\mathcal{P})$. Let

$$\hat{g} = -\overline{(-g)}^{[-\mathcal{C}]},$$

i.e., $-\hat{g}$ is the $[-\mathcal{C}]$ -envelope of $-g$. Note that

$$-\hat{g}(x) = \inf \{h(x) : h \in [-\mathcal{C}], h \geq -g\} = -\sup \{k(x) : k \in \mathcal{C}, k \leq g\}.$$

Note that under our regularity on $\mathcal{C}$, the convex cone $[-\mathcal{C}]$ satisfies all the assumptions in [Section 3](#). Moreover, since $\mathcal{C}$ is max-closed, $[-\mathcal{C}]$ is min-closed. By [Lemma 2](#), $\overline{(-g)}^{[-\mathcal{C}]} \in -\mathcal{C}$ and hence $\hat{g} \in \mathcal{C}$. We now argue that $\hat{g} = g$. Indeed, consider the maximization problem

$$V_{-g}^*(\delta_x) := \max_{\eta \leq_{[-\mathcal{C}]} \delta_x} \int [-g](x) \eta(dx).$$

Since $[-\mathcal{C}]$ is min-closed, by (a) $\implies$ (b) direction of [Theorem 1](#), we have the following strong duality: for any $x \in X$, we have

$$\max_{\eta \leq_{[-\mathcal{C}]} \delta_x} \int [-g](y) \eta(dy) = V_{-g}^*(\delta_x) = -\hat{g}(x) = \min_{h \in [-\mathcal{C}], h \geq -g} h(x).$$

For any $x \in X$, note that

$$\eta \in \Psi(\mathcal{C})_x \iff \int k(y) \eta(dy) \geq k(x) \quad \forall k \in \mathcal{C} \iff \delta_x \leq_{\mathcal{C}} \eta \iff \eta \leq_{[-\mathcal{C}]} \delta_x.$$

Therefore, since $g \in \Phi(\mathcal{P}) = \Phi(\Psi(\mathcal{C}))$, for any x , and any $P \in \Psi(\mathcal{C})$, we have

$$\int g(y)P(dy \mid x) \geq g(x),$$

which is equivalent to that for any x , we have

$$\inf_{P \in \Psi(\mathcal{C})} \int g(y)P(dy \mid x) \geq g(x),$$

which, in turn, is equivalent to

$$\inf_{\eta \in \Psi(\mathcal{C})_x} \int g(y)\eta(dy) \geq g(x).$$

For any x , since $\delta_x \in \Psi(\mathcal{C})_x$ by definition, the above holds if and only if

$$\min_{\eta \in \Psi(\mathcal{C})_x} \int g(y)\eta(dy) = g(x),$$

which holds, by our previous observation, if and only if

$$\min_{\eta \leq [-c]\delta_x} \int g(y)\eta(dy) = g(x).$$

Thus, for any x , we obtain

$$-\hat{g}(x) = \max_{\eta \leq [-c]\delta_x} \int [-g](y)\eta(dy) = - \min_{\eta \leq [-c]\delta_x} \int g(y)\eta(dy) = -g(x),$$

and hence $\hat{g}(x) = g(x)$. Thus, $g = \hat{g} \in \mathcal{C}$. Hence, $\Phi(\mathcal{P}) \subseteq \mathcal{C}$, as desired.

(b) $\implies$ (c): Since $(\mathcal{C}, \mathcal{P})$ is Blackwell-invariant, we have $\mathcal{C} = \Phi(\mathcal{P})$. It thus suffices to show that $\mathcal{P}$ is one-shot sufficient. Indeed, by Blackwell invariance, $\mathcal{P} = \Psi(\mathcal{C})$, which by [Lemma 6](#), is composition-closed and hence one-shot sufficient.

(c) $\implies$ (d): Since $\mathcal{P}$ is regular, $\text{Id} \in \mathcal{P}$, and hence $\mathcal{P} \subseteq \mathcal{P} \otimes \mathcal{P}$. To see that $\mathcal{P} \otimes \mathcal{P} \subseteq \mathcal{P}$, consider $Q \in \mathcal{P} \otimes \mathcal{P}$ and suppose that, by way of contradiction, $Q \notin \mathcal{P}$. Since $Q \notin \mathcal{P}$ and $\mathcal{P}$ is rectangular, there exists $x \in X$ such that $Q(\cdot \mid x) \notin \mathcal{P}_x$. Fix such x . Since $\mathcal{P}$ has a closed graph, $\mathcal{P}_x$ is closed. Since $\mathcal{P}_x$ is a closed convex set in $M(X)$, the set of finite signed Borel measures under the weak-* topology, which is a locally convex topological vector space, by the strict Hahn–Banach separation theorem (see, e.g., [Lukeš et al. 2009](#), Theorem A.1),

there exists a continuous function $h \in C(X)$ (since $C(X) = [M(X)]^*$) such that

$$\int h(y)Q(dy | x) > \sup_{\mu \in \mathcal{P}_x} \int h(y)\mu(dy).$$

By definition, note that we have

$$\sup_{\mu \in \mathcal{P}_x} \int h(y)\mu(dy) = \sup_{\mu \in \{P(\cdot | x) : P \in \mathcal{P}\}} \int h(y)\mu(dy) = \sup_{P \in \mathcal{P}} \int h(y)P(dy | x).$$

Meanwhile, since $Q \in \mathcal{P} \otimes \mathcal{P}$, we have

$$\int h(y)Q(dy | x) \leq \sup_{P \in \mathcal{P} \otimes \mathcal{P}} \int h(y)P(dy | x).$$

Together, these imply that

$$\sup_{P \in \mathcal{P}} \int h(y)P(dy | x) < \int h(y)Q(dy | x) \leq \sup_{P \in \mathcal{P} \otimes \mathcal{P}} \int h(y)P(dy | x),$$

contradicting $\mathcal{P}$ being one-shot sufficient. Therefore, we have $\mathcal{P} = \mathcal{P} \otimes \mathcal{P}$.

(d) $\implies$ (e): By [Lemma 6](#), $\mathcal{C} = \Phi(\mathcal{P})$ is max-closed. We show that $\Psi(\mathcal{C}) = \mathcal{P}$. By the rectangularity property of $\Psi(\mathcal{C})$ and $\mathcal{P}$, it suffices to show $\leq^{\Psi(\mathcal{C})} = \leq^{\mathcal{P}}$ by [Lemma 5](#). To show this, we will show that

$$\leq^{\Psi(\mathcal{C})} \implies \leq_{\mathcal{C}} \implies \leq^{\mathcal{P}} \implies \leq^{\Psi(\mathcal{C})}.$$

To see $\leq^{\Psi(\mathcal{C})} \implies \leq_{\mathcal{C}}$, note that if $\mu \leq^{\Psi(\mathcal{C})} \nu$, then there exists some $P \in \Psi(\mathcal{C})$ such that $\nu = P * \mu$. By the definition of $\Psi(\mathcal{C})$, for all $g \in \mathcal{C}$ and all $x \in X$, we have

$$\int g(y)P(dy | x) \geq g(x).$$

Thus, for all $g \in \mathcal{C}$, we have

$$\int g d\nu = \int g d(P * \mu) = \int \left(\int g(y)P(dy | x) \right) \mu(dx) \geq \int g d\mu,$$

and hence $\mu \leq_{\mathcal{C}} \nu$.

To see $\leq^{\mathcal{P}} \implies \leq^{\Psi(\mathcal{C})}$, note that $\mathcal{P} \subseteq \Psi(\Phi(\mathcal{P})) = \Psi(\mathcal{C})$ by [Lemma 6](#). Thus, for any $\mu, \nu \in \Delta(X)$ where $\nu = P * \mu$ for some $P \in \mathcal{P}$, we immediate have that $\nu = P * \mu$ for some

$P \in \Psi(\mathcal{C})$. Thus, $\leq^{\mathcal{P}} \implies \leq^{\Psi(\mathcal{C})}$.

Thus, to complete the argument, it suffices to show $\leq_{\mathcal{C}} \implies \leq^{\mathcal{P}}$. To this end, fix any $\mu, \nu \in \Delta(X)$ such that $\mu \leq_{\mathcal{C}} \nu$ and let

$$M(\mu) := \{P * \mu : P \in \mathcal{P}\}.$$

Suppose for contradiction that $\nu \notin M(\mu)$. By [Lemma 7](#), $M(\mu) \subseteq \Delta(X)$ is compact and convex. Since $M(\mu)$ is compact, $M(\mu)$ is closed. Since $M(\mu)$ is a closed convex set in the space of finite signed Borel measures equipped with the weak-* topology, by the strict Hahn–Banach separation theorem (see, e.g., [Lukeš et al. 2009](#), Theorem A.1), there exists a continuous linear functional that strictly separates $\{\nu\}$ from $M(\mu)$. In particular, there exists a continuous function $f \in C(X)$ such that

$$\int f \, d\nu < \inf_{\eta \in M(\mu)} \int f \, d\eta = \inf_{P \in \mathcal{P}} \int f \, d(P * \mu) = \int \left(\inf_{\eta \in \mathcal{P}_x} \int f \, d\eta \right) \mu(dx) = \int \left(\min_{\eta \in \mathcal{P}_x} \int f \, d\eta \right) \mu(dx),$$

where we have also used the rectangularity of $\mathcal{P}$, the Kuratowski–Ryll–Nardzewski measurable selection theorem, and the fact that f is continuous and $\mathcal{P}_x$ is compact for all $x \in X$ (which we have shown previously).

Now, for any $x \in X$, let

$$\underline{f}(x) := \min_{\eta \in \mathcal{P}_x} \int f(y) \eta(dy).$$

We claim that $\underline{f}$ is lower semicontinuous. Indeed, the correspondence $x \mapsto \mathcal{P}_x$ has a closed graph and hence is upper hemicontinuous. The objective is given by $-\max_{\eta \in \mathcal{P}_x} \int -f(y) \eta(dy)$ where $-f$ is continuous. Thus, by Berge’s maximum theorem, $-\underline{f}$ is upper semicontinuous, and hence $\underline{f}$ is lower semicontinuous. So $\underline{f} \in \mathcal{C}_0$.

We claim that $\underline{f} \in \Phi(\mathcal{P})$. Indeed, for any $P \in \mathcal{P}$ and any $x \in X$, we have

$$\int \underline{f}(y) P(dy | x) = \int \left(\min_{\eta \in \mathcal{P}_y} \int f(z) \eta(dz) \right) P(dy | x) = \min_{Q \in \mathcal{P}} \int f(y) (Q \circ P)(dy | x)$$

where the second equality again uses that $\mathcal{P}$ is rectangular and the Kuratowski–Ryll–Nardzewski measurable selection theorem. Moreover, for any $P \in \mathcal{P}$ and any $x \in X$, note that

$$\min_{Q \in \mathcal{P}} \int f(y) (Q \circ P)(dy | x) \geq \inf_{Q \in \mathcal{P} \otimes \mathcal{P}} \int f(y) Q(dy | x) = \inf_{Q \in \mathcal{P}} \int f(y) Q(dy | x) = \underline{f}(x),$$

where the first equality uses that $\mathcal{P}$ is composition-closed and the second equality uses

that the inf can be attained by the continuity of f and compactness of $\mathcal{P}_x$ with the value given by $\underline{f}(x)$. Together, these imply that for any $P \in \mathcal{P}$ and any $x \in X$,

$$\int \underline{f}(y)P(dy | x) \geq \underline{f}(x),$$

and hence $\underline{f} \in \Phi(\mathcal{P})$. As a result, since $\mu \leq_{\Phi(\mathcal{P})} \nu$, we must have

$$\int \underline{f} d\mu \leq \int \underline{f} d\nu.$$

Moreover, since $\text{Id} \in \mathcal{P}$, we have $\underline{f} \leq f$ by construction. Together, these imply that

$$\int \underline{f} d\nu \leq \int f d\nu < \int \underline{f} d\mu \leq \int \underline{f} d\nu,$$

where the strict inequality was shown earlier using the fact that the separating linear functional is given by f —a contradiction. Therefore, $\leq_{\mathcal{C}} \implies \leq^{\mathcal{P}}$.

Together, these imply that

$$\leq^{\Psi(\mathcal{C})} = \leq_{\mathcal{C}} = \leq^{\mathcal{P}}.$$

Then, since $\Psi(\mathcal{C})$ is rectangular by [Lemma 6](#) and $\mathcal{P}$ is rectangular as it is regular, by [Lemma 5](#), it follows that $\mathcal{P} = \Psi(\mathcal{C})$, as desired.

(e) $\implies$ (a): Suppose $\mathcal{C}$ is max-closed. We claim that $\leq_{\mathcal{C}} = \leq^{\Psi(\mathcal{C})}$. By the same argument as in the proof of (d) $\implies$ (e), we have that $\leq^{\Psi(\mathcal{C})} \implies \leq_{\mathcal{C}}$. Now, to see the other direction, fix any $\mu \leq_{\mathcal{C}} \nu$. Then, $\nu \leq_{[-\mathcal{C}]} \mu$ where the convex cone $[-\mathcal{C}]$ is min-closed. By [Theorem 1](#),²¹ there exists some P such that $\nu = P * \mu$, and $P * \delta_x \leq_{[-\mathcal{C}]} \delta_x$ for all $x \in X$. Thus, for all $x \in X$ and all $g \in \mathcal{C}$, we have

$$\int [-g(y)]P(dy | x) \leq -g(x),$$

and thus

$$\int g(y)P(dy | x) \geq g(x).$$

Therefore, $P \in \Psi(\mathcal{C})$ by definition, which in turn implies that $\mu \leq^{\Psi(\mathcal{C})} \nu$, by definition. Thus, $\leq_{\mathcal{C}} \implies \leq^{\Psi(\mathcal{C})}$. Hence, we have

$$\leq_{\mathcal{C}} = \leq^{\Psi(\mathcal{C})}.$$

²¹Note that the technical assumptions required by [Theorem 1](#) hold for $[-\mathcal{C}]$ given that $\mathcal{C}$ is regular.

Since $\mathcal{P} = \Psi(\mathcal{C})$, it follows $\leq_{\mathcal{C}} = \leq^{\mathcal{P}}$. Hence, $(\mathcal{C}, \mathcal{P})$ is Blackwell-consistent.

Uniqueness: We first show that every Blackwell-consistent pair $(\mathcal{C}, \mathcal{P})$ is maximal in the following sense: $\leq_{\mathcal{C}'} = \leq_{\mathcal{C}} \implies \mathcal{C}' \subseteq \mathcal{C}$ for any $\mathcal{C}' \subseteq \mathcal{C}_0$, and $\leq^{\mathcal{P}'} = \leq^{\mathcal{P}} \implies \mathcal{P}' \subseteq \mathcal{P}$ for any $\mathcal{P}' \subseteq \mathcal{P}_0$.

Fix any Blackwell-consistent pair $(\mathcal{C}, \mathcal{P})$. By $(a) \implies (b)$ direction, we know that $(\mathcal{C}, \mathcal{P})$ is Blackwell-invariant. Thus, we have

$$\mathcal{P} = \Psi(\mathcal{C}) \text{ and } \mathcal{C} = \Phi(\mathcal{P}).$$

First, fix any $\mathcal{C}' \subseteq \mathcal{C}_0$ such that $\leq_{\mathcal{C}'} = \leq_{\mathcal{C}}$. Fix any $g \in \mathcal{C}'$. We claim that $g \in \mathcal{C}$. Indeed, suppose for contradiction $g \notin \mathcal{C}$. Since $g \in \mathcal{C}_0$ and $\mathcal{C} = \Phi(\mathcal{P})$, for $g \notin \mathcal{C}$, there must exist some $P \in \mathcal{P}$ and some $x \in X$ such that

$$\int g(y)P(dy | x) < g(x).$$

By construction, $\delta_x \leq^{\mathcal{P}} P * \delta_x$. Moreover, by construction, $\delta_x \not\leq_{\mathcal{C}'} P * \delta_x$, which implies that $\delta_x \not\leq_{\mathcal{C}} P * \delta_x$ —contradicting to $\leq^{\mathcal{P}} = \leq_{\mathcal{C}}$. This shows that $\mathcal{C}$ is maximal.

Now, we show that $\mathcal{P}$ is maximal. Fix any $\mathcal{P}' \subseteq \mathcal{P}_0$ such that $\leq^{\mathcal{P}'} = \leq^{\mathcal{P}}$. Fix any $P \in \mathcal{P}'$. We claim that $P \in \mathcal{P}$. Indeed, suppose for contradiction $P \notin \mathcal{P}$. Since $P \in \mathcal{P}_0$ and $\mathcal{P} = \Psi(\mathcal{C})$, for $P \notin \mathcal{P}$, there must exist some $g \in \mathcal{C}$ and some $x \in X$ such that

$$\int g(y)P(dy | x) < g(x).$$

By construction, $\delta_x \leq^{\mathcal{P}'} P * \delta_x$, which implies that $\delta_x \leq^{\mathcal{P}} P * \delta_x$. Moreover, by construction, $\delta_x \not\leq_{\mathcal{C}} P * \delta_x$ —contradicting to $\leq^{\mathcal{P}} = \leq_{\mathcal{C}}$. This shows that $\mathcal{P}$ is maximal.

Now, we prove the unique representation result. Fix any Blackwell-consistent order $\leq$. Let $(\mathcal{C}, \mathcal{P}), (\mathcal{C}', \mathcal{P}')$ be two (regular) Blackwell-consistent pairs supporting the order $\leq$, i.e.,

$$\leq_{\mathcal{C}'} = \leq_{\mathcal{C}} = \leq = \leq^{\mathcal{P}} = \leq^{\mathcal{P}'}.$$

By the previous claim about the maximality of $(\mathcal{C}, \mathcal{P})$, we have $\mathcal{C}' \subseteq \mathcal{C}$ and $\mathcal{P}' \subseteq \mathcal{P}$. Now, applying the same claim to the pair $(\mathcal{C}', \mathcal{P}')$ gives $\mathcal{C} \subseteq \mathcal{C}'$ and $\mathcal{P} \subseteq \mathcal{P}'$. Therefore, we conclude $(\mathcal{C}', \mathcal{P}') = (\mathcal{C}, \mathcal{P})$. Thus, there exists a unique pair $(\mathcal{C}, \mathcal{P})$ that represents the Blackwell-consistent order $\leq$. $\square$

A.7 Proof of Proposition 4

Suppose that the constrained problem is not value-eliminating. Fix any prior x , under which the unconstrained optimal signal is unique. Clearly, if $\nu_{\text{KG}} = \delta_x$, then we immediately have that $\nu^* = \delta_x$ is the unique solution for the constrained problem since δ_x is feasible. Thus, suppose otherwise. Then, we must have $\hat{f}^{\text{cav}}(x) > f(x)$. Let

$$S = \text{conv}(\text{supp}(\nu_{\text{KG}})) \setminus \text{supp}(\nu_{\text{KG}}).$$

Fix any constrained optimal ν^* . We claim that $\nu^*(S) = 0$. Suppose for contradiction that $\nu^*(S) > 0$. Then by Corollary 5, there exists some $y \in S$ such that

$$\hat{f}^{\mathcal{C}}(y) = f(y)$$

which by the assumption that the constrained problem is not value-eliminating implies

$$\hat{f}^{\text{cav}}(y) = \hat{f}^{\mathcal{C}}(y) = f(y).$$

Now, note that since ν_{KG} is finitely supported and assigns strictly positive mass to every point in $\text{supp}(\nu_{\text{KG}})$, its barycenter x is in the relative interior of $\text{conv}(\text{supp}(\nu_{\text{KG}}))$. Then, given the above, ν_{KG} cannot be uniquely optimal since we can perturb ν_{KG} by constructing

$$\nu' := \nu_{\text{KG}} + \varepsilon \delta_y - \varepsilon \sum_i \lambda_i \delta_{x_i^{\text{KG}}},$$

for small enough $\varepsilon > 0$, where $\{x_i^{\text{KG}}\}_i := \text{supp}(\nu_{\text{KG}})$ and λ is the barycentric representation of y supported on $\text{supp}(\nu_{\text{KG}})$. Indeed, for small enough $\varepsilon > 0$, ν' is a feasible signal distinct from ν_{KG} , and, since $\hat{f}^{\text{cav}}(y) = f(y)$ and $y \in S$, the perturbed signal weakly improves the objective value and hence is optimal—a contradiction. Thus, $\nu^*(S) = 0$.

However, for ν^* to be Blackwell-dominated by ν_{KG} , it must be that

$$\nu^*(\text{conv}(\text{supp}(\nu_{\text{KG}}))) = 1,$$

which then combined with $\nu^*(S) = 0$ implies that

$$\nu^*(\text{supp}(\nu_{\text{KG}})) = 1,$$

but since ν^* 's barycenter is also x , we must have $\nu^* = \nu_{\text{KG}}$ given that ν_{KG} is the unique solution to the unconstrained problem. Thus, ν^* cannot be strictly Blackwell-dominated

by ν_{KG} . Now, consider the case of $|\Omega| = 2$. Since $\nu^*(S) = 0$, we have that

$$\text{supp}(\nu^*) \subseteq [0, 1] \setminus (x_1^{\text{KG}}, x_2^{\text{KG}}),$$

where $\{x_1^{\text{KG}}, x_2^{\text{KG}}\}$ is the support of ν^{KG} . We claim that ν^* weakly Blackwell-dominates ν^{KG} —indeed, this follows by observing that we can fuse the mass of ν^* supported on $[0, x_1^{\text{KG}}]$ into a single point x_1^* and the mass of ν^* supported on $[x_2^{\text{KG}}, 1]$ into a single point x_2^* . The fused distribution weakly Blackwell-dominates ν_{KG} and is weakly Blackwell-dominated by ν^* . Thus, the result holds. $\square$

A.8 Proof of Proposition 5

Note that $\mathcal{Q}^d$ is rectangular by definition. Moreover, since $\hat{d}(x_B, x_D, x_B) = x_D$ and $\mathcal{P}_M$ contains the identity kernel, $\mathcal{Q}^d$ contains the identity kernel as well. Lastly, since d is continuous and $\mathcal{P}_M$ has a closed graph and is convex, $\mathcal{Q}^d$ has a closed graph and is convex as well. Therefore, $\mathcal{Q}^d$ is regular for any $d \in \mathcal{D}$.

($\Leftarrow$): Suppose that d is divisible. By Theorem 3, it suffices to prove that $\mathcal{Q}^d$ is composition-closed and that $\mathcal{C}^d = \Phi(\mathcal{Q}^d)$ is regular, since these then imply that $(\mathcal{C}^d, \mathcal{Q}^d)$ is a regular pair, and moreover by Theorem 3, a Blackwell-consistent pair—and thus, $\leq^{\mathcal{Q}^d}$ is a Blackwell-consistent order.

To see that $\mathcal{Q}^d$ is composition-closed, consider any $Q_1, Q_2 \in \mathcal{Q}^d$. Fix any (x_B, x_D) . By definition, $Q_1(\cdot \mid x_B, x_D)$ is the distribution of

$$(Y, \hat{d}(x_B, x_D, Y)),$$

where $Y \sim P_1^{(x_B, x_D)}(\cdot \mid x_B)$ for some $P_1^{(x_B, x_D)} \in \mathcal{P}_M$. Moreover, $Q_2 \circ Q_1(\cdot \mid x_B, x_D)$ is the distribution of

$$(Z, \hat{d}(Y, \hat{d}(x_B, x_D, Y), Z)),$$

where $Z \sim P_2^{(x_B, x_D)}(\cdot \mid y)$ conditional on $Y = y$ for some $P_2^{(x_B, x_D)} \in \mathcal{P}_M$. Note that, for any $y, z \in X^0$, since d is divisible, by letting

$$\tilde{y} := \frac{x_D \otimes x_B \odot y}{\langle x_D \otimes x_B, y \rangle} \quad \text{and} \quad \tilde{z} := \frac{x_D \otimes x_B \odot z}{\langle x_D \otimes x_B, z \rangle},$$

we have

$$\begin{aligned}
\hat{d}(y, \hat{d}(x_B, x_D, y), z) &= d \left(d \left(x_D, \frac{x_D \otimes x_B \odot y}{\langle x_D \otimes x_B, y \rangle} \right), \frac{d \left(x_D, \frac{x_D \otimes x_B \odot y}{\langle x_D \otimes x_B, y \rangle} \right) \otimes y \odot z}{\langle d \left(x_D, \frac{x_D \otimes x_B \odot y}{\langle x_D \otimes x_B, y \rangle} \right) \otimes y, z \rangle} \right) \\
&= d \left(d(x_D, \tilde{y}), \frac{d(x_D, \tilde{y}) \otimes \tilde{y} \odot \tilde{z}}{\langle d(x_D, \tilde{y}) \otimes \tilde{y}, \tilde{z} \rangle} \right) \\
&= d_{\text{II}}(x_D, \tilde{z}; \tilde{y}) \\
&= d(x_D, \tilde{z}) \\
&= d \left(x_D, \frac{x_D \otimes x_B \odot z}{\langle x_D \otimes x_B, z \rangle} \right) \\
&= \hat{d}(x_B, x_D, z).
\end{aligned}$$

This also extends to all $y, z \in X$ by continuity of d . Moreover, since $Z \sim P_2^{(x_B, x_D)}(\cdot \mid y)$ conditional on $Y = y$, since $Y \sim P_1^{(x_B, x_D)}(\cdot \mid x_B)$, and $P_1^{(x_B, x_D)}, P_2^{(x_B, x_D)} \in \mathcal{P}_M$,

$$\mathbb{E}[Z] = \int_X \left(\int_X z P_2^{(x_B, x_D)}(dz \mid y) \right) P_1^{(x_B, x_D)}(dy \mid x_B) = \int_X y P_1^{(x_B, x_D)}(dy \mid x_B) = x_B.$$

As a result, $Q_2 \circ Q_1$ is the distribution of

$$(Z, \hat{d}(x_B, x_D, Z)),$$

for some random variable Z with $\mathbb{E}[Z] = x_B$, and hence $Q_2 \circ Q_1(\cdot \mid x_B, x_D) \in \mathcal{Q}_{(x_B, x_D)}^d$. Since (x_B, x_D) is arbitrary, we have that $Q_2 \circ Q_1 \in \mathcal{Q}^d$, and thus $\mathcal{Q}^d$ is composition-closed.

To see that $\mathcal{C}^d = \Phi(\mathcal{Q}^d)$ is regular. Note that by [Lemma 6](#), $\mathcal{C}^d$ is a convex cone that contains $1, -1$ and is closed under bounded increasing pointwise limits, with $\mathcal{C}^d \cap C(X)$ being closed under the sup-norm. Thus, it remains to prove that $\mathcal{C}^d$ can be approximated by continuous functions in $\mathcal{C}^d$ from below. Note that by Theorem 1 of [Cripps \(2018\)](#), there exists a homeomorphism $F : X \rightarrow X$ such that for all $x, y \in X^0$,

$$d(x, y) = F^{-1} \left(\frac{F(x) \otimes x \odot y}{\langle F(x) \otimes x, y \rangle} \right).$$

Recall that $d^B(x, y) = y$ denotes the Bayesian updating rule. It follows that

$$F(\hat{d}(x_B, x_D, y)) = \hat{d}^B(x_B, F(x_D), y),$$

for all $x_B, x_D, y \in X^0$.

Let $\xi : X \times X \rightarrow X \times X$ be defined as

$$\xi(x_B, x_D) := (x_B, F(x_D)),$$

for all $(x_B, x_D) \in X \times X$. Since F is a homeomorphism, ξ is a homeomorphism as well. We claim that for all $g \in \mathcal{C}_0$,

$$g \in \mathcal{C}^d \iff g \circ \xi^{-1} \in \mathcal{C}^{d^B},$$

where $\mathcal{C}^{d^B} = \Phi(\mathcal{Q}^{d^B})$. Indeed, note that, for any $g \in \mathcal{C}^d$, for any $Q \in \mathcal{Q}^d$, and for any $(x_B, x_D) \in X \times X$, letting $\eta \in (\mathcal{P}_M)_{x_B}$ be the distribution with barycenter x_B that defines $Q(\cdot \mid x_B, x_D)$, we can write

$$\int g(y_B, y_D) Q(dy_B, dy_D \mid x_B, x_D) = \int g(y, \hat{d}(x_B, x_D, y)) \eta(dy) = \int g \circ \xi^{-1}(y, \hat{d}^B(x_B, F(x_D), y)) \eta(dy).$$

Thus, $g \in \mathcal{C}^d$ if and only if for all $(x_B, x_D) \in X \times X$, we have

$$\min_{\delta_{x_B} \leq_{\text{convex}} \eta} \int g \circ \xi^{-1}(y, \hat{d}^B(x_B, F(x_D), y)) \eta(dy) = g(x_B, x_D) = g \circ \xi^{-1}(x_B, F(x_D)),$$

which, since F is a homeomorphism, is equivalent to that for all $(x_B, x'_D) \in X \times X$, we have

$$\min_{\delta_{x_B} \leq_{\text{convex}} \eta} \int g \circ \xi^{-1}(y, \hat{d}^B(x_B, x'_D, y)) \eta(dy) = g \circ \xi^{-1}(x_B, x'_D),$$

which is equivalent to $g \circ \xi^{-1} \in \Phi(\mathcal{Q}^{d^B}) = \mathcal{C}^{d^B}$, as desired.

Therefore, since ξ is a homeomorphism, it suffices to show that every $g \in \mathcal{C}^{d^B}$ can be approximated by a sequence of continuous functions from below. To prove this, let

$$\ell_{x_B, x_D}(y) := \hat{d}^B(x_B, x_D, y),$$

for all $(x_B, x_D, y) \in X \times X \times X$. Since $\hat{d}^B : X \times X \times X \rightarrow X$ is continuous and since X is compact,

$$\ell_{x_B, x_D} : X \rightarrow X$$

is continuous for all $(x_B, x_D) \in X \times X$. Let $C(X, X)$ be the set of continuous functions from X to X endowed with the sup metric

$$d_\infty(\ell_1, \ell_2) := \sup_{y \in X} \|\ell_1(y) - \ell_2(y)\|.$$

Since X is compact metric space, $(C(X, X), d_\infty)$ is also a metric space.

Now, let $\Gamma : X \times X \rightarrow C(X, X)$ be defined as

$$\Gamma(x_B, x_D) := \ell_{x_B, x_D}$$

for all $(x_B, x_D) \in X \times X$. We claim that Γ is continuous. Indeed, suppose that there is a sequence $\{(x_B^n, x_D^n)\}$ that converges to some (x_B, x_D) in $X \times X$. Since $\hat{d}^B$ is continuous on the compact set $X \times X \times X$, it is uniformly continuous. Therefore, we have

$$\lim_{n \rightarrow \infty} \sup_{y \in X} \|\hat{d}^B(x_B^n, x_D^n, y) - \hat{d}^B(x_B, x_D, y)\| = 0,$$

and hence

$$\lim_{n \rightarrow \infty} d_\infty(\ell_{x_B^n, x_D^n}, \ell_{x_B, x_D}) = 0.$$

Thus, Γ is continuous. Let

$$\Lambda := \left\{ \ell_{x_B, x_D} : (x_B, x_D) \in X \times X \right\} \subset C(X, X).$$

Since $X \times X$ is compact and Γ is continuous, its image Λ is compact in $C(X, X)$ with the sup metric. In particular, Λ is compact and metrizable.

Now note that since $\hat{d}^B$ is a divisible updating rule, as shown before, we have

$$\hat{d}^B(y, \hat{d}^B(x_B, x_D, y), z) = \hat{d}^B(x_B, x_D, z),$$

for all $x_B, x_D, y, z \in X$. It then follows that

$$[\ell_{y, \ell_{x_B, x_D}(y)}](z) = [\ell_{x_B, x_D}](z).$$

Thus, for any $\ell \in \Lambda$, we have $\ell = \ell_{x, \ell(x)}$ for all $x \in X$. As a result, for $\Theta : X \times X \rightarrow X \times \Lambda$ defined as

$$\Theta(x_B, x_D) := (x_B, \ell_{x_B, x_D}),$$

we have that Θ is a homeomorphism. Indeed, Θ is surjective since for any $(x, \ell) \in X \times \Lambda$, $\Theta(x, \ell(x)) = (x, \ell_{x, \ell(x)}) = (x, \ell)$, where $(x, \ell(x)) \in X \times X$. Meanwhile, Θ is injective because

$$\ell_{x_B, x_D}(x_B) = x_D,$$

and hence $\Theta(x_B, x_D) = \Theta(x'_B, x'_D) \implies (x_B, x_D) = (x'_B, x'_D)$. Together, since Θ is continuous, $\Theta : X \times X \rightarrow X \times \Lambda$ is a homeomorphism.

Note that for all $x_B, x_D, y \in X$

$$\Theta(y, \hat{d}^B(x_B, x_D, y)) = (y, \ell_{y, \hat{d}^B(x_B, x_D, y)}) = (y, \ell_{y, \ell_{x_B, x_D}(y)}) = (y, \ell_{x_B, x_D}),$$

where the last equality uses the divisibility of d^B .

Now, let $\mathcal{U}$ be the set of bounded lower semicontinuous functions $u : X \times \Lambda \rightarrow \mathbb{R}$ such that $x \mapsto u(x, \ell)$ is convex for all $\ell \in \Lambda$. We claim that

$$g \in \mathcal{C}^{d^B} \iff g \circ \Theta^{-1} \in \mathcal{U},$$

or, equivalently

$$u \in \mathcal{U} \iff u \circ \Theta \in \mathcal{C}^{d^B}.$$

Indeed, for any $g \in \mathcal{C}^{d^B}$, since g is bounded and lower semicontinuous and Θ is continuous, we have that $g \circ \Theta^{-1}$ is bounded and lower semicontinuous. Moreover, note that $g \in \mathcal{C}^{d^B}$ if and only if for all $(x_B, x_D) \in X \times X$ we have

$$\min_{\delta_{x_B} \preceq_{\text{convex}} \eta} \int g(y, l_{x_B, x_D}(y)) \eta(dy) = \min_{\delta_{x_B} \preceq_{\text{convex}} \eta} \int g(y, \hat{d}^B(x_B, x_D, y)) \eta(dy) = g(x_B, x_D).$$

Note that for all $(x_B, x_D) \in X \times X$, we can write

$$\int g(y, l_{x_B, x_D}(y)) \eta(dy) = \int g \circ \Theta^{-1} \circ \Theta(y, l_{x_B, x_D}(y)) \eta(dy) = \int g \circ \Theta^{-1}(y, \ell_{x_B, x_D}) \eta(dy),$$

and write

$$g(x_B, x_D) = g \circ \Theta^{-1} \circ \Theta(x_B, x_D) = g \circ \Theta^{-1}(x_B, \ell_{x_B, x_D}).$$

Thus, $g \in \mathcal{C}^{d^B}$ if and only if for all $(x_B, x_D) \in X \times X$

$$\min_{\delta_{x_B} \preceq_{\text{convex}} \eta} \int g \circ \Theta^{-1}(y, \ell_{x_B, x_D}) \eta(dy) = g \circ \Theta^{-1}(x_B, \ell_{x_B, x_D}).$$

It is immediate that if for any $\ell \in \Lambda$, $x \mapsto g \circ \Theta^{-1}(x, \ell)$ is convex, then the above holds for all $(x_B, x_D) \in X \times X$. Conversely, suppose the above holds for all $(x_B, x_D) \in X \times X$. Suppose for contradiction that there exists some $\ell \in \Lambda$ such that $x \mapsto g \circ \Theta^{-1}(x, \ell)$ is not convex. Then, there exists some $x_B \in X$ such that

$$\min_{\delta_{x_B} \preceq_{\text{convex}} \eta} \int g \circ \Theta^{-1}(y, \ell) \eta(dy) < g \circ \Theta^{-1}(x_B, \ell).$$

By definition of Λ , there exists some $(x'_B, x'_D) \in X \times X$ such that $\ell = \ell_{x'_B, x'_D}$. Now, consider

$$x_D := \hat{d}^B(x'_B, x'_D, x_B).$$

Note that, by divisibility of $\hat{d}^B$, for all $y \in X$, we have

$$\ell_{x_B, x_D}(y) = \hat{d}^B(x_B, \hat{d}^B(x'_B, x'_D, x_B), y) = \hat{d}^B(x'_B, x'_D, y) = \ell_{x'_B, x'_D}(y) = \ell(y).$$

Thus, $\ell_{x_B, x_D} = \ell$. Hence, we have

$$\min_{\delta_{x_B} \leq_{\text{convex}} \eta} \int g \circ \Theta^{-1}(y, \ell_{x_B, x_D}) \eta(dy) < g \circ \Theta^{-1}(x_B, \ell_{x_B, x_D}),$$

a contradiction. This proves that

$$g \in \mathcal{C}^{d^B} \iff g \circ \Theta^{-1} \in \mathcal{U}.$$

Since Θ is a homeomorphism, to show that any $g \in \mathcal{C}^{d^B}$ can be approximated by continuous functions in $\mathcal{C}^{d^B}$ from below, it suffices to show that this holds for all $u \in \mathcal{U}$. Recall that $u \in \mathcal{U}$ if and only if

$$u : X \times \Lambda \rightarrow \mathbb{R}$$

is bounded lower semicontinuous and satisfies that $x \mapsto u(x, \ell)$ is convex for all ℓ . Since $X \subseteq \mathbb{R}^n$ is a compact, convex, metrizable subset of a locally convex Hausdorff topological vector space, and since Λ is a metric space, by [Lemma 8](#), any $u \in \mathcal{U}$ can be approximated from below by a sequence of continuous functions $\{u_n\} \subseteq \mathcal{U}$, as desired.

($\implies$): Suppose that $\leq^{\mathcal{Q}^d}$ is a Blackwell-consistent order. Then, by [Theorem 3](#), there exists a Blackwell-consistent regular pair $(\tilde{\mathcal{C}}^d, \tilde{\mathcal{Q}}^d)$ such that $\leq^{\mathcal{Q}^d} = \leq_{\tilde{\mathcal{C}}^d} = \leq^{\tilde{\mathcal{Q}}^d}$ where $\tilde{\mathcal{Q}}^d$ is composition-closed. Moreover, by the uniqueness part of [Theorem 3](#), it follows that $\mathcal{Q}^d = \tilde{\mathcal{Q}}^d$ and thus $\mathcal{Q}^d$ is composition-closed.

Now, suppose for contradiction that d is not divisible. Then, there exist $x, y, z \in X^0$, such that $d_{\Pi}(x, z; y) \neq d(x, z)$. Since $d(x', x') = x'$ for all $x' \in X$, it must be that $x \neq y$ and $y \neq z$. Moreover, since $x, y, z \in X^0$, there exist $\varepsilon, \xi, \zeta > 0$ such that

$$y^{x, \varepsilon} := \frac{x - \varepsilon y}{1 - \varepsilon}, \quad z^{x, \xi} := \frac{x - \xi z}{1 - \xi}, \quad \text{and } z^{y, \zeta} := \frac{y - \zeta z}{1 - \zeta}$$

are in X^0 , and that $\varepsilon y + (1 - \varepsilon)y^{x, \varepsilon} = x$, $\xi z + (1 - \xi)z^{x, \xi} = x$, and $\zeta z + (1 - \zeta)z^{y, \zeta} = y$.

Now, let

$$P_1(\cdot \mid \tilde{x}) := \begin{cases} \frac{1}{2}(\varepsilon\delta_y + (1-\varepsilon)\delta_{y^{x,\varepsilon}}) + \frac{1}{2}(\xi\delta_z + (1-\xi)\delta_{z^{x,\xi}}), & \text{if } \tilde{x} = x \\ \delta_{\tilde{x}}, & \text{otherwise} \end{cases},$$

and

$$P_2(\cdot \mid \tilde{x}) := \begin{cases} \zeta\delta_z + (1-\zeta)\delta_{z^{y,\zeta}}, & \text{if } \tilde{x} = y \\ \delta_{\tilde{x}}, & \text{otherwise} \end{cases}.$$

By construction, $P_1, P_2 \in \mathcal{P}_M$. Then, let $Q_1 := Q^{P_1, d}$ and $Q_2 := Q^{P_2, d}$ where $Q^{P, d}$ is defined as the kernel that sends

$$(x_B, x_D) \mapsto (Y, \hat{d}(x_B, x_D, Y))$$

where $Y \sim P(\cdot \mid x_B)$. By construction, $Q_1, Q_2 \in \mathcal{Q}^d$.

Moreover, by construction, $Q_2 \circ Q_1(\cdot \mid x, x)$ must assign at least probability $\varepsilon/2 \cdot \zeta > 0$ on $\{(z, d_{\Pi}(x, z; y))\}$ through the belief path of $x \mapsto y \mapsto z$. Since $d_{\Pi}(x, z; y) \neq d(x, z)$, we then have

$$Q_2 \circ Q_1\left(\left\{(\tilde{x}_B, \tilde{x}_D) : \tilde{x}_D = d(x, \tilde{x}_B)\right\} \mid x, x\right) \leq 1 - \frac{\varepsilon}{2} \cdot \zeta < 1.$$

However, since $\mathcal{Q}^d$ is composition-closed, $Q_2 \circ Q_1 \in \mathcal{Q}^d$. Hence, $Q_2 \circ Q_1(\cdot \mid x, x)$ must be the distribution of $(Z, d(x, Z))$, where $Z \sim P(\cdot \mid x)$ for some $P \in \mathcal{P}_M$. In particular, it must be that

$$Q_2 \circ Q_1\left(\left\{(\tilde{x}_B, \tilde{x}_D) : \tilde{x}_D = d(x, \tilde{x}_B)\right\} \mid x, x\right) = 1,$$

a contradiction. Thus, d must be divisible. $\square$

A.9 Proof of Proposition 6

(i): Consider any ambiguity-averse preference $\leq$ defined by representation $\mathcal{V}$. Let $\tilde{\leq}$ and $\tilde{V}$ be the implied preference relation over lotteries, and its representation, respectively. That is

$$\mu_1 \tilde{\leq} \mu_2 \iff A_{\mathcal{C}}(\mu_1) \leq A_{\mathcal{C}}(\mu_2),$$

for all $\mu_1, \mu_2 \in \Delta(X)$; and

$$\tilde{V}(\mu) := \mathcal{V}(A_{\mathcal{C}}(\mu)),$$

for all $\mu \in \Delta(X)$.

We first prove the “if” part. Suppose that $\mathcal{C}$ is min-closed. By Theorem 1, for all

$\mu \in \Delta(X)$, we have

$$\begin{aligned}
\tilde{V}(\mu) &= \alpha \cdot \max_{\nu: \nu \leq_C \mu} \int u(x) \nu(dx) - (1 - \alpha) \cdot \max_{\nu: \nu \leq_C \mu} \int -u(x) \nu(dx) \\
&= \alpha \int \bar{u}(x) \mu(dx) - (1 - \alpha) \cdot \int \overline{-u}(x) \mu(dx) \\
&= \int \underbrace{\left[\alpha \cdot \bar{u}(x) - (1 - \alpha) \cdot \overline{-u}(x) \right]}_{\hat{u}(x)} \mu(dx),
\end{aligned}$$

which is an affine function of μ , and hence an expected utility representation.

For the “only if” part, suppose that $\mathcal{C}$ is not min-closed. By [Theorem 1](#), there exists $f : X \rightarrow \mathbb{R}$ such that V_f^* is not affine. Note that V_f^* corresponds to a representation $\tilde{V}$ when $u = f$ and $\alpha = 1$. Since V_f^* is not affine, it is not an expected utility representation.

(ii): Suppose that $\mathcal{C}$ is strongly non-min-closed. Let $f \in C(X)$ and $\mu_0 \in \Delta(X)$ that is absolutely continuous with a density bounded away from zero, be such that

$$\min_{g \in \mathcal{C}, g \geq f} \int g d\mu_0$$

has two distinct solutions g_1, g_2 that are not affinely related. We first show that for this f , $V_f^*(\mu)$ cannot be written as an increasing transformation of an expected utility. That is, there does not exist a strictly increasing function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ and a bounded upper semicontinuous $u : X \rightarrow \mathbb{R}$ such that

$$V_f^*(\mu) = \phi \left(\int u d\mu \right)$$

for all $\mu \in \Delta(X)$. To this end, suppose the contrary and let ϕ and u be such functions. We first claim that, for any supergradient g of V_f^* at μ_0 , we have $g \in \text{span}\{u, 1\}$ μ_0 -a.e. Indeed, consider any finite signed measure $\eta \ll \mu_0$ whose Radon–Nikodym derivative $d\eta/d\mu_0$ is essentially bounded with $\eta(X) = \int u d\eta = 0$. Then, there exists $\varepsilon > 0$ such that $\mu_1^\varepsilon := \mu_0 + \varepsilon\eta \in \Delta(X)$ and $\mu_2^\varepsilon := \mu_0 - \varepsilon\eta \in \Delta(X)$. Since $\int u d\eta = 0$, we have

$$\int u d\mu_1^\varepsilon = \int u d\mu_0 = \int u d\mu_2^\varepsilon,$$

and hence, for each $i \in \{1, 2\}$,

$$V_f^\star(\mu_i^\varepsilon) = \phi \left(\int u \, d\mu_i^\varepsilon \right) = \phi \left(\int u \, d\mu_0 \right) = V_f^\star(\mu_0).$$

Thus, since g is a supergradient of $V_f^\star$ at μ_0 ,

$$V_f^\star(\mu_0) = V_f^\star(\mu_i^\varepsilon) \leq V_f^\star(\mu_0) + \int g \, d(\mu_i^\varepsilon - \mu_0)$$

for $i \in \{1, 2\}$. Therefore,

$$\int g \, d(\mu_1^\varepsilon - \mu_0) = \varepsilon \int g \, d\eta \geq 0$$

and

$$\int g \, d(\mu_2^\varepsilon - \mu_0) = -\varepsilon \int g \, d\eta \geq 0.$$

Together, we have $\int g \, d\eta = 0$ for all supergradient g of $V_f^\star$ at μ_0 and for all such η satisfying $\int 1 \, d\eta = \int u \, d\eta = 0$. Now, consider the linear map T that maps from

$$M^\infty(\mu_0) := \left\{ \eta \in M(X) : \eta \ll \mu_0, \frac{d\eta}{d\mu_0} \in L^\infty(\mu_0) \right\}$$

to $\mathbb{R}^2$, defined as

$$T(\eta) := \left(\int 1 \, d\eta, \int u \, d\eta \right),$$

and also consider the linear functional

$$\psi_g(\eta) := \int g \, d\eta.$$

We then have that $\ker(T) \subseteq \ker(\psi_g)$. It follows that $\psi_g = l \circ T$ for some linear l . Therefore, there exist $a, b \in \mathbb{R}$ such that for all $\eta \in M^\infty(\mu_0)$,

$$\int g \, d\eta = \langle (a, b), T(\eta) \rangle = \int (bu + a) \, d\eta.$$

Thus, we have $g = bu + a$, and hence $g \in \text{span}\{u, 1\}$, μ_0 -a.e.

In the meantime, by the same argument as the proof of [Theorem 1](#) (a) $\implies$ (b), the duality gap is zero, and thus

$$V_f^\star(\mu) = \min_{g \in \mathcal{C}, g \geq f} \int g \, d\mu$$

for all $\mu \in \Delta(X)$. In particular,

$$\int g_1 d\mu_0 = \int g_2 d\mu_0 = V_f^*(\mu_0).$$

Thus, for any $\mu \in \Delta(X)$, and for each $i \in \{1, 2\}$, since

$$\int f d\nu \leq \int g_i d\nu \leq \int g_i d\mu$$

for all $\nu \leq_C \mu$, it must be that for all $\mu \in \Delta(X)$ and for $i \in \{1, 2\}$,

$$V_f^*(\mu) \leq \int g_i d\mu = V_f^*(\mu_0) + \int g_i d(\mu - \mu_0).$$

That is, both g_1 and g_2 are supergradients of V_f^* at μ_0 . Therefore, by the previous argument, we have that $g_1, g_2 \in \text{span}\{u, 1\}$ μ_0 -a.e., which implies that g_1, g_2 are affinely related since g_1 and g_2 are continuous functions and μ_0 has full support—contradicting the assumption that g_1 and g_2 are not affinely related.

Now, consider the preference $\tilde{\leq}$ defined by

$$\mu_1 \tilde{\leq} \mu_2 \iff V_f^*(\mu_1) \leq V_f^*(\mu_2),$$

for all $\mu_1, \mu_2 \in \Delta(X)$. Then, since V_f^* corresponds to a representation $\tilde{V}$ when $u = f$ and $\alpha = 1$. Since V_f^* cannot be written as a monotone transformation of an affine function, it does not represent an expected utility preference $\square$

References

- AHN, D. S. (2008): “Ambiguity Without a State Space,” *The Review of Economic Studies*, 75, 3–28.
- ALONSO, R. AND O. CÂMARA (2016): “Bayesian persuasion with heterogeneous priors,” *Journal of Economic Theory*, 165, 672–706.
- AUGIAS, V. AND L. UHE (2025): “The Economics of Convex Function Intervals,” .
- AUMANN, R. J. AND M. MASCHLER (1995): *Repeated Games with Incomplete Information*, Cambridge, MA: MIT Press, with the collaboration of Richard E. Stearns.
- AZRIELI, Y. AND R. DAS (2025): “Sequential Non-Bayesian Persuasion,” *arXiv preprint arXiv:2508.09464*.
- BALL, I. AND J.-A. ESPÍN-SÁNCHEZ (2021): “Experimental persuasion,” *Working Paper*.

- BEIGLBÖCK, M. AND M. NUTZ (2014): “Martingale Inequalities and Deterministic Counterparts,” *Electronic Journal of Probability*, 19, 1–15.
- BLACKWELL, D. (1951): “Comparison of Experiments,” in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 93–102.
- BROCAS, I. AND J. D. CARRILLO (2007): “Influence through ignorance,” *The RAND Journal of Economics*, 38, 931–947.
- CANDOGAN, O. AND P. STRACK (2021): “Optimal Disclosure of Information to a Privately Informed Receiver,” .
- CHEN, K. (2025): “Experiments in the Linear Convex Order,” *arXiv preprint arXiv:2502.06530*.
- ÇINLAR, E. (2010): *Probability and Stochastics*, Springer.
- CIOSMAK, K. J. (2023): “Applications of Strassen’s Theorem and Choquet Theory to Optimal Transport Problems, to Uniformly Convex Functions and to Uniformly Smooth Functions,” *Nonlinear Analysis*, 232, 113267.
- CRIPPS, M. (2018): “Divisible Updating,” Working Paper.
- DAYANIK, S. AND I. KARATZAS (2003): “On the optimal stopping problem for one-dimensional diffusions,” *Stochastic processes and their applications*, 107, 173–212.
- DE CLIPPEL, G. AND X. ZHANG (2022): “Non-bayesian persuasion,” *Journal of Political Economy*, 130, 2594–2642.
- DOVAL, L. AND V. SKRETA (2024): “Constrained information design,” *Mathematics of Operations Research*, 49, 78–106.
- DRUSVYATSKIY, D. AND A. S. LEWIS (2011): “Generic Nondegeneracy in Convex Optimization,” *Proceedings of the American Mathematical Society*, 139, 2519–2527.
- DWORCZAK, P. AND E. V. MUIR (2024): “A Mechanism-Design Approach to Property Rights,” Working Paper.
- DWORCZAK, P. AND A. PAVAN (2022): “Preparing for the Worst but Hoping for the Best: Robust (Bayesian) Persuasion,” *Econometrica*, 90, 2017–2051.
- DYNKIN, E. B. (1963): “Optimal choice of the stopping moment of a Markov process,” in *Doklady Akademii Nauk*, Russian Academy of Sciences, vol. 150, 238–240.
- GENTZKOW, M. AND E. KAMENICA (2014): “Costly persuasion,” *American Economic Review*, 104, 457–462.
- HE, K., F. SANDOMIRSKIY, AND O. TAMUZ (2024): “Private private information,” *arXiv preprint arXiv:2112.14356*.
- HENRY, E. AND M. OTTAVIANI (2019): “Research and the approval process: The organiza-

- tion of persuasion,” *American Economic Review*, 109, 911–955.
- KAMENICA, E. AND M. GENTZKOW (2011): “Bayesian persuasion,” *American Economic Review*, 101, 2590–2615.
- KECHRIS, A. (2012): *Classical descriptive set theory*, Springer Science & Business Media.
- KEMPERMAN, J. H. B. (1968): “The General Moment Problem, a Geometric Approach,” *The Annals of Mathematical Statistics*, 39, 93–122.
- KLEE JR., V. (1958): “Extremal Structure of Convex Sets. II,” *Mathematische Zeitschrift*, 69, 90–104.
- KLEINER, A., B. MOLDOVANU, AND P. STRACK (2021): “Extreme points and majorization: Economic applications,” *Econometrica*, 89, 1557–1593.
- KLEINER, A., B. MOLDOVANU, P. STRACK, AND M. WHITMEYER (2024): “The Extreme Points of Fusions,” *arXiv preprint arXiv:2409.10779*.
- KOBAYASHI, M. (2025): “Dynamic non-Bayesian persuasion,” *Working paper*.
- LEHMANN, E. L. (1988): “Comparing Location Experiments,” *The Annals of Statistics*, 16, 521–533.
- LI, F. AND P. NORMAN (2021): “Sequential Persuasion,” *Theoretical Economics*, 16, 639–675.
- LUKEŠ, J., J. MALÝ, I. NETUKA, AND J. SPURNÝ (2009): *Integral representation theory: applications to convexity, Banach spaces and potential theory*, vol. 35, Walter de Gruyter.
- MEYER, P. A. (1966): *Probability and potentials [by] Paul A. Meyer.*, A Blaisdell book in pure and applied mathematics, Waltham, Mass: Blaisdell Pub. Co.
- MORRIS, S. AND P. STRACK (2019): “The wald problem and the relation of sequential sampling and ex-ante information costs,” *Available at SSRN 2991567*.
- MYERSON, R. B. (1981): “Optimal Auction Design,” *Mathematics of Operations Research*, 6(1), 58–73.
- NI, B., W. SHEN, AND P. TANG (2023): “Sequential persuasion using limited experiments,” *arXiv preprint arXiv:2303.10619*.
- NIKZAD, A. (2023): “Constrained Majorization: Applications in Mechanism Design,” *Working Paper*.
- OLSZEWSKI, W. (2007): “Preferences over sets of lotteries,” *The Review of Economic Studies*, 74, 567–595.
- ROCKAFELLAR, R. T. (1970): *Convex Analysis*, vol. 28 of *Princeton Mathematical Series*, Princeton, NJ: Princeton University Press.
- ROST, H. (1971): “The stopping distributions of a Markov process,” *Inventiones mathematicae*, 14, 1–16.

- ROY, A. (1976): “Extremal dilations and their associated extremal measures,” *Journal of Functional Analysis*, 22, 8–31.
- SIMON, B. (2011): *Convexity: An Analytic Viewpoint*, Cambridge Tracts in Mathematics, Cambridge University Press.
- STRACK, P. AND K. H. YANG (2024): “Privacy-Preserving Signals,” *Econometrica*, 92, 1907–1938.
- STRASSEN, V. (1965): “The Existence of Probability Measures with Given Marginals,” *Annals of Mathematical Statistics*, 36(2), 423–439.
- VILLANI, C. (2003): *Topics in Optimal Transportation*, Graduate studies in mathematics, American Mathematical Society.
- WALD, A. (1945): “Sequential Tests of Statistical Hypotheses,” *The Annals of Mathematical Statistics*, 16, 117–186.
- WHITMEYER, M. (forthcoming): “Blackwell-monotone updating rules,” *Journal of Political Economy*.
- WU, W. (2023): “Sequential bayesian persuasion,” *Journal of Economic Theory*, 214, 105763.
- YANG, F. AND K. H. YANG (2025): “Multidimensional monotonicity and economic applications,” *arXiv preprint arXiv:2502.18876*.
- YANG, K. H. AND A. K. ZENTEFIS (2024): “Monotone function intervals: Theory and applications,” *American Economic Review*, 114, 2239–2270.