Extreme Points and Majorization

Andreas Kleiner Benny Moldovanu Philipp Strack [*]

January 7, 2026

Abstract

A key insight is that many, seemingly different, economic problems share a common mathematical structure: they all involve the maximization of a functional over sets of monotonic functions that are either majorized by, or majorize, a given function. We first present new, simpler proofs for the main characterization results of the extreme points of sets defined by monotonicity and majorization constraints obtained by Kleiner, Moldovanu, and Strack (2021). We then demonstrate how the characterization results can be fruitfully applied to a broad range of economic applications, from auction and information design to decision problems under risk such as optimal stopping. Finally, we conclude with an overview of recent, related work that extends these characterizations to settings with additional constraints, multidimensional state spaces, and alternative stochastic orders.

1 Introduction

This paper provides an overview of the applications of majorization and extreme points characterizations to the solution of various economic design problems. We begin by reviewing

[*]Kleiner and Moldovanu: University of Bonn; Strack: Yale University. Kleiner and Moldovanu acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2126/2 390838866 and under CRC-TR 224 (B01 and B02).

the concept of majorization and its role in Economics, and then present new, simpler proofs for the main characterization results of the extreme points of sets defined by monotonicity and majorization constraints obtained by Kleiner, Moldovanu, and Strack (2021). The key unifying insight is that many, seemingly different, economic problems share a common mathematical structure: they all involve the maximization of a functional over sets of monotonic functions that are either majorized by, or majorize, a given function. We demonstrate how the characterization results can be applied to auction design, contest design, information design, optimal delegation, optimal stopping, and decision problems under risk. For pedagogical reasons we also revisit famous results such as Strassen's Theorem from probability theory and show how the extreme points approach yields intuitive and simple proofs.

For each application, we explain why the problem falls within our framework and how extreme point characterizations constitute a very useful analytical tool. We conclude by surveying some of the more recent, related work that extends these characterizations to settings with additional constraints, multidimensional state spaces, and alternative stochastic orders.

2 Extreme Points and Majorization

2.1 Majorization Preliminaries

Throughout the paper, we consider the set M of real-valued, non-decreasing, integrable functions defined on the interval $[0, 1]$ endowed with the L^1-norm. [1] The monotonicity constraint arises in applications for various reasons: for example, since cumulative distribution functions are non-decreasing, the monotonicity constraint arises in information design problems where a prior distribution is given and a posterior one is sought; since incentive-compatible mechanisms often dictate non-decreasing allocations, the monotonicity constraint also arises

[1]Formally, $f \in L^1$ is an equivalence class of functions that are equal almost everywhere. An equivalence class $f \in L^1$ is non-decreasing if it contains a non-decreasing element. Whenever f is non-decreasing, it has a non-decreasing and right-continuous element, which we use as a canonical representative.

in various mechanism design problems.

We will see how an additional majorization constraint captures resource constraints on the available objects and prizes in auction and contest design, constraints on feasible information structures in information design, and constraints induced by the absence of transfers in delegation problems. Formally, for two functions $f, g \in M$ we say that f *majorizes* g, denoted by $f \succeq g$ if the following two conditions hold:

$$\begin{aligned} &\int_x^1 f(s)\,\mathrm{d}s \geq \int_x^1 g(s)\,\mathrm{d}s \quad \text{for all } x \in [0,1] \text{ and} \\ &\int_0^1 g(s)\,\mathrm{d}s = \int_0^1 f(s)\,\mathrm{d}s. \end{aligned} \tag{1}$$

We say that f *weakly majorizes* g, denoted by $f \succeq_w g$, if the first condition above holds (but not necessarily the second).

The above mentioned applications motivate the study of the subset of non-decreasing functions that are majorized by, or majorize, a given function. Let

$$\mathrm{MPS}(f) := \{g \in M \mid f \succeq g\}$$

and denote by $\mathrm{MPS}_w(f)$ the set of nonnegative, non-decreasing functions that are weakly majorized by f. Finally, let

$$\mathrm{MPC}(f) := \{g \in M \mid g \succeq f \text{ and } f(0) \leq g \leq f(1)\}.$$

An *extreme* point of a convex subset A is a point $x \in A$ that cannot be represented as a convex combination of two distinct points in A. Formally, if X is a vector space and $A \subseteq X$ is convex, $x \in A$ is an extreme point of A if $x = \alpha y + (1-\alpha)z$, for $z, y \in A$ and $\alpha \in [0,1]$ imply together that $y = x$ or $z = x$.

We characterize below the extreme points of $\mathrm{MPS}(f)$, $\mathrm{MPS}_w(f)$, and $\mathrm{MPC}(f)$. It turns out that the characterization of these extreme points is particularly tractable and that it can be used in at least three ways in various economic applications:

1. **Information about extreme points directly gives qualitative insights into solutions of optimization problems**: *Bauer's Maximum Principle* says that a convex, upper-semicontinuous functional on a non-empty, compact and convex set A of a locally convex space attains its maximum at an extreme point of A. Therefore, for many optimization problems a solution can be found at an extreme point, and properties common to all extreme points are therefore satisfied by an optimal solution.

2. **The characterization of extreme points is useful for results describing when particular such points solve a linear optimization problem**: For optimization problems where a solution can be found at an extreme point, the characterization of extreme points is often useful to use duality techniques to characterize when specific extreme points are optimal (see Section 3).

3. **Many properties of extreme points of a set are inherited by arbitrary elements of the set**: The *Krein–Milman Theorem* states that any convex and compact set A in a locally convex space is the closed, convex hull of its extreme points. Therefore, any element of such a set is a limit of convex combinations of extreme points. Alternatively, *Choquet's Theorem* gives conditions such that any element is a mixture of extreme points.[2] Therefore, if all extreme points of a given set have a common property and if this property is preserved by taking convex combinations and limits (or taking mixtures), then every element of the set has this property. We repeatedly use this insight below; for example, in an auction application we show that every extreme point of $\text{MPS}_w(f)$ is the interim allocation rule of a BIC auction and deduce from this that the set $\text{MPS}_w(f)$ characterizes the set of interim allocation rules of BIC auctions.

The following result characterizes the extreme points of $\text{MPS}(f)$:

[2] We say that an element x of a metric space is a mixture of a set A if there a probability distribution with support A whose barycenter is x.

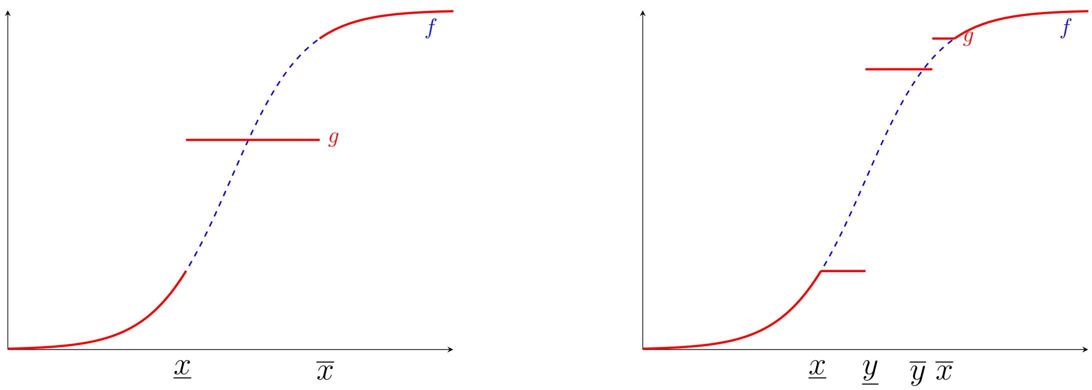

Figure 1: An extreme point g of MPS(f) *(left) and an extreme point g of* MPC(f) *(right).*

Theorem 1. *Let f be non-decreasing. Then g is an extreme point of* MPS(f) *if and only if there exists a countable collection of disjoint intervals $[\underline{x}_i, \overline{x}_i)$ indexed by $i \in I$ such that for a.e. $x \in [0,1]$*

$$g(x) = \begin{cases} f(x) & \text{if } x \notin \bigcup_{i \in I} [\underline{x}_i, \overline{x}_i) \\ \frac{\int_{\underline{x}_i}^{\overline{x}_i} f(s)\,\mathrm{d}s}{\overline{x}_i - \underline{x}_i} & \text{if } x \in [\underline{x}_i, \overline{x}_i). \end{cases} \tag{2}$$

Intuitively, if a function g is an extreme point of MPS(f) then, at any point in its domain, either the majorization constraint binds, or the monotonicity constraint binds. This implies either that $g(x) = f(x)$ or that g is constant at x. See Figure 1 for an example of an extreme point of MPS(f).

The previous result can be used to characterize the extreme points of MPS$_w(f)$. For $A \subseteq [0,1]$, denote by $\mathbf{1}_A(x)$ the indicator function of A: it equals 1 if $x \in A$ and it equals 0 otherwise.

Corollary 1. *Suppose that f is non-decreasing and nonnegative. A function g is an extreme point of* MPS$_w(f)$ *if and only if there is $\theta \in [0,1]$ such that g is an extreme point of* MPS$(f \cdot \mathbf{1}_{[\theta,1]})$ *and $g(x) = 0$ for a.e. $x \in [0, \theta)$.*

The corollary follows from Theorem 1 by observing that for any $g \in$ MPS$_w(f)$ there is $\theta \in [0,1]$ such that $g \in$ MPS$(f \cdot \mathbf{1}_{[\theta,1]})$. The extreme points of the latter set are characterized in Theorem 1.

Finally, we characterize the extreme points of MPC(f).

Theorem 2. *Let f be non-decreasing and continuous. Then $g \in \text{MPC}(f)$ is an extreme point of* MPC(f) *if and only if there exists a countable collection of intervals $[\underline{x}_i, \overline{x}_i)$, (potentially empty) sub-intervals $[\underline{y}_i, \overline{y}_i) \subset [\underline{x}_i, \overline{x}_i)$, and numbers v_i indexed by $i \in I$ such that for a.e. $x \in [0, 1]$*

$$g(x) = \begin{cases} f(x) & \text{if } x \notin \bigcup_{i \in I} [\underline{x}_i, \overline{x}_i) \\ f(\underline{x}_i) & \text{if } x \in [\underline{x}_i, \underline{y}_i) \\ v_i & \text{if } x \in [\underline{y}_i, \overline{y}_i) \\ f(\overline{x}_i) & \text{if } x \in [\overline{y}_i, \overline{x}_i). \end{cases} \tag{3}$$

To prove the above theorems, we use the following intuitive result that characterizes the extreme points of certain sets of convex functions.

Lemma 1. *(i) Let $V : [a, b] \to \mathbb{R}$ be a continuous convex function and let K_V^+ be the set of continuous convex functions U satisfying $U(a) = V(a)$ and $U(b) = V(b)$ that lie strictly above V on (a, b). If $K_V^+ \neq \emptyset$, it has a unique extreme point, which is the affine function U satisfying $U(a) = V(a)$ and $U(b) = V(b)$.*

(ii) Let $V : [a, b] \to \mathbb{R}$ be a differentiable convex function and let K_V^- be the set of continuous convex functions U satisfying $U(a) = V(a)$, $U'(a) = V'(a)$, $U(b) = V(b)$, and $U'(b) = V'(b)$ that lie strictly below V on (a, b). If $K_V^- \neq \emptyset$, the set of its extreme points equals the set of continuous convex functions that consist of at most three affine pieces and satisfy $U(c) = V(c)$ and $U'(c) = V'(c)$ for $c = a$ and $c = b$.

Proof sketch.

Part (i). For functions U that are piece-wise affine, Figure 2 illustrates that U cannot be an extreme point unless it consists of a single affine piece. More generally, whenever the left-sided derivative U'_- is not continuous, we can write U as a convex combination of $U + \varepsilon H$ and $U - \varepsilon H$, where H is a 'wedge'-function (i.e., a piecewise linear convex function consisting

of two pieces) with a kink at a discontinuity of U'_-, which implies that U is not an extreme point. On the other hand, if U is twice continuously differentiable, the second derivative will be bounded away from zero on some interval if U is not affine. We can then choose a smooth function H that vanishes outside this interval. For ε small enough, $U \pm \varepsilon H$ lie in K_V^+, implying that U is not an extreme point. For the general case, see Theorem 5.1 in Bronshtein (1978) (without proof) or the more general Lemma 9 in Augias and Uhe (2025).
Part (ii). Suppose that U is an extreme point of K_V^- and let $t_1 := \sup\{t : U'(t) = V'(a)\}$ and $t_2 := \inf\{t : U'(t) = V'(b)\}$ (see Figure 3 for an illustration). One can show that $t_1 > a$ and $t_2 < b$, so $U(t_1) < V(t_1)$ and $U(t_2) < V(t_2)$.[3]

Since U is convex, it lies above any wedge function W on $[t_1, t_2]$ formed by its supporting hyperplanes at t_1 and t_2. It follows that U is an extreme point of the set of convex functions on $[t_1, t_2]$ that lie above W and coincide with W at the endpoints. By part (i), U must be affine on $[t_1, t_2]$. Conversely, it is easy to verify that each such function U is indeed an extreme point. Q.E.D.

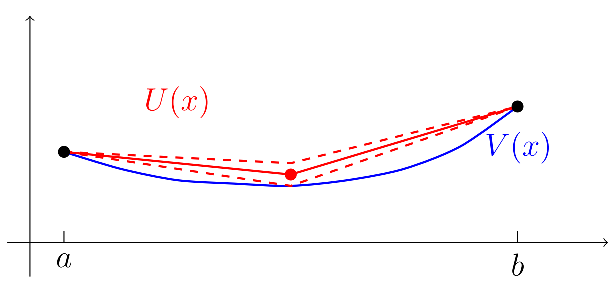

Figure 2: A piecewise-affine convex function U that is not an extreme point of K_V^+

Proof of Theorem 1. Let g be an extreme point of MPS(f) and define

$$U(x) := \int_0^x g(s)\,\mathrm{d}s$$
$$V(x) := \int_0^x f(s)\,\mathrm{d}s$$

[3] If $t_1 = a$ then, since U'_+ must be continuous at a, there exist $a < x_1 < x_2$ such that $U'_-(x_1) < U'_-(x)$ for all $x > x_1$, $U'_+(x_2) > U'_+(x)$ for all $x < x_2$, and $U'_-(x_2) > U'_+(x_1)$. However, U must be an extreme point of the set of convex functions that lie strictly above supporting hyperplanes of U at x_1 and x_2 and therefore, by part (i), affine on $[x_1, x_2]$. This contradicts $U'_-(x_2) > U'_+(x_1)$.

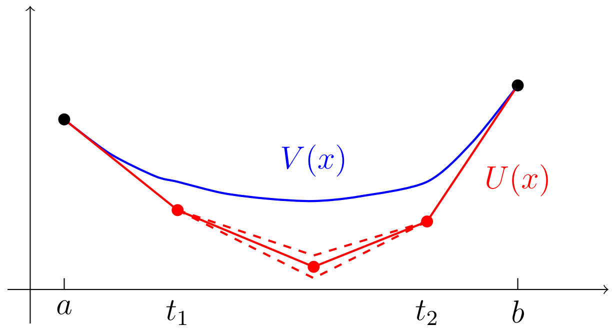

Figure 3: A piecewise-affine convex function U that is not an extreme point of K_V^-

Note that U, V are continuous convex functions that coincide at 0 and 1 and such that their slopes are bounded between $f(0)$ and $f(1)$. Moreover, U lies above V and must be an extreme point of the set of such functions.

The set $\{x \in \mathbb{R} : U(x) > V(x)\}$ is open, and is therefore a countable collection of disjoint intervals $(\underline{x}_i, \overline{x}_i)$ for $i \in I$. Restricted to each interval $[\underline{x}_i, \overline{x}_i]$, U must be an extreme point of the set of convex continuous function lying above V and coinciding with V at $\underline{x}_i$ and $\overline{x}_i$. Therefore, U must be affine on $[\underline{x}_i, \overline{x}_i]$ by Lemma 1 and g takes the form described in Theorem 1.

Conversely, it is easy to verify that U is an extreme point if it takes the postulated form and therefore that the corresponding g is an extreme point of MPS(f). Q.E.D.

Proof of Theorem 2. Let g be an extreme point of MPC(f) and define

$$U(x) := \int_0^x g(s)\,\mathrm{d}s$$
$$V(x) := \int_0^x f(s)\,\mathrm{d}s$$

Again observe that U, V are continuous, convex functions that coincide at 0 and 1 and such that their slopes are bounded between $f(0)$ and $f(1)$. Also, U lies below V and, since f is continuous by assumption, it follows that $U'(x) = V'(x)$ for $x \in \{\underline{x}_i, \overline{x}_i\}$; finally, U must be an extreme point of the set of such functions.

The set $\{x \in \mathbb{R} : U(x) < V(x)\}$ is open and is therefore a countable collection of disjoint intervals $(\underline{x}_i, \overline{x}_i)$ for $i \in I$. By Lemma 1, on each interval $[\underline{x}_i, \overline{x}_i]$, U must be piecewise affine with $U'(x) = V'(x)$ for $x \in \{\underline{x}_i, \overline{x}_i\}$. Therefore, g takes the form described in Theorem 2.

Conversely, it is easy to verify that U is an extreme point if it takes the postulated form and therefore that the corresponding g is an extreme point of $\mathrm{MPC}(f)$. Q.E.D.

3 Optimization under Majorization Constraints

We now consider optimization problems where the objective is a linear (or possibly convex) functional, and where the constraint set is defined by majorization and monotonicity constraints. Many problems in Economics naturally have a linear objective since those correspond to maximization of expected utility over some space of distributions. The next elegant result, due to Fan and Lorentz (1954) is very useful for applications because it provides conditions on the objective function such that a maximum over majorization sets determined by a function f is attained either at f itself (highest variability), or at a particular function g with at most two steps (lowest variability)

Theorem 3 (Fan and Lorentz). *Let* $K : [0,1] \times [0,1] \to \mathbb{R}$ *. Then*

$$\int_0^1 K(f(t), t)\, \mathrm{d}t \leq \int_0^1 K(g(t), t)\, \mathrm{d}t$$

holds for any two non-decreasing functions $f, g : [0,1] \to [0,1]$ *such that* $f \preceq g$ *if and only if the function* $K(u,t)$ *is convex in* u *and super-modular in* (u,t)*.*

3.1 Maximizing a Linear Functional on $\mathrm{MPS}(f)$

Given a non-decreasing function f and a bounded function c consider the problem

$$\max_{h \in \mathrm{MPS}(f)} \int_0^1 c(x) h(x)\, \mathrm{d}x. \tag{4}$$

There are three cases:

1. If c is non-decreasing, an optimal g will be as large as possible for high values of x. From the majorization constraint it follows that f itself is a solution to the optimization problem.

2. If c is non-increasing, an optimal g will be as large as possible for low values of x. From the monotonicity constraint it follows that the constant function $g(x) = \int_0^1 f(s)\,\mathrm{d}s$ for all x is a solution.

3. If c is not monotone, other extreme points of $\mathrm{MPS}(f)$ may be optimal. Proposition 2 in Kleiner, Moldovanu, and Strack (2021) characterizes the conditions under which an arbitrary extreme point is optimal. Roughly speaking, the ironing technique, originally used in Myerson (1981) for an optimization problem formulated without majorization constraints, can be used if the constraint set is MPS.

3.2 Maximizing a Linear Functional on $\mathrm{MPC}(f)$

We now analyze the problem

$$\max_{h \in \mathrm{MPC}(f)} \int_0^1 c(x) h(x)\,\mathrm{d}x\,. \tag{5}$$

Again, there are three cases:

1. If c is non-increasing, an optimal solution is as large as possible for small values of x. It follows from the majorization constraint that f solves this problem.

2. If c is non-decreasing, an optimal solution is as large as possible for high values of x. It follows from the majorization constraint and the boundary constraint that an optimum is obtained at the step function g defined by

$$g(x) = \begin{cases} f(0) & \text{for } x < \overline{x} \\ f(1) & \text{for } x \geq \overline{x}, \end{cases}$$

where $\overline{x}$ solves

$$\int_0^{\overline{x}} f(0)\,\mathrm{d}s + \int_{\overline{x}}^1 f(1)\,\mathrm{d}s = \int_0^1 f(s)\,\mathrm{d}s$$

3. When c is not monotone, Proposition 3 in Kleiner, Moldovanu, and Strack (2021) use the extreme points characterization together with duality results from Dworczak and Martini (2019) to determine when particular extreme points are optimal.

3.3 How the Two Problems Differ

The two problems (4) and (5) differ in a fundamental way and distinct tools are needed to solve them: for problem (4) with constraint set MPS, we can use an ironing procedure to find an explicit solution, whereas for problem (5) with constraint set MPC we need to use a guess-and-verify approach. To illustrate these differences, we contrast the two problems:

$$\max_{g\in \mathrm{MPS}(f)} \int_0^1 c(x)g(x)\,\mathrm{d}x \qquad\qquad \max_{g\in \mathrm{MPC}(f)} \int_0^1 c(x)g(x)\,\mathrm{d}x$$

Note that we can impose without loss of generality that $g(1) = f(1)$ for any solution.[4] Defining $C(x) = -\int_0^x c(s)\,\mathrm{d}s$ and using integration by parts, and eliminating the terms $C(1)f(1)$ (which are constant in g), these problems can be written as:

$$\max_{g\in \mathrm{MPS}(f)} \int_0^1 C(x)\,\mathrm{d}g(x) \qquad\qquad \max_{g\in \mathrm{MPC}(f)} \int_0^1 C(x)\,\mathrm{d}g(x)$$

The corresponding dual linear programs are given by (see, for example, Shapiro, 2010; Dworczak and Martini, 2019):

$$\min_{p \text{ concave}} \int_0^1 p(x)\,\mathrm{d}f(x) \quad \text{s.t. } p \geq C \qquad\qquad \min_{p \text{ convex}} \int_0^1 p(x)\,\mathrm{d}f(x) \quad \text{s.t. } p \geq C$$

Moreover, strong duality holds, and therefore g is an optimal solution to the primal

[4]Because $g(1) \leq f(1)$ and changing g at a single point will not change the objective value.

problem if and only if there exists a feasible solution to the corresponding dual problem that achieves the same value.

Note that the dual problem on the left is easy to solve: We aim to find a concave function that lies above a given function C and minimizes the integral. The concave envelope $\hat{C}$ of C is the pointwise smallest concave function that lies above C and therefore solves the dual problem. It follows that $g \in \mathrm{MPS}(f)$ solves the primal problem if and only if both problems attain the same value: $\int C(s)\,\mathrm{d}g(s) = \int \hat{C}(s)\mathrm{d}f(s)$. This implies that g, when interpreted as a probability measure, puts zero mass on points where $\hat{C}(s) < C(s)$. Letting $\{(\underline{x}_i, \overline{x}_i) : i \in I\}$ denote a maximal collection of maximal intervals on which $\hat{C} < C$ and observing that $\hat{C}$ is affine when restricted to $[\underline{x}_i, \overline{x}_i]$, it follows that a solution is given by the extreme point

$$g(x) = \begin{cases} f(x) & \text{if } x \notin \bigcup_{i \in I}[\underline{x}_i, \overline{x}_i) \\ \frac{\int_{\underline{x}_i}^{\overline{x}_i} f(s)\,\mathrm{d}s}{\overline{x}_i - \underline{x}_i} & \text{if } x \in [\underline{x}_i, \overline{x}_i). \end{cases}$$

Hence, an ironing technique analogous to Myerson provides a constructive method to obtain a solution.

Contrast the above situation to the optimization problem with constraint set MPC: The dual problem is given on the right and is not as immediate to solve. In general, if C is not convex, there is no pointwise smallest convex function above C. This explains why no procedure analogous to ironing is known to solve problems with the constraint set MPC. Linear programming duality can still be used to verify whether a particular extreme point solves a given problem, as illustrated in Kleiner, Moldovanu, and Strack (2021). However, this method does not provide a direct construction of the solution to a specific problem.

4 Applications

4.1 Allocation Problems

The problem of allocating heterogeneous goods to privately informed agents lies at the heart of mechanism design theory. This framework, pioneered by Myerson (1981) for single-object auctions, has been extended to encompass increasingly complex allocation environments. A fundamental question in this literature is: what allocation rules can be implemented when agents have private information about their preferences over the objects?

The seminal work of Border (1991) provided a complete solution to the above question for single-object settings by characterizing feasible interim allocation rules.[5] We illustrate in this section how the methodology developed above can be used to obtain a simple characterization of feasible allocation rules: these are precisely the allocations that are weakly majorized by the efficient allocation. This argument extends Border's (1991) result to settings with objects of heterogeneous qualities and provides a simple proof.

There are n agents. For each agent i, her type t_i is drawn uniformly and independently from $[0, 1]$. Each agent values at most one object, and the value of an agent with type t is $v = V(t)$, where $V : [0, 1] \to \mathbb{R}$ is strictly increasing and bounded.[6]

There are n objects with qualities $0 \leq q_1 \leq q_2 \leq \ldots \leq q_n = 1$, and we define $\mathcal{A} \subset \{0, q_1, q_2 \ldots, q_n\}^n$ to be the set of *feasible* allocations that allocate each object at most once, and allocate to each agent at most one object. If agent i with type t_i receives an object with quality q and pays p for it, her payoff is $v(t_i)q - p$.

Fix an allocation rule $a : [0, 1]^n \to \Delta(\mathcal{A})$ that depends on the agents' types $t_1, \ldots, t_n$ and on the outcome ω of a randomization device. For each i, define the *interim allocation rule* by

$$A_i(t_i) = \mathbb{E}_{t_{-i}}[a_i(t_i, t_{-i})|t_i]$$

[5] See also Maskin and Riley (1984) and Matthews (1984). This is not the original formulation. For connections to majorization see Hart and Reny (2015).

[6] Equivalently, values are distributed I.I.D. according to the CDF $F := V^{-1}$.

This represents the expected quality obtained by agent i with type t_i. The interim allocation rule $A_i(t_i)$ captures which quality agent i expects to receive based solely on her own type, averaging over all possible type realizations of other agents. It is straightforward to show that an allocation a is part of a Bayesian incentive compatible (BIC) mechanism if and only if each induced interim allocation A_i is non-decreasing.

Denote by a^* the *assortative allocation* of agents to objects where the highest type gets highest quality, etc. and where ties are broken by fair randomization. In our symmetric model, assortative matching a^* is incentive compatible, and induces the symmetric interim allocation

$$A_i^*(t_i) = \sum_{k=1}^{n} q_k \left[\frac{(n-1)!}{(k-1)!(n-k)!}(t_i)^{k-1}(1-t_i)^{n-k} \right].$$

A vector of interim allocations $A = (A_1, \ldots, A_n)$ where $A_i : [0,1] \to \mathbb{R}$ is *feasible* if there exists an allocation rule a that induces A as its set of interim allocations. We restrict attention to non-decreasing and symmetric interim allocation rules where $A_i = A_j$ for $i, j = 1, 2, \ldots, n$.

To understand which allocation rules are implementable, consider first the simplest case: allocating a single object among n agents (i.e., $q_n = 1$ and $q_k = 0$ for $k < n$). The efficient (first-best) allocation gives the object to the highest type, yielding the interim allocation $A^*(t_i) = t_i^{n-1}$. This is just the probability that agent i has the highest type among all n agents. Note that $A \preceq_w A^*$ if and only if $\int_t^1 A(s)\mathrm{d}s \leq \int_t^1 s^{n-1}\mathrm{d}s = \frac{1}{n}[1-t^n]$ for all t. In our terminology, Border's theorem says that a non-decreasing symmetric interim allocation A is feasible if and only if A is weakly majorized by the efficient (and assortative) allocation A^*. This condition is clearly necessary since no rule can allocate with higher probability to high types than the assortative allocation rule. It is not obvious that this condition is in fact sufficient.

This majorization perspective becomes even more powerful when we move beyond single objects to our setting with heterogeneous qualities:

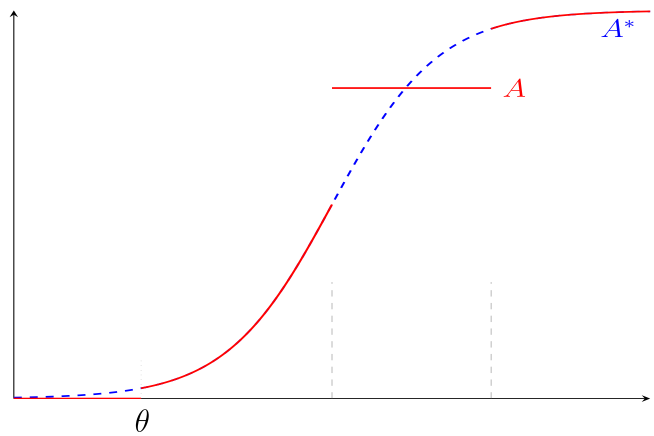

Figure 4: An extreme point A of $\mathrm{MPS}_w(A^*)$. The allocation rule A (solid red) is weakly majorized by the efficient allocation A^* (dashed blue). Below threshold θ, no allocation occurs. On intervals like $[0.5, 0.75]$, types are pooled with constant allocation preserving the average of A^*. This extreme point can be implemented by assigning identical scores to pooled types.

Theorem 4. *A non-decreasing symmetric allocation rule A is feasible if and only if $A \preceq_w A^*$.*

Theorem 4 establishes a novel connection between feasibility and efficiency as it establishes that the interim feasibility constraints are completely determined by the efficient allocation. This connection is fruitful since it holds beyond the multi-unit allocation problem. For example, Kleiner, Moldovanu, and Strack (2021) discuss an allocation problem with group-specific quotas and argue that the feasible set is characterized as those interim allocation rules that are weakly majorized by the most-efficient allocation rule subject to the quotas.

Proof. The key insight is that any feasible allocation rule can be decomposed into a mixture of simple 'pooling' rules. To see why, consider first extreme points of $\mathrm{MPS}_w(A^*)$.

By Corollary 1, any extreme point A has a threshold θ below which no allocation occurs. Above this threshold, Theorem 1 tells us that A alternates between two behaviors: either it matches A^* exactly, or it pools types within intervals, giving them all the same expected allocation (see Figure 4 for an illustration).

This structure has a natural implementation: assign to each type t the score $A(t)$, so that all types in a given pooling interval receive identical scores, then allocate objects assortatively to scores with uniform tie-breaking, and do not allocate any object to types below the cutoff θ. The majorization constraint ensures that the average allocation within each pooled interval equals what those types would receive under efficient allocation, which shows that A is feasible.

If A is a convex combination of extreme points, we can implement A by a convex combination of the allocation rules that implement these extreme points. More generally, by Choquet's theorem any $A \in \mathrm{MPS}_w(A^*)$ is a mixture of extreme points, which using Lemma 3, can be implemented by a mixture of the corresponding allocation rules. Q.E.D.

Another simple but powerful consequence of the above argument is the equivalence between Bayesian and dominant-strategy implementation in symmetric settings (first noted by Manelli and Vincent, 2010 for auctions with one object and by Gershkov et al., 2013 for general social choice settings with independent valuations and with transfers).

Corollary 2. *For any symmetric, BIC mechanism there exists an equivalent, symmetric DIC mechanism that yields all agents the same interim utility, and that creates the same social surplus.*

This follows because, as shown above, the designer can implement any BIC interim allocation by randomizing over mechanisms such that each one of them implements an extreme points of $\mathrm{MPS}_w(A^*)$. Those mechanisms allocate according to the efficient (and assortative) allocation if an agent's type is not in a pooling interval. If the agent's type is in a pooling interval, they first redraw uniformly at random a "virtual" type from all the types in the pooling interval and then allocate according to the efficient allocation as if the agent's virtual types were their true types. These operations preserve the ex-post monotonicity of the assortative allocation and therefore the mechanisms are implementable in dominant strategies. Finally the proof is completed because dominant strategy incentive compatibility is preserved under randomization over mechanisms and thus any implementable interim allocation can be implemented in dominant strategies by randomizing over the simple mechanisms described above.

4.1.1 Revenue Maximization

Consider now incentive-compatible mechanisms where the utility of the lowest type is zero (as required by individual rationality and revenue optimality). Denote by $J(v) = v - \frac{1-V^{-1}(v)}{V^{-1\prime}(v)}$ the "virtual value" function.[7] Then the expected revenue generated by a symmetric mechanism with interim allocation rule A equals $n\int_0^1 J(V(t))A(t)\,\mathrm{d}t$. Thus, by Theorem 4, the revenue maximization problem becomes

$$\max_{A\in \mathrm{MPS}_w(A^*)} n\int_0^1 J(V(t))A(t)\,\mathrm{d}t$$

Since the objective is linear, a maximum is attained at an extreme point of $\mathrm{MPS}_w(A^*)$. The characterization of extreme points therefore immediately yields qualitative insights into the nature of optimal allocation rules, e.g., into the solution of the classical revenue maximization problem for the allocation of one object analyzed by Myerson and by Riley and Samuelson. Results on the solution to linear problems subject to majorization constraints (e.g., the optimization of a weighted average of revenue and the agents' welfare) can be used to obtain finer properties of solutions to specific problems.

4.1.2 Optimal Dynamic Allocation

Gershkov and Moldovanu (2009) and Gershkov and Moldovanu (2010) study revenue and welfare maximization in a dynamic assignment model where a fixed inventory of objects of different qualities are assigned to impatient, agents that arrive over time and have different values for quality. The main trade-off concerns the allocation of each object to either an arriving agent or to a future arriving one with a potentially higher value for quality. In optimal mechanisms the set of types is divided in distinct intervals and an agent with type

[7] Recall that the probability with which a given agent has a valuation below v is $F(v) = V^{-1}(v)$. Substituting this into J recovers the standard formulation of virtual values.

in the i-th interval is assigned upon arrival the i-th best still available object. These authors directly show how the optimal allocation is majorized by the efficient one and use this as a measure of the inefficiency that is inherent in the dynamic model because the allocation cannot be delayed and because agents do not arrive all at the same time. Ashlagi, Monachou, and Nikzad (2025) allow for both objects and agents to randomly arrive over time and use the above described majorization approach in order to design optimal queuing systems. In their model agents and objects of different qualities arrive over time. Agents' utilities are given by a supermodular function that takes into account an agent's preference for quality (her type θ) and the quality ω of the object she receives (which may be stochastic). An agent who waits t units of time until allocation incurs a cost $c(t)$. The designer's goal is to maximize a linear combination of allocative efficiency and the agent' welfare (which includes their waiting costs).

The authors derive the feasible allocation plans (these are mean-preserving spreads of the efficient, assortative matching) and use the extreme points characterization to compute an optimal, direct revelation mechanism which belongs to the class of *monotone disjoint queue mechanisms.* Such a mechanism features a finite number of queues: agents are assigned to a specific queue if and only if their type falls in a certain interval. Every arriving object is sent to each of the queues with a constant rate (possibly zero) and is then offered to the earliest arrived agent assigned to that queue (First In First Out queue discipline).[8] Roughly speaking, a disjoint queue mechanism is monotone if for any two objects with distinct qualities, either the two objects are sent to the same queue, or the higher quality object is sent to the queue corresponding to higher types of the agents. Such mechanisms parallel the coarse matching schemes discussed below.

[8] In a disjoint queue mechanism, the arrival rate of agents to each queue equals the arrival rate of objects to that queue and therefore the queues' lengths and the time a newly arrived agent has to wait until allocation in a particular queue remain constant over time.

4.1.3 A Continuum of Agents and Objects

We now analyze a canonical allocation problem with a continuum of agents and objects of different qualities, often studied in the matching and contest literatures. We index agents by their quantile and let the t-th quantile agent have a valuation $V(t)$. Similarly, we index prizes on $[0,1]$ and let the prize with index s have quality $Q(s)$. We assume here for simplicity that V and Q are strictly increasing. If an agent with index t obtains prize s and pays p, her utility is given by $V(t)Q(s) - p$.

A symmetric allocation rule A maps an agent's index t to her expected quality $A(t)$. The assortative allocation therefore satisfies $A^*(t) = Q(t)$ and the same reasoning as above yields the following useful characterization of feasible allocations:

Proposition 1. *A symmetric, non-decreasing allocation A is feasible if and only if $A \preceq_w Q$.*

As an example, consider a contest where each agent with type t makes an effort (or submits a bid), and where agents are matched to prizes according to their bids. The assortative allocation is given by Q. and is strictly increasing. It is implemented by the strictly increasing bidding equilibrium

$$b(t) = V(t)\,Q(t) - \int_0^t Q(\tau)V'(\tau)d\tau$$

It is well-known that agents' welfare from the physical allocation of prizes is maximized by the assortative scheme,[9] but agents need to waste resources (e.g., signaling costs, payments to a designer) in order to achieve the needed separation. Another feasible scheme is *random* matching where, independently of bids, everyone gets a prize equal to the expected value of the prize distribution μ_Q. Expected welfare from the physical allocation is smaller than under assortative matching, but random matching can be implemented without costs. Intermediate schemes can be obtained by *coarse* matching: for example, an agent with a bid in given quantile is randomly matched to a prize in the same quantile, i.e., he expects to obtain the

[9] This follows from the rearrangement inequality of Hardy, Littlewood, and Pólya (1934).

average prize in that quantile.

The Proposition below generalizes and complements several well-known, existing results in the contest and matching literature (see Damiano and Li, 2007, Hoppe, Moldovanu, and Sela, 2009, Condorelli, 2012 and Krishna et al., 2025). These are obtained as immediate consequences of our theoretical insights described above together with the Fan-Lorenz Theorem.

Proposition 2.

1. *Assume that the distribution of types F is convex. Then each type of the agent prefers random matching to any other scheme.*

2. *Random matching (assortative matching) maximizes the agents' average utility if the distribution of types F has an increasing (decreasing) Failure Rate.*

3. *If F has an increasing failure Rate, the revenue (i.e., average bid) to a designer is maximized by assortative matching.*

Proof. See Kleiner, Moldovanu, and Strack (2021).

4.1.4 Allocation with Externalities: Priority Services and Status Prizes

Gershkov and Winter (2023) consider a model of priority service and study the implications of the allocation of priorities and their pricing on the consumer welfare. Xiao (2024) shows how their model can be readily analyzed through the lens of majorization.

In their benchmark model a monopolist faces a mass 1 of consumers with heterogeneous costs of waiting denoted by t, where t is distributed according to a distribution F on a closed interval. The monopolist can serve (at zero cost) a mass m of consumers in m units of time. A customer with type t who gets service at time $\tau \in [0, 1]$ while paying a price $p \geq 0$ has a utility of $-p - t\tau$. A direct mechanism consists of a payment function $p(t)$ and a (potentially random) expected waiting time $\tau(t)$—the latter must be non-increasing in an incentive compatible mechanism. Note that an arbitrary non-increasing waiting time function $\tau(t)$ is not necessarily feasible here because of the abundance or scarcity of different costs t—this is the information encoded by the distribution F. Let the function $s(t) = 1-\tau(t)$ quantify the value of priority by the time saved compared to being served last.[10] The key observation here is the feasibility condition, again centered around the efficient allocation (assortative matching) $s(t) = 1 - \tau(t) = F(t)$. Therefore one obtains that a non-increasing expected waiting time function $\tau(t)$ is feasible if and only if $s(t) \in \text{MPS}(F)$.[11] The intuition is that the monopolist can induce full separation by offering infinitely many priority levels and serving agents in descending order of their costs yielding $s(t) = F(t)$. Any pooling of types in the same priority level is a mean preserving spread of F in the quantile space.

An analogous feasibility condition also exists for the contest over status prizes considered by Moldovanu, Sela, and Shi (2007) where agents value the mass of agents with status below their own, but suffer from the mass of agents with status above their own (see Xiao, 2024). Under full separation in a model with a continuum of agents and status levels,[12] the status value of an agent with type t is given by $F(t) - (1 - F(t)) = 2F(t) - 1$ where F denotes the distribution of types. A non-decreasing status allocation $s(t)$ is here feasible if and only if $s(t) \in \text{MPS}(2F(t) - 1)$.

Based on the above observations about feasibility and maximization under majorization constraints, various problems can be solved, e.g., finding the optimal mechanism that maximizes some linear combination of the revenue obtained when agents pay for priority or make efforts to earn status and the agents' welfare (prizes minus costs).

4.2 Mean-Preserving Spreads

The concave order is a basic concept in decision theory under risk. It captures the idea that risk-averse agents who have a concave utility function prefer one distribution over another if the former's outcomes are less dispersed than the latter's. The connection between the

[10]The fixed value per consumer for being served at all is normalized here to zero.

[11]If exclusion is allowed then the condition involves only weak majorization.

[12]Moldovanu, Sela, and Shi (2007) only considers the case with a finite number of agents.

concave order and majorization provides a useful geometric interpretation: in one dimension, the concave order is equivalent to comparing the areas under cumulative distribution functions.

Formally, a cumulative distribution function F dominates a cumulative distribution function G in the concave order, $F \geq_{cv} G$, if $\int u(x)\,\mathrm{d}F(x) \geq \int u(x)\,\mathrm{d}G(x)$ for any concave, continuous function u such that both integrals exist.

Proposition 3. *$F \geq_{cv} G$ if and only if $F \succeq G$.*

The condition $F \succeq G$ (i.e., $\int_s^\infty [F(x) - G(x)]dx \geq 0$) has a clear geometric interpretation: the area between the two CDFs, measured from any point s to infinity, shows that G accumulates more probability mass in the upper tail than F. This excess mass in the tails is precisely what makes G 'riskier' than F. The proof of this result, which is standard (e.g. Shaked and Shanthikumar, 2007), uses integration by parts and the fact that the extremal functions of the set of concave functions are all of the form $\min\{x, c\}$ for $c \in \mathbb{R}$.

Proof. Using integration by parts, we can write for any s,

$$\begin{aligned}\int_s^\infty [F(x) - G(x)]dx &= -s[F(s) - G(s)] - \int_s^\infty x\,\mathrm{d}[F - G](x)\\ &= \int_{-\infty}^\infty - \max\{s, x\} d[F - G](x).\end{aligned}$$

Suppose $F \geq_{cv} G$. Since $-\max\{s, x\}$ is concave in x, the right-hand side is nonnegative. Moreover, since $u(x) = x$ and $u(x) = -x$ are concave, F and G have the same mean, which implies $\int_{-\infty}^\infty [F(x) - G(x)]dx = 0$. Therefore, $F \succeq G$.

For the converse, suppose there is a concave continuous function u such that $\int u(x)\,\mathrm{d}F(x) < \int u(x)\,\mathrm{d}G(x)$. We can approximate u by a piecewise affine concave function $\bar{u}$ such that $\int \bar{u}(x)\,\mathrm{d}F(x) < \int \bar{u}(x)\,\mathrm{d}G(x)$. Moreover, there is an affine function a such that $\bar{u}$ is a conic combination of a and concave functions of the form $-\max\{s_i, x\}$ for $s_i \in \mathbb{R}$. But then $F \succeq G$ implies $\int \bar{u}(x)\,\mathrm{d}F(x) \geq \int \bar{u}(x)\,\mathrm{d}G(x)$, a contradiction. Q.E.D.

The next result relates the concave order to mean-preserving spreads. This fundamental insight goes back to Hardy, Littlewood, and Pólya (1934) in the context of inequalities. It was generalized by Blackwell, Sherman, Stein, and Cartier. Strassen (1965) gave the most general formulation, showing that the existence of mean-preserving spreads is equivalent to a certain martingale coupling.

If G is a mean-preserving spread of F, it is standard to use Jensen's inequality to show that $F \geq_{cv} G$, which in turn implies $F \succeq G$. The converse is more difficult to establish. One needs to show that if $F \succeq G$ then G is a *mean-preserving spread* of F: there exists a (measurable) kernel $K : \mathbb{R} \to \Delta(\mathbb{R})$ such that, for every x, the expected value of $K(x)$ is x, and such that K carries F to G: $\int K(x)\, \mathrm{d}F(x) = G$. Intuitively, a mean-preserving spread takes probability mass from each point of a distribution and spreads it toward the tails while keeping the mean constant. This creates a "riskier" distribution in the sense that outcomes become more dispersed.

Figure 5 illustrates the key insight: when G is an extreme point of $\mathrm{MPS}(F)$, it is easy to construct the required kernels. In particular, if G is an extreme point of $\mathrm{MPS}(F)$, it equals F everywhere except on intervals where G is constant. On each such interval $[\underline{x}, \overline{x})$, the mass that F distributes on the interval is concentrated by G at the endpoints. The kernel K that we construct splits the point mass at each x in the interval between $\underline{x}$ and $\overline{x}$ in proportions that preserve the mean. For general $G \in \mathrm{MPS}(F)$, we mix over such kernels to show that G is a mean-preserving spread.[13]

Proposition 4. *Let F and G have bounded support. Then $F \succeq G$ if and only if G is a mean-preserving spread of F.*

Proof. Suppose first that G is an extreme point of $\mathrm{MPS}(F)$. By Theorem 1, there exists a collection of disjoint intervals $[\underline{x}_i, \overline{x}_i)$ indexed by $i \in I$ such that G is constant on each interval and coincides with F outside these intervals. Without loss of generality, we can

[13]Since our extreme point characterization applies to functions with a bounded domain, we restrict attention to distributions with bounded support in the following result.

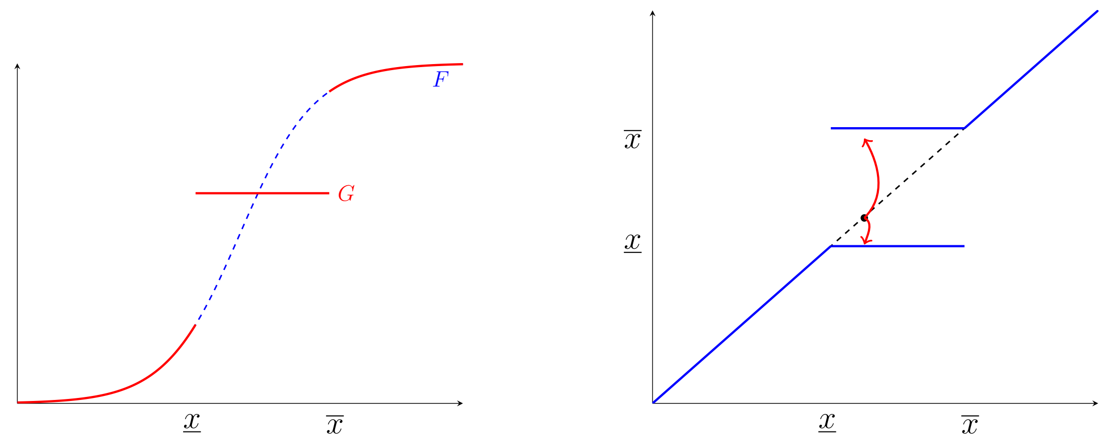

Figure 5: An extreme point G of MPS(F) *(left) and the corresponding mean-preserving spread (right).*

assume that there are no other intervals on which G is constant. We can define a kernel as follows: if $x \in (\underline{x}_i, \overline{x}_i)$ for some $i \in I$, $K(x)$ is the probability measure with discrete support $\{\underline{x}_i, \overline{x}_i\}$ and expected value x; if $x \notin \bigcup_i (\underline{x}_i, \overline{x}_i)$ then $K(x) = \delta_x$. Because conditional on each interval, the expected values of F and G coincide, we get $G = \int K(x)\,\mathrm{d}F(x)$. It follows that G is a mean-preserving spread of F.

If G is a convex combination of extreme points of MPS(F) then the corresponding convex combination of kernels shows that G is also a mean-preserving spread. More generally, by Choquet's theorem any $G \in \text{MPS}(F)$ is a mixture of extreme points. The corresponding mixture of kernels is again a kernel, showing that any $G \in \text{MPS}(F)$ is a mean-preserving spread of F.[14]

For the converse, suppose that G is a mean-preserving spread of F and let K denote a corresponding kernel. Jensen's inequality implies for any concave function h that

$$\int h(x)\,\mathrm{d}G(x) = \int\int h(y)K(x,\mathrm{d}y)dF(x) \leq \int h(x)\,\mathrm{d}F(x)$$

[14]Formally, Choquet's theorem implies there is a probability measure μ supported on the extreme points of MPS(F) that represents G. Define the function $K : \mathbb{R} \times \text{ext}(\text{MPS}(F)) \to \mathbb{R}$ by letting, for each $H \in \text{ext}(\text{MPS}(F))$, $K(\cdot, H)$ denote the kernel constructed as above. It can be verified that for each $x \in \mathbb{R}$, the function $K(x,\cdot)$ is continuous, hence K is jointly measurable. We can then define the (measurable) kernel $\bar{K} : \mathbb{R} \to \mathbb{R}$ by $\bar{K}(x) := \int K(x,H)\,\mathrm{d}\mu(H)$. Since the expected value of $\bar{K}(x) = x$ and $\int \bar{K}(x)\,\mathrm{d}F(x) = G$, it follows that G is a mean-preserving spread of F.

Hence $F \geq_{cv} G$. In turn, Proposition 3 implies $F \succeq G$. Q.E.D.

4.2.1 Decisions under Risk

As a quick application, recall the utility functionals with rank-dependent assessments of probabilities a la Quiggin (1982) and Yaari (1987):

$$U(F) = \int_0^1 v(t)\, \mathrm{d}(g \circ F)(t),$$

where F is the distribution of a random variable on the interval $[0,1]$, $v : [0,1] \to \mathbb{R}$ is differentiable, strictly increasing and bounded, and where $g : [0,1] \to [0,1]$ is strictly increasing, continuous and onto. The function v represents a transformation of monetary payoffs, while the function g represents a transformation of probabilities. Letting $g(x) = x$ yields the classical von-Neumann and Morgenstern expected utility model where risk-aversion is equivalent to v being concave. Letting $v(x) = x$ yields Yaari's (1987) dual utility theory, where risk aversion is equivalent to g being concave. Because of the possible interactions between v and g, it is not a priori clear what properties yield risk aversion in the general rank-dependent model.

To apply the Fan-Lorentz framework, we use integration by parts to obtain a representation of the form $U(F) = \int_0^1 K(F(t), t)\, dt + \text{const}$, where $K(u,t) = -g(u)v'(t)$. Note that the Fan-Lorentz conditions (convexity and supermodularity of K) are satisfied here if and only if g and v are concave. As a consequence, the utility functional $U(F)$ is then Schur-concave (and hence monotonic with respect to the concave stochastic order), and the agent whose preferences are represented by U is *risk averse*.[15]

[15] The equivalence between the concavity of the functions v and g, and risk-aversion has been pointed out by Hong, Karni, and Safra (1987), who built on Machina (1982).

Dybvig (1988) studies a simplified version of the following problem:

$$\min_X \mathbb{E}[\ XY\ ]$$
$$\text{s.t. } X \geq_{cv} Z,$$

where Y and Z are given random variables (see also Beare, 2023). Y represents here the distribution of a pricing function over the states of the world, and the goal is to choose, given Y, the cheapest contingent claim X that is less risky than a given claim Z.[16] We obtain that:

$$\mathbb{E}[XY] \geq \int_0^1 F_Y^{-1}(1-t)F_X^{-1}(t)\,\mathrm{d}t \geq \int_0^1 F_Y^{-1}(1-t)F_Z^{-1}(t)\,\mathrm{d}t$$

where the first inequality follows by the rearrangement inequality of Hardy, Littlewood and Polya (the anti-assortative part!), and where the second inequality follows by the Fan-Lorentz Theorem.

Choosing a random variable X that has the same distribution as Z and that is anti-comonotonic with Y,[17] attains the lower bound of $\int_0^1 F_Y^{-1}(1-t)F_Z^{-1}(t)\,\mathrm{d}t$ and such a choice solves the portfolio choice problem.[18] If $Y' \leq_{cv} Y$, we also obtain by the Fan-Lorentz inequality (now applied to the functional with argument F_Y^{-1}) that

$$\inf_{X \succ_{cv} Z} \mathbb{E}[XY] = \int_0^1 F_Y^{-1}(1-t)F_Z^{-1}(t)\,\mathrm{d}t \geq \int_0^1 F_{Y'}^{-1}(1-t)F_Z^{-1}(t)\,\mathrm{d}t = \inf_{X \succ_{cv} Z} \mathbb{E}[XY']$$

In other words, a decision maker that becomes more informed (in the Blackwell sense) about the pricing distribution will achieve a better result.

[16] To make the problem well-defined, Y needs to be essentially bounded and X, Z must be integrable.

[17] This can always be done if the underlying probability space is non-atomic. A random vector (X, Y) is anticomonotonic if there exists a random variable W and non-decreasing functions h_1, h_2 such that $(X, Y) =^{dist} (h_1(W), -h_2(W))$.

[18] For more details on this problem see Dana (2005) and the literature cited there. It does not use the Fan-Lorentz inequality.

4.3 Optimal Stopping

Optimal stopping problems arise in various economic contexts: for example, when a firm considers when to exercise a real option. The mathematical structure of these problems turns out to be connected to majorization. For an overview of this and related problems, see Obłój (2004).

Consider an agent observing a stochastic process $(M_t : t \geq 0)$ that evolves as a continuous martingale starting from an initial distribution $M_0 \sim F$. The agent must decide when to stop the process and receive the realized value. This setup captures many economic situations where information arrives gradually and where the decision to stop is irreversible—once you stop, you cannot restart the process. The assumption that the process is a martingale is without loss of generality for a time-homogeneous diffusion as we can (under standard assumptions) consider a transformation $(\phi(M_t))_t$ of the original process that is a martingale.

The key insight is that the set of achievable outcome distributions through stopping coincides with the set of mean-preserving spreads of the initial distribution. To make this precise, we allow for randomized stopping strategies in order to obtain a convex problem. Formally, a randomized stopping time is a function $\tau : \Omega \times [0, 1] \to \mathbb{R}_+$ that is measurable with respect to the natural filtration augmented by an independent, uniform random variable.

Proposition 5. *Let (M_t) be a continuous martingale with $M_0 \sim F$ and unbounded quadratic variation. For any probability measure G on $\mathbb{R}$ with bounded support, the following are equivalent:*

1. *There exists a stopping time τ such that $M_\tau \sim G$ and the stopped process $M_{t\wedge\tau}$ is uniformly integrable.*

2. *$G \in \text{MPS}(F)$.*

Proof Sketch. The forward direction follows from Doob's optional stopping theorem: for any stopping time τ such that $M_{t\wedge\tau}$ is uniformly integrable, it holds that $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$, and for any convex function ϕ we have $\int \phi(r)dG(r) = \mathbb{E}[\mathbb{E}[\phi(M_\tau)|M_0]] \geq \mathbb{E}[\phi(M_0)] = \int \phi(r)dF(r)$

by Jensen's inequality. Thus, G is a mean preserving spread of F.

For the reverse direction, we use our characterization of extreme points. If G is an extreme point of $\text{MPS}(F)$ with the interval structure from Theorem 1, the optimal stopping rule is transparent: stop at the first time the process exits the union of open intervals $\bigcup_{i \in I}(\underline{x}_i, \overline{x}_i)$. Since the martingale has unbounded quadratic variation, it will eventually exit any bounded interval, ensuring $M_{t \wedge \tau}$ is uniformly integrable. The resulting distribution is precisely G because the process stops only at the boundary points $\{\underline{x}_i, \overline{x}_i\}$ with the correct probabilities to preserve the mean.

For general $G \in \text{MPS}(F)$, express it as a mixture of extreme points and use the corresponding randomized mixture of stopping strategies. Q.E.D.

4.4 Persuasion with Preferences over the Posterior Mean

We consider here the persuasion problem studied by Kolotilin (2018), Gentzkow and Kamenica (2016), and Dworczak and Martini (2019).

The state of the world $\omega \in [0, 1]$ is distributed according to a continuous distribution $F : [0, 1] \to [0, 1]$, and a sender can reveal information about the state to an uninformed receiver. The sender chooses a signal (or Blackwell experiment) π that consists of a signal realization space S and a family of distributions $(\pi_\omega)_\omega$ over S, conditional on each state. By Bayes' rule each signal induces a distribution of posteriors, and hence a distribution of posterior means. The receiver observes the choice of signal and the signal realization, and then chooses an optimal action that depends on the mean of the posterior, denoted here by x. The sender's indirect utility v is state independent and only depends on the posterior mean x.[19] The following result (see Blackwell, 1953; Strassen, 1965; Kolotilin, 2018) characterizes which distributions of posterior means are feasible:

Proposition 6. *G is feasible if and only if $F \preceq G$.*

[19]This allows for the sender's payoff to depend on the action taken by the receiver.

Any signal is a "garbling" of the prior, and thus, for any signal π, the prior F is a mean-preserving spread of the generated distribution of posterior means G_π, i.e. $G_\pi \succeq F$. For the converse, suppose G is an extreme point of MPC(F). Then we can construct a signal that induces G: either you learn the state, there is pooling on an interval, or there is an interval on which two realizations are sent (by inverting the mean-preserving spread). If G is a convex combination of extreme points, we take the convex combination of the relevant signals as the space of Blackwell experiments is closed under randomization; more generally, arbitrary mixtures of experiments are also experiments.

Hence, formally, the sender's problem is to choose a distribution over posterior mean beliefs of the receiver G that solves:

$$\max_{G \in \text{MPC}(F)} \int_0^1 v(x) \, \mathrm{d}G(x) \, .$$

As the objective is linear, a maximum is attained at one of the extreme points characterized in Theorem 2. This immediately implies that an optimal signal structure partitions the states in intervals such that, in each interval:

1. Either all states are perfectly revealed.

2. Or states are pooled, so that only one (deterministic) signal is sent for all states in this interval.

3. Or at most two different (potentially random) signals are sent for states in that interval, inducing two possible posterior means on this interval.

Policies as the above have been called "bi-pooling" by Arieli et al. (2023). Thus, an information policy corresponding to an extreme point is quite simple: it partitions the state space into two sets, one of which is a union of disjoint open intervals $\cup_i(\underline{x}_i, \overline{x}_i)$. In the complement of the union of intervals, the sender completely reveals the state. Whenever the state is in one of the intervals $(\underline{x}_i, \overline{x}_i)$, the sender reveals which interval it is in and provides an additional (possibly uninformative) binary signal. While this binary signal is in general not unique, there always exists a simple partitional signal where, for each interval $(\underline{x}_i, \overline{x}_i)$, the designer picks a sub-interval in which he sends one signal, and sends the other signal in the complement.

4.5 Optimal Delegation

Delegation problems arise whenever a principal must rely on a better-informed agent to make decisions. The challenge is that the agent's preferences may not align with the principal's, creating a fundamental trade-off between information extraction and control. The canonical delegation model, developed by Holmström (1984) and refined by Melumad and Shibano (1991), Alonso and Matouschek (2008), and Amador and Bagwell (2013), elegantly captures this tension. Recent work has revealed that this problem shares the same mathematical structure as our earlier applications: it reduces to choosing functions subject to majorization constraints (Kleiner, Moldovanu, and Strack, 2021; Kolotilin and Zapechelnyuk, 2025). This observation has led to new insights into which mechanisms optimally resolve the trade-off between information extraction and control.

A principal must choose an action $a \in \mathbb{R}$ but the optimal choice depends on a state $\theta \in [0, 1]$ known only to the agent, with prior distribution $F : \mathbb{R} \to [0, 1]$. The principal can commit to a mechanism that maps the agent's reports into (possibly random) actions. Both parties have quadratic preferences centered at different bliss points; specifically, the agent's utility is $u_A(\theta, a) = \theta a - \frac{1}{2}a^2$ and the principal's utility is $u_P(\theta, a) = \gamma(\theta)a - \frac{1}{2}a^2$, where $\gamma : [0, 1] \to \mathbb{R}$ is continuous and captures the misalignment in preferences—if $\gamma(\theta) = \theta$, preferences would be perfectly aligned. The principal believes that the state is distributed according to a density f, and both players have expected utility preferences.

A mechanism $m : \mathbb{R} \to \Delta(\mathbb{R})$ assigns a lottery over actions to each report. By the revelation principle, we can focus on incentive-compatible mechanisms where truthful reporting is optimal.

To understand the connection to majorization, it is instructive to work with indirect utility functions instead of working with mechanisms directly. The agent's *indirect utility* from mechanism m is

$$U_m(\theta) := \max_{\theta'} \ \mathbb{E}_{a\sim m(\theta')}\left[u_A(\theta,a)\right]$$

This function captures the payoff that each type expects from optimal reporting. The crucial insight is that the set of achievable indirect utilities has a simple characterization:

Lemma 2. *A function U is the indirect utility of a mechanism if and only if U is convex and $U(\theta) \leq h(\theta) := \frac{1}{2}\theta^2$ for all $\theta \in \mathbb{R}$, where $h(\theta)$ is type-θ's payoff under full discretion.*

Proof sketch: It is easy to see that if U is the indirect utility of some mechanism, then U is convex as a maximum of affine functions and that each type receives a weakly lower utility than if she could choose her favorite action, so $U \leq h$.

To develop intuition for the converse, observe that for any extreme point U of the set of convex functions below h, we can easily construct a mechanism that has U as its indirect utility: as observed in the proof of Theorem 2, there is a collection of disjoint open interval $(\underline{x}_i, \overline{x}_i)$ such that U coincides with h outside these intervals, and on each interval U consists of at most three affine pieces, two of which are tangential to h at $\underline{x}_i$ and $\overline{x_i}$, respectively. This indirect utility is induced by allowing the agent to choose any action in the complement of $\bigcup_i(\underline{x}_i, \overline{x}_i)$, and adding to this choice set at most one additional lottery over actions that the agent could choose per each interval.

Because the set of convex functions below h is not compact, this argument does not imply directly that every convex function below h is the indirect utility of some mechanism. However, a direct argument that does not rely on the extreme point characterization is easily available, see Lemma 1 in Kleiner (2025). Q.E.D.

To see the explicit majorization structure, consider mechanisms satisfying $U(\theta) = h(\theta)$ for all θ outside some interval $[\theta_1, \theta_2]$. If θ_1 and θ_2 are chosen small and large enough, this is a mild assumption (recall that types outside $[0, 1]$ have zero probability), one that is shown to be without loss of optimality in Kolotilin and Zapechelnyuk (2025) and Kleiner, Moldovanu, and Strack (2021). This assumption ensures that the derivative of the indirect utility of any mechanism satisfies $U' \succeq h'$; since U' is also non-decreasing, $U' \in \mathrm{MPC}(h')$. Conversely, for every function in $\mathrm{MPC}(h')$, its antiderivative is the indirect utility of an incentive-compatible mechanism.

Using the assumption on the principal's preferences, the objective becomes linear once formulated in terms of indirect utilities. We obtain the following result:

Proposition 7. *The principal's problem can be stated as choosing a function in $MPC(h')$ to maximize a linear objective function.*

We mention three implications of this result: First, it implies optimal mechanisms are relatively simple. In particular, optimal mechanisms involve at most one lottery per interval of "excluded" actions. Second, it implies that the optimization results in Kleiner, Moldovanu, and Strack (2021) can be employed to characterize optimal mechanisms in delegation models. Finally, the result implies that there is a close connection between this delegation problem and the persuasion problem described in Section 4.4 because both share the same mathematical structure. Kolotilin and Zapechelnyuk (2025) show that, as long as types and actions are one-dimensional, the two problems are in fact equivalent even under more general preferences than the ones considered here.

This equivalence breaks down in multiple dimensions, where the delegation problem no longer admits a majorization formulation. Nevertheless, Kleiner (2025) shows that even then, optimal delegation reduces to choosing convex functions below a given bound—the natural multidimensional generalization of the framework described above.

4.6 Additional Applications

As an example of the more recent applications that have appeared in the literature, consider Bergemann et al. (2022) who combine a classical allocation problem with an information design problem. They derive the revenue-maximizing information structure in the classical second-price auction where bidders' values follow the symmetric, independent values paradigm, and where the seller can provide signals about the bidders' values such that each bidder observes only information about his/her own value and such that there is no common source of randomization in the signals. Finally, their seller is restricted to symmetric information structures. The seller faces then a trade-off: providing more information improves allocative efficiency but also induces higher information rents for bidders. Technically the authors solve the following optimization problem:

$$\max_{G \in \mathrm{MPC}(F)} \int_0^{\infty} t \, \mathrm{d}[nG^{n-1}(t)(1 - G(t)) + G^n(t)]$$

where F is the prior distribution of values for each bidder, where n is the number of bidders and where G is the posterior distribution induced by the seller's information policy. Note that

$$nG^{n-1}(t)(1 - G(t)) + G^n(t)$$

represents here the distribution of the second order statistic out n random variables distributed according to G. The above maximization problem is not linear (and neither convex nor concave) in G, and therefore it seems outside the scope of our methods. Nevertheless, recalling that $G \in \mathrm{MPC}(F) \Leftrightarrow G^{-1} \in \mathrm{MPS}(F^{-1})$ a change of variable from G to G^{-1} transforms the optimization problem into

$$\max_{G^{-1} \in \mathrm{MPS}(F^{-1})} \int_0^1 S_n'(q) G^{-1}(q) dq$$

where $S_n(q) = nq^{n-1}(1 - q) + q^n$ is the quantile distribution of the second order statistic (observe that this independent of G). The above transformed problem is now linear in G^{-1} and therefore it can be readily solved by the methods described above. The information disclosure policy that maximizes the revenue of the seller is to fully reveal low values where competition is relatively high but to pool high values where competition is relatively low.

4.7 Additional Constraints

In some economic applications, other constraints are relevant, in addition to the monotonicity and majorization constraints. Nikzad (2022) analyzes the extreme points of the set of functions that are either in MPS(f) or in MPC(f) and satisfy finitely many additional linear constraints (see also Candogan and Strack, 2023). These extreme points are closely related to the extreme points characterized above: they can have up to one additional jump discontinuity per each extra constraint. Nikzad (2022) then applies his results to auction problems in which the designer has to satisfy additional constraints (e.g., the welfare generated by the auction has to lie above some bound), to a redistribution problem based on Dworczak, Kominers, and Akbarpour (2021), and to a procurement problem based on Gershkov et al. (2021).

Taking a different but related approach, Augias and Uhe (2025) consider extreme points of the set of convex functions that lie between two given convex functions and satisfy that their subgradients lie in a given interval. To connect this to the majorization results above, note that the two given convex functions can be taken to be the integrals of non-decreasing functions f and g, where $f \succeq g$. Thus, Augias and Uhe (2025) obtain as a special case of their analysis a characterization of the extreme points of the set MPS(f) $\cap$ MPC(g) where $f \succeq g$. In words, these authors characterize extreme points of the set of functions that are simultaneously mean-preserving spreads of some function f and mean-preserving contractions of some function g. They apply their results to delegation and persuasion problems with outside options.

4.8 Multidimensional Extensions

While majorization elegantly captures various economic design problems in one-dimensional spaces, the resulting notions in multidimensional spaces are more complex. One example is the pair of delegation and persuasion problems that give rise to the same constraint set MPC. To illustrate the differences between the corresponding problems in multidimensional settings, observe that our characterization above implies that the extreme points of the set of convex functions that lie below a given convex function have a tractable structure in one dimension. In contrast, in multidimensional spaces the set of extremal convex functions is very large and is, in fact, dense in the set of convex functions (Johansen, 1974; Bronshtein, 1978; Lahr and Niemeyer, 2024). This implies that a mere characterization of the extreme points loses many of the advantages we highlighted above, and shows that different analytical tools are required for multidimensional problems of this form (e.g., results on when particular solutions are optimal instead of results on which solutions can be optimal). In a principal-agent model with a multidimensional state space, Lahr and Niemeyer point to an isomorphism between DIC mechanisms and delegation sets. The latter are convex subsets of the set of lotteries called menus. In particular, a mechanism is an extreme point if and only if the respective menu is an extreme point. Hence, properties of extreme mechanisms—that appear as solutions to linear or convex optimization problems—can be directly translated into geometric properties of the extreme points of the set of menus. Their main result characterizes the extreme points of the set of DIC mechanisms for screening problems with linear utility: such mechanisms are exhaustive, meaning that their menus cannot be scaled and translated to make additional constraints binding. The characterization is of a geometric nature, in terms of the combination of two properties called indecomposability and maximality that were first studied by Gale (1954).

On the other hand, the characterization of the extreme points of the set of probability distributions that are mean-preserving contractions of a given probability distribution can be extended to multidimensional spaces. Moreover, these extreme points retain their relatively simple structure that we observed in one dimension: every extreme point with finite support is characterized by a partition of the state space into convex sets such that the original distribution is contracted, on each partition element, to a probability distribution with affinely independent support (Kleiner et al., 2024). An interesting feature that cannot be gleaned from the unidimensional case is that convex partitions corresponding to exposed extreme points has an additional structure that turns it into a Laguerre (or power) diagram of the type encountered in the theory of semi-discrete optimal transportation. These insights also illustrate a fundamental difference between persuasion and delegation problems in multidimensional spaces.

4.9 Alternative Orders

Moving beyond the majorization order, recent work has explored alternative ways to order functions that arise naturally in economic applications. K. H. Yang and Zentefis (2024) characterize the extreme points of the set of non-decreasing functions that lie pointwise between two given non-decreasing functions. They apply their results to gerrymandering, quantile-based persuasion (see also Kolotilin and Wolitzky, 2024), overconfidence, and security design. Gershkov et al. (2025) use a similar extreme point characterization in order to find the optimal monopolistic insurance contracts offered to privately informed agents who have dual utility functions a la Yaari (1987).

F. Yang and K. H. Yang (2025) analyze non-decreasing functions $f : [0,1]^n \to [0,1]$ and their extreme points. They also analyze tuples of non-decreasing functions $(q_1, \ldots, q_n)$, where $q_i : [0,1] \to [0,1]$, such that there is a function $f : [0,1]^n \to [0,1]$ whose ith marginal is q_i (that is, $\int f(x_i, x_{-i}) \, \mathrm{d}x_{-i} = q_i(x_i)$) and characterize the extreme points of this set. They apply their results to analyze interim efficient mechanisms in bilateral trade, asymmetric reduced form auctions, public good problems, and information design with privacy constraints.

These extensions demonstrate both the versatility of the extreme point approach and the rich variety of constraint structures arising in modern economic design problems.

Appendix

Lemma 3. *Let X, Y be completely metrizable topological vector spaces and let $T : X \to Y$ be affine, continuous, and onto. Then the inverse of T has a measurable selection. Moreover, for any $\mu \in \Delta(Y)$ there exists $\nu \in \Delta(X)$ such that $\int g \circ T \, \mathrm{d}\nu = \int g \, \mathrm{d}\mu$ for all g such that either integral exists.*

Proof. T is an open mapping by Thm. 5.18 in Aliprantis and Border (2006). Therefore, its inverse, denoted by S, is lower hemicontinuous (Thm. 17.7 in Aliprantis and Border, 2006). It follows that S is weakly measurable (Def. 18.1 in Aliprantis and Border, 2006). Since T is continuous and onto, S is closed-valued and nonempty-valued. By Thm. 18.13 in Aliprantis and Border (2006), it has a measurable selection, denoted by s. We define the pushforward of μ under s by $\nu(E) := \mu(s^{-1}(E))$ for any Borel $E \subseteq X$. The claim follows from the change of variable formula for the pushforward measure. Q.E.D.

Proof of Proposition 5. Let T be a stopping time such that $M_{T \wedge t}$ is an uniformly integrable martingale and let G denote its distribution. For any convex h, $(h(M_t))$ is a submartingale and Doob's optional stopping theorem implies $\int h \, \mathrm{d}G = \mathbb{E}h(M_T) \geq \mathbb{E}h(M_0) = \int h \, \mathrm{d}F$. Hence $G \in \mathrm{MPS}(F)$.

For the converse, let G be an extreme point of $\mathrm{MPS}(F)$. By Theorem 1, there exists a countable family $\{(\underline{x}_i, \overline{x}_i)\}_{i \in I}$ such that

$$G(x) = \begin{cases} F(x) & \text{if } x \notin \bigcup_{i \in I} [\underline{x}_i, \overline{x}_i) \\ \frac{\int_{\underline{x}_i}^{\overline{x}_i} F(s) \, \mathrm{d}s}{\overline{x}_i - \underline{x}_i} & \text{if } x \in [\underline{x}_i, \overline{x}_i). \end{cases} \tag{6}$$

Let $A := \mathbb{R} \setminus [\bigcup_i (\underline{x}_i, \overline{x}_i)]$. Then the following stopping time T satisfies $M_T \sim G$: $T(\omega, u) = \inf\{t : M_t(\omega) \in A\}$, that is, stop as soon as the process enters A.

If G is a convex combination of extreme points, then the corresponding convex combination of stopping times induces G as the distribution of the stopped process. For arbitrary

$G \in \mathrm{MPS}(F)$ there exists a probability distribution μ supported on the extreme points of $\mathrm{MPS}(F)$ that represents G. By Lemma 3, there is a corresponding distribution over stopping times, and we can define the corresponding randomized stopping time. (Alternatively, we can consider a sequence of convex combination of extreme points of $\mathrm{MPS}(F)$ that converges to G. The corresponding sequence of stopping times converges along a subsequence (Baxter and Chacon, 1977), and the limits induces distribution G.) Q.E.D.

References

Aliprantis, Charalambos D and Kim C Border (2006). *Infinite dimensional analysis: a hitchhiker's guide*. Springer Science & Business Media (cit. on p. 36).

Alonso, Ricardo and Niko Matouschek (2008). "Optimal Delegation". In: *Review of Economic Studies* 75.1, pp. 259–293. (Cit. on p. 29).

Amador, Manuel and Kyle Bagwell (2013). "The Theory of Optimal Delegation with an Application to Tariff Caps". In: *Econometrica* 81.4, pp. 1541–1599 (cit. on p. 29).

Arieli, Itai, Yakov Babichenko, Rann Smorodinsky, and Takuro Yamashita (2023). "Optimal persuasion via bi-pooling". In: *Theoretical Economics* 18.1, pp. 15–36 (cit. on p. 28).

Ashlagi, Itai, Faidra Monachou, and Afshin Nikzad (2025). "Optimal allocation via waitlists: Simplicity through information design". In: *Review of Economic Studies* 92.1, pp. 40–68 (cit. on p. 17).

Augias, Victor and Lina Uhe (2025). *The Economics of Convex Function Intervals* (cit. on pp. 6, 33).

Baxter, John R and Rafael V Chacon (1977). "Compactness of stopping times". In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 40, pp. 169–181 (cit. on p. 37).

Beare, Brendan K. (2023). "Optimal measure preserving derivatives revisited". In: *Mathematical Finance* 33.2, pp. 370–388 (cit. on p. 25).

Bergemann, Dirk, Tibor Heumann, Stephen Morris, Constantine Sorokin, and Eyal Winter (2022). "Optimal information disclosure in classic auctions". In: *American Economic Review: Insights* 4.3, pp. 371–388 (cit. on p. 31).

Blackwell, David (1953). "Equivalent comparisons of experiments". In: *The annals of mathematical statistics*, pp. 265–272 (cit. on p. 27).

Border, Kim C (1991). "Implementation of reduced form auctions: A geometric approach". In: *Econometrica: Journal of the Econometric Society*, pp. 1175–1187 (cit. on p. 12).

Bronshtein, Efim Mikhailovich (1978). "Extremal convex functions". In: *Siberian Mathematical Journal* 19.1, pp. 6–12 (cit. on pp. 6, 34).

Candogan, Ozan and Philipp Strack (2023). "Optimal disclosure of information to privately informed agents". In: *Theoretical Economics* 18.3, pp. 1225–1269 (cit. on p. 33).

Condorelli, Daniele (2012). "What money can't buy: Efficient mechanism design with costly signals". In: *Games and Economic Behavior* 75.2, pp. 613–624 (cit. on p. 19).

Damiano, Ettore and Hao Li (2007). “Price discrimination and efficient matching”. In: *Economic Theory* 30.2, pp. 243–263 (cit. on p. 19).
Dana, Rose-Anne (2005). “A representation result for concave Schur concave functions”. In: *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 15.4, pp. 613–634 (cit. on p. 25).
Dworczak, Piotr, Scott Duke Kominers, and Mohammad Akbarpour (2021). “Redistribution through markets”. In: *Econometrica* 89.4, pp. 1665–1698 (cit. on p. 33).
Dworczak, Piotr and Giorgio Martini (2019). “The Simple Economics of Optimal Persuasion”. In: *Journal of Political Economy* 127.5, pp. 1993–2048 (cit. on pp. 10, 27).
Dybvig, Philip H (1988). “Distributional analysis of portfolio choice”. In: *Journal of Business*, pp. 369–393 (cit. on p. 25).
Fan, Ky and Georg Gunther Lorentz (1954). “An integral inequality”. In: *The American Mathematical Monthly* 61.9, pp. 626–631 (cit. on p. 8).
Gale, David (1954). “Irreducible convex sets”. In: *Proc. Intern. Congr. Math., Amsterdam* 2, pp. 217–218 (cit. on p. 34).
Gentzkow, Matthew and Emir Kamenica (2016). “A Rothschild-Stiglitz approach to Bayesian persuasion”. In: *American Economic Review* 106.5, pp. 597–601 (cit. on p. 27).
Gershkov, Alex, Jacob Goeree, Alexey Kushnir, Benny Moldovanu, and Xianwen Shi (2013). “On the equivalence of Bayesian and dominant strategy implementation”. In: *Econometrica* 81.1, pp. 197–220 (cit. on p. 15).
Gershkov, Alex and Benny Moldovanu (2009). “Dynamic revenue maximization with heterogeneous objects: A mechanism design approach”. In: *American Economic Journal: Microeconomics* 1.2, pp. 168–198 (cit. on p. 16).
— (2010). “Efficient sequential assignment with incomplete information”. In: *Games and Economic Behavior* 68.1, pp. 144–154 (cit. on p. 16).
Gershkov, Alex, Benny Moldovanu, Philipp Strack, and Mengxi Zhang (2021). “A theory of auctions with endogenous valuations”. In: *Journal of Political Economy* 129.4, pp. 1011–1051 (cit. on p. 33).
— (2025). “Optimal security design for risk-averse investors”. In: *American Economic Review* 115.6, pp. 2050–2092 (cit. on p. 35).
Gershkov, Alex and Eyal Winter (2023). “Gainers and losers in priority services”. In: *Journal of Political Economy* 131.11, pp. 3103–3155 (cit. on p. 19).
Hardy, Godfrey H, John E Littlewood, and George Pólya (1934). *Inequalities.* Cambridge University Press (cit. on pp. 18, 22).
Hart, Sergiu and Philip J Reny (2015). “Implementation of reduced form mechanisms: a simple approach and a new characterization”. In: *Economic Theory Bulletin* 3.1, pp. 1–8 (cit. on p. 12).
Holmström, Bengt (1984). “On the Theory of Delegation”. In: *Bayesian Models in Economic Theory.* Ed. by M Boyer and R Kihlstrom. North Holland, pp. 115–141 (cit. on p. 29).
Hong, Chew Soo, Edi Karni, and Zvi Safra (1987). “Risk aversion in the theory of expected utility with rank dependent probabilities”. In: *Journal of Economic theory* 42.2, pp. 370–381 (cit. on p. 24).
Hoppe, Heidrun C, Benny Moldovanu, and Aner Sela (2009). “The theory of assortative matching based on costly signals”. In: *The Review of Economic Studies* 76.1, pp. 253–281 (cit. on p. 19).

Johansen, Søren (1974). “The extremal convex functions”. In: *Mathematica Scandinavica* 34.1, pp. 61–68 (cit. on p. 34).

Kleiner, Andreas (2025). “Optimal delegation in a multidimensional world”. In: *arXiv preprint arXiv:2208.11835* (cit. on pp. 30, 31).

Kleiner, Andreas, Benny Moldovanu, and Philipp Strack (2021). “Extreme points and majorization: Economic applications”. In: *Econometrica* 89.4, pp. 1557–1593 (cit. on pp. 0, 1, 9–11, 14, 19, 29–31).

Kleiner, Andreas, Benny Moldovanu, Philipp Strack, and Mark Whitmeyer (2024). “The Extreme Points of Fusions”. In: *arXiv preprint arXiv:2409.10779* (cit. on p. 34).

Kolotilin, Anton (2018). “Optimal Information Disclosure: A Linear Programming Approach”. In: *Theoretical Economics* 13.2, pp. 607–635 (cit. on p. 27).

Kolotilin, Anton and Alexander Wolitzky (2024). “Distributions of posterior quantiles via matching”. In: *Theoretical Economics* 19.4, pp. 1399–1413 (cit. on p. 35).

Kolotilin, Anton and Andriy Zapechelnyuk (2025). “Persuasion meets delegation”. In: *Econometrica* 93.1, pp. 195–228 (cit. on pp. 29–31).

Krishna, Kala, Sergey Lychagin, Wojciech Olszewski, Ron Siegel, and Chloe Tergiman (2025). “Pareto improvements in the contest for college admissions”. In: *Review of Economic Studies*, rdaf033 (cit. on p. 19).

Lahr, Patrick and Axel Niemeyer (2024). *Extreme Points in Multi-Dimensional Screening* (cit. on p. 34).

Machina, Mark J (1982). “” Expected utility” analysis without the independence axiom”. In: *Econometrica: Journal of the Econometric Society*, pp. 277–323 (cit. on p. 24).

Manelli, Alejandro M and Daniel R Vincent (2010). “Bayesian and dominant-strategy implementation in the independent private-values model”. In: *Econometrica* 78.6, pp. 1905–1938 (cit. on p. 15).

Maskin, Eric and John Riley (1984). “Optimal auctions with risk averse buyers”. In: *Econometrica: Journal of the Econometric Society*, pp. 1473–1518 (cit. on p. 12).

Matthews, Steven A (1984). “On the implementability of reduced form auctions”. In: *Econometrica: Journal of the Econometric Society*, pp. 1519–1522 (cit. on p. 12).

Melumad, Nahum and Toshiyuki Shibano (1991). “Communication in Settings with No Transfers”. In: *RAND Journal of Economics* 22.2, pp. 173–198 (cit. on p. 29).

Moldovanu, Benny, Aner Sela, and Xianwen Shi (2007). “Contests for status”. In: *Journal of political Economy* 115.2, pp. 338–363 (cit. on p. 20).

Myerson, Roger B (1981). “Optimal auction design”. In: *Mathematics of operations research* 6.1, pp. 58–73 (cit. on pp. 9, 12).

Nikzad, Afshin (2022). “Constrained Majorization: Applications in Mechanism Design” (cit. on p. 33).

Obłój, Jan (2004). “The Skorokhod embedding problem and its offspring.” In: *Probability Surveys [electronic only]* 1, pp. 321–392 (cit. on p. 26).

Quiggin, John (1982). “A theory of anticipated utility”. In: *Journal of economic behavior & organization* 3.4, pp. 323–343 (cit. on p. 24).

Shaked, Moshe and J George Shanthikumar (2007). *Stochastic orders*. Springer (cit. on p. 21).

Shapiro, Alexander (2010). “On duality theory of conic linear problems”. In: *Semi-Infinite Programming: Recent Advances*. Ed. by M. A. Goberna and M. A. Lopez. Springer, U.S.A. (cit. on p. 10).

Strassen, Volker (1965). “The existence of probability measures with given marginals”. In: *The Annals of Mathematical Statistics* 36.2, pp. 423–439 (cit. on pp. 22, 27).
Xiao, Peiran (2024). “Allocating Positional Goods: A Mechanism Design Approach”. In: *arXiv preprint arXiv:2411.06285* (cit. on pp. 19, 20).
Yaari, Menahem E (1987). “The dual theory of choice under risk”. In: *Econometrica: Journal of the Econometric Society*, pp. 95–115 (cit. on pp. 24, 35).
Yang, Frank and Kai Hao Yang (2025). “Multidimensional Monotonicity and Economic Applications”. In: *arXiv preprint arXiv:2502.18876* (cit. on p. 35).
Yang, Kai Hao and Alexander K. Zentefis (2024). “Monotone Function Intervals: Theory and Applications”. In: *American Economic Review* 114.8, pp. 2239–70 (cit. on p. 35).