Extreme Points and Majorization: Economic Applications*

Andreas Kleiner Benny Moldovanu Philipp Strack†

April 8, 2020

Abstract

We characterize the set of extreme points of monotonic functions that are either majorized by a given function f or themselves majorize f and show that these extreme points play a crucial role in many economic design problems. Our main results show that each extreme point is uniquely characterized by a countable collection of intervals. Outside these intervals the extreme point equals the original function f and inside the function is constant. Further consistency conditions need to be satisfied pinning down the value of an extreme point in each interval where it is constant. Finally, we apply these insights to a varied set of economic problems: equivalence and optimality of mechanisms for auctions and (matching) contests, Bayesian persuasion, optimal delegation, and decision making under uncertainty.

1 Introduction

The *majorization* relation, due to Hardy, Littlewood and Polya (1929), embodies an elegant notion of "variability" and defines a partial order among vectors in Euclidean space, or among integrable functions.[1] In this paper we show that, somewhat surprisingly, many well-known optimal design and decision problems have a basic common

*We wish to thank Alex Gershkov for very helpful remarks and to seminar participants at Tel-Aviv and at the Bonn-Mannheim CRC conference for their comments. Moldovanu acknowledges financial support from the German Science Foundation.

†Kleiner: Arizona State University; Moldovanu: University of Bonn; Strack: Yale University

[1] In Economics, a related order has been popularized and applied, most famously to the theory of choice under risk, under the name *second-order stochastic dominance.*

structure: all these problems can be reduced to the choice of an optimal element - that maximizes a given functional - from the set of monotonic functions that are either majorized by, or majorize a given monotonic function f. Examples are the determination of feasible and optimal auctions and matching contests, of optimal delegation, Bayesian persuasion, and of optimal risky choice for non-expected utility decision makers.

Our main results characterize the extreme points of the sets of monotonic functions that are majorized by, or majorize a given monotonic function f. The monotonicity constraint, a novel feature of our work, is not standard in the mathematical literature: the set of extreme points that respect monotonicity is quite different from the set of extreme points obtained without imposing it (see Ryff (1967) for the latter). In addition, every extreme point is *exposed*, i.e., it can be obtained as the unique maximizer of some linear functional. Hence, no extreme point can be a-priori dismissed as potentially irrelevant for maximization.

The majorization constraint is not always explicit in the description of the economic problems, and it arises for different reasons. For example, in the theory of auctions it stems from a feasibility condition related to the availability of a limited supply (i.e., reduced-form auctions), in the theory of optimal delegation it is a consequence of incentive-compatibility, and in Bayesian persuasion it is induced by information garbling together with Bayesian consistency. The monotonicity constraint also arises for various reasons, for example because of incentive compatibility constraints, or because a cumulative distribution function is non-decreasing.

Our characterization of extreme points is useful in applications since, by Choquet's theorem[2], any feasible element in a relevant majorization set can be expressed as an integral with respect to a measure that is supported on the extreme points of that set. In particular, any linear or convex functional will attain a maximum on an extreme point.

Information about the extreme points is very useful besides their role for maximization: any property that is satisfied by the extreme points and that is preserved under averaging will also be satisfied by all elements of a majorization set. Since the sets of extreme points of majorization sets are much smaller than the original sets, and since they can be easily parametrized (see Theorem 1 and 2), the integral representation drastically simplifies the task of establishing a given property for the original set. In Section 4.1 we illustrate this methodology in the classical context of auctions: our insights almost immediately imply both a generalized version of Border's Theorem about reduced auctions, and the equivalence of Bayesian and Dominant Strategy incentive compatible mechanisms in the symmetric case. The characterization of extreme points

[2]See Phelps (2001) for an excellent introduction.

also immediately yields the revenue maximizing mechanism and the welfare meaximizing mechanisms in multi-prize contests where agents waste resources in order to obtain prizes..

Consider the set of non-decreasing functions that majorize, or are majorized by, a non-decreasing function f. Roughly speaking, each extreme point of this set is characterized by its specific, countable collection of intervals. Outside these intervals an extreme point must equal f, and inside each interval the extreme point is a step function that takes at most three different values determined by specific, local "equal-areas"consistency conditions (such that the majorization constraints become tight). We relate these flat areas to the classical *ironing* procedure, and show how our majorization/extreme points focus illuminate it and its uses in applications.

We also identify more specialized conditions on the objective functional such as super-modularity that allow us to infer further features of the particular extreme point where the objective functional will attain its maximum. A functional that respects the majorization order (or its converse) will have an optimum on an element that is the least variable (most variable) in a given set. Thus, under conditions that are often present in applications and that can be easily checked, the optimum is either achieved at the a-priori fixed function f or at a step function g with at most two steps. This is a consequence of an elegant theorem due to Fan and Lorentz (1954) that identifies necessary and sufficient conditions for a large class of convex functionals to respect the majorization order.

We conclude the paper with a varied array of illustrations. We cover feasibility, equivalence and optimality of mechanisms for auctions and (matching) contests in Section 4.1. In Sections 4.2-4.4 we show how our majorization approach illuminates and simplifies optimal delegation and optimal Bayesian persuasion, while making transparent the close connections between these two models and their respective optimal mechanisms. In particular, our approach implies that the equivalence between delegation and persuasion mechanisms obtained for a subset of mechanisms by Kolotilin and Zapechelnyuk (2019) extends to all randomized mechanisms, and provides novel results on the optimality of stochastic delegation mechanisms. We also observe that the different instances of the same general application, e.g. Bayesian persuasion, may require both types of maximization described in our paper - on majorizing and on majorized sets of functions, respectively. In recent, independent work, Arieli et al. (2020) also study a Bayesian persuasion problem via an extreme points approach and consider maximization on a majorizing set of functions.[3] Finally, in Section 4.5 we apply our approach to decision making under uncertainty, including portfolio choice, for agents

[3] See Sections 2 and 4.3 for more details. We thank Itai Arieli for bringing this paper to our attention.

with expected and non-expected utility.

Our main goal in this last part is to reveal the common underlying role of majorization, and to offer a unified treatment to well-known but complex problems that have been previously attacked by separate, "ad-hoc" methods. We also show that several classical results in the relevant literatures are straightforward corollaries of our findings.

The rest of the paper is organized as follows: In the next Subsection we present some majorization preliminaries and relations to other concepts. Section 2 contains the representation result a la Choquet, and the characterizations of extreme points of the majorization sets. In Section 3 we study the optimization of objective functionals with special properties such as convexity, linearity and super-modularity. A major role is played by the Fan-Lorentz integral inequality. Section 4 contains the applications: auctions and contests (4.1), optimal delegation (4.2), Bayesian persuasion (4.3 and 4.4), decision making under uncertainty with expected and non-expect utility (4.5). Section 5 concludes. Several proofs are gathered in an Appendix.

1.1 Majorization Preliminaries

Throughout, we consider functions that map the unit interval $[0,1]$ into the real numbers, and we identify functions that are equal almost everywhere. For two non-decreasing functions $f,g \in L^1(0,1)$ we say that f *majorizes* g, denoted by $g \prec f$, if the following two conditions hold:

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

We say that f *weakly majorizes* g, denoted by $g \prec_w f$, if the first condition above holds (but not necessarily the second). For non-monotonic functions f,g majorization is defined analogously by comparing their non-decreasing *rearrangements* f^*,g^*, i.e. f majorizes g if $g^* \prec f^*$.

Majorization is closely related to other concepts from Economics and Statistics. Let X_F and X_G be random variables with distributions F and G, respectively, defined on the interval $[0,1]$. Define also

$$G^{-1}(x) = \sup\{s : G(s) \leq x\},\ x \in [0,1]$$

to be the generalized inverse (or quantile function) of G, and analogously for F. It follows from Shaked and Shanthikumar (2005, Section 3.A) that

$$G \prec F \;\Leftrightarrow X_G \leq_{cv} X_F \Leftrightarrow X_F \leq_{cx} X_G \Leftrightarrow F^{-1} \prec G^{-1},$$

where *cv* (*cx*) denotes the *concave (convex) stochastic order* among random variables. We also have

$$G \prec F \;\Leftrightarrow X_G \leq_{ssd} X_F \text{ and } \mathbb{E}[X_G] = \mathbb{E}[X_F],$$

where *ssd* denotes the standard *second-order stochastic dominance*.[4] This implies that F majorizes G if and only if G is a *mean preserving spread* of F, i.e., one can construct random variables X, Y, jointly distributed on some probability space, such that $X \sim F$, $Y \sim G$ and such that $Y = \mathbb{E}[X|Y]$.[5]

2 Extreme Points and Majorization

An *extreme* point of a convex set A is a point $x \in A$ that cannot be represented as a convex combination of two other points in A.[6] 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. In particular, such a set has extreme points. The interest in extreme points from an optimization point of view stems from *Bauer's Maximum Principle*: 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.

Let $L^1(0,1)$ denote the real-valued and integrable functions defined on $[0,1]$ and recall that $f \in L^1(0,1)$ is in fact an equivalence class of functions that are equal almost everywhere.[7] Given $f \in L^1(0,1)$, let the *orbit* of f, $\Omega(f)$, be the set of all functions that are majorized by f:

$$\Omega(f) = \{g \in L^1(0,1) \mid g \prec f\}.$$

Ryff (1967) has shown that $g \in \Omega(f)$ is an extreme point of this set if and only if $g = f \circ \Psi$

[4] A non-decreasing density $f = F'$ majorizes another non-decreasing density $g = G'$ if and only if the associated distribution F dominates G in *first-order stochastic dominance.*

[5] See Strassen (1965).

[6] Formally $x \in A$ 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 $z = x$.

[7] Addition (scalar multiplication) in L^1 is defined by pointwise addition (scalar multiplication) of arbitrary representatives of the equivalence classes. An equivalence class g is an extreme point of a convex set $M \subset L^1$ if $g \in M$ and there do not exist equivalence classes $f_1, f_2 \in M$ different from g and $\alpha \in (0,1)$ such that $g = \alpha f_1 + (1-\alpha) f_2$. Finally, an equivalence class 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 canonical representative.

where Ψ is a *measure preserving* transformation of $[0,1]$ into itself. This generalizes the discrete case analyzed by Hardy, Littlewood and Polya where the extreme points correspond, by the *Birkhoff-von Neumann Theorem*, to permutation matrices.

In economic applications we are often interested in functional maximizers that are non-decreasing, e.g., a cumulative distribution function in Bayesian persuasion, or an incentive compatible allocation in mechanism design. Thus, we are led to the study of the subset of non-decreasing functions in the orbit $\Omega(f)$,

$$\Omega_m(f) = \{g \in L^1(0,1) \mid g \text{ non-decreasing such that } g \prec f\}.$$

Similarly, we denote by $\Omega_{m,w}(f)$ the set of non-decreasing functions that are weakly majorized by f. Finally, let

$$\Phi_m(f) = \{g \in L^1(0,1) \mid g \text{ non-decreasing such that } g \succ f \text{ and } f(0_+) \le g \le f(1_-)\}.$$ [8]

Proposition 1 (Representation)**.**

1. *Let $f \in L^1(0,1)$ be non-decreasing. Then, the sets $\Omega_m(f)$, $\Omega_{m,w}(f)$, and $\Phi_m(f)$ are convex and compact in the norm topology, and hence the respective sets of extreme points are not empty.*[9]
2. *For any $g \in \Omega_m(f)$ there exists a probability measure λ_g supported on the set of extreme points of $\Omega_m(f)$, $\operatorname{ext}\Omega_m(f)$, such that $g = \int_{\operatorname{ext}\Omega_m(f)} h \,\mathrm{d}\lambda_g(h)$ (and analogously for any $g \in \Omega_{m,w}(f)$ and $g \in \Phi_m(f)$).*[10]

The second part of the Proposition is a consequence of Choquet's celebrated theorem that constitutes a powerful strengthening of the Krein-Milman insight. Immediate implications are a generalized Jensen inequality and the Bauer's Maximum Principle for the respective majorization sets.

While applications of Choquet's result in infinite-dimensional function spaces are often hampered by the difficulty to identify all extreme points of a given set, we offer below relatively simple characterizations of the relevant extreme points.

[8] The additional constraint $f(0_+) := \lim_{x \searrow 0} f(x) \le g \le \lim_{x \nearrow 1} f(x) =: f(1_-)$ ensures compactness, and is suitable for our applications below.

[9] For linear maximization purposes it is enough to establish compactness in the weak topology. We need here the stronger result in order to apply Choquet's Theorem.

[10] This holds if and only if $V(g) = \int V(h)\,\mathrm{d}\mu(h)$ for any continuous, linear functional V. The integral in the proposition's statement is a *Bochner integral*. See Aliprantis and Border (2006) for an introductory exposition.

Theorem 1. *Let f be non-decreasing. Then g is an extreme point of $\Omega_m(f)$ if and only if there exists a 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 $\Omega_m(f)$ 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.

Our next result establishes that all extreme points of $\Omega_m(f)$ are *exposed*. This implies that we cannot a-priori exclude any extreme point from consideration when maximizing a linear functional. Recall that an element x of a convex set A is exposed if there exists a linear functional that attains its maximum on A uniquely at x.[11] Every exposed point must be extreme, but the converse need not be true in general.

Corollary 1. *Every extreme point of $\Omega_m(f)$ is exposed.*

Following the approach in Horsley and Wrobel (1987) (who, like Ryff, did not impose monotonicity), we can extend our characterization of extreme points to the set of weakly majorized functions. 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 2. *Suppose that f is non-decreasing and non-negative. A function g is an extreme point of $\Omega_{m,w}(f)$ if and only if it is an extreme point of the orbit $\Omega_m(f \cdot \mathbf{1}_{[\theta,1]})$ for some $\theta \in [0, 1]$.*

Finally, we characterize the extreme points of the set of non-decreasing functions that majorize f and that have the same range as f, denoted by $\Phi_m(f)$.

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

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

[11] Formally, x is exposed if there exists a supporting hyperplane H such that $H \cap A = \{x\}$.

and such that the following three conditions are satisfied:

$$(\overline{y}_i - \underline{y}_i)v_i = \int_{\underline{x}_i}^{\overline{x}_i} f(s)\,\mathrm{d}s - f(\underline{x}_i)(\underline{y}_i - \underline{x}_i) - f(\overline{x}_i)(\overline{x}_i - \overline{y}_i) \tag{3}$$

$$f(\underline{x}_i)(\overline{y}_i - \underline{x}_i) + f(\overline{x}_i)(\overline{x}_i - \overline{y}_i) \le \int_{\underline{x}_i}^{\overline{x}_i} f(s)\,\mathrm{d}s \le f(\underline{x}_i)(\underline{y}_i - \underline{x}_i) + f(\overline{x}_i)(\overline{x}_i - \underline{y}_i)\,. \tag{4}$$

If $v_i \in (f(\underline{y}_i), f(\overline{y}_i))$ then for an arbitrary point m_i satisfying $f(m_i) = v_i$ it must hold that

$$\int_{m_i}^{1} f(s)\,\mathrm{d}s \le v_i(\overline{y}_i - m_i) + f(\overline{x}_i)(\overline{x}_i - \overline{y}_i). \tag{5}$$

Condition (3) in the Theorem ensures that g and f have the same integrals for each sub-interval $[\underline{x}_i, \overline{x}_i)$, analogously to the condition imposed in Theorem 1. Condition (4) ensures that $v_i \in (f(\underline{x}_i), f(\overline{x}_i))$. If f crosses g in the interval $[\underline{y}_i, \overline{y}_i]$ then there is $m_i \in [\underline{y}_i, \overline{y}_i]$ such that $f(m_i) = v_i$. In this case, Condition (5) ensures that $\int_s^{\overline{x}_i} f(t)\,\mathrm{d}t \le \int_s^{\overline{x}_i} f(t)\,\mathrm{d}t$ for all $s \in [\underline{x}_i, \overline{x}_i)$ and thus that $f \prec g$. If $v_i \notin (f(\underline{y}_i), f(\overline{y}_i))$ Condition (4) is enough to ensure that $f \prec g$ and thus Condition (5) is not necessary.

We note here that the instance of Bayesian persuasion studied by Arieli et al. (2020) corresponds to a maximization exercise over a set of majorizing functions of the form Φ_m (see also Section 4.3 for details). Analogously to our Theorem 2, these authors identify the extreme points in their problem and further show that all extreme points are exposed.

Extreme Points: An Intuitive Description Consider the case where f is a cumulative distribution function (CDF) and recall that a CDF admits a jump at a given value if the distribution assigns a mass-point to that value. As h majorizes g if and only if g is a mean-preserving spread of h, it follows that $\Omega_m(f)$ if the set of *mean preserving spreads* of f and $\Phi_m(f)$ is the set of *mean preserving contractions* of f.

These properties are also reflected in the extreme points: Each extreme point $g \in \Omega_m(f)$ is obtained by taking the mass in each interval $[\underline{x}_i, \overline{x}_i]$ and spreading it out into two mass-points at the boundaries of the interval, $\underline{x}_i$ and $\overline{x}_i$ (see Figure 1). There is a unique way to do so while preserving the mean determined by (2). In contrast, each extreme point $g \in \Phi_m(f)$ is obtained by contracting the mass in each interval $[\underline{x}_i, \overline{x}_i]$ into two mass-points placed at $\underline{y}_i$ and $\overline{y}_i$. If $v_i \in (f(\underline{y}_i), f(\overline{y}_i))$ the CDFs g and f intersect at $m_i \in (\underline{y}_i, \overline{y}_i)$. Mass to the left of $m_i = f^{-1}(v_i)$ is moved to $\underline{y}_i$ and mass to the right of m_i is moved to $\overline{y}_i$ (see Figure 1).[12] Condition (3) determines the mass

[12]If f is not strictly increasing, then f is constant on the interval $\{s\colon f(s) = v\}$, which implies that the distribution assigns no mass to that interval. Thus, any choice of m in that interval will lead to mass being moved in the same way.

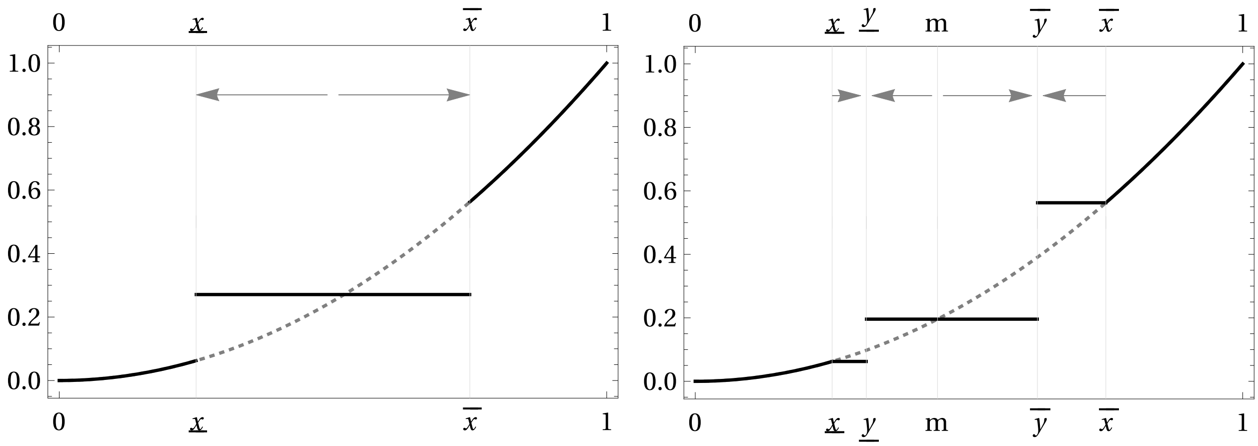

Figure 1: This figure illustrates the differences between the extreme points of $\Omega_m(f)$ and $\Phi_m(f)$. Here $f(s) = s^2$, and there is a single interval $[\underline{x}, \overline{x}] = [1/4, 3/4]$ with $[\underline{y}, \overline{y}] = [5/16, 10/16]$. On the left is the corresponding extreme point in $\Omega_m(f)$ and on the right is the corresponding extreme point in $\Phi_m(f)$. The arrows indicate how mass is moved to transform f into the extreme point.

at these mass-points, and ensures that the mean is preserved on the interval $[\underline{x}_i, \overline{x}_i]$. Condition (4) ensures that g can be obtained from f by moving mass, and Condition (5) ensures that g is a contraction of f.[13]

The main insight of Theorems 1 and 2 is that the mean-preserving spreads (or contractions) of f described there *cannot* be represented as convex combinations of other functions in Ω_m and in Φ_m, respectively, and that these are the *only* functions with this property.[14]

3 Special Objective Functionals

Our previous characterizations of extreme points determines all functions that can arise as a maximizer of some convex functional over a set described by monotonicity and majorization constraints. None of these maximizers can be a-priori ruled out, even if one restricts attention to linear functionals. However, in many applications, further

[13]A simpler characterization where each interval $[\underline{x}_i, \overline{x}_i]$ is split into the two intervals $[\underline{x}_i, m]$ and $[m, \overline{x}_i]$ each containing only a single mass-point is not valid since the mean on these sub-intervals need not be preserved by an extreme point.

[14]A related intuition for why the extreme points involve only two mass-points in each interval appears in Winkler (1988). He shows that every extreme point of a set of probability measures characterized by n constraints is the sum of at most $n+1$ mass-points. In the case of a unique constraint on the mean, this implies that any extreme point is a sum of at most two mass-points. Winkler's characterization does not hold here sinc ee impose uncountably many majorization constraints.

monotonicity or super-modularity conditions are either naturally satisfied or can be imposed on the objective function. We show below how such conditions can be used to further shrink the set of relevant extreme points.

3.1 Convex, Super-modular Functionals

A functional $V : L^1(0,1) \to \mathbb{R}$ that is monotonic with respect to the majorization order is called *Schur-concave*. Chan et al. (1987) showed that a law-invariant[15], Gâteaux-differentiable functional $V : L^1(0,1) \to \mathbb{R}$ is Schur-concave if and only if its *Gâteaux-derivatives* in directions that uniformly increase the value in a lower interval and uniformly decrease the value in a higher interval by the same total amount are non-positive (for details see Section B.1 in the Online Appendix). This criterion is not always easy to check in practice, but a very useful integral inequality, due to Fan and Lorentz (1954) identifies a large set of convex and Schur-concave functionals.[16]

Theorem 3 (Fan and Lorentz 1954)**.** *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* $f \prec g$ *if and only if the function* $K(u,t)$ *is convex in* u *and super-modular in* (u,t)*.*

Theorem 3 is extremely useful for the applications below since it provides conditions on the objective function such that the maximum over majorization sets determined by a function f is attained either at f itself (highest variability), or at a particular step function g with at most two jumps (lowest variability).

3.2 Linear Optimization under Majorization Constraints

We now consider optimization problems where the objective is a linear functional, and where the constraint set is defined by majorization and by monotonicity. The classical *Riesz Representation Theorem* says that, for every continuous, linear functional V on $L^1(0,1)$, there exists a unique, essentially bounded function $c \in L^\infty(0,1)$ such that for every $f \in L^1(0,1)$

$$V(f) = \int_0^1 c(x) f(x)\,\mathrm{d}x \tag{6}$$

[15]This means that the functional is constant over the equivalence class of functions with the same distribution (or non-decreasing re-arrangement). This requirement replaces symmetry in the discrete formulation - see Appendix.

[16]In the Appendix we provide an intuition based on the Chan et al. (1987) result.

Thus, we can confine here attention to the maximization of this kind of integrals. Note that a linear kernel of the form $K(f,x) = c(x)f(x)$ is super-modular (sub-modular) in (f,x) (and hence the linear functional given in (6) is Schur-concave (convex)) if and only if c is non-decreasing (non-increasing). We repeatedly apply this observation below.

3.2.1 Maximizing a Linear Functional on $\Omega_m(f)$

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

$$\max_{h\in\Omega_m(f)} \int c(x)h(x)\,\mathrm{d}x. \tag{7}$$

There are three cases to consider:

1. If the function c is non-decreasing, f itself is the solution for the optimization problem.
2. If c is non-increasing, then the solution for the optimization problem is the overall constant function g that is equal to $\mu_f = \int_0^1 f(x)\,\mathrm{d}x$. This follows since $h \succ g$ for any $h \in \Omega_m(f)$.
3. If c is not monotonic, other extreme points of $\Omega_m(f)$ may be optimal.

The next result essentially characterizes the conditions under which an arbitrary extreme point is optimal. The ironing technique, originally used in Myerson (1981) (see also Toikka 2011) for an optimization problem formulated without majorization constraints, can be used if the constraint set includes all non-decreasing functions in a given orbit.

Define

$$C(x) = \int_0^x c(s)\,\mathrm{d}s$$

and let $\operatorname{conv} C$ denote the convex hull of C, i.e., the largest convex function that lies below C.

Proposition 2. *Let g be an extreme point of $\Omega_m(f)$, and let $\{[\underline{x}_i, \overline{x}_i)|i \in I\}$ be the collection of intervals described in Theorem 1. If $\operatorname{conv} C$ is affine on $[\underline{x}_i, \overline{x}_i)$ for each $i \in I$ and if $\operatorname{conv} C = C$ otherwise, then g solves problem* (7)*. Moreover, if f is strictly increasing then the converse holds.*

3.2.2 Maximizing a Linear Functional on $\Phi_m(f)$

We now analyze the problem

$$\max_{h\in\Phi_m(f)} \int c(x)h(x)\,\mathrm{d}x\,. \tag{8}$$

Again, there are three cases:

1. If the function c is non-increasing then f solves this problem.
2. If c is non-decreasing, then the 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
$$\int_0^{\overline{x}} f(0_+)\,\mathrm{d}s + \int_{\overline{x}}^1 f(1_-)\,\mathrm{d}s = \int_0^1 f(s)\,\mathrm{d}s$$
Indeed, it holds that $g \in \Phi_m(f)$ and that $g \succ h$ for all $h \in \Phi_m(f)$. Therefore, the Fan-Lorentz Theorem 3 implies that g is optimal in this case.
3. If c is non-monotonic we cannot directly use the Fan-Lorentz result, but the following observations suggests an approach to solve the problem:[17]

Lemma 1. *Let*

$$C(x) = \int_0^x c(s)\,\mathrm{d}s.$$

A function $g \in \Phi_m(f)$ is optimal if there exists a concave function $\overline{C}(x) \leq C(x)$ such that:

1. $\int \overline{C}(x)\,\mathrm{d}g(x) = \int C(x)\,\mathrm{d}g(x)$, $\overline{C}(0) = C(0)$ *and* $\overline{C}(1) = C(1)$ *and*
2. $\int_0^1 \overline{C}'(x)g(x)\,\mathrm{d}x = \int_0^1 \overline{C}'(x)f(x)\,\mathrm{d}x$.

In general, there is no pointwise largest concave function below a given function. In order to verify that g is optimal, one therefore has to construct a concave function $\overline{C}$ that is specific to g. This contrasts the situation in the previous Subsection, where the convex hull provided a largest convex function below a given function.

4 Economic Applications

We now apply the above gained theoretical insights to various economic settings. We show how seemingly different and well-known problems share a common structure: they

[17] The proof of Lemma 1 is provided in the Online Appendix.

all involve maximization of functionals over majorization sets.

4.1 The Ranked-Item Auction and Contest Models

There are N agents with types $\theta_1, \ldots, \theta_N$ that are independently and identically distributed on $[0,1]$ according to a common distribution F, with bounded density $f > 0$. Each agent wants at most one object.

There are $M \leq N$ objects with qualities $0 \leq q_1 \leq q_2 \leq \ldots \leq q_M = 1$. If agent i with type θ_i receives an object with quality q_m and pays t for it, then his utility is given by $\theta_i q_m - t$. As we can always add objects with zero quality, it is without loss of generality to assume that $M = N$.

Let Π denote the set of doubly sub-stochastic $N \times N$-matrices. An *allocation* rule $\boldsymbol{\alpha} : [0,1]^N \to \Pi$ represents a (possibly random) allocation of objects as a function of types:[18] $\alpha_{ij}(\theta_i, \theta_{-i})$ denotes the probability with which agent i obtains the object with quality j. We denote by $\boldsymbol{\alpha}_i$ its i-th row vector, and also denote $\mathbf{q} = (q_1, \ldots, q_N)$. For an allocation $\boldsymbol{\alpha}$ and for each i, let

$$\varphi_i(\theta_i) = \int_{[0,1]^{N-1}} [\boldsymbol{\alpha}_i(\theta_i, \theta_{-i}) \cdot \mathbf{q}] \, f_{-i}(\theta_{-i}) \, \mathrm{d}\theta_{-i}.$$

denote the expected quality obtained by agent i, conditional on his type - this is also called the *interim allocation rule.* It is straightforward to show that an allocation $\boldsymbol{\alpha}$ is part of a *Bayesian incentive compatible* (IC) mechanism if and only if each induced *interim allocation* φ_i is non-decreasing.[19]

It is useful to also consider the quantile transformations $s_i = F(\theta_i)$, and to define the *interim quantile allocation* functions

$$\psi_i(s_i) = \varphi_i(F^{-1}(s_i))$$

These are also non-decreasing for IC allocations.[20]

Denote by $\boldsymbol{\alpha}^* : [0,1]^N \to \Pi$ the *assortative matching* of agents to objects (highest

[18] These are non-negative matrices with row- and column-sums weakly less than 1. It follows from Budish et al. (2013) that any such matrix corresponds to a randomization over feasible deterministic allocations.

[19] See for example Gershkov and Moldovanu (2010) who use discrete majorization in a dynamic mechanism design framework with several qualities.

[20] Note that, seen as a random variable, s_i is uniformly distributed.

type gets highest quality, etc.) with ties broken by fair randomization:

$$\alpha^*_{ik} = \begin{cases} \frac{1}{|\{j:\, \theta_j = \theta_i\}|} & \text{if } |\{j:\, \theta_j < \theta_i\}| \leq k-1 \leq |\{j:\; \theta_j \leq \theta_i\}| \\ 0 & \text{else} \end{cases}.$$

In our symmetric model, assortative matching $\boldsymbol{\alpha}^*$ is incentive compatible, and induces the symmetric interim allocation

$$\varphi^*_i(\theta_i) = \varphi^*(\theta_i) = \sum_{k=1}^{N} q_k \left[\frac{(N-1)!}{(k-1)!(N-k)!} F(\theta_i)^{k-1} (1 - F(\theta_i)^{N-k} \right]$$

and the symmetric interim quantile allocation

$$\psi^*_i(s_i) = \psi^*(s_i) = \sum_{k=1}^{N} q_k \left[\frac{(N-1)!}{(k-1)!(N-k)!} (s_i)^{k-1} (1 - s_i)^{N-k} \right]$$

4.1.1 A Generalization of Border's Feasibility Condition

We first show how our results can be used to prove a generalization of the symmetric version of Border's theorem for the above model. We say that a set of interim allocations $\{\varphi_i\}_{i=1}^N$ where $\varphi_i : [0,1] \to \mathbb{R}$, $i = 1, 2, \ldots, N$ is *feasible* if there exists an allocation rule $\boldsymbol{\alpha}$ that induces $\{\varphi_i\}_{i=1}^N$ as its set of interim allocations, conditional on type. We restrict attention below to symmetric interim allocation rules where $\varphi_i = \varphi$, $i = 1, 2, \ldots, N$ and thus $\psi_i = \psi, i = 1, 2, \ldots, N$.

In our terminology, Border's Theorem (1991) for the single object case, i.e., $q_N = 1$ and $q_k = 0$ for $k < N$, says that a symmetric and monotonic interim allocation φ is feasible if and only if the associated quantile interim allocation satisfies $\psi \prec_w s^{N-1}$. [21] In this case, the assortative matching interim allocation $\varphi^*(\theta_i) = [F(\theta_i)]^{N-1}$ is the *efficient* allocation and hence $\psi^*(s_i) = (s_i)^{N-1}$.The generalization to our present model is:

Theorem 4. *In the ranked-items auction model, a symmetric and monotonic interim allocation rule φ is feasible if and only if its associated quantile interim allocation $\psi(s) = \varphi(F^{-1}(s))$ satisfies*

$$\psi \prec_w \psi^*$$

[21] See also Maskin and Riley (1984) and Matthews (1984). This is not the original formulation. For connections to majorization see Hart and Reny (2015) for the one-object case, and Gershkov et al. (2019) for the identical objects case. Hart and Reny's proof method is direct. Gerskov et al.' use a result by Che et al. (2013) that is based on a network-flow approach.

where ψ^* is the quantile interim allocation generated by the assortative matching allocation.

Proof: We first show that $\psi \prec_w \psi^*$ is necessary for feasibility. Consider a monotonic and symmetric quantile interim allocation rule ψ generated by $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$. As switching to the assortative rule takes high-quality objects from lower types and gives them to higher types, we have that

$$\mathbb{E}[\boldsymbol{\alpha}_i(\theta) \cdot \mathbf{q} \mid \theta_i \geq \tau] \leq \mathbb{E}[\boldsymbol{\alpha}_i^*(\theta) \cdot \mathbf{q} \mid \theta_i \geq \tau]$$

for each agent i and for every $\tau \in [0,1]$. Note that

$$\begin{aligned}\mathbb{E}[\boldsymbol{\alpha}_i(\theta) \cdot \mathbf{q} \mid \theta_i \geq \tau] &= \frac{1}{1-F(\tau)} \int_\tau^1 \left[\int_{[0,1]^{n-1}} \boldsymbol{\alpha}_i(\theta_i, \theta_{-i}) \cdot \mathbf{q} f_{-i}(\theta_{-i}) \,\mathrm{d}\theta_{-i} \right] f(\theta_i) \,\mathrm{d}\theta_i \\ &= \frac{1}{1-F(\tau)} \int_\tau^1 \varphi(\theta_i) f(\theta_i) \,\mathrm{d}\theta_i = \frac{1}{1-s} \int_s^1 \psi(t_i) \,\mathrm{d}t_i\end{aligned}$$

where $s = F(\tau)$. Since this holds for any $\tau \in [0,1]$, we obtain that $\psi \prec_w \psi^*$.

For the converse, recall that, by Corollary 2, every extreme point ψ of $\Omega_{m,w}(\psi^*)$ is described by $\widetilde{s}_i \in [0,1]$ and by a collection of intervals $[\underline{s}_i, \overline{s}_i) \subseteq [\widetilde{s}_i, 1]$ such that

$$\psi(s_i) = \begin{cases} \psi^*(s_i) & \text{if } s_i \geq \widetilde{s}_i \text{ and } s_i \notin \bigcup_{i \in I} [\underline{s}_i, \overline{s}_i) \\ \frac{\int_{\underline{s}_i}^{\overline{s}_i} \psi^*(t_i)\,\mathrm{d}t_i}{\overline{s}_i - \underline{s}_i} & \text{if } s_i \in [\underline{s}_i, \overline{s}_i) \\ 0 & \text{if } s_i < \widetilde{s}_i \end{cases}.$$

Any such extreme point is feasible as it is implemented by the allocation rule that does not allocate to types below $\tilde{\theta}_i = F^{-1}(\widetilde{s}_i)$, uses fair randomization to determine the allocation in each interval $[\underline{\theta}_i, \overline{\theta}_i) = [F^{-1}(\underline{s}_i), F^{-1}(\overline{s}_i))$, and is otherwise assortative. Formally,

$$\alpha_{ik}(\theta) = \begin{cases} \frac{\mathbf{1}_{\{\theta_i > \tilde{\theta}\}}}{|\{j \colon m(\theta_j) = m(\theta_i)\}|} & \text{if } |\{j \colon m(\theta_j) < m(\theta_i)\}| \leq k-1 \leq |\{j \colon m(\theta_j) \leq m(\theta_i)\}| \\ 0 & \text{else} \end{cases}, \tag{9}$$

where $m : [0,1] \to [0,1]$ equals

$$m(\theta) = \begin{cases} \theta & \text{if } \theta \notin \bigcup_{i \in I} [\underline{\theta}_i, \overline{\theta}_i) \\ \frac{\overline{\theta}_i - \underline{\theta}_i}{2} & \text{if } \theta \in [\underline{\theta}_i, \overline{\theta}_i) \end{cases}.$$

We note that the allocation α can be implemented in *dominant strategies* by assigning the objects assortatively outside of $\bigcup_{i\in I}[\underline{\theta}_i, \overline{\theta}_i)$, and by using random serial dictatorship to determine the allocation of objects among agents whose values lie in the same interval $[\underline{\theta}_i, \overline{\theta}_i)$.

Let P be the mapping that assigns to any allocation rule $\boldsymbol{\alpha}$ its induced interim quantile allocation rule, and note that P is a bounded linear operator. Also, let $T : \operatorname{ext}\Omega_{m,w}(\psi^*) \to L^1([0,1]^N, \mathbb{R}^{N\times N})$ be a measurable function that assigns to any extreme point a corresponding allocation rule (which exists by Lemma 3 in the Online Appendix) and observe that $P(T(\psi)) = \psi$.

It follows from Proposition 1 that for any $\psi \in \Omega_{m,w}(\psi^*)$ there exists a probability measure λ_ψ supported on $\operatorname{ext}\Omega_{m,w}(\psi^*)$ such that $\psi = \int_{\operatorname{ext}\Omega_{m,w}(\psi^*)} \widetilde{\psi}\, \mathrm{d}\lambda_\psi(\widetilde{\psi})$. Let ν_ψ be the *pushforward* measure of λ_ψ under T [22] and define the allocation

$$\boldsymbol{\alpha} = \int \widetilde{\boldsymbol{\alpha}}\, \mathrm{d}\nu_\psi(\widetilde{\boldsymbol{\alpha}}).$$

This allocation rule induces the interim quantile allocation ψ because

$$P(\boldsymbol{\alpha}) = \int P(\widetilde{\boldsymbol{\alpha}})\, \mathrm{d}\nu_\psi(\widetilde{\boldsymbol{\alpha}}) = \int P(T(\widetilde{\psi}))\, \mathrm{d}\lambda_\psi(\widetilde{\psi}) = \int \widetilde{\psi}\, \mathrm{d}\lambda_\psi(\widetilde{\psi}) = \psi.$$

The first equality follows from Lemma 11.45 in Aliprantis and Border (2006), the second by the change-of-variable formula for pushforward measures, the third since, by definition, $P(T(\widetilde{\psi})) = \widetilde{\psi}$, and the final equality follows from the definition of λ_ψ. We conclude that any $\psi \in \Omega_{m,w}(\psi^*)$ is feasible. □

4.1.2 BIC - DIC Equivalence

We now derive an equivalence result between symmetric Bayesian IC (BIC) mechanisms and symmetric Dominant Strategy IC (DIC) mechanisms (see Manelli and Vincent (2010) for an analysis of the one-object auction case, and Gerhskov et al (2013) for general social choice problems[23]).

Theorem 5. *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.*

The proof of Theorem 5 (given in the Online Appendix) is similar to the proof of Theorem 4: every extreme point can be implemented by a DIC mechanism, and,

[22] Defined by $\nu(B) = \mu(T^{-1}(B))$ for any Borel subset B.

[23] Both papers also treat the asymmetric case. Manelli and Vincent use the weaker Krein-Milman Theorem and an approximation argument. Gershkov et al. use a result from probability theory about measures with monotonic marginals.

by Choquet's theorem, the interim allocation associated with every BIC mechanism can be represented as a mixture over extreme points; the result follows since DIC incentive compatibility is equivalent to a monotonicity condition that is preserved under averaging[24].

4.1.3 The Revenue Maximizing Ranked-Item Auction

We now turn to revenue maximization. Consider an allocation $\boldsymbol{\alpha}$ that is part of an IC mechanism, i.e., the associated interim allocation $\{\varphi_i\}_{i=1}^N$ are non-decreasing. Assume also that $\boldsymbol{\alpha}$ is individually rational, and that the utility of the lowest type is zero (as required by revenue optimality). By standard methods, it is straightforward to show that expected revenue generated by $\boldsymbol{\alpha}$ is

$$\int_{[0,1]^N} \sum_{i=1}^{N} \left(\left[\theta_i - \frac{1 - F(\theta_i)}{f(\theta_i)} \right] [\boldsymbol{\alpha}_i(\theta_i, \theta_{-i}) \cdot \mathbf{q}] \right) f(\theta_1) ... f(\theta_N) \, \mathrm{d}\theta_1 ... \mathrm{d}\theta_N \, .$$

For a symmetric mechanism, the above expression becomes

$$N \int_0^1 \left[\theta_1 - \frac{1 - F(\theta_1)}{f(\theta_1)} \right] \varphi(\theta_1) \, f(\theta_1) \, \mathrm{d}\theta_1 = N \int_0^1 \left[F^{-1}(s_1) - \frac{1 - s_1}{f(F^{-1}(s_1))} \right] \psi(s_1) \, \mathrm{d}s_1 \, .$$

Thus, by Theorem 4, the revenue maximization problem becomes

$$\max_{\psi \in \Omega_{m,w}(\psi^*)} \int_0^1 \left[F^{-1}(s_1) - \frac{1 - s_1}{f(F^{-1}(s_1))} \right] \psi(s_1) \, \mathrm{d}s_1$$

where ψ^* is the interim quantile function induced by assortative matching. Corollary 2 shows that the optimal solution is an extreme point of $\Omega_m(\psi^* \cdot \mathbf{1}_{[\widehat{s}_1, 1]})$ for some $\widehat{s}_1 \in [0, 1]$. Assuming that the virtual value function $\theta_1 - \frac{1-F(\theta_1)}{F'(\theta_1)}$ is increasing, it is straightforward to see that the type $\widehat{\theta}_1 = F^{-1}(\widehat{s}_1)$ must solve the equation $\theta_1 - \frac{1-F(\theta_1)}{F'(\theta_1)} = 0$. Also, the objective function is then super-modular, and thus the Fan-Lorenz Theorem 3 immediately yields that the optimal allocation $\widehat{\psi}$ satisfies[25]

$$\widehat{\psi}(s_1) = \begin{cases} \psi^*(s_1) & \text{for } s_1 \geq \widehat{s}_1 \\ 0 & \text{otherwise} \end{cases} .$$

[24] An argument similar to the ones used in Theorem 4 and 5 also shows that, for any convex objective function, there exists an optimal mechanism that is non-randomized.

[25] See also Gershkov et al. (2019) who look at a revenue maximization problem with several identical goods where the objective is convex rather than linear. The convexity stems there from investments undertaken prior to the auction.

This allocation can be implemented by a standard matching auction with a reserve price (say pay-your-bid or all-pay) where the highest bidder gets the highest quality, and so on. If the virtual value is not increasing, other extreme points may be optimal, corresponding to the outcome of an "ironing procedure", as described in Proposition 2.

4.1.4 Matching Contests

We now analyze the same basic model as above, but assume that there is a continuum of agents and prizes. Let F denote the distribution of types on the interval $[0, 1]$, and let G denote the distribution of prizes awarded, also on $[0, 1]$. For simplicity, we assume that both F and G are strictly increasing, and we look here at allocation schemes where all prizes are distributed. If an agent with type θ obtains prize q and pays t for it, then her utility is given by $\theta q - t$.[26]

We consider contests where each agent makes an effort (or submits a bid), and where agents are matched to prizes according to their bids. The assortative matching allocation of prizes to agents is given here by $\varphi^*(\theta) = G^{-1}(F(\theta))$, and is strictly increasing. It is implemented by the strictly increasing bidding equilibrium

$$t(\theta) = \theta\varphi^*(\theta) - \int_0^\theta \varphi^*(s)\,\mathrm{d}s$$

The induced interim quantile allocation is given here by

$$\psi^*(s) = \varphi^*(F^{-1}(s)) = G^{-1}(F(F^{-1}(s)) = G^{-1}(s)$$

While matching output is maximized by the assortative scheme,[27] agents waste resources (e.g., signaling costs, payments to a designer) in order to achieve it.

Another feasible scheme is *random* matching where, independently of bids, everyone gets a prize equal to the expected value of the prize distribution μ_G. Output is smaller than that in assortative matching, but random matching can be implemented without any bidding costs. The induced quantile distribution of prizes is given by

$$G_r(x) = \begin{cases} 0 & \text{if } x \leq \mu_G \\ 1 & \text{otherwise} \end{cases}$$

and thus $G_r \succ G \Leftrightarrow G_r^{-1} \prec G^{-1}$.

[26]This standard formulation is easily generalized to other multiplicative, supermodular production functions and also (at least for some questions) to non-linear costs.

[27]This follows from the famous rearrangement inequality of Hardy, Littlewood and Polya (1929).

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 average prize in that quantile. Coarse matching schemes balance output and bidding costs in less extreme ways than random or assortative matching, and have the potential to be superior for some objectives.

The Proposition below generalizes and complements several well-known, existing results in the contest and matching literature (see Hoppe, Moldovanu, Sela [HMS] (2009), Damiano and Li (2007) and Olszewski and Siegel (2018)). These are obtained as immediate consequences of our theoretical insights together with the Fan-Lorenz Theorem[28]:

Proposition 3.

1. *A matching scheme is feasible and incentive compatible if and only if the induced distribution of prizes G_{ic} satisfies $G_{ic}^{-1} \prec G^{-1}$.*
2. *Assume that the distribution of types F is convex. Then each type of the agent prefers random matching to any other scheme.*[29]
3. *Random matching (assortative matching) maximizes the agents' welfare if the distribution of types F has an Increasing (Decreasing) Failure Rate.*[30]
4. *If F has an Increasing Failure Rate, the revenue (i.e., average bid) to a designer is maximized by assortative matching.*[31]

4.2 Optimal Delegation

We now study a model of optimal delegation.[32] The state of the world θ is distributed according to a distribution F with support $[0,1]$ and with density f. Its realization is privately observed by an agent. The action space is the real line.

The agent's Bernoulli utility from a deterministic action a in state θ is given by $U_A(\theta, a) = -(\theta - a)^2$, and the principal's Bernoulli utility is given by $U_P(\theta, a) =$

[28]We provide a proof of Proposition 3 in the Online Appendix.

[29]F being convex on $[0,1]$ implies, in particular, that F first-order stochastically dominates the uniform distribution on this interval. Thus, the present result generalizes the one in HMS (2009), who did not consider intermediate schemes. See also Olszewski and Siegel (2018)) for a derivation that includes coarse matching. If F is concave, there is a uniquely defined interval $[\theta^*, 1]$ such that all types in this interval prefer assortative matching while all types in $[0, \theta^*)$ prefer random matching (see HMS (2009)).

[30]This generalizes one of the main results of HMS (2009) who compared the two extreme cases (random and assortative), but did not consider intermediate schemes. A converse also holds: random matching (assortative matching) minimizes average welfare if the distribution of types F has a Decreasing (Increasing) Failure rate.

[31]See also Damiano and Li (2007).

[32]Variants of this model have been analyzed, for example, by Holmström (1984), Melumad and Shibano (1991), Alonso and Matouschek (2008), and Amador and Bagwell (2013).

$-(\gamma(\theta)-a)^2$, where $\gamma:[0,1]\to\mathbb{R}$ is bounded.[33] We denote by $\Lambda=\sup_{\theta\in[0,1]}|\theta-\gamma(\theta)|$ the maximal disagreement between the agent and the principal. Both agent and principal have expected utility preferences.

A *direct mechanism* $M:[0,1]\to\Delta(\mathbb{R})$ assigns to each report made by the agent a lottery over actions with finite variance. The principal can implement any incentive compatible (IC) direct mechanism by offering a menu of lotteries, out of which the agent chooses a preferred one; conversely, any menu of lotteries induces an IC direct mechanism.[34]

For a direct mechanism M denote by $\mu_M:[0,1]\to\mathbb{R}$ its type-dependent mean action function and by $\sigma^2_M:[0,1]\to\mathbb{R}_+$ its type-dependent variance. Since both the agent's and the principal's indirect utilities can be expressed as a function of μ_M and σ^2_M,

$$U_A(\theta)=-(\theta-\mu_M(\theta))^2-\sigma^2_M(\theta)\,,$$
$$U_P(\theta)=-(\gamma(\theta)-\mu_M(\theta))^2-\sigma^2_M(\theta),$$

we shall identify each mechanism with its induced mean and variance functions $M=(\mu_M,\sigma^2_M)$.

In general, the set of IC mechanisms cannot be satisfactorily characterized by majorization.[35] But, we show below that it is without loss of generality for maximizing the principal's utility to only consider a subset of IC mechanisms that can be characterized in this way.

We call a mechanism *undominated* if there does not exists a mechanism where the set of actions is a singleton, and that yields a higher utility for the principal. In Lemma 4 in the Online Appendix we prove that, in every IC, undominated mechanism, the utility of every type of the agent is bounded from below by $-2Var(\gamma(\theta)+2\Lambda^2)$.

Proposition 4. *Define an interval of actions* $[\underline{a},\overline{a}]$ *by*

$$[\underline{a},\overline{a}]=[-\sqrt{2Var(\gamma(\theta)+2\Lambda^2)},1+\sqrt{2Var(\gamma(\theta)+2\Lambda^2)}]\,.$$

[33]Our approach can easily be extended to more general utility functions. In particular, we obtain analogous results if $U_A(\theta,a)=\theta a+b(a)$ and $U_P(\theta,a)=\gamma(a)a+b(a)$ for a strongly concave function b. Closely related utility functions have been used in the literature, e.g. Amador and Bagwell (2013) and Kolotilin and Zapechelnyuk (2019).

[34]This argument is the familiar *taxation principle*, but note that there are no monetary transfers here.

[35]For example, the mechanism (μ_0,σ_0) that always implements the deterministic action 0 and the one (μ_1,σ_1) that always implements the deterministic action 1 satisfy $\int_0^1\mu_0(\theta)\,\mathrm{d}F(\theta)=\int_0^1 0\,\mathrm{d}F(s)=0\neq 1=\int_0^1 1\,\mathrm{d}F(\theta)=\int_0^1\mu_1(\theta)\,\mathrm{d}F(\theta)$. Thus, μ_0 and μ_1 are *not* comparable to any other function by majorization simultaneously.

A (potentially randomized) undominated mechanism $M = (\mu_M, \sigma^2_M)$ *is incentive compatible if and only if there exists an extension*[36] $(\mu_{\tilde{M}}, \sigma^2_{\tilde{M}})$ *of the functions* μ_M, σ^2_M *to the interval* $[\underline{a}, \overline{a}]$ *such that* $\mu_{\tilde{M}}(\underline{a}) = \underline{a}$, $\mu_{\tilde{M}}(\overline{a}) = \overline{a}$, $\sigma^2_{\tilde{M}}(\underline{a}) = \sigma^2_{\tilde{M}}(\overline{a}) = 0$, *and such that:*

1. $\mu_{\tilde{M}} \in \Phi_m(a^*)$ *where* $a^* : [\underline{a}, \overline{a}] \to [\underline{a}, \overline{a}]$ *is the identity, and*
2. $\sigma^2_{\tilde{M}}(\theta) = -(\mu_{\tilde{M}}(\theta) - \theta)^2 - 2\int_{\underline{a}}^{\theta} (\mu_{\tilde{M}}(s) - s)\,\mathrm{d}s$ *for all* $\theta \in [\underline{a}, \overline{a}]$.

Proof: Necessity. Let $M = (\mu_M, \sigma^2_M)$ be an undominated IC mechanism. Define a new mechanism on the extended type space $[\underline{a}, \overline{a}]$ by the menu that consists of all options $(\mu_M(\theta), \sigma^2_M(\theta))_{\theta \in [0,1]}$ available in the original mechanism M and, in addition, the two deterministic actions $\underline{a}, \overline{a}$. Any such menu induces an IC direct mechanism $\tilde{M} = (\mu_{\tilde{M}}, \sigma^2_{\tilde{M}})$ that assigns to every agent in the extended type space $[\underline{a}, \overline{a}]$ his most preferred option.

By Lemma 4 in the Appendix, the agent's utility in M is bounded from below by $-2Var(\gamma(\theta)) - 2\Lambda^2$. This implies that any original type θ prefers the allocation assigned to her in M to the deterministic actions $\underline{a}$ and $\overline{a}$, and thus that $\mu_{\tilde{M}}(\theta) = \mu_M(\theta)$ for any $\theta \in [0, 1]$. Clearly, it is also optimal for an agent of type $\underline{a}$ $(\overline{a})$ to pick the deterministic action $\underline{a}$ $(\overline{a})$ in $\tilde{M}$, and hence, $\mu_{\tilde{M}}(\underline{a}) = \underline{a}$, $\mu_{\tilde{M}}(\overline{a}) = \overline{a}$ and $\tilde{\sigma}_M(\underline{a}) = \tilde{\sigma}_M(\overline{a}) = 0$. As a consequence, an agent with hypothetical type $\underline{a}$ $(\overline{a})$ obtains utility 0 in $\tilde{M}$.

Since type $\underline{a}$ obtains utility 0, it follows from the envelope theorem and from the super-modularity of the agent's utility in (θ, μ) that the mechanism $\tilde{M} = (\mu_{\tilde{M}}, \sigma^2_{\tilde{M}})$ is IC if and only if $\mu_{\tilde{M}}$ is non-decreasing and satisfies the envelope condition for all $\theta \in [\underline{a}, \overline{a}]$:

$$- (\theta - \mu_{\tilde{M}}(\theta))^2 - \sigma^2_{\tilde{M}}(\theta) = 2\int_{\underline{a}}^{\theta} [\mu_{\tilde{M}}(s) - s]\,\mathrm{d}s \tag{10}$$

Since

$$-(\theta - \mu_{\tilde{M}}(\theta))^2 - \sigma^2_{\tilde{M}}(\theta) \leq 0,$$

the envelope condition (10) implies that

$$\int_{\underline{a}}^{\theta} \mu_{\tilde{M}}(s)\,\mathrm{d}s \leq \int_{\underline{a}}^{\theta} a^*(s)\,\mathrm{d}s,$$

where $a^*(s) = s$. Since $\mu_{\tilde{M}}(\overline{a}) = \overline{a}$ and $\sigma^2_{\tilde{M}}(\overline{a}) = 0$, we obtain by (10) that

$$\int_{\underline{a}}^{\overline{a}} [\mu_{\tilde{M}}(s) - a^*(s)]\,\mathrm{d}s = 0.$$

[36] A function $\tilde{g} : [\underline{a}, \overline{a}] \to \mathbb{R}$ is an extension of a function $g : [0, 1] \to \mathbb{R}$ to the interval $[\underline{a}, \overline{a}]$ if $\tilde{g}(\theta) = g(\theta)$ for all $\theta \in [0, 1]$.

We conclude that $\mu_{\tilde{M}} \in \Phi_m(a^*)$. Thus, $(\mu_{\tilde{M}}, \sigma_{\tilde{M}})$ is an extension of (μ_M, σ_M^2) to $[\underline{a}, \overline{a}]$ with the desired properties.

Sufficiency. Conversely, suppose that $(\mu_{\tilde{M}}, \sigma_{\tilde{M}}^2)$ are such that $\mu_{\tilde{M}} \in \Phi_m(a^*)$ and such that $\sigma_{\tilde{M}}^2$ satisfies the condition of the Proposition. Then, we can define a stochastic mechanism $M = (\mu_M, \sigma_M^2)$ by the restriction of $(\mu_{\tilde{M}}, \sigma_{\tilde{M}}^2)$ to the set of types $[0, 1]$. This mechanism is well-defined since its variance is non-negative:

$$\begin{aligned}\sigma_M^2(\theta) &= -(\mu_{\tilde{M}}(\theta) - \theta)^2 - 2\int_{\underline{a}}^{\theta} (\mu_{\tilde{M}}(s) - s)\,\mathrm{d}s \\ &= -2\int_{\theta}^{\tilde{\mu}_M(\theta)} (\mu_{\tilde{M}}(\theta) - s)\mathrm{d}s - 2\int_{\underline{a}}^{\theta} (\mu_{\tilde{M}}(s) - s)\,\mathrm{d}s \\ &\geq -2\int_{\underline{a}}^{\tilde{\mu}_M(\theta)} (\mu_{\tilde{M}}(s) - s)\,\mathrm{d}s \geq 0,\end{aligned}$$

where the first inequality follows since $\mu_{\tilde{M}}$ is non-decreasing, and the second follows since $\mu_{\tilde{M}} \succ a^*$. Since $\mu_{\tilde{M}}$ is non-decreasing and $(\mu_{\tilde{M}}, \sigma_{\tilde{M}}^2)$ satisfies the envelope condition by assumption, it follows that the mechanism M is IC. □

Kovac and Mylovanov (2009) characterized IC mechanisms by: 1) monotonicity of the mean action function; 2) the envelope condition determining the variance functions, and 3) a non-negativity constraint on the variance. This imposes a joint constraint on the mean action function and the variance of the lowest type. In contrast, our condition $\mu_{\tilde{M}} \in \Phi_m(a^*)$ encompasses the monotonicity constraint on the mean action function, and ensures that the variance derived by the envelope conditions is non-negative for all types if $\sigma_{\tilde{M}}^2(\underline{a}) = 0$. This new formulation allows us to reduce the problem to a linear maximization problem where we optimize only over mean action functions subject to the majorization constraint.

Similarly to the revenue equivalence result in auction theory (see also Section 4.1.4), we can now use Proposition 4 to show that the value of the principal in different, undominated, IC delegation mechanisms depends only on the implemented mean action function[37]:

Proposition 5 (Value Equivalence)**.** *Fix an arbitrary undominated, IC delegation mechanism $M = (\mu_M, \sigma_M^2)$ and let $\mu_{\tilde{M}}, \sigma_{\tilde{M}}^2$ be an extension satisfying the conditions of Proposition 4. The principal's expected utility in M is only a function of $\mu_{\tilde{M}}$ and is*

[37] The characterization of IC delegation mechanisms formally resembles that of IC allocations via reduced form auctions But, the majorization constraint in the delegation application is the opposite of that the auction context, and the envelope condition characterizing the variance is non-linear due to the agent's quadratic utility.

given by

$$V_P(\mu_{\tilde{M}}) = 2\int_{\underline{a}}^{\overline{a}} J(\theta)\,\mu_{\tilde{M}}(\theta)\,\mathrm{d}\theta + C\,, \tag{11}$$

where the "virtual value" $J : [\underline{a}, \overline{a}] \to \mathbb{R}$ *is defined as*

$$J(\theta) = \begin{cases} 1 & \textit{for } \theta \in [\underline{a}, 0) \\ 1 - F(\theta) + (\gamma(\theta) - \theta)f(\theta) & \textit{for } \theta \in [0, 1] \\ 0 & \textit{for } \theta \in (1, \overline{a}] \end{cases}$$

and where

$$C = \int_0^1 (\theta^2 - \gamma(\theta)^2)f(\theta) - 2\theta(1 - F(\theta))\,\mathrm{d}\theta + \underline{a}^2\,.$$

Proposition 5 follows from Proposition 4 by applying a partial integration argument to the extension $\mu_{\tilde{M}}$ (see the Online Appendix). What is remarkable about the above "virtual value" characterization is that the objective of the principal (i) does not depend on the choice of the extension $\mu_{\tilde{M}}$ (as long as it satisfies the conditions in Proposition 4) and (ii) becomes linear in the extension of the mean allocation rule $\mu_{\tilde{M}}$ despite the fact that the original objective of the principal was strictly concave in μ_M.

Corollary 3. *The principal's problem is given by*

$$\max_{\mu_{\tilde{M}} \in \Phi_m(a^*)} V_P(\mu_{\tilde{M}})$$

and therefore an extreme point of $\Phi_m(a^*)$ *must be optimal.*

We can thus apply our earlier results to this particular problem, starting with some insights into the nature of optimal delegation mechanisms:

1) Recall that an extreme point $\mu_{\tilde{M}}$ of $\Phi_m(a^*)$ is characterized by a collection of intervals $[\underline{\theta}_i, \overline{\theta}_i)$ with sub-intervals $[\underline{y}_i, \overline{y}_i)$ indexed by $i \in I$ such that:

1. If, for some $i \in I$, $\theta \in [\underline{\theta}_i, \overline{\theta}_i)$ and $\underline{y}_i = \overline{y}_i$ then

$$\mu_{\tilde{M}}(\theta) = \begin{cases} \underline{\theta}_i & \text{for } \theta < \frac{\overline{\theta}_i + \underline{\theta}_i}{2} \\ \overline{\theta}_i & \text{for } \theta > \frac{\overline{\theta}_i + \underline{\theta}_i}{2} \end{cases}.$$

2. If, for some $i \in I$, $\theta \in [\underline{\theta}_i, \overline{\theta}_i)$ and $\underline{y}_i < \overline{y}_i$ then

$$\mu_{\tilde{M}}(\theta) = \begin{cases} \underline{\theta}_i & \text{for } \theta < \underline{y}_i \\ v_i & \text{for } \theta \in [\underline{y}_i, \overline{y}_i) \\ \overline{\theta}_i & \text{for } \theta > \overline{y}_i \end{cases},$$

where v_i is defined in equation (3).

3. If $\theta \notin \bigcup_{i \in I} [\underline{\theta}_i, \overline{\theta}_i)$ then $\mu_{\tilde{M}}(\theta) = \theta$.

Such a mechanism is implemented by letting the agent choose any action $a \in [\underline{a}, \overline{a}] \setminus \bigcup_{i \in I} (\underline{\theta}_i, \overline{\theta}_i)$ and, for each $i \in I$ such that $\underline{y}_i < \overline{y}_i$, adding to the agent's choice set an additional option with mean v_i and variance $(\underline{\theta}_i - \underline{y}_i)^2 - (\underline{y}_i - v_i)^2$. In particular, note that a delegation mechanism corresponding to an extreme point is deterministic if $\underline{y}_i = \overline{y}_i$ for each $i \in I$.

Optimal delegation mechanisms sometimes involve deliberate randomization by the principal (see Kovac and Mylovanov (2009) and Alonso and Matouschek (2008) for examples). But, our result above significantly reduce the class of uniquely optimal stochastic mechanisms: any extreme (and thus exposed) point will use at most one non-degenerate lottery on each of the intervals $(\underline{\theta}_i, \overline{\theta}_i)$, and any stochastic extreme point will have a discontinuous mean-action function.

2) Certain Bayesian persuasion problems give rise to the same class of optimization problems (see Section 4.3 below), and this allows us to extend the equivalence observed in Kolotilin and Zapechelnyuk (2019) to stochastic delegation and to general persuasion mechanisms. As an illustration, we now provide a sufficient condition for a deterministic delegation mechanism to be optimal by applying a result in Dworczak and Martini (2019) about the optimality of *monotone partitional signals* in Bayesian persuasion[38]

Corollary 4. *Suppose that there are $a_1, a_2 \in [\underline{a}, \overline{a}]$ such that J is non-increasing on the intervals $[\underline{a}, a_1]$ and $[a_2, \overline{a}]$, and non-decreasing on the interval $[a_1, a_2]$. Then a deterministic mechanism is optimal.*

Proof: Using integration by parts for the Riemann-Stieltjes integral,[39] the principal's objective becomes

$$\max_{\mu_{\tilde{M}} \in \Phi_m(a^*)} \int_{\underline{a}}^{\overline{a}} \left(- \int_{\underline{a}}^{\theta} J(s)\, \mathrm{d}s \right) \mathrm{d}\mu_{\tilde{M}}(\theta).$$

[38]Our result also extends a result by Kovac and Mylovanov (2009). Recently, Kartik et al. (2020) provided sufficient conditions for the optimality of deterministic mechanisms in a related veto bargaining model.

[39]Note that $\mu_{\tilde{M}}(\theta)$ is non-decreasing and hence has bounded variation.

The assumption implies that the integrand, as a function of θ, is convex on $[\underline{a}, a_1]$ and on $[a_2, \overline{a}]$, and concave on $[a_1, a_2]$. It is therefore an *affine-closed function* (see Definition 2 in Dworczak and Martini (2019)). Their Theorem 3 implies then that the principal's problem is solved by an extreme point such that, in the notation of our Theorem 2, $\underline{y}_i = \overline{y}_i$ for all $i \in I$ (see also Section 4.3). As explained above, any such mechanism corresponds to a deterministic delegation mechanism. □

3) Finally, one can use Lemma 1 to analyze under what conditions a particular extreme point is optimal. In particular, the Fan-Lorentz Theorem (Theorem 3) immediately yields a result obtained by Kovac and Mylovanov (2009) who used a rather different approach:

Corollary 5. *Full delegation, i.e., allowing the agent to chose any action in* $[0, 1]$ *is optimal if* $J(\theta) = 1 - F(\theta) + (\gamma(\theta) - \theta)f(\theta)$ *is non-increasing on* $[0, 1]$, *and if* $\gamma(0) \leq 0$ *and* $\gamma(1) \geq 1$.

Proof: The assumptions imply that J is non-increasing on $[\underline{a}, \overline{a}]$, and thus the objective is linear, sub-modular and thus Schur-convex. The function a^* itself is then a maximizer over $\Phi_m(a^*)$ for any such functional. As a consequence, each type gets a mean allocation equal to his type $\mu_M(\theta) = \theta$. In turn, Proposition 4 implies that the variance for each type, $\sigma_M(\theta)$, equals zero. □

4.3 Persuasion with Preferences over the Posterior Mean

We consider here the persuasion problem studied by Kolotilin (2018) and Dworczak and Martini (2019). The state of the world ω is distributed according to a continuous distribution F on the interval $[0, 1]$, 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 the 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.[40]

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. $G_\pi \succ F$. Conversely, it is well known that, for any G such that $G \succ F$, there exists a

[40]Note that this allows for the senders payoff to depend on the action taken by the receiver.

signal π such $G_\pi = G$. Hence, formally, the sender's problem is to choose a distribution over posterior mean beliefs of the receiver G that solves:

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

As the above objective is linear, the 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 two different (potentially random) signals are sent for states in that interval, inducing two possible posterior means on this interval.

A signal structure is called *monotone partitional* if it partitions the state space into intervals such that each interval is either of type 1 or type 2; such an information structure either reveals the state perfectly, or sends the same signal for all states in an interval. While other information structures may be optimal, our result implies that the optimal signal structure can still be implemented in a simply way by sending at most two signals on each interval. Arieli et al. (2020) independently obtained the same result - they call signal structures of type 3 *bi-pooled*.[41]

Equivalence to Optimal Delegation

Our majorization/extreme points approach highlights the close connections between the above described problem of Bayesian persuasion and the delegation problem studied in Subsection 4.2. Although the delegation problem is a-priori non-linear, we have shown that both exercises can be reduced to a maximization of a linear functional over a set Φ_m of majorizing functions. Hence, the basic structure of their respective optimal mechanisms is identical.

Kolotilin and Zapechelnyuk (2019) have recently established a formal equivalence between optimal delegation and Bayesian persuasion for the case where the set of policies for the principal was exogenously restricted to deterministic delegation mechanisms and to monotone partitional signals, respectively. Our majorization characterization

[41] The optimality of such a structure in a particular example has already been established Gentzkow and Kamenica (2016).

immediately implies that this equivalence holds without **any** restrictions on the policy space: optimal signal structures for Bayesian persuasion that are not monotone partitional correspond to randomized optimal delegation mechanisms.

4.4 Persuasion with a Publicly Informed Receiver

Consider now a different Bayesian persuasion problem where there are two states $\theta \in \{0, 1\}$. Denote by x the posterior probability assigned to the high state after observing a signal that is equal to the posterior expectation. The sender's payoff v is state independent and only depends on x.

There is a population of receivers, each starting out with a prior $X_0 \sim H$. Assume that the prior of a given receiver is observable to the sender. For each prior, the sender picks a distribution $G(\cdot|x_0) : [0, 1] \to \mathbb{R}_+$ subject to the constraint that the posterior expectation must be consistent with the prior

$$\int_0^1 s \, \mathrm{d}G(s|x_0) = x_0 \, .$$

Denote by G the average distribution over posteriors induced for the receiver

$$G(s) = \int G(s|x_0) \, \mathrm{d}H(x_0) \, .$$

Denote by Z a random variable with conditional distribution $G(\cdot|x_0)$. Letting $X = X_0 + Z$, we note that $X \sim G$ and that X is a mean preserving spread of X_0. This is equivalent to $G \prec H$. Hence the sender's problem becomes

$$\max_{G \in \Omega_m(H)} \int_0^1 v(x) \, \mathrm{d}G(x) \, .$$

In contrast to the previous Bayesian persuasion problem, here agents are ex-ante informed and the sender can only generate more information. As the above objective is linear, the maximum is attained at an extreme point as characterized in Theorem 1. This immediately implies that an optimal signal structure partitions the prior beliefs of the receiver into intervals such that, in each interval, either:

1. No information is provided to the receiver, or
2. Two different signals are sent on an interval, that move the posterior belief of each receiver to the boundary points of the interval.

4.5 Decision-Making Under Uncertainty

We now briefly illustrate how our insights can be applied in order to understand how agents with non-expected utility preferences choose among risky prospects.

4.5.1 Rank-Dependent Utility and Choquet Capacities

Quiggin (1982) and Yaari (1987) axiomatically derived utility functionals with rank-dependent assessments of probabilities of the form[42]

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

where F is the distribution of a random variable on the interval $[0,1]$, $v : [0,1] \to R$ is continuous, 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[43].

The case $g(x) = x$ yields the classical von-Neumann and Morgenstern expected utility model where risk-aversion is equivalent to v being concave. The case $v(x) = x$ 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 clear what properties yield risk aversion in the general rank-dependent model. Using integration by parts, we can also write:

$$\begin{aligned} U(F) &= \int_0^1 v(t)\,\mathrm{d}(g \circ F)(t) = v(1) - \int_0^1 v'(t)(g \circ F)(t)\,\mathrm{d}t \\ &= v(1) + \int_0^1 K(F(t), t)\,\mathrm{d}t \end{aligned}$$

where

$$K(F, t) = -v'(t)(g \circ F)$$

and where we used $g(0) = 0$ and $g(1) = 1$. Then

$$\frac{\partial^2 K(F,t)}{\partial F \partial t} = -g'(F(t))v''(t) \geq 0$$

[42] Their theory is a bit more general (for example it allows a more general domain for the functions v and F) . For the sake of consistency, we keep here a framework that is compatible with the rest of the paper.

[43] For the sake of brevity we assume below that both g an v are twice differentiable. Since the Fan-Lorentz result does not require differentiability, the observations below generalize.

for all t if and only if v is concave. Similarly

$$\frac{\partial^2 K(F,t)}{\partial^2 F} = -g^{''}(F(t))v'(t) \geq 0$$

for all t if and only if g is concave.

Hence, the Fan-Lorentz conditions are satisfied if and only if $v'' \leq 0$ and $g" \leq 0$. As a consequence, the utility functional $U = \int_0^1 v(t)\, \mathrm{d}(g \circ F)(t)$ is Schur-concave, and the agent whose preferences are represented by U is *risk averse* , exactly as under standard expected utility[44].

Another important strand of the literature on non-expected utility considers ambiguity aversion.[45] The main tool is the *Choquet integral* with respect to a (convex) *capacity* - note that this is unrelated to the Choquet representation used above! Analogously to the derivations above, it can be shown that the Choquet integral yields a Schur-concave functional if and only if it is computed with respect to a convex capacity.

4.5.2 A Portfolio Choice Problem

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

$$X \geq_{cv} Z \Leftrightarrow F_X \succ F_Z \Leftrightarrow F_X^{-1} \prec F_Z^{-1}.$$

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 (1929) (the anti-assortative part!), and where the second inequality follows

[44]The equivalence between the concavity of the functions v and g, and risk-aversion has been pointed out by Hong et al (1987), who build on Machina (1982), and Yaari (1987).

[45]See Schmeidler (1989) and Wakker (1990) for the relations between rank-dependent utility and Choquet integrals.

by the Fan-Lorentz Theorem.

By choosing a random variable X that has the same distribution as Z and that is anti-comonotonic with Y,[46] the lower bound $\int_0^1 F_Y^{-1}(1-t)F_Z^{-1}(t)\,\mathrm{d}t$ is attained, and hence such a choice solves the portfolio choice problem.[47]

If $Y' \leq_{cv} Y$, we obtain by the Fan-Lorentz inequality (now applied to the functional with argument F_Y^{-1}) that

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

In other words, a decision maker that becomes more informed (in the Blackwell sense) will bear a lower cost.

5 Conclusion

We provided characterizations of the extreme points of the sets of all monotonic functions that are either majorized by, or themselves majorize a given function. We have also shown that many well-known optimization exercises in Economics can be rephrased as maximizing a convex functional over such sets. Hence, a maximum must be attained at one of the extreme points.

Together with an integral representation result due to Choquet, the characterization of extreme points directly imply many results, both novel and well-known. For example, in the context of auctions it implies both, a new generalization of Border's Theorem and the known equivalence between Bayes and dominant strategy incentive compatible mechanisms. For optimal delegation and Bayesian persuasion, our results imply that it is without loss of generality to restrict attention to a small class of mechanisms, and reveal a novel, general equivalence result between these two problems and their (possibly randomized) solutions.

An interesting question for future research is if an analogous extreme point characterization could be obtained for notions of multivariate majorization. Such a result would be potentially useful in various other applications, e.g., information revelation in auctions where the state is naturally multi-dimensional.

[46] This can always be done if the underlying probability space is non-atomic.

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

A Appendix

Proof of Proposition 1: We first establish that $\Omega_m(f)$ is a compact subset of L^1 in the norm topology. Since f is non-decreasing, it has a non-decreasing and right-continuous representative that we also denote by f.[48] For any element of $\Omega_m(f)$, we use its non-decreasing and right-continuous representative that is left-continuous at 1. Then, for any $g \in \Omega_m(f)$, $f(0) \leq g(x) \leq f(1)$, and the total variation of g is uniformly bounded by $f(1) - f(0)$. *Helly's Selection Theorem* therefore implies that any sequence $\{g_n\}$ in $\Omega_m(f)$ has a subsequence that converges pointwise, and in L^1, to some function g with bounded variation. Since $\int_x^1 g_n(s)\mathrm{d}s \leq \int_x^1 f(s)\mathrm{d}s$, we obtain that $\int_x^1 g(s)\mathrm{d}s \leq \int_x^1 f(s)\mathrm{d}s$, with equality for $x = 0$. Also, since each g_n is non-decreasing, g is non-decreasing and we conclude that $\Omega_m(f)$ is compact in the topology induced by the L^1-norm. Analogous arguments establish compactness of $\Omega_{m,w}(f)$ and $\Phi_m(f)$.

It is clear from the definitions that the sets $\Omega_m(f)$, $\Omega_{m,w}(f)$ and $\Phi_m(f)$ are convex. It then follows from Choquet's theorem that, for any $g \in \Omega_m(f)$, there is a probability measure λ_g that puts measure 1 on the extreme points of $\Omega_m(f)$ such that $g = \int h \mathrm{d}\lambda_g(h)$. The same argument applies to $\Omega_{m,w}(f)$ and $\Phi_m(f)$. □

Preparations for the Proof of Theorem 1.

Fix $g \in \Omega_m(f)$. Since $f, g \in L^1$ are non-decreasing, they contain non-decreasing and right-continuous representatives, which we also denote by f and g, respectively. Let $f(x_-) = \lim_{x' \uparrow x} f(x')$ and $f(x_+) = \lim_{x' \downarrow x} f(x')$. Given $s_1, s_2 \in [0,1]$ such that $s_1 < s_2$ and given $y \in [g(s_1), g(s_2)]$, define

$$u(s) := \mathit{median}\{g(s) - g(s_1),\ g(s) - g(s_2),\ y - g(s)\} \text{ for } s \in [s_1, s_2] \text{ and } u(s) = 0 \text{ else.} \tag{12}$$

Lemma 2.

1. $g \pm u$ is non-decreasing, and $g(s_1) \leq (g \pm u)(s) \leq g(s_2)$ for all $s \in [s_1, s_2]$.
2. If $g(s_1) < g(s)$ for all $s > s_1$, then $u \not\equiv 0$.
3. If $g(s) < g(s_2)$ for all $s < s_2$ and if g is continuous at s_2, then $u \not\equiv 0$.
4. There exists $y \in [g(s_1), g(s_2)]$ such that $\int_{s_1}^{s_2} u(s)\mathrm{d}s = 0$.

Proof of Lemma 2:

[48] A nondecreasing function $f : [0,1] \to \mathbb{R}$ has at most countably many discontinuities, and limits from the right are defined for each $x \in [0,1)$.

(1) Let

$$s_a := \inf\left\{x \;\mid\; g(x) \geq \frac{g(s_1)+y}{2}\right\} = \inf\{\, x \;\mid\; g(x)-g(s_1) \geq y - g(x)\}$$

and

$$s_b := \inf\left\{x \;\mid\; g(x) \geq \frac{g(s_2)+y}{2}\right\} = \inf\{\, x \;\mid\; g(x)-g(s_2) \geq y - g(x)\}$$

It follows that

$$u(s) = \begin{cases} g(s)-g(s_1) & \text{for } s \in (s_1, s_a) \\ y - g(s) & \text{for } s \in (s_a, s_b) \\ g(s) - g(s_2) & \text{for } s \in (s_b, s_2). \end{cases}$$

and hence that

$$(g+u)(s) = \begin{cases} 2g(s)-g(s_1) & \text{for } s \in [s_1, s_a) \\ y & \text{for } s \in [s_a, s_b) \\ 2g(s) - g(s_2) & \text{for } s \in [s_b, s_2). \end{cases}$$

By the definition of s_a, and because $g+u$ is right-continuous, we obtain

$$(g+u)(s_a^-) = 2g(s_a^-) - g(s_1) \leq y = (g+u)(s_a)$$

Similarly,

$$(g+u)(s_b^-) = y \leq 2g(s_b^+) = (g+u)(s_b)$$

by definition of s_b. Since, in addition, $u(s_1) = u(s_2) = 0$ we conclude that $g+u$ is non-decreasing. Similar arguments show that $g-u$ is non-decreasing as well. Since $u(s) = 0$ for $s \notin (s_1, s_2)$ the inequalities follow.

(2) Note that the first argument of the median function in (12) is strictly positive for $s > s_1$ since, by assumption, $g(s_1) < g(s)$ for all $s > s_1$.

If $y = g(s_1)$ then the third argument in the definition of u is strictly negative for $s > s_1$, and the second argument is also strictly negative for a sufficiently small interval $s \in (s_1, s_1 + \delta)$. Hence, $u \neq 0$ on a set of positive measure and therefore $u \not\equiv 0$.

If $y > g(s_1)$ then the right-continuity of g implies that there exists $\delta > 0$ such that the third argument is strictly positive on $[s_1, s_1 + \delta]$; similarly, there exists $\delta' > 0$ such that the second term is strictly negative on $[s_1, s_1 + \delta']$. Hence, $u \neq 0$ on a set of positive measure and therefore $u \not\equiv 0$.

(3) If $y = g(s_2)$ then the third argument in the definition of u is strictly positive for

$s < s_2$ since $g(s) < g(s_2)$ for all $s < s_2$; if $y < g(s_2)$ then continuity of g at s_2 implies that there is $\delta > 0$ such that the third argument is strictly positive on $[s_2 - \delta, s_2]$; the second argument is strictly negative for $s < s_2$; and continuity of g at s_2 implies that there is $\delta' > 0$ such that the first argument is strictly positive on $[s_2 - \delta', s_2]$. Hence, $u \neq 0$ on a set with positive measure and therefore $u \not\equiv 0$.

(4) In order to emphasize the fact that the definition of u in (12) depends on the parameter y we write $u(s, y)$ in this part. Note that, for all s, the function $u(s, y)$ is continuous in y, and that, for all $y \in [g(s_1), g(s_2)]$, $u(\cdot, y)$ is integrable in s and uniformly bounded. Hence, $\int_0^1 u(s, y)\mathrm{d}s$ is continuous in y. If $y = g(s_1)$ then $u(s, y) \leq 0$ for all s; if $y = g(s_2)$ then $u(s, y) \geq 0$ for all s. The intermediate value theorem implies therefore that there exists $y \in [g(s_1), g(s_2)]$ such that $\int_0^1 u(s, y)\mathrm{d}s = 0$. □

Proof of Theorem 1: "⇒": Suppose that g is an extreme point, and denote its non-decreasing and right-continuous representative also by g. The proof proceeds in two steps: Step 1 shows that, if g is non-constant in an interval around x, then $f(x) = g(x)$. Step 2 argues that, if g constant on an interval around x, then it has the same average as f on this interval.

Step 1: Fix an arbitrary $s_1 \in [0, 1)$ and suppose that $g(s_1) < g(s)$ for all $s > s_1$. Since g is right-continuous, if $g(s_1) < f(s_1)$, then there exists $s_2 > s_1$ such that $g(s_2) < f(s_1)$. Define u according to (12) such that $\int_{s_1}^{s_2} u(s)\mathrm{d}s = 0$. Then $(g \pm u)(s) < f(s)$ holds on $[s_1, s_2]$ as

$$g(s) \pm u(s) \leq g(s_2) < f(s_1) \leq f(s).$$

Also, $\int_{s_2}^1 f(s) - g(s)\mathrm{d}s \geq 0$ holds since $f \succ g$. This implies that $\int_x^1 f(s) - (g \pm u)(s)ds \geq 0$ for all x, and hence that $g \pm u \in \Omega_m(f)$. Lemma 1 (ii) implies then that $u \not\equiv 0$, contradicting the assumption that g is an extreme point of $\Omega_m(f)$.

Similarly, if $g(s_1) > f(s_1)$ then there exists $s_2 > s_1$ such that $f(s_2) < g(s_1)$. Define u according to (12) such that $\int_{s_1}^{s_2} u(s)\mathrm{d}s = 0$. Then $(g \pm u)(s) > f(s)$ holds on $[s_1, s_2]$. Since

$$\int_{s_1}^1 (f(s) - g(s))\mathrm{d}s = \int_{s_1}^1 [f(s) - (g \pm u)(s)]\mathrm{d}s \geq 0$$

we conclude that $\int_x^1 [f(s) - (g \pm u)(s)]\mathrm{d}s \geq 0$ for all x. Hence, $g \pm u \in \Omega_m(f)$. Lemma 2 (ii) implies that $u \not\equiv 0$, contradicting the assumption that g is an extreme point of $\Omega_m(f)$. We conclude that, if for an arbitrary $x \in [0, 1)$ the inequality $g(x) < g(s)$ holds for all $s > x$, then $g(x) = f(x)$.

Step 2: It follows from Step 1 that, for any $x \in [0, 1)$ such that $f(x) \neq g(x)$, there exists an interval containing x where g is constant. Hence, there exists a countable collection of non-degenerate intervals $\{[\underline{x}_i, \overline{x}_i) | i \in \mathcal{I}\}$ such that, for each i, $g(s) = g(\underline{x}_i)$ for $s \in [\underline{x}_i, \overline{x}_i)$, $g(s) < g(\underline{x}_i)$ for $s < \underline{x}_i$, $g(s) > g(\underline{x}_i)$ for $s > \overline{x}_i$, and $f(x) = g(x)$ for $x \neq 1$ with $x \notin \bigcup_i [\underline{x}_i, \overline{x}_i)$.

Suppose now that $\int_{\underline{x}_i}^{\overline{x}_i} (f(s) - g(s)) \mathrm{d}s < 0$ for some $i \in \mathcal{I}$. This implies that $\int_{\overline{x}_i}^{1} (f(s) - g(s)) \mathrm{d}s > 0$ and, since g is constant on $[\underline{x}_i, \overline{x}_i)$, that $f(\underline{x}_i) < g(\underline{x}_i)$. If $g(\underline{x}_i^-) = g(\underline{x}_i)$ we can choose $s_2 = \underline{x}_i$ and $s_1 < s_2$ large enough such that u defined according to (12) satisfies $g \pm u \in \Omega_m(f)$ and $u \not\equiv 0$, contradicting that g is an extreme point. Hence, $g(\underline{x}_i^-) < g(\underline{x}_i)$. Also, if $g(s) > g(\overline{x}_i)$ for all $s > \overline{x}_i$ we can choose $s_1 = \overline{x}_i$ and $s_2 > s_1$ small enough such that u defined according to (12) satisfies $g \pm u \in \Omega_m(f)$ and $u \not\equiv 0$, contradicting that $\mathbf{g}$ is an extreme point. Hence, g is constant to the right of $\overline{x}_i$. Let $b = \sup\{\ x \ \mid \ g(x) = g(\overline{x}_i)\}$. There are two cases to consider:

Case 1: $\int_b^1 (f(s) - g(s)) \mathrm{d}s > 0$. If g is continuous at b, then we can choose $s_1 = b$ and $s_2 > s_1$ small enough such that u satisfies $g \pm u \in \Omega_m(f)$ and $u \not\equiv 0$. Hence, $g(b^-) < g(b)$. We can therefore choose $\varepsilon > 0$ and $\delta > 0$ such that

$$g \pm (\varepsilon \mathbf{1}_{[\underline{x}_i, \overline{x}_i)} - \delta \mathbf{1}_{[\overline{x}_i, b)}) \in \Omega_m(f),$$

contradicting the fact that g is an extreme point.

Case 2: $\int_b^1 (f(s) - g(s)) \mathrm{d}s = 0$. Since, by assumption, $\int_{\underline{x}_i}^{\overline{x}_i} (f(s) - g(s)) \mathrm{d}s < 0$ and $\int_{\underline{x}_i}^{1} (f(s) - g(s)) \mathrm{d}s \geq 0$ are true, we obtain $\int_{\overline{x}_i}^{1} (f(s) - g(s)) \mathrm{d}s > 0$. This implies that $\int_{\overline{x}_i}^{b} (f(s) - g(s)) \mathrm{d}s > 0$, and hence that $g(b^-) < f(b)$. Since $\int_b^1 (f(s) - g(s)) \mathrm{d}s = 0$, $f(b) > g(b)$ would imply $\int_{b+\varepsilon}^{1} (f(s) - g(s)) \mathrm{d}s < 0$ for $\varepsilon > 0$ small enough, which contradicts $f \succ g$. Therefore, $g(b^-) < f(b) \leq g(b)$. We can therefore choose $\varepsilon > 0$ and $\delta > 0$ such that

$$g \pm (\varepsilon \mathbf{1}_{[\underline{x}_i, \overline{x}_i)} - \delta \mathbf{1}_{[\overline{x}_i, b)}) \in \Omega_m(f)$$

contradicting the fact that g is an extreme point.

We can conclude that $\int_{\underline{x}_i}^{\overline{x}_i} (f(s) - g(s)) \mathrm{d}s \geq 0$ for all $i \in \mathcal{I}$. Since $\int_0^1 (f(s) - g(s)) \mathrm{d}s = 0$ and $f(s) = g(s)$ for $s \notin \bigcup_i [\underline{x}_i, \overline{x}_i)$, we obtain $\int_{\underline{x}_i}^{\overline{x}_i} (f(s) - g(s)) \mathrm{d}s = 0$ for all $i \in \mathcal{I}$.

"$\Leftarrow$": Suppose that g has the form in the statement, and suppose that there exists $u \in L^1$ such that $g \pm u \in \Omega_m(f)$. Let $\{[\underline{x}_i, \overline{x}_i) \mid i \in \mathcal{I}\}$ be a countable collection of intervals such that $g$is constant on each of the intervals, and such that $f(x) = g(x)$ for $x \notin \bigcup_{i \in \mathcal{I}} [\underline{x}_i, \overline{x}_i)$. Since $g \pm \mathbf{u}$ is nondecreasing, for every $i \in \mathcal{I}$there is a constant function that equals $u$for almost every $x \in [\underline{x}_i, \overline{x}_i)$. Also, the properties of $g$imply that $\int_x^1 (f(s) - g(s)) \mathrm{d}s = 0$for $x \notin \bigcup_i (\underline{x}_i, \overline{x}_i)$, and therefore $\int_x^1 u(s) \mathrm{d}s = 0$must hold for $x \notin \bigcup_i (\underline{x}_i, \overline{x}_i)$. Together this implies $u(s) = 0$for almost every $s \in \bigcup_i (\underline{x}_i, \overline{x}_i)$and therefore $\int_x^1 u(s) \mathrm{d}s = 0$holds for all $x$and hence $u(x) = 0$for a.e. x. We conclude that

$g$is an extreme point of $\Omega_m(f)$. □

Proof of Corollary 1: Fix an arbitrary $h \in \Omega_m(f)$and define

$$c(x) = \begin{cases} x & \text{if } x \notin \bigcup_{i\in I}[\underline{x}_i, \overline{x}_i) \\ \overline{x}_i - x + \underline{x}_i & \text{if } x \in [\underline{x}_i, \overline{x}_i). \end{cases}$$

Moreover, let

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

and

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

It follows that

$$\int_0^1 (g(x) - h(x))c(x)\mathrm{d}x \geq \int_0^1 (g(x) - \overline{h}(x))c(x)\mathrm{d}x = \int_0^1 (g(x) - \overline{h}(x))\overline{c}(x)\mathrm{d}x \quad (13)$$

$$= \int_0^1 \int_x^1 g(s) - \overline{h}(s)\mathrm{d}s \; \mathrm{d}\overline{c}(x) = \int_0^1 \int_x^1 f(s) - \overline{h}(s)\mathrm{d}s \; \mathrm{d}\overline{c}(x) \geq 0, \quad (14)$$

where the first inequality and the first equality follow from Chebyshev's inequality, the second equality follows from integration by parts, the third follows from Theorem 1, and the final inequality holds since $\overline{h} \prec f$. Consequently, c determines a supporting hyperplane for $\Omega_m(f)$ through g. Moreover, this hyperplane contains no other point in $\Omega_m(f)$: equality holds in (13) only if h is constant on each of the intervals $[\underline{x}_i, \overline{x}_i)$ (see Fink and Jodeit (1984)), which yields $h = \overline{h}$. Equality holds in (14) only if $\int_x^1 \overline{h}(s)\mathrm{d}s = \int_x^1 f(s)\mathrm{d}s$ for all $x \notin \bigcup_i [\underline{x}_i, \overline{x}_i)$. This implies that

$$\overline{h}(x) = f(x) = g(x) \text{ for all } x \notin \bigcup_{i\in I} [\underline{x}_i, \overline{x}_i) \text{ and}$$

$$\overline{h}(x) = \frac{\int_{\underline{x}_i}^{\overline{x}_i} f(s)\mathrm{d}s}{\overline{x}_i - \underline{x}_i} = g(x) \text{ for } x \in [\underline{x}_i, \overline{x}_i)$$

Therefore, g is the only point of $\Omega_m(f)$ contained in the hyperplane. □

Proof of Corollary 2: Observe first that $\Omega_{m,w}(f) = \bigcup_{\theta\in[0,1]} \Omega_m(f \cdot \mathbf{1}_{[\theta,1]})$: Since f is non-negative, $g \prec f \cdot \mathbf{1}_{[\theta,1]}$ implies that $g \prec_w f$. Conversely, if $g \prec_w f$ then there exists $\theta \in [0,1]$ such that

$$\int_0^1 g(x)\mathrm{d}x = \int_0^1 f(x) \cdot \mathbf{1}_{[\theta,1]}(x)\mathrm{d}x$$

and therefore $g \prec f \cdot \mathbf{1}_{[\theta,1]}$.

It follows that, for any extreme point g of $\Omega_{m,w}(f)$, there exists $\theta \in [0,1]$ such that $g \in \Omega_m(f \cdot \mathbf{1}_{[\theta,1]})$. Since $\Omega_m(f \cdot \mathbf{1}_{[\theta,1]}) \subseteq \Omega_{m,w}(f)$, g must be an extreme point of this set.

Conversely, suppose that g is an extreme point of $\Omega_m(f \cdot \mathbf{1}_{[\theta,1]})$ with representative g, and suppose that there exists $u \in L^1$ such that $g \pm u \in \Omega_{m,w}(f)$. It follows that $g(x) = 0$ for almost every $x \in [0,\theta)$ and, since $g(x) \pm u(x) \geq 0$ for almost every x, we obtain that $u(x) = 0$ for almost every $x \in [0,\theta)$. Therefore, $g \pm u \in \Omega_{m,w}(f \cdot \mathbf{1}_{[\theta,1]})$. Also, since

$$\int_\theta^1 g(s)\mathrm{d}s = \int_\theta^1 f(s) \cdot \mathbf{1}_{[\theta,1]}(s)\mathrm{d}x$$

, we obtain $\int_\theta^1 u(s)\mathrm{d}s = 0$. We conclude that

$$\int_\theta^1 (g \pm u)(s)\mathrm{d}s = \int_\theta^1 f(s) \cdot \mathbf{1}_{[\theta,1]}\mathrm{d}s$$

and therefore that $g \pm u \in \Omega_m(f \cdot \mathbf{1}_{[\theta,1]})$. Since g is an extreme point of $\Omega_m(f)$, $u \equiv 0$, and hence g is an extreme point of $\Omega_{m,w}(f)$. □

Proof of Theorem 2: "⇒": Let g be an extreme point of $\Phi_m(f)$, and denote its non-decreasing and right-continuous representative that is left-continuous at $x = 1$ also by g. Recall that f is continuous by assumption.

Step 1: Fix any x such that $g(x) < g(s)$ for all $s > x$. If $\int_x^1 (g(s) - f(s))\mathrm{d}s > 0$, then we can choose $s_1 = x$ and $s_2 > s_1$ small enough such that u defined in (12) satisfies $g \pm u \in \Phi_m(f)$ and $u \not\equiv 0$, a contradiction; hence, $\int_x^1 (g(s) - f(s))\mathrm{d}s \leq 0$ for any such x.

Now if $g(x) > f(x)$, then right-continuity of g and of f implies that there exists $\varepsilon > 0$ such that $\int_{x+\varepsilon}^1 (g(s) - f(s))\mathrm{d}s < 0$, which contradicts $g \succ f$. Therefore, $g(x) \leq f(x)$.

If $g(x) < f(x)$,[49] then this inequality holds on $[x, x+\varepsilon)$ for some $\varepsilon > 0$, and hence we can choose $s_1 = x$ and $s_2 > s_1$ small enough such that $g \pm u \in \Phi_m(f)$, and such that $u \not\equiv 0$, contradicting that g is an extreme point.

We conclude that, if $g(x) < g(s)$ for all $s > x$, then $g(x) = f(x)$.

Step 2: Hence, for all x, either $g(x) = f(x)$ or there exists $y > x$ such that g is constant on $[x, y]$. Since $[x, y]$ contains a rational number, there is a countable collection of intervals I_j such that g is constant on I_j for each j, and such that $f = g$ outside of

[49] If $x = 1$ then left-continuity of f and g at 1 imply that there is $\varepsilon > 0$ such that $\int_{1-\varepsilon}^1 f(s) - g(s)\mathrm{d}s > 0$, contradicting $g \succ f$. Hence, $x < 1$.

$\bigcup_j I_j$. Let

$$Y = \{y \in \bigcup_j cl(I_j) \mid \int_y^1 (f(s) - g(s))\mathrm{d}s = 0\}$$

and observe that, since f is strictly increasing, the collection of sets Y is countable. Then Y defines a partition of $\bigcup_j I_j$ into non-degenerate intervals. Consider an arbitrary such interval, say $[\underline{x}_i, \overline{x}_i)$. We have

$$\begin{aligned} \int_{\underline{x}_i}^1 (f(s) - g(s))\mathrm{d}s &= 0 \;, \int_{\overline{x}_i}^1 (f(s) - g(s))\mathrm{d}s = 0, \text{ and} \\ \int_x^1 (f(s) - g(s))\mathrm{d}s &< 0 \quad \text{for all } x \in (\underline{x}_i, \overline{x}_i) \text{ (since } g \succ f), \end{aligned}$$

and g is piece-wise constant on $[\underline{x}_i, \overline{x}_i)$.

We now prove that g consists of either two or three pieces on this interval. Indeed, if $[\underline{x}_i, \overline{x}_i)$ is partitioned into more than three intervals, then there are non-empty intervals $[a, b)$ and $[c, d)$ with $a > \underline{x}_i$ and $d < \overline{x}_i$ such that g is constant on these intervals and increases strictly at a, b, c, d (i.e., $g(a) > g(s)$ for all $s < a$, $g(s) > g(a)$ for all $s > b$, $g(c) > g(s)$ for all $s < c$, and $g(s) > g(c)$ for all $s > d$). Moreover, since $\int_x^1 (f(s) - g(s))\mathrm{d}s$ is continuous in the variable x, it achieves its maximum on $[a, d]$, which is strictly negative by assumption. Now if g were continuous at $x \in \{a, b, c, d\}$ we could choose s_1 and s_2 such that u defined by (12) satisfies $g \pm u \in \Phi_m(f)$ and $u \not\equiv 0$ (Lemma 2), a contradiction. Hence, g must have a discrete jump at $x \in \{a, b, c, d\}$. But, this implies that we can choose $\delta, \varepsilon > 0$ small enough such that u defined by

$$u(s) = \delta \mathbf{1}_{[a,b)}(s) - \varepsilon \mathbf{1}_{[c,d)}$$

satisfies $g \pm u \in \Phi_m(f)$, contradicting the assumption that g is an extreme point.

Finally, we show that $\lim_{s \uparrow \overline{x}_i} g(s) = f(\overline{x}_i)$. Observe that $g(\overline{x}_i) \le f(\overline{x}_i)$ since the right-continuity of g and f would otherwise imply that $\int_y^1 (g(s) - f(s))\mathrm{d}s < 0$ for some y whenever $\overline{x}_i < 1$, and that $g(1) \le f(1)$ since g is continuous at 1 and $g(x) \le f(1)$ a.e. by assumption. By an analogous argument, it must hold that $f(\overline{x}_i) \le \lim_{s \uparrow \overline{x}_i} g(s)$. Since $\lim_{s \uparrow \overline{x}_i} g(s) \le g(\overline{x}_i)$, we obtain

$$\lim_{s \uparrow \overline{x}_i} g(s) \le g(\overline{x}_i) \le f(\overline{x}_i) \le \lim_{s \uparrow \overline{x}_i} g(s)$$

and thus all terms are equal. Similar arguments establish that $g(\underline{x}_i) = f(\underline{x}_i)$.

"$\Leftarrow$": Suppose that g satisfies the conditions in the statement of the theorem. Then g is non-decreasing and $g \succ f$, hence $g \in \Phi_m(f)$. Now suppose that $u \in L^1$ satisfies

$g \pm u \in \Phi_m(f)$. Then, for all $i \in \mathcal{I}$ it must hold that

$$\int_{\underline{x}_i}^{1} u(s)\mathrm{d}s = 0 \text{ and } \int_{\overline{x}_i}^{1} u(s)\mathrm{d}s = 0.$$

If $i \in \mathcal{I}$ is such that

$$g(x) = \begin{cases} f(\underline{x}_i) & \text{if } x \in [\underline{x}_i, \underline{y}_i) \\ f(\overline{x}_i) & \text{if } x \in [\underline{y}_i, \overline{x}_i) \end{cases}$$

then u is constant on $(\underline{x}_i, \underline{y}_i)$ and on $[\underline{y}_i, \overline{x}_i)$. If $\underline{x}_i = 0$ then $u = 0$ on $[\underline{x}_i, \underline{y}_i)$ since $g(0) = f(0)$ and $(g \pm u)(0) \geq f(0)$. So suppose $\underline{x}_i > 0$ and $u < 0$ on $[\underline{x}_i, \underline{y}_i)$ (and otherwise consider $-u$). Then

$$(g + u)(\underline{x}_i) < f(\underline{x}_i)$$

Since f is continuous and since $g + u$ is non-decreasing, we obtain for some $\varepsilon > 0$ that

$$\int_{\underline{x}_i - \varepsilon}^{1} [(g + u)(s) - f(s)]\mathrm{d}s < 0$$

which yields a contradiction. Therefore, $u = 0$ on $[\underline{x}_i, \underline{y}_i)$ and since $\int_{\underline{x}_i}^{\overline{x}_i} u(s)\mathrm{d}s = 0$ we obtain that $u = 0$ on $[\underline{x}_i, \overline{x}_i)$.

On the other hand, if $i \in \mathcal{I}$ is such that g satisfies

$$g(x) = \begin{cases} f(\underline{x}_i) & \text{if } x \in [\underline{x}_i, \underline{y}_i) \\ v_i & \text{if } x \in [\underline{y}_i, \overline{y}_i) \\ f(\overline{x}_i) & \text{if } x \in [\overline{y}_i, \overline{x}_i) \end{cases}$$

for some v_i then u is constant on $[\underline{x}_i, \underline{y}_i)$, on $[\underline{y}_i, \overline{y}_i)$ and on $[\overline{y}_i, \overline{x}_i)$. The same arguments as in the preceding paragraph imply that $u = 0$ on $[\underline{x}_i, \underline{y}_i)$. Now assume $u < 0$ on $[\overline{y}_i, \overline{x}_i)$ (otherwise consider $-u$). Then $(g + u)(\overline{y}_i) < f(\overline{x}_i)$ and there exists $\varepsilon > 0$ such that

$$\int_{\overline{x}_i - \varepsilon}^{1} (g + u)(s) - f(s)\mathrm{d}s < 0,$$

a contradiction. We conclude that $u = 0$ on $[\overline{y}_i, \overline{x}_i)$ and since $\int_{\underline{x}_i}^{\overline{x}_i} u(s)\mathrm{d}s = 0$ we obtain that $u = 0$ on $[\underline{x}_i, \overline{x}_i)$.

Observe that $\int_x^1 (f(s) - g(s))\mathrm{d}s = 0$ for $x \notin \bigcup [\underline{x}_i, \overline{x}_i)$ and hence $\int_x^1 u(s)\mathrm{d}s = 0$ for $x \notin \bigcup [\underline{x}_i, \overline{x}_i)$. Since $u(x) = 0$ for all $x \in \bigcup_i [\underline{x}_i, \overline{x}_i)$, we conclude that $\int_x^1 u(s)\mathrm{d}s = 0$ for all $x \in [0, 1]$, and therefore that $u \equiv 0$.

We need to show that Conditions (3), (4), (5) are equivalent to $\int_{\underline{x}_i}^{\overline{x}_i} f(s)-g(s)\mathrm{d}s = 0$, $v_i \in [f(\underline{y}_i), f(\overline{y}_i)]$, $g \succ f$, respectively. We begin by showing that (3) is equivalent to $\int_{\underline{x}_i}^{\overline{x}_i} f(s) - g(s)\mathrm{d}s = 0$. Plugging in the definition of g yields that this condition is equivalent to

$$0 = \int_{\underline{x}_i}^{\overline{x}_i} f(s)\mathrm{d}s - f(\underline{x}_i)(\underline{y}_i - \underline{x}_i) - f(\overline{x}_i)(\overline{x}_i - \overline{y}_i) - v_i(\overline{y}_i - \underline{y}_i)$$

and thus equivalent to (3).

We next show that (4) is equivalent to $v_i \in [f(\underline{x}_i), f(\overline{x}_i)]$ and thus to the monotonicity of g. It follows from (3) that $v_i \leq f(\overline{x}_i)$ is equivalent to

$$\int_{\underline{x}_i}^{\overline{x}_i} f(s)\mathrm{d}s - f(\underline{x}_i)(\underline{y}_i - \underline{x}_i) - f(\overline{x}_i)(\overline{x}_i - \overline{y}_i) \leq f(\overline{x}_i)(\overline{y}_i - \underline{y}_i).$$

Adding $f(\underline{x}_i)(\underline{y}_i - \underline{x}_i) - f(\overline{x}_i)(\overline{x}_i - \overline{y}_i)$ yields

$$\int_{\underline{x}_i}^{\overline{x}_i} f(s)\mathrm{d}s \leq f(\underline{x}_i)(\underline{y}_i - \underline{x}_i) + f(\overline{x}_i)(\overline{x}_i - \underline{y}_i)\,.$$

The other side of the inequality follows from an analogous argument for $f(\underline{x}_i) \leq v_i$ and we thus have that (4) is equivalent to $v_i \in [f(\underline{x}_i), f(\overline{x}_i)]$.

Finally, we show that (5) ensures that $g \succ f$ if $v_i \in (f(\underline{y}_i), f(\overline{y}_i))$ and that $g \succ f$ is automatically satisfied if $v_i \notin (f(\underline{y}_i), f(\overline{y}_i))$. As $\int_x^1 f(s)-g(s)\mathrm{d}s = 0$ for all $x \notin \bigcup[\underline{x}_i, \overline{x}_i)$ it suffices to show that $\int_x^1 f(s) - g(s)\mathrm{d}s \leq 0$ for all $x \in [\underline{x}_i, \overline{x}_i)$.

Consider the case where $v_i \in (f(\underline{y}_i), f(\overline{y}_i))$. Since f is continuous and since $v_i \in [f(\underline{y}_i), f(\overline{y}_i)]$, there exists a point $m_i \in (\underline{y}_i, \overline{y}_i)$ such that $f(m_i) = v_i$. As $g(x) \leq f(x)$ for $x \in [\underline{x}_i, \underline{y}_i]$, we obtain for all $x \in [\underline{x}_i, \underline{y}_i]$ that

$$0 = \int_{\underline{x}_i}^1 f(s) - g(s)\mathrm{d}s \geq \int_x^1 f(s) - g(s)\mathrm{d}s\,.$$

Furthermore, as $g(x) \geq f(x)$ for $x \in [\underline{y}_i, m]$ we get that for all $x \in [\underline{y}_i, m]$

$$\int_x^1 f(s) - g(s)\mathrm{d}s \leq \int_m^1 f(s) - g(s)\mathrm{d}s\,.$$

A symmetric argument shows that the same conclusion holds for all $x \in [m, \overline{y}_i]$ and that $\int_x^1 f(s) - g(s)\mathrm{d}s \leq 0$ for all $x \in [\overline{y}_i, \overline{x}_i)$. We thus have that $\int_x^1 f(s) - g(s)\mathrm{d}s \leq 0$ for all $x \in [\underline{x}_i, \overline{x}_i)$ if and only if $\int_m^1 f(s) - g(s)\mathrm{d}s \leq 0$, which is equivalent to (5).

If $v_i \notin (f(\underline{y}_i, \overline{y}_i))$ then $x \mapsto \int_x^1 f(s) - g(s)\mathrm{d}s$ is quasi-concave on the interval $[\underline{x}_i, \overline{x}_i]$

and thus maximized at either $\underline{x}_i$ or $\overline{x}_i$. Condition (3) ensures that this integral equals zero at both points and thus $f \prec g$. $\square$

Proof of Proposition 2: To simplify notation, let $\overline{C}$ denote the convex hull of C. Note that, since C is continuous, $\overline{C}(0) = C(0)$ and $\overline{C}(1) = C(1)$. Also, if $\overline{C}(x) < C(x)$ for all $x \in (a,b) \subset [0,1]$, then $\overline{C}$ is affine on (a,b).

For every non-decreasing function h that satisfies $h \prec f$ we obtain[50]

$$\int c(x)h(x)\mathrm{d}x = C(1)h(1) - \int_0^1 C(x)\mathrm{d}h(x) \leq \overline{C}(1)h(1) - \int_0^1 \overline{C}(x)\mathrm{d}h(x) \tag{15}$$

$$= \int_0^1 \overline{C}'(x)h(x)\mathrm{d}x \leq \int_0^1 \overline{C}'(x)f(x)\mathrm{d}x, \tag{16}$$

where the equalities follow from integration by parts for the Riemann-Stieltjes integral, where the first inequality follows since $\overline{C}(x) \leq C(x)$, and where the final inequality follows from the Fan-Lorentz Theorem since $\overline{C}'$ is non-decreasing.

Since, by assumption, $\overline{C}(x) = C(x)$ for $x \notin \bigcup_{i \in I}[\underline{x}_i, \overline{x}_i)$ and since g is constant on $[\underline{x}_i, \overline{x}_i)$, we obtain that

$$\int_0^1 C(x)\mathrm{d}g(x) = \int_0^1 \overline{C}(x)\mathrm{d}g(x);$$

and hence, (15) holds as an equality for $h = g$. Also, since $f(x) = g(x)$ for $x \notin \bigcup_{i \in I}[\underline{x}_i, \overline{x}_i)$, since $\overline{C}$ is affine on $[\underline{x}_i, \overline{x}_i)$, and since g is constant on $[\underline{x}_i, \overline{x}_i)$ with $g(x) = \int_{\underline{x}_i}^{\overline{x}_i} f(s)\mathrm{d}s$, we obtain

$$\int_0^1 \overline{C}'(x)g(x)\mathrm{d}x = \int_0^1 \overline{C}'(x)f(x)\mathrm{d}x.$$

Hence, setting $h = g$ also satisfies (16) as an equality, and we conclude that g is optimal.

For the converse, assume that f is strictly increasing. Observe first that there is $h \in \Omega_m(f)$ that satisfies (15) as an equality: Let $\{\ [\underline{y}_j, \overline{y}_j] \mid j \in J\}$ be a minimal collection of intervals such that $\overline{C}$ is affine on $[\underline{y}_j, \overline{y}_j]$ for each $j \in J$ and such that $\overline{C}(x) = C(x)$ for all $x \notin \bigcup_{j \in J}[\underline{y}_j, \overline{y}_j]$. Define h to be constant on $[\underline{y}_j, \overline{y}_j]$ for each j with

$$\int_{\underline{y}_j}^{\overline{y}_j} h(s)\mathrm{d}s = \int_{\underline{y}_j}^{\overline{y}_j} f(s)\mathrm{d}s,$$

and set $h(x) = f(x)$ for $x \notin \bigcup_{j \in J}[\underline{y}_j, \overline{y}_j]$. It follows from the previous step that h

[50] Since $\overline{C}(x)$ is convex, $\overline{C}'(x)$ exists a.e. and we extend its definition by right-continuity to all x.

satisfies (15) and (16) with equality.

If $\overline{C}$ is not affine on $[\underline{x}_i, \overline{x}_i)$ for some $i \in I$, then $\overline{C}'$ is non-decreasing and it is not constant on $[\underline{x}_i, \overline{x}_i)$. Since f is strictly increasing and g is constant on $[\underline{x}_i, \overline{x}_i)$, an application of Chebyshev's inequality (see Theorem 1 in Fink and Jodeit (1984)) yields

$$\int_{\underline{x}_i}^{\overline{x}_i} 1 \mathrm{d}x \int_{\underline{x}_i}^{\overline{x}_i} \overline{C}'(x) f(x) \mathrm{d}x > \int_{\underline{x}_i}^{\overline{x}_i} f(x) \mathrm{d}x \int_{\underline{x}_i}^{\overline{x}_i} \overline{C}'(x) \mathrm{d}x$$
$$= \int_{\underline{x}_i}^{\overline{x}_i} g(x) \mathrm{d}x \int_{\underline{x}_i}^{\overline{x}_i} \overline{C}'(x) \mathrm{d}x = \int_{\underline{x}_i}^{\overline{x}_i} 1 \mathrm{d}x \int_{\underline{x}_i}^{\overline{x}_i} g(x) \overline{C}'(x) \mathrm{d}x.$$

Hence, g satisfies (15) with strict inequality, and therefore g cannot be optimal.

If $\overline{C}(x) < C(x)$ for some $x \notin \bigcup_{i \in I} [\underline{x}_i, \overline{x}_i)$ then there is $\varepsilon > 0$ such that $\overline{C}(z) < C(z)$ for all $z \in [x, x+\varepsilon]$ and $g(x) < g(x+\varepsilon)$ (since f is strictly increasing). Hence,

$$\int_x^{x+\varepsilon} \overline{C}(s) \mathrm{d}g(x) < \int_x^{x+\varepsilon} C(s) \mathrm{d}g(x)$$

and g satisfies (15) as a strict inequality and therefore g cannot be optimal. □

References

[2006] Aliprantis, C. D., & Border, K.C (2006). *Infinite Dimensional Analysis*, 3rd edition. Berlin: Springer Verlag.

[2008] Alonso, R., & Matouschek, N. (2008). Optimal Delegation. *The Review of Economic Studies* **75**(1), 259–293.

[2013] Amador, M., & Bagwell, K. (2013). The Theory of Optimal Delegation with an application to Tariff Caps. *Econometrica* **81**(4), 1541–1599.

[2020] Arieli, I., Babichenko, Y., Smorodinsky, R. and Yamashita, T. (2020) "Optimal Persuasion via Bi-Pooling", discussion paper.

[1991] Border, K. (1991): "Implementation of Reduced Form Auctions: A Geometric Approach," *Econometrica* **59**(4), 1175–1187.

[2013] Budish, E., Che, Y.-K., Kojima, F. & Milgrom, P. (2013): "Designing Random Allocation Mechanisms: Theory and Applications," *American Economic Review* **103**(2), 585–623.

[2013] Che, Y.K., Kim, J. and Mierendorff, K. (2013): "Generalized Reduced Form Auctions: A Network-Flow Approach," *Econometrica* **81**(6), 2487–2520.

[1987] Chan, W., Proschan, F., & Sethuraman, J. (1987). “Schur–Ostrowski Theorems for Functionals on L_1(0,1),” *SIAM Journal on Mathematical Analysis*, **18**(2), 566–578.

[2007] Damiano, E., & Li, H. (2007). “Price Discrimination and Efficient Matching,” *Economic Theory*, **30**(2), 243–263.

[2005] Dana, R. A. (2005). “A Representation Result for Concave Schur Concave Functions,” *Mathematical Finance* **15**(4), 613–634

[2019] Dworczak, P. & Martini, G. (2019). “The Simple Economics of Optimal Persuasion,” *Journal of Political Economy*, **127**(5),1993–2048

[1988] Dybvig, P. H. (1988). “Distributional Analysis of Portfolio Choice,” *Journal of Business,* 369–393.

[1954] Fan, K. and Lorentz, G.G. (1954): “An Integral Inequality,” *The American Mathematical Monthly* **61**(9), 626–631

[1984] Fink A. M. and Jodeit Jr, M. (1984): “On Chebyshev’s Other Inequality,” *Lecture Notes-Monograph Series*, Vol. 5, Inequalities in Statistics and Probability

[2016] Gentzkow, M. & Kamenica, E. (2016). “A Rothschild-Stiglitz Approach to Bayesian Persuasion,” *American Economic Review: Papers & Proceedings*, **106** (5), 597–601.

[2010] Gershkov, A., & Moldovanu, B. (2010). “Efficient Sequential Assignment with Incomplete Information,” *Games and Economic Behavior* **68**(1), 144–154.

[2013] Gershkov, A., Goeree J., Kushnir, A., Moldovanu, B. and Shi, X. (2013): “On the Equivalence of Bayesian and Dominant Strategy Implementation,” *Econometrica* **81**(1), 197–220.

[2019] Gershkov, A., Moldovanu, B., Strack, P. and Zhang, M. (2019). ”A Theory of Auctions with Endogenous Valuations”, discussion paper, University of Bonn.

[1929] Hardy, G. H., Littlewood J. E., and Polya G. (1929): “Some Simple Inequalities Satisfied by Convex Functions,” *Messenger Math* **58**, 145–152.

[2015] Hart, S. and Reny, P (2015): “Implementation of Reduced Form Mechanisms: a Simple Approach and a New Characterization,” *Economic Theory Bulletin* **3**(1), 1–8.

[1984] Holmström, B. (1984). On the Theory of Delegation," in: *Bayesian Models in Economic Theory*. Ed. by M. Boyer, and R. Kihlstrom. North-Holland, New York.

[1987] Hong, C. S., Karni, E., & Safra, Z. (1987). "Risk Aversion in the Theory of Expected Utility with Rank Dependent Probabilities," *Journal of Economic Theory*, **42**(2), 370–381.

[2009] Hoppe, H. C., Moldovanu, B., & Sela, A. (2009)." The Theory of Assortative Matching Based on Costly Signals," *The Review of Economic Studies*, **76**(1), 253–281.

[1987] Horsley A. & Wrobel, A. J. (1987). "The Extreme Points of Some Convex Sets in the Theory of Majorization," *Investigationes Mathematicae (Proceedings A)*, **90**(2), 171–176.

[2020] Kartik, N., Kleiner, A.& Van Weelden, R. (2020). "Delegation in Veto Bargaining," *Working Paper*.

[2018] Kolotilin, A. (2018). "Optimal Information Disclosure: A Linear Programming Approach," *Theoretical Economics*, **13**, 607–635.

[2019] Kolotilin, A. & Zapechelnyuk, A. (2019). "Persuasion Meets Delegation," preprint arXiv:1902.02628.

[2009] Kovac, E. & Mylovanov, T. (2009). "Stochastic Mechanisms in Settings Without Monetary Transfers: The Regular Case," *Journal of Economic Theory*, **144**, 1373–1395.

[1982] Machina, M. J. (1982). ""Expected Utility" Analysis without the Independence Axiom," *Econometrica*, 277–323.

[2010] Manelli, A. and Vincent, D. (2010): "Bayesian and Dominant Strategy Implementation in the Independent, Private Values Model," *Econometrica* **78**(6), 1905–1939.

[1984] Maskin, E. S. and Riley, J. (1984), "Optimal Auctions with Risk-Averse Buyers," *Econometrica* **52**, 1473–1518.

[1984] Matthews, S. A. (1984), "On the Implementability of Reduced Form Auctions," *Econometrica* **52**, 1519–1522.

[1991] Melumad, N. D., and Shibano, T. (1991), "Communication in Settings with No Transfers." *The RAND Journal of Economics*, 173–198.

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

[2018] Olszewski, W., & Siegel, R. (2018). "Pareto Improvements in the Contest for College Admissions," Working Paper, Northwestern University

[2001] Phelps, R. R. (2001). *Lectures on Choquet's Theorem*. Springer Science & Business Media.

[1982] Quiggin, J. (1982). "A Theory of Anticipated Utility," J*ournal of Economic Behavior & Organization* **3**(4), 323–343.

[1967] Ryff, J.V. (1967): " Extreme Points of Some Convex Subsets of $L_1(0, 1)$," *Proceedings of the American Mathematical Society* **18**(6), 1026–1034.

[1989] Schmeidler, D. (1989). "Subjective Probability and Expected Utility without Additivity". *Econometrica ,* 571–587.

[2007] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*, New York: Springer

[1965] Strassen, V. (1965): "The Existence of Probability Measures with Given Marginals," *Mathematical Statistics* **36**, 423–439.

[2011] Toikka, J. (2011). "Ironing without Control," *Journal of Economic Theory* **146**,2510–2526.

[1990] Wakker, P. (1990). "Under Stochastic Dominance Choquet-Expected Utility and Anticipated Utility Are Identical," *Theory and Decision* **29**(2), 119–132.

[1988] Winkler, Gerhard. (1988). "Extreme Points of Moment Sets," *Mathematics of Operations Research* **13**(4),581–587.

[1987] Yaari, M. E. (1987). "The Dual Theory of Choice under Risk," *Econometrica*, 95–115.

B Online Appendix

B.1 Schur-Convex Functions and Functionals

Consider X_F and X_G to be uniform, discrete random variables, each taking n values $x_F = (x_F^1, ..., x_F^n)$ and $x_G = (x_G^1, .., x_G^n)$, respectively. Then

$$x_F \prec_{dm} x_G \Leftrightarrow F^{-1} \prec G^{-1} \Leftrightarrow G \prec F$$

where $\prec_{dm}$ denotes the classical discrete majorization relation due to Hardy, Littlewood and Polya. Thus, discrete majorization is equivalent to the present majorization relation applied to quantile functions.

A function $V : \mathbb{R}^n \to \mathbb{R}$ is *Schur-convex* (*concave*) if $V(\mathbf{x}) \geq V(\mathbf{y})$ $(V(\mathbf{x}) \leq V(\mathbf{y}))$ whenever $\mathbf{x} \succ_{dm} \mathbf{y}$. If V is a symmetric function, and if all its partial derivatives exist, then the *Schur-Ostrovski criterion* says that V is *Schur-convex (concave)* if and only if

$$(x_i - x_j)\left(\frac{\partial V}{\partial x_i} - \frac{\partial V}{\partial x_j}\right) \geq (\leq)\ 0 \text{ for all } x.$$

It is useful to have a similar characterization for continuous majorization. Chan et al. (1987) showed that a law-invariant[51], Gâteaux-differentiable functional $V : L^1(0,1) \to \mathbb{R}$ respects the majorization relation on $L^1(0,1)$, if and only if its *Gâteaux-derivatives* in specially defined directions are non-positive. The considered directions are of the form

$$h = \lambda_1 \mathbf{1}_{(a,b)} + \lambda_2 \mathbf{1}_{(c,d)}$$

with $0 \leq a < b < c < d \leq 1$ and $\lambda_1 \geq 0 \geq \lambda_2$ such that $\lambda_1(b-a) + \lambda_2(d-c) = 0$. Note that the function h takes at most two values that are different from zero, and is decreasing on $[a,b] \cup [c,d]$. Moreover, $\int_0^1 h(t)\,\mathrm{d}t = 0$.

This result also yields a simple intuition for the Fan Lorentz Theorem in the case where K is differentiable. Consider a monotonic f and note that, for any direction h, the Gâteaux-derivative of the functional $V(f) = \int_0^1 K(f(t),t)\,\mathrm{d}t$ is given by

$$\delta V(f,h) = \frac{d}{d\varepsilon}\int_0^1 K(f(t) + \varepsilon h(t), t)\,\mathrm{d}t\,|_{\varepsilon=0} = \int_0^1 K_f(f(t),t)h(t)\,\mathrm{d}t,$$

where the last equality follows by interchanging the order of differentiation and inte-

[51]This means that the functional is constant over the equivalence class of functions with the same distribution (or non-decreasing re-arrangement). This requirement replaces the symmetry in the discrete formulation.

gration.[52] The Fan-Lorentz conditions imply together that

$$\frac{\mathrm{d}K_f}{\mathrm{d}t} = f_t \cdot K_{ff} + K_{ft} \geq 0.$$

For a direction h such that $\int_0^1 h(t)\,\mathrm{d}t = 0$, and such that h is a decreasing two-step function as defined above, we obtain that

$$\delta V(f,h) = \int_0^1 K_f(f(t),t)h(t)\,\mathrm{d}t \leq 0.$$

Hence the Fan-Lorentz functional $V(f) = \int_0^1 K(f(t),t)\,\mathrm{d}t$ is Schur-concave by the result of Chan et al. (1987).

B.2 Proofs

Proof of Lemma 1: For any $h \in \Phi_m(f)$,

$$\int_0^1 c(x)h(x)\mathrm{d}x = C(1)h(1) - \int_0^1 C(x)\mathrm{d}h(x)$$
$$\leq \overline{C}(1)h(1) - \int_0^1 \overline{C}(x)\mathrm{d}h(x) = \int_0^1 \overline{C}'(x)h(x)\mathrm{d}x \leq \int_0^1 \overline{C}'(x)f(x)\mathrm{d}x,$$

where the final inequality follows from the Fan-Lorentz inequality because $\overline{C}'(x)$ is non-increasing. Since g satisfies these inequalities with equality, we conclude that g is optimal. □

Lemma 3. *There is a measurable mapping* $T : \operatorname{ext}\Omega_{m,w}(\psi^*) \to L^1([0,1]^N, \mathbb{R}^{N\times N})$ *that assigns to any extreme point* ψ *a DIC allocation rule that implements* ψ*.*

Proof: The mapping $P : L^1([0,1]^N, \mathbb{R}^{N\times N}) \to L^1([0,1], \mathbb{R})$ that assigns to any allocation rule its induced interim quantile allocation rule is linear, surjective, and continuous, and hence an open mapping (Theorem 5.18 in Aliprantis and Border (2006)). Its inverse is therefore a lower-semicontinuous correspondence (Theorem 17.7 in Aliprantis and Border). Since the set of DIC allocation rules is closed, the correspondence $\mathcal{T}$ that assigns to any extreme point its set of implementing DIC allocation rules is therefore weakly measurable (see Definition 18.1 in Aliprantis and Border). Since $\mathcal{T}$ is closed-valued and since the construction in the text establishes that it is non-empty valued, the *Kuratowski-Ryll-Nardzewski Selection Theorem* (see Theorem 18.13 in Aliprantis and Border (2006)) implies that it admits a measurable selector T. □

[52]This is allowed since K is convex in f.

Proof of Theorem 5: Consider a symmetric, BIC mechanism with induced quantile interim allocation function ψ. Then, by the result in the previous subsection, $\psi \prec_w \psi^*$ where ψ^* is induced by the assortative matching allocation. By Proposition 1-2, there exists a probability measure λ_ψ, supported on $\operatorname{ext}\Omega_{m,w}(\psi^*)$, such that

$$\psi = \int_{\operatorname{ext}\Omega_{m,w}(\psi^*)} \widetilde{\psi}\, \mathrm{d}\lambda_\psi(\widetilde{\psi})\,.$$

For any $\widetilde{\psi}$ in $\operatorname{ext}\Omega_{m,w}(X^*)$ recall that $T(\widetilde{\psi}) = \boldsymbol{\alpha}^{\widetilde{\psi}}$ denotes the allocation that generates $\widetilde{\psi}$ as defined in (9) (see Lemma 3 in the Appendix) and recall hat $\boldsymbol{\alpha}^{\widetilde{\psi}}$ can be chosen to be part of a symmetric, DIC mechanism In other words $\boldsymbol{\alpha}_i^{\widetilde{\psi}}$ is symmetric, and for every i and for every θ_{-i}, the function $\boldsymbol{\alpha}_i^{\widetilde{\psi}}(\theta_i, \theta_{-i}) \cdot \mathbf{q}$ is monotonic in θ_i. Define then an allocation $\boldsymbol{\alpha}$ by

$$\boldsymbol{\alpha} = \int \widetilde{\boldsymbol{\alpha}}\, \mathrm{d}\nu_\psi(\widetilde{\boldsymbol{\alpha}}).$$

where ν_ψ is the pushforward measure of λ_ψ under T. This allocation is a randomization over allocations belonging to DIC mechanisms, and hence $\boldsymbol{\alpha}_i(\theta_i, \theta_{-i}) \cdot \mathbf{q}$ is itself monotonic, and thus part of a DIC mechanism. Moreover, this DIC mechanism generates an interim expected allocation equal to the original ψ, yielding the wished equivalence.

□

Proof of Proposition 3: 1) This follows from the feasibility Theorem 4 and generalizes **all** matching schemes considered in the literature.[53]

2) Assuming that the distribution of prizes is G_{ic}, the expected utility of the agent with type θ in the contest is given by

$$U(\theta) = \int_0^\theta G_{ic}^{-1}(F(\tau))\, \mathrm{d}\tau$$

This is the standard payoff-equivalence result a la Myerson. Let us first maximize $U(1)$, the utility of the highest type. Substituting $F(\theta) = s$, yields the problem

$$\max_{G_{ic}^{-1} \in \Omega_m(G^{-1})} \int_0^1 G_{ic}^{-1}(s) f(s)\, \mathrm{d}s\,.$$

We immediately obtain from the Fan-Lorentz Theorem 3 that the maximizer is G_r^{-1} (G^{-1}) if the density f is non-increasing (non-decreasing), i.e. if F is convex (concave).[54]

[53]See for example the schemes considered by Olszewski and Siegel (2018)) - these are in fact extreme points of the majorization set, and our result shows that the restriction to these is without loss for determining Pareto optimal allocations.

[54]If the distribution of types F is uniform, then the highest type is indifferent among all feasible schemes

Thus, the highest type prefers the random allocation if the distribution of types is convex. But, then it is easy to see that all types prefer the random allocation.

3) Consider now the average contestant utility (welfare) given by

$$\int_0^1 U(\theta)f(\theta)\,\mathrm{d}\theta = \int_0^1 \left(\int_0^\theta G_{ic}^{-1}(F(\tau))\,\mathrm{d}\tau\right) f(\theta)\,\mathrm{d}\theta =$$
$$\int_0^1 G_{ic}^{-1}(F(\theta))(1-F(\theta))\,\mathrm{d}\theta = \int_0^1 G_{ic}^{-1}(s)(1-s)\,\mathrm{d}F^{-1}(s)$$

where the second equality follows by integration by parts, and the last equality by substituting $s = F(\theta)$.

Observe that $F^{-1}(s) = -\ln(1-s)$ and that $(1-s)\,\mathrm{d}F^{-1}(s) = 1$ for the *exponential distribution.* We obtain by Theorem 3 that random matching (assortative matching) maximizes average welfare if the distribution of types F is more convex (concave) on its domain than the exponential distribution, which yields the result.

4) If F has an increasing failure rate, the revenue (i.e., average bid) to a designer is maximized by assortative matching because assortative matching maximizes aggregate output while, by the above result, it also minimizes the agents' welfare. □

Lemma 4. *A mechanism is* undominated *if there does not exists a mechanism where the set of actions is a singleton that yields a higher utility for the principal. The utility of the agent in any undominated, IC mechanism satisfies* $U_A(\theta) \geq -2Var(\gamma(\theta)) - 2\Lambda^2$

Proof: A first observation is that, in any undominated mechanism M, the utility of the principal is bounded from below by the utility she obtains in the mechanism where she takes the ex-ante optimal action $\mathbb{E}[\gamma(\theta)]$, and where she does not ask the agent to report. The principal's utility in that mechanism is given by $-Var(\gamma(\theta))$. Hence, in mechanism M there must exist at least one type $\hat{\theta}$ such that $U_p(\hat{\theta}) \geq -Var(\gamma(\theta))$. As the agent can always pretend to be of type $\hat{\theta}$, a lower bound on the utility of the agent is given by[55]

$$\begin{aligned} U_A(\theta) &\geq -(\theta - \mu_M(\hat{\theta}))^2 - \sigma_M^2(\hat{\theta}) = -([\gamma(\theta) - \mu_M(\hat{\theta})] + [\theta - \gamma(\theta)])^2 - \sigma_M^2(\hat{\theta}) \\ &\geq -2(\gamma(\theta) - \mu_M(\hat{\theta}))^2 - \sigma_M^2(\hat{\theta}) - 2(\theta - \gamma(\theta))^2 \\ &\geq -2U_P(\hat{\theta}) - 2\Lambda^2 \geq -2Var(\gamma(\theta)) - 2\Lambda^2\,. \end{aligned}$$

□

Proof of Proposition 5: The principal's expected utility from using an IC mechanism

since his utility is $\int_0^1 G_{ic}^{-1}(s)\,\mathrm{d}s = \mu_G$.

[55]The second inequality follows as for all $a, b \in \mathbb{R}$ we have $(a-b)^2 \geq 0 \Leftrightarrow \frac{a^2}{2} + \frac{b^2}{2} \geq ab \Leftrightarrow 2a^2 + 2b^2 \geq (a+b)^2$.

M can be written as

$$V_P(\mu_{\tilde{M}}) = \int_{\underline{a}}^{\overline{a}} \left[-\gamma(\theta)^2 + 2\gamma(\theta)\mu_{\tilde{M}}(\theta) - \mu_{\tilde{M}}(\theta)^2 - \sigma^2_{\tilde{M}}(\theta) \right] \, \mathrm{d}F(\theta).$$

Substituting for $\sigma^2_{\tilde{M}}(\theta)$ by the characterization of incentive compatibility, we obtain that the principal's utility becomes:

$$V_P(\mu_{\tilde{M}}) = \int_{\underline{a}}^{\overline{a}} \left[-\gamma(\theta)^2 + 2\gamma(\theta)\mu_{\tilde{M}}(\theta) - \mu_{\tilde{M}}(\theta)^2 - \left(-(\mu_{\tilde{M}}(\theta) - \theta)^2 - 2\int_{\underline{a}}^{\theta} (\mu_{\tilde{M}}(s) - s) \, \mathrm{d}s \right) \right] \, \mathrm{d}F(\theta).$$

Integration by parts yields

$$\begin{aligned}
\int_{\underline{a}}^{\overline{a}} \int_{\underline{a}}^{\theta} (\mu_{\tilde{M}}(s) - s) \, \mathrm{d}s f(\theta) \, \mathrm{d}\theta &= \left[\int_{\underline{a}}^{\theta} (\mu_{\tilde{M}}(s) - s) \, \mathrm{d}s F(\theta) \right]_{\theta=\underline{a}}^{\theta=\overline{a}} - \int_{\underline{a}}^{\overline{a}} (\mu_{\tilde{M}}(\theta) - \theta) F(\theta) \, \mathrm{d}\theta \\
&= \int_{\underline{a}}^{\overline{a}} (\mu_{\tilde{M}}(\theta) - \theta)(1 - F(\theta)) \, \mathrm{d}\theta \, .
\end{aligned}$$

Plugging this back into the above equation and simplifying yields

$$\begin{aligned}
V_P(\mu_{\tilde{M}}) &= \int_{\underline{a}}^{\overline{a}} \left[-\gamma(\theta)^2 f(\theta) + 2(\gamma(\theta) - \theta) f(\theta) \mu_{\tilde{M}}(\theta) + \theta^2 f(\theta) + 2(\mu_{\tilde{M}}(\theta) - \theta)(1 - F(\theta)) \right] \, \mathrm{d}\theta \\
&= \int_{\underline{a}}^{\overline{a}} \left[2\left((\gamma(\theta) - \theta) f(\theta) + (1 - F(\theta)) \right) \mu_{\tilde{M}}(\theta) + f(\theta)(\theta^2 - \gamma(\theta)^2) - 2\theta(1 - F(\theta)) \right] \, \mathrm{d}\theta \, .
\end{aligned}$$

This establishes the result. □