Extreme Points and Majorization: Economic Applications*

Andy Kleiner   Benny Moldovanu   Philipp Strack†
March 9, 2020

Abstract

We characterize the set of extreme points of monotone functions that are either majorized by a given function $f$ or themselves majorize $f$. Any feasible element in a majorization set can be expressed as an integral with respect to a measure supported on the extreme points of that set. We show that these extreme points play a crucial role in mechanism design, Bayesian persuasion, optimal delegation and many other models of decision making with expected and non-expected utility. 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 the extreme points in each interval where it is constant. Finally, we apply these insights to a varied set of economic problems.

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.\footnote{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.}

\footnote{Kleiner: Arizona State University; Moldovanu: University of Bonn; Strack: Yale University}
\footnote{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.}
In this paper we show that, somewhat surprisingly, many well-known optimal design and decision problems - examples are the determination of optimal auctions and matching contests, of optimal delegation, Bayesian persuasion and optimal risky choice for non-expected utility decision makers - have a basic common 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 $f$.

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). We also show that any 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. Note that 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 be also 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 5.1 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

\(^2\)See Choquet (1960) and also Phelps (2001) for an excellent introduction.
Strategy incentive compatible mechanisms in the symmetric case.

Roughly speaking, each extreme point is characterized by its specific, countable collection of intervals. Outside these intervals an extreme point must equal the original, fixed function $f$, and inside each interval the extreme point is a step function with at most three different values that are 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 supermodularity 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 certain 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, Bayesian persuasion, optimal delegation, decision making under uncertainty for agents with expected and non-expected utility. We also note that the different instances of the same application, e.g. Bayesian persuasion, may require both type of maximizations on majorization sets described in our paper.

Our main goal in this part last part is to reveal the common underlying role of majorization, and to offer a unified treatment to otherwise complex and well known 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 Section 2 we present some majorization preliminaries and relations to other concepts. Section 3 contains the representation result a la Choquet and the characterizations of extreme points of the majorization sets. In Section 4 we study the optimization of objective functionals with special properties such as convexity, linearity and supermodularity. A major role is played by the Fan-Lorentz integral inequality. Section 5 contains the applications: auctions and contests (5.1), Bayesian persuasion (5.2 and 5.3), optimal delegation (5.4), decision under uncertainty with expected and non-expect utility (5.5). Several proofs are gathered in an Appendix.
2 Majorization Preliminaries

We first recall several concepts and results from the theory of majorization. Throughout, we consider functions that map the unit interval \([0, 1]\) into itself, 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:

\[
\int^1_x g(s) ds \leq \int^1_x f(s) ds \quad \text{for all } x \in [0, 1] \quad (1)
\]

\[
\int^1_0 g(s) ds = \int^1_0 f(s) ds.
\]

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 the non-decreasing rearrangements \(f^*, g^*\), i.e. \(f\) majorizes \(g\) if \(g^* \prec f^*\).

**Relation to Other Orders** Majorization is closely related to other concepts from Economics and Statistics. Let \(X_F\) and \(X_G\) be random variables with distributions \(F\) and \(G\) 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 \(F\). By the derivations in Section 3 and Theorem 3.A.5 of Shaked and Shanthikumar (? ) we obtain

\[
G \prec F \iff X_G \leq_{cv} X_F \iff X_F \leq_{cx} X_G \iff F^{-1} \prec G^{-1},
\]

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

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

where \(ssd\) denotes the standard second-order stochastic dominance.\(^3\) 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, jointly distributed on some probability space, \(X, Y\), such that \(X \sim F, Y \sim G\) and

---

\(^3\)Note also that 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.
such that $Y = \mathbb{E}[X|Y]^4$.

Finally, 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

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

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

3 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 $A$. The classical Krein–Milman Theorem (1940) states that if $A$ is convex and compact set in a locally convex space, then $A$ 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 (1958): 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. We denote by the plain letter $f \in f$ a typical element. 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 $g = f \circ \Psi$ where $\Psi$ is a measure preserving transformation of $[0,1]$ into itself. This generalizes the discrete case, where the corresponding result is due to Hardy, Littlewood and Polya:

---

4See Strassen (1965).

5Formally $x \in A$ is an extreme point if $x = \alpha y + (1 - \alpha)z$, for $z, y \in A$ and $\alpha \in [0,1]$ imply together that $y = x$ or $z = x$.

6Addition (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.
the measure preserving transformations are then represented by doubly-stochastic matrices, and the extreme points correspond, by the \textit{Birkoff-von Neumann Theorem}, to permutation matrices.

As we shall see below, 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 $\Omega(f)$,

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

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

$$\Phi_m(f) = \{ g \in L^1(0,1) | g \text{ non-decreasing with } g \succ f \text{ and with } \essinf f \leq g \leq \esssup f \}.$$

Our first result establishes that, for a non-decreasing $f$, the sets $\Omega_m(f)$, $\Omega_{m,w}(f)$ and $\Phi_m(f)$ are compact, and that every point in these sets can be represented as the expectation of a probability measure that is supported \textit{only} on the extreme points.

**Proposition 1** (Representation Theorem).

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.$^8$

2. For any $g \in \Omega_m(f)$ there exists a probability measure $\mu_g$ supported on the set of extreme points of $\Omega_m(f)$, $\text{ext} \Omega_m(f)$, such that $g = \int_{\text{ext} \Omega_m(f)} h d\mu_g(h)$ (and analogously for any $g \in \Omega_{m,w}(f)$ and $g \in \Phi_m(f)$).$^9$

The second part of the Proposition is a consequence of Choquet’s celebrated theorem (Choquet (1960)) 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.

$^7$The additional constraint $\essinf f \leq g \leq \esssup f$ ensures compactness, and is suitable for our applications below.

$^8$For 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.

$^9$This holds if and only if $V(g) = \int V(h) d\mu(h)$ for any continuous, linear functional $V$. The integral in the Proposition’s statement is a \textit{Bochner integral} of a vector-valued function. See Aliprantis and Border (2006) for an introductory exposition.
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.

**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 \([x_i, \pi_i)\) indexed by \( i \in I \) and \( g \in g \) such that

\[
g(x) = \begin{cases} 
  f(x) & \text{if } x \notin \bigcup_{i \in I} [x_i, \pi_i) \\
  \frac{\int_{x_i}^{x} f(s) \, ds}{\pi_i - x_i} & \text{if } x \in [x_i, \pi_i).
\end{cases}
\]  

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 \( x \).\(^{10}\) 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 \( 1_A(x) \) 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 \( \Omega_m(f 1_{[\theta, 1]}) \) 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 \([x_i, \pi_i)\) and (potentially empty) sub-intervals

\(^{10}\)Formally, \( x \) is exposed if there exists a supporting hyperplane \( H \) such that \( H \cap A = \{x\} \).
\([y_i, \bar{y}_i) \subset [x_i, \bar{x}_i)\) indexed by \(i \in I\) and \(g \in \mathbf{g}\) such that

\[
g(x) = \begin{cases} 
  f(x) & \text{if } x \notin \bigcup_{i \in I} [x_i, \bar{x}_i) \\
  f(x_i) & \text{if } x \in [x_i, y_i) \\
  v_i & \text{if } x \in [y_i, \bar{y}_i) \\
  f(\bar{x}_i) & \text{if } x \in [\bar{y}_i, \bar{x}_i)
\end{cases}
\]

and such that the following three conditions are satisfied:

\[
v_i = \frac{1}{\bar{y}_i - y_i} \left( \int_{x_i}^{\bar{x}_i} f(s)ds - f(x_i)(y_i - x_i) - f(\bar{x}_i)(\bar{y}_i - \bar{x}_i) \right) \quad (3)
\]

\[
f(x_i)(\bar{y}_i - x_i) + f(\bar{x}_i)(\bar{y}_i - \bar{x}_i) \leq \int_{x_i}^{\bar{x}_i} f(s)ds \leq f(x_i)(y_i - x_i) + f(\bar{x}_i)(\bar{y}_i - \bar{x}_i) \quad (4)
\]

\[
\int_{m_i}^{1} f(s)ds \leq v_i(\bar{y}_i - m) + f(\bar{x}_i)(\bar{y}_i - \bar{x}_i) \quad (5)
\]

where \(m_i \in [y_i, \bar{y}_i)\) is an arbitrary point with \(f(m_i) = v_i\).

Condition (3) in the Theorem ensures that \(g\) and \(f\) have the same integrals for each sub-interval \([x_i, \bar{x}_i)\), analogously to the condition imposed in Theorem 1. Condition (4) ensures that \(v \in [f(y_i), f(\bar{y}_i)]\), and hence that \(f\) crosses \(g\) in the interval \([y_i, \bar{y}_i)\) - this implies that \(m\) is well defined. Condition (5) ensures that \(\int_{x_i}^{\bar{x}_i} f(t)dt \leq \int_{x_i}^{\bar{x}_i} f(t)dt\) for all \(s \in [x_i, \bar{x}_i)\) and thus that \(f \prec g\). We illustrate the differences between the extreme points of \(\Omega_m(f), \Phi_m(f)\) in Figure 1. To intuitively understand these differences consider the case where \(f\) is a cumulative distribution function (CDF). 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 reflected in the extreme points of \(\Omega_m\) and \(\Phi_m\). Recall that a CDF \(g\) admits a jump at a value if the distribution assigns a mass-point to that value. Each extreme point \(g \in \Omega_m(f)\) is obtained by taking the mass in each interval \([x_i, \bar{x}_i)\] and spreading it out into two mass points at the boundaries values of the interval \(x_i\) and \(\bar{x}_i\) (see Figure 1). There is a unique way of doing so while preserving the mean determined by (2). In contrast, each extreme point \(g \in \Phi_m(f)\) is obtained by contracting the the mass in each interval \([x_i, \bar{x}_i]\) into two mass points at \(y_i\) and \(\bar{y}_i\). Mass to the left of \(m = f^{-1}(v)\) is moved to \(y_i\) and mass to the right of \(m\) is moved to \(\bar{y}_i\) (see Figure 1).\(^{11}\) Condition (3) determines the mass at these mass points and ensures that the

\(^{11}\)If \(f\) is not strictly increasing, the function \(f\) is constant on on interval \(\{s : f(s) = v\}\), which implies that
mean is preserved. Condition (4) ensures that \( g \) can be obtained from \( f \) by moving mass. Condition (5) ensures that \( g \) is a contraction of \( f \).

While it is relatively immediate that the functions described in Theorem 1 and 2 are mean-preserving spreads (or contractions) of \( f \), the main insight of Theorem 1 and 2 is that these functions can not be represented as convex combinations of other functions in \( \Omega_m(f) \) (or \( \Phi_m(f) \)) and are in fact the only functions with these properties.\(^{12}\)

### 4 Special Objective Functionals

Our previous characterization of extreme points determines all functions that can arise as a maximizer of some convex functional over a set described by majorization constraints. None of these maximizers can be a-priori ruled out even if one restricts to linear functionals. However, in applications, further monotonicity or super-modularity conditions are often naturally when interpreted as a CDF, the distribution assigns no mass to that interval and thus any choice of \( m \) in that interval will lead mass to be moved in the same way.

\(^{12}\)A related intuition for why the extreme points involve only two mass points in each interval comes from the result by Winkler (1988) which states 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 just a mean constraint that implies that any extreme point is a sum of at most two mass points. Note, however, that Winkler’s result is not applicable in our case and his characterization does not hold as we impose the majorization constraint (which in Winkler’s language corresponds to uncountably many constraints).
satisfied or imposed on the objective function. In this section we show how such conditions can be used to further shrink the set of relevant extreme points.

4.1 Schur-Concave Functionals

Recall that a function $V : \mathbb{R}^n \to \mathbb{R}$ is Schur-convex (concave) if $V(x) \geq V(y)$ ($V(x) \leq V(y)$) whenever $x \succ_{dm} y$. 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 will be very useful for our applications below to have a similar characterization for continuous majorization. Chan et al. (1987) showed that a law-invariant 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

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

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, $\int_0^1 h(t)dt = 0$.

The above criterion is not easy to check in practice, but a very useful integral inequality, due to Ky Fan and G.G. Lorentz (1954) identifies a large set of convex and Schur-concave functionals (see below for numerous applications):

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

$$\int_0^1 K(f(t),t)dt \leq \int_0^1 K(g(t),t)dt$$

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 $(u,t)$.

For a simple intuition in the case where $K$ is differentiable, consider a monotonic $f$ and note that, for any direction $h$, the Gâteaux-derivative of the functional $V(f) = \int_0^1 K(f(t),t)dt$

\footnote{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.}
is given by
\[ \delta V(f, h) = \left. \frac{d}{d\alpha} \int_0^1 K(f(t) + \alpha h(t), t) \, dt \right|_{\alpha=0} = \int_0^1 K_f(f(t), t) h(t) \, dt, \]
where the last equality follows by interchanging the order of differentiation and integration.\footnote{This is allowed since \( K \) is convex in \( f \).}
The Fan-Lorentz conditions imply together that
\[ \frac{dK_f}{dt} = f_t \cdot K_{ff} + K_{ft} \geq 0. \]
For a direction \( h \) such that \( \int_0^1 h(t) \, dt = 0 \), and such that \( h \) is a decreasing two-step function as defined above, we obtain that
\[ \delta V(f, h) = \int_0^1 K_f(f(t), t) h(t) \, dt \leq 0. \]
Hence the Fan-Lorentz functional \( V(f) = \int_0^1 K(f(t), t) \, dt \) is Schur-concave by the result of Chan et al. (1987).

### 4.2 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 (see for example Brezis 2010) 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
\[ V(f) = \int_0^1 c(x) f(x) \, dx \]
for every \( f \in L^1(0, 1) \). Thus, we can confine here attention to the maximization of this kind of integrals.

Note that a linear \( K(f, x) = c(x) f(x) \) is supermodular (submodular) in \( (f, x) \) and hence, by Theorem 3, the functional \( V(f) = \int_0^1 c(x) f(x) \, dx \) is Schur-concave (convex) if and only if \( c \) is non-increasing (non-decreasing). We repeatedly apply this observation below.
4.2.1 Maximizing a Linear Functional on $\Omega_m(f)$

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

$$\max_{h \in L^1(0,1)} \int c(x)h(x)dx \quad (6)$$

s. t. $h \in \Omega_m(f)$

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 $\mu_f = \int_0^1 f(x)dx$. 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. We now analyze when a given extreme point is optimal. To do so, define

$$C(x) = \int_0^x c(s)ds$$

and let $\text{conv} C$ 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 $\{(x_i, \bar{x}_i)|i \in I\}$ be the collection of intervals described in Theorem 1.

If $\text{conv} C$ is affine on $[x_i, \bar{x}_i]$ for each $i \in I$ and if $\text{conv} C = C$ otherwise, then $g$ solves problem (6). Moreover, if $f$ is strictly increasing then the converse holds.

Essentially, this result characterizes the conditions under which an arbitrary extreme point is optimal. We show that 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.

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

We now analyze the problem

$$\max_{h} \int c(x)h(x)dx \quad (7)$$

s. t. $h \in \Phi_m(f)$
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

$$g(x) = \begin{cases} 
\text{ess inf } f & \text{for } x < \bar{x} \\
\text{ess sup } f & \text{for } x \geq \bar{x}, 
\end{cases}$$

where $\bar{x}$ solves

$$\int_0^\bar{x} \text{ess inf } f ds + \int_{\bar{x}}^1 \text{ess sup } f ds = \int_0^1 f(s) ds = \mu_f$$

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:

**Lemma 1.** Let

$$C(x) = \int_0^x c(s) ds.$$ 

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

1. $\overline{C}(x) = C(x)$ holds for $g$-almost every $x$ and for $x = 1$ and
2. $\int_0^1 \overline{C}'(x) g(x) dx = \int_0^1 \overline{C}'(x) f(x) dx$.

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.

## 5 Economic Applications

In this section we apply the theoretical insights gained above to various economic problems. We show how seemingly different and well-known problems share a common structure: they all involve maximization of functionals over majorization sets.

### 5.1 The Ranked-Item Auction and Contest Models

In this Subsection we analyze an auction/contest model where several commonly ranked, non-identical objects (or prizes) are allocated to several agents. We show that several classical
results about feasibility of Bayesian Incentive Compatible (BIC) mechanisms, the equivalence of these to Dominant Strategy Incentive Compatible (DIC) mechanisms, and the determination of optimal mechanisms (welfare, revenue) are all corollaries of our above characterization of extreme points.

There are $N$ agents with types $\theta_1, \ldots, \theta_N$ that are independently 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 $0 \leq y_1 \leq y_2 \leq \ldots \leq y_M = 1$. If agent $i$ with type $\theta_i$ receives an object with quality $y_m$ and pays $t$ for it, then his utility is given by $\theta_i y_m - t$. As we can always add objects with zero quality, it is without loss of generality to assume that $M = N$.

Let $\Pi$ denote the set of doubly sub-stochastic $N \times N$-matrices. An allocation rule $\alpha : [0, 1]^N \rightarrow \Pi$ represents a (possibly random) allocation of objects as a function of types.\footnote{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.} $\alpha_{ij}(\theta_i, \theta_{-i})$ denotes the probability with which agent $i$ obtains the object with quality $j$. We denote by $\alpha_i$ its $i$-th row vector, and also denote $\mathbf{y} = (y_1, \ldots, y_N)$. For an allocation $\alpha$ and for each $i$, let

$$\varphi_i(\theta_i) = \int_{\theta_i} \left[ \alpha_i(\theta_i, \theta_{-i}) \cdot \mathbf{y} \right] f_{-i}(\theta_{-i}) d\theta_{-i}. $$

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 $\alpha$ is part of a Bayesian incentive compatible mechanism if and only if each induced interim allocation $\varphi_i$ is non-decreasing.\footnote{See for example Gershkov and Moldovanu (2010) who use discrete majorization in a dynamic mechanism design framework with several qualities.}

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 incentive compatible allocations.\footnote{Note that, seen as a random variable, $s_i$ is uniformly distributed.}

Denote by $\alpha^* : [0, 1]^N \rightarrow \Pi$ the assortative matching of agents to objects (highest 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 \(\alpha^*\) induces the symmetric interim allocation given by

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

and the symmetric interim quantile allocation

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

The assortative matching allocation is incentive compatible.

### 5.1.1 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 \(\alpha\) 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 can be now formulated as:\(^{18}\)

**Theorem 4.** (Border 1991)\(^{19}\) Assume that there is only one object, i.e., \(y_N = 1\) and \(y_k = 0\) for \(k < N\). A symmetric and monotonic interim allocation \(\varphi\) is feasible if and only if the associated quantile interim allocation \(\psi(s) = \varphi(F^{-1}(s))\) satisfies

\[
\psi \prec_w s^{N-1}
\]

\(^{18}\)This is not the original formulation. See also Hart and Reny (2015) and Gershkov et al. (2019) for the one-object case, and for identical objects case, respectively. Hart and Reny’s proof is direct. Gerskov et al.’s proof use a result by Che at al. (2013) based on a network-flow approach.

\(^{19}\)See also Maskin and Riley (1984) and Matthews (1984).
Note that in the one-object case, the assortative matching interim allocation becomes 
\[ \varphi^*(\theta_i) = [F(\theta_i)]^{N-1}, \] 
the efficient allocation of the single available object, and hence \( \psi^*(s_i) = (s_i)^{N-1} \). The generalization to our present model is:

**Theorem 5.** In the ranked-items auction model, a symmetric and monotonic interim allocation rule \( \varphi \) is feasible if and only if its associated quantile interim allocation \( \psi(s) = \varphi(F^{-1}(s)) \) satisfies 
\[ \psi <_w \psi^* \]
where \( \psi^* \) is the quantile interim allocation generated by the assortative matching allocation.

**Proof:** We first show that \( \psi <_w \psi^* \) is necessary for feasibility. Consider a monotonic and symmetric quantile interim allocation rule \( \psi \) generated by \( \alpha \neq \alpha^* \). As switching to the assortative rule takes high-quality objects from lower types and gives them to higher types we have that 
\[ \mathbb{E}[\alpha_i(\theta) \cdot y | \theta_i \geq \tau] \leq \mathbb{E}[\alpha_i^*(\theta) \cdot y | \theta_i \geq \tau] \]
for each agent \( i \) and for every \( \tau \in [0, 1] \). Note that 
\[ \mathbb{E}[\alpha_i(\theta) \cdot y | \theta_i \geq \tau] = \frac{1}{1 - F(\tau)} \int_{\tau}^{1} \left[ \int_{[0,1]^{n-1}} \alpha_i(\theta, \theta_{-i}) \cdot y f_{-i}(\theta_{-i}) d\theta_{-i} \right] f(\theta_i) d\theta_i = \frac{1}{1 - F(\tau)} \int_{\tau}^{1} \varphi(\theta_i) f(\theta_i) d\theta_i = \frac{1}{1 - s} \int_{s}^{1} \psi(t_i) dt_i \]
where \( s = F(\tau) \). Since this holds for any \( \tau \in [0, 1] \), we obtain that \( \psi <_w \psi^* \).

For the converse, recall that, by Corollary 2, every extreme point \( \psi \) of \( \Omega_{m,w}(\psi^*) \) is described by \( \tilde{s}_i \in [0, 1] \) and by a collection of intervals \( [\underline{s}_i, \bar{s}_i] \subseteq [\tilde{s}_i, 1] \) such that 
\[ \psi(s_i) = \begin{cases} 
\psi^*(s_i) & \text{if } s_i \geq \tilde{s}_i \text{ and } s_i \notin \cup_{i \in I} [\underline{s}_i, \bar{s}_i] \\
\int_{\underline{s}_i}^{\bar{s}_i} \psi^*(t_i) dt_i & \text{if } s_i \in [\underline{s}_i, \bar{s}_i] \\
0 & \text{if } s_i < \tilde{s}_i 
\end{cases} \]
Any such extreme point is feasible as it is implemented by the allocation rule that does not allocate to types below \( \tilde{\theta}_i = F^{-1}(\tilde{s}_i) \), uses fair randomization to determine the allocation in each interval \( [\theta_i, \tilde{\theta}_i] = [F^{-1}(\underline{s}_i), F^{-1}(\bar{s}_i)] \), and is otherwise assortative. Formally, 
\[ \alpha_{ik}(\theta) = \begin{cases} 
\frac{1}{|\{j: m(\theta_j) > m(\theta_i)\}|} & \text{if } |\{j: m(\theta_j) < m(\theta_i)\}| \leq k - 1 \leq |\{j: m(\theta_j) \leq m(\theta_i)\}|, \\
0 & \text{else} 
\end{cases}, \quad (8) \]
where $m : [0, 1] \rightarrow [0, 1]$ equals

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

We note that the allocation $\alpha$ can be implemented in dominant strategies by assigning the objects assortatively outside of $\bigcup_{i \in I} [\theta_i, \bar{\theta}_i)$ and using random serial dictatorship to determine the allocation of objects among agent whose value lies in the same interval $[\theta_i, \bar{\theta}_i)$.

Let $P$ be the mapping that assigns to any allocation rule $\alpha$ its induced interim quantile allocation rule, and note that $P$ is a bounded linear operator. Also, let $T : \text{ext } \Omega_{m,w}(\psi^*) \rightarrow L^1([0,1]^N, \mathbb{R}^{N \times N})$ be a measurable function that assigns to any extreme point a corresponding allocation rule (which exists by Lemma 4 in the 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 $\mu$ supported on $\text{ext } \Omega_{m,w}(\psi^*)$ such that $\psi = \int_{\text{ext } \Omega_{m,w}(\psi^*)} \tilde{\psi} d\mu(\tilde{\psi})$. Let $\nu$ be the pushforward measure of $\mu$ under $T$\footnote{Defined by $\nu(B) = \mu(T^{-1}(B))$ for any Borel subset $B$.} and define the allocation

$$\alpha = \int \tilde{\alpha} d\nu(\tilde{\alpha}).$$

This allocation rule induces the interim quantile allocation $\psi$ because

$$P(\alpha) = \int P(\tilde{\alpha}) d\nu(\tilde{\alpha}) = \int P(T(\tilde{\psi})) d\mu(\tilde{\psi}) = \int \tilde{\psi} d\mu(\tilde{\psi}) = \psi.$$ 

The first equality follows from Lemma 11.45 in Aliprantis and Border (2006), the second by the change-of-variable formula for pushforward measures (Lemma III.10.8 in Dunford and Schwartz (1957)), the third since, by definition, $P(T(\tilde{\psi})) = \tilde{\psi}$, and the final equality follows from the definition of $\mu$. We conclude that any $\psi \in \Omega_{m,w}(\psi^*)$ is feasible, as desired.

### 5.1.2 BIC - DIC Equivalence

As another straightforward consequence of Proposition 1, we now derive an equivalence result between symmetric Bayesian Incentive Compatible (BIC) mechanisms and symmetric Dominant Strategy Incentive Compatible (DIC) mechanisms (see Manelli and Vincent (2010) for an analysis of the one-object auction case, and Gerhskov et al (2013) for general social...
Theorem 6. 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.

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

\[
\psi = \int_{\text{ext} \Omega_{m,w}(\psi^*)} \tilde{\psi} d\mu(\tilde{\psi}).
\]

For any \( \tilde{\psi} \) in \( \text{ext} \Omega_{m,w}(X^*) \) recall that \( T(\tilde{\psi}) = \alpha^{\tilde{\psi}} \) denotes the allocation that generates \( \tilde{\psi} \) as defined in (8). Note that \( \alpha^{\tilde{\psi}} \) can be chosen to be part of a symmetric, DIC mechanism (see Lemma 4 in the Appendix). In other words \( \alpha_i^{\tilde{\psi}} \) is symmetric, and for every \( i \) and for every \( \theta_{-i} \), the function \( \alpha_i^{\tilde{\psi}}(\theta_i, \theta_{-i}) \cdot y \) is monotonic in \( \theta_i \). Define then an allocation \( \alpha \) by

\[
\alpha = \int \tilde{\alpha} d\nu(\tilde{\alpha}).
\]

where \( \nu \) is the pushforward measure of \( \mu \) under \( T \). This allocation is a randomization over allocations belonging to DIC mechanisms, and hence \( \alpha_i(\theta_i, \theta_{-i}) \cdot y \) is itself monotonic, and thus part of a DIC mechanism. Moreover, this DIC mechanism generates an interim expected allocation equal to the original \( \psi \), yielding the wished equivalence.

Remark 1. An argument similar to the ones used in Theorem 5 and 6 can be used to show that for any convex objective function there exists an optimal mechanism that is non-randomized. This follows as there always exists an optimal allocation which is an extreme point and all extreme points can be implemented in non-randomized mechanisms.

5.1.3 The Revenue Maximizing Ranked-Item Auction

Consider an allocation \( \alpha \) that is part of an incentive compatible mechanism, i.e., the associated interim allocation \( \{\varphi_i\}_{i=1}^N \) are non-decreasing. Assume also that \( \alpha \) is individually

---

21Both 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.
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 $\alpha$ is

$$\int_{[0,1]^N} \sum_{i=1}^N \left( \theta_i - \frac{1 - F(\theta_i)}{f(\theta_i)} \right) [\alpha_i(\theta_i, \theta_{-i}) \cdot \mathbf{y}] f(\theta_1) \ldots f(\theta_N)d\theta_1 \ldots d\theta_N.$$  

For a symmetric mechanism, the above expression becomes

$$N \int_0^1 \left[ \theta_1 - \frac{1 - F(\theta_1)}{f(\theta_1)} \right] \varphi(\theta_1) f(\theta_1)d\theta_1 = N \int_0^1 \left[ F^{-1}(s_1) - \frac{1 - s_1}{f(F^{-1}(s_1))} \right] \psi(s_1) ds_1.$$  

Thus, by Theorem 5, the revenue maximization problem becomes

$$\max_{\psi} \int_0^1 \left[ F^{-1}(s_1) - \frac{1 - s_1}{f(F^{-1}(s_1))} \right] \psi(s_1) ds_1$$

s.t. $\psi \in \Omega_{m,w}(\psi^*)$

where $\psi^*$ is the interim quantile function induced by assortative matching. Corollary 2 shows that the optimal solution is an extreme point of $\Omega_{m,w}(\psi^*1[\hat{s}_1,1])$ for some $\hat{s}_1 \in [0, 1]$. 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 $\hat{\theta}_1 = F^{-1}(\hat{s}_1)$ must solve the equation $\theta_1 - \frac{1 - F(\theta_1)}{F'(\theta_1)} = 0$. Also, the objective function is then supermodular, and thus the Fan-Lorenz Theorem 3 immediately yields that the optimal allocation $\hat{\psi}$ satisfies\(^{22}\)

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

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.

### 5.1.4 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 $G$

\(^{22}\)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.
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 $\theta$ obtains prize $y$ and pays $t$ for it, then her utility is given by $\theta y - t$.\(^{23}\)

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

$$t(\theta) = \theta \varphi^*(\theta) - \int_0^\theta \varphi^*(s)ds$$

Note that 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\(^{24}\), 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 $\mu_G$. 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 $G_r \succ G \iff G_r^{-1} \prec G^{-1}$.

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 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

\(^{23}\)This standard formulation is easily generalized to other multiplicative, supermodular production functions and also (at least for some questions) to non-linear costs.

\(^{24}\)This follows from the famous rearrangement inequality of Hardy, Littlewood and Polya(1929))
in the contest and matching literature (see Hoppe, Moldovanu, Sela (2009), Damiano and Li (2007) and Olszewski and Siegel (2016)). These are obtained as immediate consequences of our theoretical insights:

**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 agent prefers random matching to any other scheme.\(^{25}\)

3. Random matching (assortative matching) maximizes the agents’ welfare if the distribution of types $F$ has an Increasing (Decreasing) Failure Rate.\(^{26}\)

4. If $F$ has an Increasing Failure Rate, the revenue (i.e., average bid) to a designer is maximized by assortative matching.\(^{27}\)

**Proof:** 1) This follows from the feasibility Theorem 5 and generalizes all matching schemes considered in the literature.\(^{28}\)

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

$$U(\theta) = \int_{0}^{\theta} G_{e}^{-1}(F(\tau)) d\tau$$

This is the standard payoff-equivalence result a la Myerson. Let us first maximize $U(1)$, the

\(^{25}\) $F$ 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 Hoppe, Moldovanu, Sela (HMS, 2009), who did not consider intermediate schemes. See also Olszewski and Siegel (2016)) 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 Hoppe et al (2009)).

\(^{26}\) 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.

\(^{27}\) See also Damiano and Li (2007).

\(^{28}\) See for example the schemes considered by Olszewski and Siegel (2016)) - 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.
utility of the highest type. Substituting $F(\theta) = s$, yields the problem

$$
\max_{G^{-1}_{ic}} \int_0^1 G^{-1}_{ic}(s)f(s)ds
$$

s.t. $G^{-1}_{ic} \prec G^{-1}$

We immediately obtain from the Fan-Lorentz Theorem 3 that the maximizer is $G^{-1}_{iv}(G^{-1})$ if the density $f$ is non-increasing (non-decreasing), i.e. if $F$ is convex (concave).\(^{29}\) 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

$$
\int_0^1 U(\theta)f(\theta)d\theta = \int_0^1 \left( \int_0^\theta G^{-1}_{ic}(F(\tau))d\tau \right) f(\theta)d\theta =
$$

$$
\int_0^1 G^{-1}_{ic}(F(\theta))(1 - F(\theta))d\theta = \int_0^1 G^{-1}_{iv}(s)(1 - s)dF^{-1}(s)
$$

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 $(1 - s)dF^{-1}(s) = 1$ 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.

5.2 Bayesian Persuasion 1

We consider here the persuasion problem studied by Dworczak and Martini (2019) (see also Kolotilin (2018)). Suppose that the state of the world $\omega$ is distributed according to

\(^{29}\)If the distribution of types $F$ is uniform, then the highest type is indifferent among all feasible schemes since his utility is $\int_0^1 G^{-1}_{ic}(s)ds = \mu_G$. 

22
a continuous distribution $F$ on the interval $[0, 1]$, and that an informed sender can reveal information about the state to an uninformed receiver.

The sender chooses a signal $\pi$ that consists of a signal realization space $S$ and a family of distributions $\{\pi_\omega\}$ over $S$. 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 expected value of the posterior, denoted here by $x$. The sender’s payoff $v$ is state independent and only depends on $x$.

Any signal is a “garbling” of the prior, and thus, for any signal $\pi$, the prior $F$ is a mean-preserving spread of the generated distribution of posterior means $G_\pi$, i.e. $G_\pi \succ F$. Conversely, it is well known that, for any $G$ such that $G \succ F$, there exists a signal $\pi$ such $G_\pi = G$.

Hence, formally, the sender’s problem is to choose a distribution $G$:

$$\max_G \int_0^1 v(x) dG(x)$$

s.t. $G \succ F$

Theorem 2 immediately implies that an optimal signal structure is a combination where either:

1. The state is perfectly revealed.
2. All states in an interval are pooled, so that only one signal is sent on this interval.
3. Two different signals are sent on an 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 will either reveal the state perfectly, or send the same signal for all states on an interval. While other information structures can be optimal,\textsuperscript{30} 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.

\textsuperscript{30}Gentzkow and Kamenica (2016) construct an example in which the optimal signal structure is not monotone partitional.
5.3 Bayesian Persuasion 2

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
\[
\int_0^1 s dG(s | x_0) = x_0.
\]

Denote by \( G \) the average distribution over posteriors induced for the receiver
\[
G(s) = \int G(s | x_0) dH(x_0).
\]

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
\[
\max_G \int_0^1 v(x) dG(x) \\
\text{s.t. } G \prec H
\]

In contrast to the previous problem, here agents are ex-ante informed and the sender can only generate more information. Theorem 1 immediately implies that an optimal signal structure is a combination where either:

1. The state is perfectly revealed.
2. Two different signals are sent on an interval.

5.4 Optimal Delegation

We now apply our results to a model of optimal delegation, variants of which have been analyzed, for example, by Holmström (1984), Melumad and Shibano (1991), Alonso and
Matouschek (2008), and Amador and Bagwell (2013).

The state of the world $x$ is distributed according to a density $\delta$ with support $[0, 1]$, and its realization is privately observed by the agent. A principal chooses an action $y \in Y$, where $Y \subset \mathbb{R}$ is compact. His utility function is given by $u(x, y)$, where $y$ is the chosen action and $x$ is the state of the world. We assume that $u$ is concave in $y$ for each $x$, and attains its maximum for each $x$. The agent’s utility from action $y$ in state $x$ is given by $u_A(y - y_A(x), x)$, where $y_A(x)$ is continuously differentiable and strictly increasing, and where $u_A(\cdot, x)$ is single-peaked and symmetric around 0 for each $x$. Using a normalization, we can assume that $y_A(x) \equiv x$.

The principal uses a direct, incentive-compatible and deterministic mechanism, $g : [0, 1] \to Y$, that maps the reported state to an action. Equivalently, this can be interpreted as the principal delegating the decision to the agent, but restricting the agent’s choice to a closed delegation set $D \subseteq Y$.

Formally, the principal’s problem is given by

$$
\max_g \int u(x, g(x)) \delta(x) dx \\
\text{s.t. } g \text{ being incentive-compatible}
$$

The following Lemma shows that any incentive-compatible mechanism majorizes a particularly simple mechanism. To state the result, we extend $g$ to a function

$$
g : [\min\{0, g(0)\}, \max\{1, g(1)\}] \to \mathbb{R}
$$

by setting $g(x) = g(0)$ for $x < 0$ and $g(x) = g(1)$ for $x > 1$.

**Lemma 2.** Suppose that $g(x)$ is incentive-compatible. Then $g$ is an extreme point of $\Phi_m(\overline{g})$ where

$$
\overline{g}(x) = \begin{cases} 
  g(0) & \text{if } x < g(0) \\
  x & \text{if } g(0) \leq x \leq g(1) \\
  g(1) & \text{if } g(1) < x.
\end{cases}
$$

**Proof:** Suppose that $g$ is incentive-compatible. Standard arguments imply that $g$ is non-decreasing. Now suppose that there exists $x$ such that $g(x) < \overline{g}(x)$, which implies $g(x) < x$. Let $\underline{s} = g(x)$ and $s^* = \sup\{s | g(s) = \underline{s}\}$. By incentive compatibility, $g(s) = \underline{s}$ for all $s \in [\underline{s}, s^*)$ and type $s^*$ must be indifferent between $\underline{s}$ and $\overline{s} = \lim_{s \downarrow s^*} g(s)$.\footnote{Note that $g$ is not constant on $[\underline{s}, 1]$ because $\underline{s} < \overline{g}(x) \leq g(1)$. Hence, $s^* < \max\{1, g(1)\}$ because if $g$ jumps at 1 incentive compatibility implies $g(1) > 1$. Therefore, $\lim_{s \downarrow s^*} g(s)$ is well-defined.} If $\underline{s} = \overline{s}$ then type $s^*$ has
an incentive to report a slightly higher type, and \( g \) is not incentive-compatible. Hence, \( s < \bar{s} \) and therefore \( s^* - s = \bar{s} - s^* = 1/2(\bar{s} - s) \). Incentive compatibility also implies that \( g(s) = \bar{s} \) for all \( s \in (s^*, \bar{s}] \) and hence

\[
\int_{s}^{\bar{s}} g(s)ds = (s^* - s)s + (\bar{s} - s^*)\bar{s} = 1/2\bar{s}^2 - 1/2s^2 = \int_{s}^{\bar{s}} sds = \int_{s}^{\bar{s}} \bar{g}(s)ds.
\]

Analogous arguments apply if there exists \( x \) such that \( g(x) > \bar{g}(x) \). We conclude that either \( g(x) = \bar{g}(x) \), or there exists an interval \([s, \bar{s}]\) such that \( g \) is piecewise-constant with one jump on this interval and \( \int_{s}^{\bar{s}} g(s)ds = \int_{s}^{\bar{s}} \bar{g}(s)ds \). Theorem 2 implies that \( g \) is an extreme point of \( \Phi_m(\bar{g}) \).

In particular, if \( g \) is incentive compatible, it is an extreme point for which all intervals \([y_i, \bar{y}_i)\) are empty, so \( g \) has two steps on each pooling interval. Incentive compatible rules share this feature with partitional signal structures, which we discussed in Section 5.2. This provides another perspective on the equivalence between monotone Bayesian persuasion and certain delegation problems, which is discussed in more detail in Kolotilin and Zapechelnyuk (2019).

A mechanism of the form

\[
\bar{g}(x) = \begin{cases} 
  a & \text{for } x < a \\
  x & \text{for } a \leq x \leq b \\
  b & \text{for } b < x
\end{cases}
\]

(9)

can be implemented by delegating the decision to the agent, and by allowing her to freely choose any action in \([a, b]\). These mechanisms are called interval delegation. Given their simplicity, such mechanisms seem particularly relevant for practical applications and have indeed been a focus of the relevant literature. Holmström (1984) solved for the optimal mechanism within the class of interval delegation mechanisms, and the later literature has identified conditions under which a particular interval delegation mechanism is optimal within the class of all incentive-compatible mechanisms (for example, Alonso and Matouschek (2008), Amador and Bagwell (2013)).

An application of the Fan-Lorentz inequality provides a simple sufficient condition for interval delegation to be optimal. We note here that the submodularity condition used below is stronger than needed for this result, but that it allows for a particularly simple proof.

**Corollary 3.** If \( u(x, y)\delta(x) \) is submodular in \((x, y)\) then interval delegation is optimal.
Proof: A solution to the problem always exists (Holmström, (1984)) and we denote it by \(g^\ast\). Let \(a = g^\ast(0), b = g^\ast(1)\), and let \(\bar{g}\) be defined as in (9). By Lemma 2, \(g^\ast\) is an extreme point of \(\Phi_m(\bar{g})\) and hence, \(g^\ast \succ \bar{g}\). Since \(u(x, y)\delta(x)\) is concave in \(y\) and submodular in \((x, y)\), the Fan-Lorentz inequality implies that
\[
\int u(x, \bar{g}(x))\delta(x)dx \geq \int u(x, g^\ast(x))\delta(x)dx.
\]
and we conclude that \(\bar{g}\) is optimal. \(\square\)

5.5 Decision-Making Under Uncertainty

The main purpose of this Section is to illustrate how our insights can be applied in order to understand how agents with non-expected utility preferences choose among risky prospects. All models below are very well-known.

5.5.1 Rank-Dependent Utility

Quiggin (1982) and Yaari (1987) axiomatically derived utility functionals with rank-dependent assessments of probabilities of the form\(^{32}\)
\[
U(F) = \int_0^1 v(t)d(g \circ F)(t)
\]
where \(F\) is the distribution of a random variable on the interval \([0, 1]\), \(v : [0, 1] \rightarrow R\) is continuous, strictly increasing and bounded, and where \(g : [0, 1] \rightarrow [0, 1]\) is strictly increasing, continuous and onto. The function \(v\) represents a transformation of monetary payoffs, while the function \(g\) represents a transformation of probabilities. For the sake of a brief treatment we assume below that both \(g\) an \(v\) are twice differentiable. But, aince the Fan-Lorentz result does not require differentiability, the observations below generalize.

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.\(^{33}\)

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

\(^{32}\)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.

\(^{33}\)Yaari (1987) works with the decumulative distribution function, so his probability transformations are required to be convex in order to yield risk aversion.
write:

\[ U(F) = \int_0^1 v(t) d(g \circ F)(t) = v(1) - \int_0^1 v'(t)(g \circ F)(t) dt \]

\[ = v(1) + \int_0^1 K(F(t), t) dt \]

where

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

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

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

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 \( U = \int_0^1 v(t) d(g \circ F)(t) \) is Schur-concave, and the agent whose preferences are represented by \( U \) is risk averse, exactly as under standard expected utility.

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).34

5.5.2 Choquet Capacities

Another important strand of the literature on non-expected utility considers ambiguity aversion.35 The main tool is the Choquet integral with respect to a (convex) capacity - note that this is unrelated to the Choquet representation used above.

Let \( f : [0, 1] \to [0, 1] \) be a convex, increasing function with \( f(0) = 0 \) and \( f(1) = 1 \), and let \( X \) be a bounded random variable. The Choquet integral of \( X \) with respect to the capacity

\[ 34 \text{If } g \text{ is convex (or linear) but } v \text{ is not necessarily concave, then our basic optimization problems treated above are still convex, and the corresponding maxima will be attained on an extreme point. If } g \text{ is concave, then the maximization problems are amenable to standard variational techniques.} \]

\[ 35 \text{See for example Schmeidler (1989)} \]
defined by $f$ is:

$$E_f(X) = \int_{-\infty}^{0} (f(\mathbb{P}[X > t]) - 1)dt + \int_{0}^{\infty} f(\mathbb{P}[X > t])dt$$

It can be shown that the above is equivalent to:

$$E_f(X) = \int_{0}^{1} f'(1-t)F_X^{-1}(t)dt$$

Letting

$$K(F^{-1}, t) = -F^{-1}f'(1-t)$$

we obtain that the kernel $K$ is linear in $F^{-1}$. Note also that $K$ is supermodular in $(F^{-1}, t)$ if and only if $f$ is convex. Hence, by the Fan-Lorentz Theorem (1954) $-E_f(X) \leq -E_f(Y)$ if $F_X^{-1} \prec F_Y^{-1}$. Since $F_X^{-1} \prec F_Y^{-1}$ is equivalent to $F_X \succ F_Y$, we obtain

$$E_f(X) \geq E_f(Y) \text{ if } F_X \succ F_Y$$

In other words, the Choquet integral yields a Schur-concave functional if and only if it is computed with respect to a convex capacity.

5.5.3 A Portfolio Choice Problem

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

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

where $Y$ and $Z$ are given random variables. $Y$ represents 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:

\[ \mathbb{E}[XY] \geq \int_0^1 F_Y^{-1}(1 - t)F_X^{-1}(t)dt \geq \int_0^1 F_Y^{-1}(1 - t)F_Z^{-1}(t)dt \]

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 by the Fan-Lorentz Theorem.

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

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

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

In other words, a decision maker that becomes more informed will bear a lower cost.

6 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 representative \( f \). For any \( g \in \Omega_m(f) \), let \( g \) be a non-decreasing representative that is left-continuous at 1. Then \( f(0) \leq g(x) \leq f(1) \) and the total variation of \( g \) is uniformly bounded by \( f(1) - f(0) \).

Helly’s Selection Theorem (see for example Kolmogorov and Fomin (1975)) 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 \( \int_x^1 g_n(s)ds \leq \int_x^1 f(s)ds \), we obtain that \( \int_x^1 g(s)ds \leq \int_x^1 f(s)ds \) 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 \( \Omega_{mw}(f) \) and \( \Phi_m(f) \).

It is clear from the definitions that the sets \( \Omega_m(f) \), \( \Omega_{mw}(f) \) 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 \( \mu \) that puts measure 1 on the extreme points of \( \Omega_m(f) \) such that \( g = \int h d\mu(h) \). The same

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

\[ 37 \text{ For more details on this problem see Dana(2005) and the literature cited there. Note that it does not use the Fan-Lorentz inequality.} \]
Preparations for the Proof of Theorem 1.

Fix \( g \in \Omega_m(f) \). Since \( f \) and \( g \) are non-decreasing, they contain non-decreasing and right-continuous representatives \( f \) and \( g \). 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

\[
u(s) := \text{median}\{g(s) - g(s_1), g(s) - g(s_2), y - g(s)\} \quad \text{for} \quad s \in [s_1, s_2] \quad \text{and} \quad \nu(s) = 0 \quad \text{else},
\]

and let \( u \in L^1(0, 1) \) denote the corresponding equivalence class.

**Lemma 3.**

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 \( \int_{s_1}^{s_2} u(s) \, ds = 0 \).

**Proof of Lemma 3:**

(1) Let

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

and

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

It follows that

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

---

\[38\] A nondecreasing function \( f : [0, 1] \to \mathbb{R} \) has at most countably many discontinuities and limits from the right are defined for each \( x \in [0, 1] \).
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 also 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 (10) 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 \(u \neq 0\).

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 \(u \neq 0\).

(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 \(u \neq 0\).

(4) In order to emphasize the fact that the definition of \(u\) in (10) 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, \(\int_0^1 u(s, y)ds\) 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 \(\int_0^1 u(s, y)ds = 0\).
Proof of Theorem 1: \(\Rightarrow\): Suppose that \(g\) is an extreme point, and let \(g\) be a non-decreasing and right-continuous representative. 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 (10) such that \(\int_{s_1}^{s_2} u(s)ds = 0\). 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, \(\int_{s_1}^{1} f(s) - g(s)ds \geq 0\) holds since \(f > g\). This implies that \(\int_{s}^{1} f(s) - (g \pm u)(s)ds \geq 0\) for all \(x\), and hence that \(g \pm u \in \Omega_m(f)\). Lemma 1 (ii) implies then that \(u \neq 0\), 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 (10) such that \(\int_{s_1}^{s_2} u(s)ds = 0\). Then \((g \pm u)(s) > f(s)\) holds on \([s_1, s_2]\). Since

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

we conclude that \(\int_{s}^{1} [f(s) - (g \pm u)(s)]ds \geq 0\) for all \(x\). Hence, \(g \pm u \in \Omega_m(f)\). Lemma 3 (ii) implies that \(u \neq 0\), 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 \(\{[x_i, \overline{x}_i]|i \in \mathcal{I}\}\) such that, for each \(i\), \(g(s) = g(x_i)\) for \(s \in [x_i, \overline{x}_i)\), \(g(s) < g(x_i)\) for \(s < x_i\), \(g(s) > g(x_i)\) for \(s > \overline{x}_i\), and \(f(x) = g(x)\) for \(x \neq 1\) with \(x \notin \bigcup_i [x_i, \overline{x}_i]\).

Suppose now that \(\int_{x_i}^{\overline{x}_i} (f(s) - g(s))ds < 0\) for some \(i \in \mathcal{I}\). This implies that \(\int_{x_i}^{1} (f(s) - g(s))ds > 0\) and, since \(g\) is constant on \([x_i, \overline{x}_i)\), that \(f(x_i) < g(x_i)\). If \(g(x_i^-) = g(x_i)\) we can choose \(s_2 = x_i\) and \(s_1 < s_2\) large enough such that \(u\) defined according to (10) satisfies \(g \pm u \in \Omega_m(f)\) and \(u \neq 0\), contradicting that \(g\) is an extreme point. Hence, \(g(x_i^-) < g(x_i)\).

Also, if \(g(s) > g(x_i)\) 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 (10) satisfies \(g \pm u \in \Omega_m(f)\) and \(u \neq 0\), contradicting that \(g\) is an extreme point. Hence, \(g\) is constant to the right of \(\overline{x}_i\). Let \(b = \sup \{x | g(x) = g(\overline{x}_i)\}\). There are two cases to consider:
We can therefore choose \( \varepsilon > 0 \) such that
\[
\varepsilon > \Omega u
\]
is a constant function that equals \( \Omega u \), contradicting the fact that \( g \) is an extreme point.

Case 2: \( \int_b^1 (f(x) - g(x)) \)ds > 0. Since, by assumption, \( \int_{\underline{x}_i}^{\overline{x}_i} (f(x) - g(x)) \)ds < 0 and \( \int_{\underline{x}_i}^{\overline{x}_i} (f(x) - g(x)) \)ds ≥ 0 are true, we obtain \( \int_{\underline{x}_i}^{\overline{x}_i} (f(x) - g(x)) \)ds > 0. This implies that \( \int_{\underline{x}_i}^{\overline{x}_i} (f(x) - g(x)) \)ds > 0, and hence that \( g(b^-) < g(b) \). Since \( \int_b^1 (f(s) - g(s)) \)ds = 0, \( f(b) > g(b) \) would imply \( \int_{b+\varepsilon}^1 (f(s) - g(s)) \)ds < 0 for \( \varepsilon > 0 \) small enough, which contradicts \( f > g \). Therefore, \( g(b^-) < f(b) \leq g(b) \). We can therefore choose \( \varepsilon > 0 \) and \( \delta > 0 \) such that
\[
\varepsilon > 0 \text{ for } \delta > 0 \text{ such that }
g \pm (\varepsilon 1_{[\underline{x}_i, \overline{x}_i]} - \delta 1_{[\underline{x}_i, b)}) \in \Omega_m(f)
\]
contradicting the fact that \( g \) is an extreme point.

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

Therefore, \( f \) is an extreme point of \( \Omega_m(f) \).

Proof of Corollary 1: Fix an arbitrary \( h \in \Omega_m(f) \) and define
\[
c(x) = \begin{cases} 
x & \text{if } x \notin \bigcup_{i \notin f}[\underline{x}_i, \overline{x}_i] \\
\overline{x}_i - x + \underline{x}_i & \text{if } x \in [\underline{x}_i, \overline{x}_i].
\end{cases}
\]
Moreover, let
\[
\bar{c}(x) = \begin{cases} 
  c(x) & \text{if } x \notin \bigcup_{i \in I} [x_i, \overline{x}_i) \\
  \int_{x_i}^{x} \frac{c(s)ds}{\overline{x}_i - x_i} & \text{if } x \in [x_i, \overline{x}_i) 
\end{cases}
\]
and
\[
\bar{h}(x) = \begin{cases} 
  h(x) & \text{if } x \notin \bigcup_{i \in I} [x_i, \overline{x}_i) \\
  \int_{x_i}^{x} \frac{h(s)ds}{\overline{x}_i - x_i} & \text{if } x \in [x_i, \overline{x}_i) 
\end{cases}
\]
It follows that
\[
\int_{0}^{1} (g(x) - h(x))c(x)dx \geq \int_{0}^{1} (g(x) - \bar{h}(x))c(x)dx = \int_{0}^{1} (g(x) - \bar{h}(x))\bar{c}(x)dx \tag{11}
\]
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 \(\bar{h} \prec f\). 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 (11) only if \(h\) is constant on each of the intervals \([x_i, \overline{x}_i)\) (see Fink and Jodeit (1984)), which yields \(h = \bar{h}\). Equality holds in (12) only if \(\int_{x}^{1} \bar{h}(s)ds = \int_{x}^{1} f(s)ds\) for all \(x \notin \bigcup_{i \in I} [x_i, \overline{x}_i)\). This implies that
\[
\bar{h}(x) = f(x) = g(x) \quad \text{for all } x \notin \bigcup_{i \in I} [x_i, \overline{x}_i) \text{ and}
\]
\[
\bar{h}(x) = \int_{x_i}^{x} \frac{f(s)ds}{\overline{x}_i - x_i} = g(x) \quad \text{for } x \in [x_i, \overline{x}_i)
\]
Therefore, \(g\) is the only point of \(\Omega_m(f)\) contained in the hyperplane. \(\square\)

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

It follows that, for any extreme point \(g\) of \(\Omega_{m,w}(f)\), there exists \(\theta \in [0,1]\) such that \(g \in \Omega_m(f 1_{[\theta,1]}).\) Since \(\Omega_m(f 1_{[\theta,1]}) \subseteq \Omega_{m,w}(f)\), \(g\) must be an extreme point of this set.
Conversely, suppose that \( g \) is an extreme point of \( \Omega_m(f \mathbb{1}_{[\theta,1]}) \) 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, \( g \pm u \in \Omega_{m,w}(f \mathbb{1}_{[\theta,1]}) \). Also, since

\[
\int_0^1 g(s)ds = \int_0^1 f(s)\mathbb{1}_{[\theta,1]}(s)dx
\]

we obtain \( \int_0^1 u(s)ds = 0 \). We conclude that

\[
\int_0^1 (g \pm u)(s)ds = \int_0^1 f(s)\mathbb{1}_{[\theta,1]}ds
\]

and therefore that \( g \pm u \in \Omega_m(f \mathbb{1}_{[\theta,1]}) \). 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:** \( \Rightarrow \): Let \( g \) be an extreme point of \( \Phi_m(f) \), and denote by \( g \) a non-decreasing and right-continuous representative that is left-continuous at \( x = 1 \). Also, let \( f \) be the continuous representative of \( f \).

**Step 1:** Fix any \( x \) such that \( g(x) < g(s) \) for all \( s > x \). If \( \int_x^1 (g(s) - f(s))ds > 0 \), then we can choose \( s_1 = x \) and \( s_2 > s_1 \) small enough such that \( u \) defined in (10) satisfies \( g \pm u \in \Phi_m(f) \) and \( u \not\equiv 0 \), a contradiction; hence, \( \int_x^1 (g(s) - f(s))ds \leq 0 \) for any such \( x \).

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

If \( g(x) < f(x) \), 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 \( \bigcup_j I_j \). Let

\[
Y = \{y \in \bigcup_j cl(I_j) \mid \int_y^1 (f(s) - g(s))ds = 0\}
\]

\(^{39}\)If \( x = 1 \) then left-continuity of \( f \) and \( g \) at \( 1 \) imply that there is \( \varepsilon > 0 \) such that \( \int_{1-\varepsilon}^1 f(s) - g(s)ds > 0 \), contradicting \( g \succ f \). Hence, \( x < 1 \).
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 \([\bar{x}_i, \bar{x}_i]\). We have

\[
\int_{\bar{x}_i}^1 (f(s) - g(s)) ds = 0, \quad \int_{\bar{x}_i}^1 (f(s) - g(s)) ds = 0, \quad \text{and} \\
\int_x^1 (f(s) - g(s)) ds < 0 \quad \text{for all } x \in (\bar{x}_i, \bar{x}_i) \quad \text{(since } g > f),
\]

and \( g \) is piece-wise constant on \([\bar{x}_i, \bar{x}_i]\).

We now prove that \( g \) consists of either two or three pieces on this interval. Indeed, if \([\bar{x}_i, \bar{x}_i]\) is partitioned into more than three intervals, then there are non-empty intervals \([a, b)\) and \([c, d)\) with \( a > \bar{x}_i \) and \( d < \bar{x}_i \) 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 \( \int_x^1 (f(s) - g(s)) ds \) 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 (10) satisfies \( g \pm u \in \Phi_m(f) \) and \( u \neq 0 \) (Lemma 3), a contradiction. Hence, \( g \) must have a discrete jump at \( x \in \{a, b, c, d\} \). But, this implies that we can choose \( \epsilon, \delta > 0 \) small enough such that \( u \) defined by

\[
u(s) = \delta 1_{[a, b)}(s) - \epsilon 1_{[c, d)}
\]

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

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

\[
\lim_{s \uparrow \bar{x}_i} g(s) \leq g(\bar{x}_i) \leq f(\bar{x}_i) \leq \lim_{s \uparrow \bar{x}_i} g(s)
\]

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

“\( \Leftarrow \)” Suppose that \( g \) satisfies the conditions in the statement of the theorem. Then \( g \) is non-decreasing and \( g > f \), hence \( g \in \Phi_m(f) \). Now suppose that \( u \in L^1 \) satisfies \( g \pm u \in \Phi_m(f) \), and let \( u \) be a representative such that \( g \pm u \) is non-decreasing. Then, for all \( i \) it must hold that

\[
\int_{\bar{x}_i}^1 u(s) ds = 0 \quad \text{and} \quad \int_{\bar{x}_i}^1 u(s) ds = 0.
\]
If \( i \in \mathcal{I} \) is such that
\[
g(x) = \begin{cases} 
  f(x_i) & \text{if } x \in [x_i, y_i] \\
  f(x) & \text{if } x \in [y_i, \bar{x}_i] 
\end{cases}
\]
then \( u \) is constant on \((x_i, y_i)\) and on \([y_i, \bar{x}_i]\). If \( x_i = 0 \) then \( u = 0 \) on \([x_i, y_i]\) since \( g(0) = f(0) \) and \((g \pm u)(0) \geq f(0)\). So suppose \( x_i > 0 \) and \( u < 0 \) on \([x_i, y_i]\) (and otherwise consider \(-u\)). Then
\[
(g + u)(x_i) < f(x_i)
\]
Since \( f \) is continuous and since \( g + u \) is non-decreasing, we obtain for some \( \varepsilon > 0 \) that
\[
\int_{x_i - \varepsilon}^{1} [(g + u)(s) - f(s)] ds < 0
\]
which yields a contradiction. Therefore, \( u = 0 \) on \([x_i, y_i]\) and since \( \int_{x_i}^{x_i} u(s) ds = 0 \) we obtain that \( u = 0 \) on \([x_i, \bar{x}_i]\).

On the other hand, if \( i \in \mathcal{I} \) is such that \( g \) satisfies
\[
g(x) = \begin{cases} 
  f(x_i) & \text{if } x \in [x_i, y_i] \\
  v_i & \text{if } x \in [y_i, \bar{y}_i] \\
  f(x) & \text{if } x \in [\bar{y}_i, \bar{x}_i] 
\end{cases}
\]
for some \( v_i \) then we can choose a representative \( u \in \mathfrak{u} \) which is constant on \([x_i, y_i]\), on \([y_i, \bar{y}_i]\) and on \([\bar{y}_i, \bar{x}_i]\). The same arguments as in the preceding paragraph imply that \( u = 0 \) on \([x_i, y_i]\). Now assume \( u < 0 \) on \([y_i, x_i]\) (otherwise consider \(-u\)). Then \((g + u)(y_i) < f(\bar{x}_i)\) and there exists \( \varepsilon > 0 \) such that
\[
\int_{x_i - \varepsilon}^{1} (g + u)(s) - f(s) ds < 0
\]
a contradiction. We conclude that \( u = 0 \) on \([\bar{y}_i, \bar{x}_i]\) and since \( \int_{x_i}^{x_i} u(s) ds = 0 \) we obtain that \( u = 0 \) on \([x_i, \bar{x}_i]\).

Observe that \( \int_{x_i}^{1} (f(s) - g(s)) ds = 0 \) for \( x \notin \bigcup_{i} [x_i, \bar{x}_i] \) and hence \( \int_{x_i}^{1} u(s) ds = 0 \) for \( x \notin \bigcup_{i} [x_i, \bar{x}_i] \). Since \( u(x) = 0 \) for all \( x \in \bigcup_{i} [x_i, \bar{x}_i] \), we conclude that \( \int_{x_i}^{1} u(s) ds = 0 \) for all \( x \in [0, 1] \), and therefore that \( u \equiv 0 \).

We need to show that Conditions (3), (4), (5) are equivalent to \( \int_{x_i}^{x_i} f(s) - g(s) ds = 0 \), \( v_i \in [f(y_i), f(\bar{y}_i)] \), \( g \succ f \), respectively.

We begin by showing that (3) is equivalent to \( \int_{x_i}^{x_i} f(s) - g(s) ds = 0 \). Plugging in the
definition of \( g \) yields that this condition is equivalent to

\[
0 = \int_{\overline{x}_i}^{\overline{y}_i} f(s) ds - f(x_i)(y_i - x_i) - f(\overline{x}_i)(\overline{y}_i - \overline{x}_i) - v_i(\overline{y}_i - y_i)
\]

and thus equivalent to (3).

We next show that (4) is equivalent to \( v_i \in [f(y_i), f(\overline{y}_i)] \). It follows from (3) that \( v_i \leq f(\overline{y}_i) \) is equivalent to

\[
\int_{\overline{x}_i}^{\overline{y}_i} f(s) ds - f(x_i)(y_i - x_i) - f(\overline{x}_i)(\overline{y}_i - \overline{x}_i) \leq f(\overline{y}_i)(\overline{y}_i - y_i)
\]

Adding \( f(x_i)(y_i - x_i) - f(\overline{x}_i)(\overline{y}_i - \overline{x}_i) \) yields

\[
\int_{\overline{x}_i}^{\overline{y}_i} f(s) ds \leq f(x_i)(y_i - x_i) + f(\overline{x}_i)(\overline{y}_i - \overline{x}_i).
\]

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

Finally, we show that (5) is equivalent to \( \mathbf{g} \succ \mathbf{f} \). As \( \int_{x}^{1} f(s) - g(s) ds = 0 \) for all \( x \notin \bigcup [\overline{x}_i, \overline{y}_i] \) it suffices to show that \( \int_{x}^{1} f(s) - g(s) ds \leq 0 \) for all \( x \in [\overline{x}_i, \overline{y}_i] \). Since \( f \) is continuous and since \( v_i \in [f(y_i), f(\overline{y}_i)] \), there exists a point \( m \in [y_i, \overline{y}_i] \) such that \( f(m) = v_i \). As \( g(x) \leq f(x) \) for \( x \in [x_i, y_i] \), we obtain for all \( x \in [x_i, y_i] \) that

\[
0 = \int_{x_i}^{1} f(s) - g(s) ds \geq \int_{x}^{1} f(s) - g(s) ds.
\]

Furthermore, as \( g(x) \geq f(x) \) for \( x \in [y_i, m] \) we get that \( \int_{x_i}^{1} f(s) - g(s) ds \leq 0 \) for all \( x \in [\overline{y}_i, \overline{x}_i] \). We thus have that

\[
\int_{x}^{1} f(s) - g(s) ds \leq \int_{m}^{1} f(s) - g(s) ds.
\]

A symmetric argument shows that the same conclusion holds for all \( x \in [m, \overline{y}_i] \) and that \( \int_{x}^{1} f(s) - g(s) ds \leq 0 \) for all \( x \in [\overline{y}_i, \overline{x}_i] \). We thus have that \( \int_{x}^{1} f(s) - g(s) ds \leq 0 \) for all \( x \in [x_i, \overline{x}_i] \) if and only if

\[
\int_{m}^{1} f(s) - g(s) ds \leq 0
\]

which is equivalent to (5). \( \square \)
Proof of Proposition 2: Note that \( \overline{C}(1) = C(1) \) since \( C \) is continuous. 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\(^{40}\)

\[
\int c(x)h(x)dx = C(1)h(1) - \int_0^1 C(x)dh(x) \leq \overline{C}(1)h(1) - \int_0^1 \overline{C}(x)dh(x) \quad (13)
\]

\[
= \int_0^1 \overline{C}'(x)h(x)dx \leq \int_0^1 \overline{C}'(x)f(x)dx, \quad (14)
\]

where the equalities follow from integration by parts, where the first inequality follows since \( \overline{C}(x) \leq C(x) \), and where the final inequality follows from the Fan-Lorentz Theorem 3 since \( \overline{C}' \) is non-decreasing.

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

\[
\int_0^1 C(x)dg(x) = \int_0^1 \overline{C}(x)dg(x);
\]

hence, (13) holds as an equality for \( h = g \). Also, since \( f(x) = g(x) \) for \( x \notin \bigcup_{i \in I} [x_i, \overline{x}_i] \), since \( \overline{C} \) is affine on \( [x_i, \overline{x}_i] \), and since \( g \) is constant on \( [x_i, \overline{x}_i] \) with \( g(x) = \int_{x_i}^{\overline{x}_i} f(s)ds \), we obtain

\[
\int_0^1 \overline{C}'(x)g(x)dx = \int_0^1 \overline{C}'(x)f(x)dx.
\]

Hence, setting \( h = g \) also satisfies (14) 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 (13) as an equality: Let \( \{ [y_j, \overline{y}_j] \mid j \in J \} \) be a minimal collection of intervals such that \( \overline{C} \) is affine on \( [y_j, \overline{y}_j] \) for each \( j \in J \) and such that \( \overline{C}(x) = C(x) \) for all \( x \notin \bigcup_{j \in J} [y_j, \overline{y}_j] \).

Define \( h \) to be constant on \( [y_j, \overline{y}_j] \) for each \( j \) with

\[
\int_{y_j}^{\overline{y}_j} h(s)ds = \int_{y_j}^{\overline{y}_j} f(s)ds,
\]

and set \( h(x) = f(x) \) for \( x \notin \bigcup_{j \in J} [y_j, \overline{y}_j] \). It follows from the previous step that \( h \) satisfies (13) and (14) with equality.

If \( \overline{C} \) is not affine on \( [x_i, \overline{x}_i] \) for some \( i \in I \), then \( \overline{C}' \) is non-decreasing and it is not constant on \( [x_i, \overline{x}_i] \). Since \( f \) is strictly increasing and \( g \) is constant on \( [x_i, \overline{x}_i] \), an application

\(^{40}\)Since \( \overline{C}(x) \) is convex, \( \overline{C}'(x) \) exists a.e. and we extend its definition by right-continuity to all \( x \).
of Chebyshev’s inequality (see Theorem 1 in Fink and Jodeit (1984)) yields
\[
\int_{x_i}^{x_j} dx \int_{x_i}^{x_j} C'(x)f(x)dx > \int_{x_i}^{x_j} f(x)dx \int_{x_i}^{x_j} C(x)dx
\]
\[
= \int_{x_i}^{x_j} g(x)dx \int_{x_i}^{x_j} C'(x)dx = \int_{x_i}^{x_j} dx \int_{x_i}^{x_j} g(x)C'(x)dx.
\]
Hence, \(g\) satisfies (13) with strict inequality, and therefore \(g\) cannot be optimal.

If \(C(x) < C(x)\) for some \(x \notin \bigcup_{i \in I}(x_i, x_i)\) then there is \(\varepsilon > 0\) such that \(C(z) < C(z)\) for all \(z \in [x, x + \varepsilon]\) and \(g(x) < g(x + \varepsilon)\) (since \(f\) is strictly increasing). Hence,
\[
\int_{x}^{x+\varepsilon} C(s)dg(x) < \int_{x}^{x+\varepsilon} C(s)dg(x)
\]
and \(g\) satisfies (13) as a strict inequality and therefore \(g\) cannot be optimal. \(\square\)

**Proof of Lemma 1:** For any \(h \in \Phi_m(f)\),
\[
\int_0^1 c(x)h(x)dx = C(1)h(1) - \int_0^1 C(x)dh(x)
\]
\[
\leq C(1)h(1) - \int_0^1 C(x)dh(x) = \int_0^1 C'(x)h(x)dx \leq \int_0^1 C'(x)f(x)dx,
\]
where the final inequality follows from the Fan-Lorentz inequality because \(C'(x)\) is non-increasing. Since \(g\) satisfies these inequalities with equality, we conclude that \(g\) is optimal. \(\square\)

**Lemma 4.** There is a measurable mapping \(T : \text{ext } \Omega_{mw}(\psi^*) \to L^1([0, 1]^N, \mathbb{R}^{N \times N})\) that assigns to any extreme point \(\psi\) a DIC allocation rule that implements \(\psi\).

**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 \(T\) 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 \(T\) 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 $T$. □

References


*Mathematical Finance* **15**(4), 613-634

Publishers.

*Journal of Political Economy*, **127**(5).


