COMPARATIVE STATICS, INFORMATIVENESS, AND
THE INTERVAL DOMINANCE ORDER

By John K.-H. Quah and Bruno Strulovici

Abstract: We identify a natural way of ordering functions, which we call the *interval dominance order*, and show that this concept is useful in the theory of monotone comparative statics and also in statistical decision theory. This ordering on functions is weaker than the standard one based on the single crossing property (Milgrom and Shannon, 1994) and so our monotone comparative statics results apply in some settings where the single crossing property does not hold. For example, they are useful when examining the comparative statics of optimal stopping time problems. We also show that certain basic results in statistical decision theory which are important in economics - specifically, the complete class theorem of Karlin and Rubin (1956) and the results connected with Lehmann's (1988) concept of informativeness - generalize to payoff functions that obey the interval dominance order.

Keywords: single crossing property, interval dominance order, supermodularity, comparative statics, optimal stopping time, complete class theorem, statistical decision theory, informativeness.

JEL Classification Numbers: C61, D11, D21, F11, G11.

Authors' Emails: john.quah@economics.ox.ac.uk bruno.strulovici@economics.ox.ac.uk

Acknowledgments: We would like to thank Ian Jewitt for many stimulating conversations. We also received many helpful comments from seminar and conference participants in Budapest, Exeter, Kos, Michigan, Northwestern, Oxford, Singapore, and Warwick. In particular, we would like to thank Rabah Amir, Alan Beggs, Eddie Dekel, Juan-Jose Ganuza, Paul Milgrom, Leonard Mirman, Andrea Patacconi, Herakles Polemarchakis, Edward Schlee, and Aleksey Tetenov.

1. Introduction

A basic problem in the theory of monotone comparative statics concerns the behavior of an optimal solution as the objective function changes.[1] Consider a family of real-valued functions $\{f(\cdot, s)\}_{s\in S}$, defined on the domain $X \subseteq R$ and parameterized by s in $S \subseteq R$. Under what conditions can we guarantee that[2]

$$\text{argmax}_{x\in X} f(x, s'') \geq \text{argmax}_{x\in X} f(x, s') \;\text{ whenever } s'' > s' \;? \tag{1}$$

In an influential paper, Milgrom and Shannon (1994) showed that (1) holds if the family of functions $\{f(\cdot, s)\}_{s\in S}$ obeys the *single crossing property* (SCP). Apart from guaranteeing (1), the single crossing property has other features that makes it an easily applicable concept for comparative statics, but it is not a necessary condition for (1). Indeed, Milgrom and Shannon show that it is necessary and sufficient to guarantee that

$$\text{for all } Y \subseteq X, \;\text{argmax}_{x\in Y} f(x, s'') \geq \text{argmax}_{x\in Y} f(x, s') \;\text{ whenever } s'' > s'. \tag{2}$$

This leaves open the possibility that there may be other useful concepts for comparative statics in situations where the modeler is principally interested in comparative statics on the domain X (rather than all subsets of X).

The first objective of this paper is to introduce a new way of ordering functions that is weaker than the single crossing property but is still sufficient to guarantee (1). We call this new order the *interval dominance order* (IDO). We show that the family $\{f(\cdot, s)\}_{s\in S}$ obeys the interval dominance order if and only if

$$\text{for all intervals } Y \subseteq X, \;\text{argmax}_{x\in Y} f(x, s'') \geq \text{argmax}_{x\in Y} f(x, s') \;\text{ whenever } s'' > s'. \tag{3}$$

[1] Early contributions to this literature include Topkis (1978), Milgrom and Roberts (1990), Vives (1990), and Milgrom and Shannon (1994). A textbook treatment can be found in Topkis (1998). Ashworth and Bueno de Mesquita (2006) discusses applications in political science.

[2] The sets in (1) are ordered according to the strong set order, which we define in Section 2. It reduces to the standard order on the real numbers when the sets are singletons.

It is clear from this characterization that IDO is weaker than SCP.[3]

For IDO to be useful in applications, it helps if there is a simple way of checking for the property. A sufficient condition for a family of differentiable functions $\{f(\cdot,s)\}_{s\in S}$ to obey IDO is that, for any two functions $f(\cdot,s'')$ and $f(\cdot,s')$ with $s''>s'$, there is a nondecreasing positive function h (which may depend on s' and s'') such that

$$\frac{df}{dx}(x,s'') \geq h(x)\frac{df}{dx}(x,s').$$

We give applications where the function h arises naturally.

An important feature of the single crossing property is that it is, in some sense, robust to the introduction of uncertainty. Suppose $\{f(\cdot,s)\}_{s\in S}$ is an SCP family and interpret s as the state of the world, which is unknown to the agent when he is choosing x. Assuming that the agent is an expected utility maximizer, he will choose x to maximize $\int_{s\in S} f(x,s)\,\lambda(s)ds$, where λ is his subjective probability over states. Since the optimal choice of x increases with s if s is known, one expects the agent's decision under uncertainty to have the same pattern, i.e., the optimally chosen x should increase when higher states are more likely. It turns out that SCP does indeed possess this comparative statics property (see Athey (2002)); formally,

$$\text{argmax}_{x\in X}\int_{s\in S} f(x,s)\gamma(s)ds \geq \text{argmax}_{x\in X}\int_{s\in S} f(x,s)\lambda(s)ds \tag{4}$$

whenever $\{f(\cdot,s)\}_{s\in S}$ obeys SCP and γ dominates λ by the monotone likelihood ratio.[4] An important feature of the interval dominance order is that, even though it is a weaker property than SCP, it is still robust to the introduction of uncertainty; i.e., (4) holds whenever $\{f(\cdot,s)\}_{s\in S}$ obeys IDO.

The second objective of this paper is to bridge the gap between the literature on monotone comparative statics and the closely related literature in statistical decision theory on informativeness. In that setting, the agent takes an action x (in X) before

[3] $Y\subset X$ is an *interval of* X if any $x\in X$ is also in Y whenever there is x' and x'' in Y such that $x'\leq x\leq x''$. For example, if $X=\{1,2,3\}$, then $\{1,2\}$ (but not $\{1,3\}$) is an interval of X.

[4] Note that γ dominates λ by the monotone likelihood ratio if $\gamma(s)/\lambda(s)$ is increasing in s.

the state is known but after observing a signal z (in $Z \subset R$) which conveys information on the true state. The *information structure* H refers to the family of distribution functions $\{H(\cdot|s)\}_{s\in S}$, where $H(\cdot|s)$ is the distribution of z conditional on the state s. Assuming that the agent has a prior given by the density function λ on S, the value of H for the payoff function f is $\mathcal{V}(H, f) \equiv \max_{\phi\in D} \left[\int_{s\in S} \int_{z\in Z} f(\phi(z), s) dH(z|s)\, \lambda(s) ds\right]$, where D is the set of all decision rules (which are maps from Z to X). So $\mathcal{V}(H, f)$ is the agent's ex ante expected payoff from an optimally chosen decision rule.

Lehmann identifies an intuitive condition under which H can be thought of as *more informative* than (another information structure) G. He goes on to show that if H is more informative than G, then

$$\mathcal{V}(H, f) \geq \mathcal{V}(G, f) \tag{5}$$

whenever f obeys the following property: $f(\cdot, s)$ is a quasiconcave function of the action x, and (1) holds. (In other words, the peaks of the quasiconcave functions $f(\cdot, s)$ are moving right with increasing s.) This restriction imposed on f by Lehmann implies that $\{f(\cdot, s)\}_{s\in S}$ obeys the interval dominance order, but significantly, it *need not* obey the single crossing property. We extend Lehmann's result by showing that, if H is more informative than G, then (5) holds whenever $\{f(\cdot, s)\}_{s\in S}$ obeys IDO (and thus, in particular, SCP). In this way, we have found a single condition on the payoff function that is useful for *both* comparative statics and comparative informativeness, so results in one category extend seamlessly into results in the other.[5]

Our final major result uses Lehmann's informativeness theorem to identify conditions under which a decision maker, including one who is not Bayesian, will pick a decision rule where the action is increasing in the signal. (In statistical terminology,

[5]Economic applications of Lehmann's concept of informativeness can be found in Persico (2000), Athey and Levin (2001), Levin (2001), Bergemann and Valimaki (2002), and Jewitt (2006). The distinction between Lehmann's restriction on the payoff function and the single crossing property was first highlighted in Jewitt (2006), which also discusses the significance of this distinction in economic applications.

increasing decision rules form an *essentially complete class.*) Our result generalizes the complete class theorem of Karlin and Rubin (1956). Karlin and Rubin assume that the payoff functions obey the same restriction as the one employed by Lehmann; we generalize it to an IDO class of payoff functions.

The paper contains several applications, which serve primarily to illustrate the use of the IDO property and the results relating to it, but may also be of interest in themselves. The following are the main applications. (1) We show that for any optimal stopping problem, a lower discount rate delays the optimal stopping time and raises the value of the problem. (2) The IDO property can also be used to examine a basic issue in optimal growth theory: we show that a lower discount rate leads to capital deepening, i.e., the optimal capital stock is higher at all times. Both of these results are shown under general conditions, providing significant extensions to existing results. (3) We illustrate the use of the informativeness results by applying them to a portfolio problem. Consider a group of investors who pool their funds with a single fund manager; the manager chooses a portfolio consisting of a safe and a risky asset, and each investor's return is proportional to their contribution to the fund. We show that a fund manager who is more informed than another in the sense of Lehmann can choose a portfolio that gives higher ex ante utility to *every* investor. (4) Finally, we consider a treatment response problem studied in Manski (2005). We show that our generalization of Karlin and Rubin's complete class theorem allows for more realistic payoff functions and fractional treatment rules.[6]

The paper is organized as follows. The next section introduces the interval dominance order and identifies some of its basic properties. Section 3 is devoted to applications. The IDO property for decision-making under uncertainty is studied in Section 4. In Section 5, we introduce Lehmann's notion of informativeness and extend his result while Section 6 proves a complete class theorem.

[6]The first two examples are found in Section 3, the third in Section 5, and the last in Section 6.

2. The Interval Dominance Order

Let X be a subset of the real line (denoted by R) and f and g two real-valued functions defined on X. We say that g dominates f by the *single crossing property* (which we denote by $g \succeq_{SC} f$) if for all x'' and x' such that $x'' > x'$, the following holds:

$$f(x'') - f(x') \geq (>)\, 0 \implies g(x'') - g(x') \geq (>)\, 0. \tag{6}$$

A family of real-valued functions $\{f(\cdot, s)\}_{s \in S}$, defined on X and parameterized by s in $S \subset R$ is referred to as an *SCP family* if the functions are ordered by the single crossing property (SCP), i.e., whenever $s'' > s'$, we have $f(\cdot, s'') \succeq_{SC} f(\cdot, s')$. Note that for any $x'' > x'$, the function $\Delta : S \to R$ defined by $\Delta(s) = f(x'', s) - f(x', s)$ crosses the horizontal axis at most once, which gives the motivation for the term 'single crossing'.

The crucial role played by the single crossing property when comparing the solutions to optimization problems was highlighted by Milgrom and Shannon (1994). Since the solution to an optimization problem is not necessarily unique, before we state their result, we must first define an ordering on sets. Let S' and S'' be two subsets of R. We say that S'' dominates S' in the *strong set order* (see Topkis (1998)), and write $S'' \geq S'$ if for any for x'' in S'' and x' in S', we have $\max\{x'', x'\}$ in S'' and $\min\{x'', x'\}$ in S'.[7] It follows immediately from this definition that if $S'' = \{x''\}$ and $S' = \{x'\}$, then $x'' \geq x'$. More generally, suppose that both sets contain their largest and smallest elements. Then it is clear that the largest (smallest) element in S'' is larger than the largest (smallest) element in S'.[8]

[7] Note that this definition of the strong set order makes sense on any lattice (see Topkis (1998)).

[8] Throughout this paper, when we say that something is 'greater' or 'increasing', we mean to say that it is greater or increasing in the *weak* sense. Most of the comparisons in this paper are weak, so this convention is convenient. When we are making a strict comparison, we shall say so explicitly, as in 'strictly higher', 'strictly increasing', etc.

Theorem (Milgrom and Shannon (1994)): *Suppose that f and g are real-valued functions defined on $X \subset R$. Then* $\text{argmax}_{x \in Y} g(x) \geq \text{argmax}_{x \in Y} f(x)$ *for any* $Y \subseteq X$ *if and only if* $g \succeq_{\text{SC}} f$.[9]

Note that the necessity of the single crossing property is obvious since we are requiring monotonicity of the optimal solution for *all* subsets Y of X. In particular, we can choose $Y = \{x', x''\}$, in which case $\text{argmax}_{x \in Y} g(x) \geq \text{argmax}_{x \in Y} f(x)$ implies (6). In fact, SCP is *not* necessary for monotone comparative statics if we only require $\text{argmax}_{x \in Y} g(x) \geq \text{argmax}_{x \in Y} f(x)$ for $Y = X$ or for Y belonging to a particular subcollection of the subsets of X. Consider Figures 1 and 2: in both cases, we have $\text{argmax}_{x \in X} g(x) \geq \text{argmax}_{x \in X} f(x)$; furthermore, $\text{argmax}_{x \in Y} g(x) \geq \text{argmax}_{x \in Y} f(x)$ where Y is any closed interval contained in X. In Figure 1, SCP is satisfied (specifically, (6) is satisfied) but this is not true in Figure 2. In Figure 2, we have $f(x') = f(x'')$ but $g(x'') < g(x')$, violating SCP.

This type of violation of SCP can arise naturally in an economic setting, as the following simple example shows. We shall return to this example at various points in the paper to illustrate our results.

Example 1. Consider a firm producing some good whose price we assume is fixed at 1 (either because of market conditions or for some regulatory reason). It has to decide on the production capacity (x) of its plant. Assume that a plant with production capacity x costs Dx, where $D > 0$. Let s be the state of the world, which we identify with the demand for the good. The unit cost of producing the good in state s is $c(s)$. We assume that, for all s, $D + c(s) < 1$. The firm makes its capacity decision *before* the state of the world is realized and its production decision *after* the state is revealed.

Suppose it chooses capacity x and the realized state of the world (and thus realized demand) is $s \geq x$. In this case, the firm should produce up to its capacity, so that its

[9]Milgrom and Shannon's result is situated in a lattice space. The theorem stated here is its one-dimensional analog.

profit $\Pi(x,s) = x - c(s)x - Dx$. On the other hand, if $s < x$, the firm will produce (and sell) s units of the good, giving it a profit of $\Pi(x,s) = s - c(s)s - Dx$. It is easy to see that $\Pi(\cdot, s)$ is increasing linearly for $x \leq s$ and thereafter declines, linearly with a slope of $-D$. Its maximum is achieved at $x = s$, with $\Pi(s,s) = (1 - c(s) - D)s$. Suppose $s'' > s'$ and $c(s'') > c(s')$; in other words, the state with higher demand also has higher unit cost. Then it is possible that $\Pi(s'', s'') < \Pi(s'', s')$; diagrammatically, this means that the peak of the $\Pi(\cdot, s'')$ curve (achieved at $x = s''$) lies below the $\Pi(\cdot, s')$ curve. If this occurs, we have the situation depicted in Figure 2, with $f(\cdot) = \Pi(\cdot, s')$ and $g(\cdot) = \Pi(\cdot, s'')$.

We wish to find a way of ordering functions that is useful for monotone comparative statics and also weaker than the single crossing property. In particular, we want the ordering to allow us to say that g dominates f (with respect to this new order) whenever both functions are quasiconcave and $\text{argmax}_{x \in X}\, g(x) \geq \text{argmax}_{x \in X}\, f(x)$ (as in Figure 2). To this end, it is useful to look again at Figure 2 and to notice that violations of (6) can only occur if we compare points x' and x'' on opposite sides of the maximum point of f. This suggests that a possible way of weakening SCP, while retaining comparative statics, is to require (6) to hold only for a certain collection of pairs $\{x', x''\}$, rather than all possible pairs.

The set J is an *interval of* X if, whenever x' and x'' are in J, any element x in X such that $x' \leq x \leq x''$ is also in J.[10] Let f and g be two real-valued functions defined on X. We say that g dominates f by the *interval dominance order* (or, for short, g I-dominates f, with the notation $g \succeq_I f$) if (6) holds for x'' and x' such that $x'' > x'$ and $f(x'') \geq f(x)$ for all x in the interval $[x', x''] = \{x \in X : x' \leq x \leq x''\}$. Clearly,

[10]Note that X need not be an interval in the conventional sense, i.e., X need not be, using our terminology, an *interval of* R. Furthermore, the fact that J is an interval of X does not imply that it is an interval of R. For example, if $X = \{1, 2, 3\}$ then $J = \{1, 2\}$ is an interval of X but of course neither X nor J are intervals of R.

the interval dominance order (IDO) is weaker than ordering by SCP. For example, in Figure 2, g I-dominates f but g does not dominate f by SCP.

In visualizing the relationship between f and g, it may be helpful to note that we could re-write the definition in the following manner: $g \succeq_I f$ if

$$x'' \in \text{argmax}_{x\in[x',x'']} f(x) \text{ (and } x' \notin \text{argmax}_{x\in[x',x'']} f(x))$$
$$\implies x'' \in \text{argmax}_{x\in[x',x'']} g(x) \text{ (and } x' \notin \text{argmax}_{x\in[x',x'']} g(x)).$$

With the definition presented in this manner, it should come as no surprise that $g \succeq_I f$ guarantees that $\text{argmax}_{x\in X} g(x) \geq \text{argmax}_{x\in X} f(x)$. We state this formally in Theorem 1, which gives the precise sense in which IDO is both sufficient and necessary for monotone comparative statics. For the necessity part of the result, we shall impose a mild regularity condition on the objective function. A function $f : X \to R$ is said to be *regular* if $\text{argmax}_{x\in[x',x'']} f(x)$ is nonempty for any points x' and x'' with $x'' > x'$. Suppose the set X is such that $X \cup [x', x'']$ is always closed, and thus compact, in R (with the respect to the Euclidean topology). This is true, for example, if X is finite, if it is closed, or if it is a (not necessarily closed) interval. Then f is regular if it is upper semi-continuous with respect to the relative topology on X.

THEOREM 1: *Suppose that f and g are real-valued functions defined on $X \subset R$ and $g \succeq_I f$. Then the following property holds:*

($\star$) $\text{argmax}_{x\in J} g(x) \geq \text{argmax}_{x\in J} f(x)$ *for any interval J of X.*

Furthermore, if property ($\star$) holds and g is regular, then $g \succeq_I f$.

Proof: Assume that g I-dominates f and that x'' is in $\text{argmax}_{x\in J} f(x)$ and x' is in $\text{argmax}_{x\in J} g(x)$. We need only consider the case where $x'' > x'$. Since x'' is in $\text{argmax}_{x\in J} f(x)$, we have $f(x'') \geq f(x)$ for all x in $[x', x''] \subseteq J$. Since $g \succeq_I f$, we also have $g(x'') \geq g(x')$; thus x'' is in $\text{argmax}_{x\in J} g(x)$. Furthermore, $f(x'') = f(x')$ so that x' is in $\text{argmax}_{x\in J} f(x)$. If not, $f(x'') > f(x')$ which implies (by the fact that $g \succeq_I f$) that $g(x'') > g(x')$, contradicting the assumption that g is maximized at x'.

To prove the other direction, we assume that there is an interval $[x', x'']$ such that $f(x'') \geq f(x)$ for all x in $[x', x'']$. This means that x'' is in $\text{argmax}_{x\in[x',x'']} f(x)$. There are two possible violations of IDO. One possibility is that $g(x') > g(x'')$; in this case, by the regularity of g, the set $\text{argmax}_{x\in[x',x'']}g(x)$ is nonempty but does not contain x'', which violates $(\star)$. Another possible violation of IDO occurs if $g(x'') = g(x')$ but $f(x'') > f(x')$. In this case, the set $\text{argmax}_{x\in[x',x'']}g(x)$ either contains x', which violates $(\star)$ since $\text{argmax}_{x\in[x',x'']}f(x)$ does not contain x', or it does not contain x'', which also violates $(\star)$. QED

While Theorem 1 is a straightforward result that follows easily from our definition of the interval dominance order, it is worth stating as a theorem because it provides the basic motivation for the concept. However, for the IDO concept to be genuinely useful, we need to demonstrate that it has other attractive features. For the purposes of application, it will be helpful if there is a simple way of checking that the property holds. The next result provides such a condition.

PROPOSITION 1: *Suppose that X is an interval of R, the functions f, $g : X \to R$ are absolutely continuous on compact intervals in X (and thus f and g are differentiable a.e.), and there is an increasing and positive function $\alpha : X \to R$ such that $g'(x) \geq \alpha(x)f'(x)$ a.e. Then $g \succeq_I f$. More specifically, if $f(x'') \geq f(x)$ for all $x \in [x'x'']$, then*

$$g(x'') - g(x') \geq \alpha(x')(f(x'') - f(x')). \tag{7}$$

If the function α in Proposition 1 is a constant $\bar{\alpha}$, then we obtain $g(x'') - g(x') \geq \bar{\alpha}(f(x'') - f(x'))$, which implies $g \succeq_{SC} f$. When α is not a constant, the functions f and g in Proposition 1 need not be related by SCP, as the following example shows.

Let $f : [0, M] \to R$ be a differentiable and quasiconcave function, with $f(0) = 0$ and a unique maximum at x^* in $(0, M)$. Let $\alpha : [0, M] \to R$ be given by $\alpha(x) = 1$ for $x \leq x^*$ and $\alpha(x) = 1 + (x - x^*)$ for $x > x^*$. Consider $g : [0, M]$ satisfying $g(0) = f(0) = 0$ with $g'(x) = \alpha(x)f'(x)$ (as in Proposition 1). Then it is clear that $g(x) = f(x)$ for $x \leq x^*$ and $g(x) < f(x)$ for $x > x^*$. In other words, g coincides with f

up to the point $x = x^*$; thereafter, g falls more steeply than f. The function g is also quasiconcave with a unique maximum at x^* and g I-dominates f (weakly), but g does not dominate f by SCP. To see this, choose x' and x'' on either side of x^* such that $f(x') = f(x'')$. Then we have a violation of (6) since $g(x') = f(x') = f(x'') > g(x'')$.

Proposition 1 is a consequence of the following lemma.

LEMMA 1: *Suppose $[x', x'']$ is a compact interval of R and α and h are real-valued functions defined on $[x', x'']$, with h integrable and α increasing (and thus integrable as well). If $\int_x^{x''} h(t)dt \geq 0$ for all x in $[x', x'']$, then*

$$\int_{x'}^{x''} \alpha(t)h(t)dt \geq \alpha(x') \int_{x'}^{x''} h(t)dt. \tag{8}$$

Proof: We confine ourselves to the case where α is an increasing and differentiable function. If we can establish (8) for such functions, then we can extend it to all increasing functions α since any such function can be approximated by an increasing and differentiable function.

The function $H(t) \equiv \alpha(t) \int_t^{x''} h(z)dz$ is absolutely continuous and thus differentiable a.e. By the fundamental theorem of calculus, $H(x'') - H(x') = \int_{x'}^{x''} H'(t)dt$; furthermore, $H(x') = \alpha(x') \int_{x'}^{x''} h(t)dt$, $H(x'') = 0$, and by the product rule,

$$H'(t) = \alpha'(t) \int_t^{x''} h(z)dz - \alpha(t)h(t). \tag{9}$$

Therefore,

$$\begin{aligned}
-\alpha(x') \int_{x'}^{x''} h(t)dt &= -H(x') \\
&= \int_{x'}^{x''} H'(t)\, dt \\
&= \int_{x'}^{x''} \alpha'(x) \left(\int_t^{x''} h(z)dz \right) dt - \int_{x'}^{x''} \alpha(t)h(t)\, dt,
\end{aligned}$$

where the last equality follows from (9). The term $\int_{x'}^{x''} \alpha'(x) \left(\int_t^{x''} h(z)dz \right) dt \geq 0$ by assumption and so (8) follows. QED

Proof of Proposition 1: Consider x'' and x' in X such that $x'' > x'$ and assume that $f(x) \leq f(x'')$ for all x in $[x', x'']$. Since f is absolutely continuous on $[x', x'']$, $f(x'') - f(x) = \int_x^{x''} f'(t)dt$ (with an analogous expression for g). We then have

$$\int_{x'}^{x''} g'(t)dt \geq \int_{x'}^{x''} \alpha(t) f'(t)dt \geq \alpha(x') \int_{x'}^{x''} f'(t)dt,$$

where the second inequality follows from Lemma 1. This leads to (7), from which we obtain $g(x'') \geq (>) g(x')$ if $f(x'') \geq (>) f(x')$. QED

In this section, we introduced the interval dominance order and showed how it is relevant to monotone comparative statics. The next section is devoted to applying the theoretical results obtained so far: we illustrate the use (and usefulness) of the IDO property through a number of simple examples. We shall focus on individual decision problems, though our methods can also be applied in a game-theoretic context.[11]

Our theory could be fruitfully developed in different directions. For example, the interval dominance order can be extended to a multidimensional setting. Notions like quasisupermodularity (Milgrom and Shannon (1994)) and $\mathcal{C}$-quasisupermodularity (Quah (2007)), which are important for multidimensional comparative statics, are essentially variations on the single crossing property; therefore, like the single crossing property, they can also be generalized. We explore these issues in a companion paper (see Quah and Strulovici (2007)).

In this paper, we shall instead focus on the development of the theory for decision-making under uncertainty. In particular, we show in Section 4 that the interval dominance order shares an important feature with the SCP order - in some precise

[11] Monotone comparative statics based on the single crossing property has been used, amongst other things, to guarantee that a player's strategy is increasing in the strategy of his opponent (see Milgrom and Roberts (1990)) and, in Bayesian games, to guarantee that a player's strategy is increasing in the signal he receives (Athey (2001)). An IDO-based theory can serve the same purpose and, since our results are more general, the restrictions on payoff functions to guarantee monotonicity are potentially less stringent.

sense, the order is preserved when one moves from a non-stochastic to a stochastic environment. Section 4 also provides the foundation for the results on informativeness covered in Sections 5 and 6. Readers who are keen on exploring those developments can skip the next section and go straight to Section 4.

3. Applications of the IDO property

Example 2. A very natural application of Proposition 1 is to the comparative statics of optimal stopping time problems. We consider a simple deterministic problem here; in Quah and Strulovici (2007) we show that the results in the next proposition extend naturally to a stochastic optimal stopping time problem.

Suppose we are interested in maximizing $V_\delta(x) = \int_0^x e^{-\delta t} u(t)dt$ for $x \geq 0$, where $\delta > 0$ and the function $u : R_+ \to R$ is bounded on compact intervals and measurable. So x may be interpreted as the stopping time, δ is the discount rate, $u(t)$ the cash flow or utility of cash flow at time t (which may be positive or negative), and $V_\delta(x)$ is the discounted sum of the cash flow (or its utility) when x is the stopping time.

We are interested in how the optimal stopping time changes with the discount rate. It seems natural that the optimal stopping time will rise as the discount rate δ falls. This intuition is correct but it cannot be proved by the methods of concave maximization since V_δ need not be a quasiconcave function. Indeed, it will have a turning point every time u changes sign and its local maxima occur when u changes sign from positive to negative. Changing the discount rate does not change the times at which local maxima are achieved, but it potentially changes the time at which the *global* maximum is achieved, i.e., it changes the optimal stopping time. The next result gives the solution to this problem.

Proposition 2: *Suppose that $\delta_1 > \delta_2 > 0$. Then the following holds:*
(i) $V_{\delta_2} \succeq_I V_{\delta_1}$*;*
(ii) $\text{argmax}_{x\geq 0} V_{\delta_2}(x) \geq \text{argmax}_{x\geq 0} V_{\delta_1}(x)$*; and*
(iii) $\max_{x\geq 0} V_{\delta_2}(x) \geq \max_{x\geq 0} V_{\delta_1}(x)$.

Proof: The functions V_{δ_2} and V_{δ_1} are absolutely continuous and thus differentiable a.e.; moreover,

$$V'_{\delta_2}(x) = e^{-\delta_2 x}u(x) = e^{(\delta_1-\delta_2)x}V'_{\delta_1}(x).$$

Note that the function $\alpha(x) = e^{(\delta_1-\delta_2)x}$ is increasing, so $V_{\delta_2} \succeq_I V_{\delta_1}$ (by Proposition 1) and (ii) follows from Theorem 1. For (iii), let us suppose that $V_{\delta_1}(x)$ is maximized at $x = x^*$. Then for all x in $[0, x^*]$, $V_{\delta_1}(x) \leq V_{\delta_1}(x^*)$. Note that $V_{\delta_1}(0) = V_{\delta_2}(0) = 0$ and $\alpha(0) = 1$. Thus, applying the inequality (7) (with $x' = 0$ and $x'' = x^*$) we obtain $V_{\delta_2}(x^*) \geq V_{\delta_1}(x^*)$. Since $\max_{x\geq 0} V_{\delta_2}(x) \geq V_{\delta_2}(x^*)$, we obtain (iii). QED

Arrow and Levhari (1969) has a version of Proposition 2(iii) (but not (ii)). They require u to be a continuous function; with this assumption, they show that the value function $\bar{V}$, defined by $\bar{V}(\delta) = \max_{x\geq 0} V_\delta(x)$ is right differentiable and has a negative derivative. This result is the crucial step (in their proof) guaranteeing the existence of a *unique* internal rate of return for an investment project, i.e., a unique δ^* such that $\bar{V}(\delta^*) = 0$. It is possible for us to extend and apply Proposition 2 to prove something along these lines, but we shall not do so in this paper.[12]

We should point out that we cannot strengthen Proposition 2(i) to say that $V_{\delta_2} \succeq_{SC} V_{\delta_1}$. To see this, suppose $u(t) = 1 - t$ and choose x' and x'' (with x' and x'' smaller and bigger than 1 respectively) such that $V_{\delta_1}(x') = V_{\delta_1}(x'')$. It follows from this that the function F, defined by $F(\delta) = V_\delta(x'') - V_\delta(x') = \int_{x'}^{x''} e^{-\delta t}(1-t)\,dt$ satisfies

$$F(\delta_1) = \int_{x'}^{x''} e^{-\delta_1 t}(1-t)\,dt = 0. \tag{10}$$

Note that $F'(\delta_1) = -\int_{x'}^{x''} e^{-\delta_1 t}t(1-t)\,dt > 0$ because of (10). Therefore, if δ_2 is close to and smaller than δ_1, we have $F(\delta_2) < 0$; equivalently, $V_{\delta_2}(x') > V_{\delta_2}(x'')$. So we obtain a violation of SCP.

Put another way, if $u(t) = 1 - t$ and the *only* stopping times available to the agent were x' and x'', then while the agent will be indifferent between them at $\delta = \delta_1$, she will strictly prefer x' at $\delta = \delta_2$ – so a lower discount rate leads to earlier stopping. On

[12]We would like to thank H. Polemarchakis for pointing out Arrow and Levhari's result.

the other hand, if she can stop at any time $x \geq 0$, then it is clear that, whatever the discount rate, it is optimal to stop just before u turns negative, i.e. at $x = 1$. This is reflected in Proposition 2.

Example 3. Consider a firm that chooses output x to maximize profit, given by $\Pi(x) = xP(x) - C(x)$, where P is the inverse demand function and C is the cost function. Imagine that there is a change in market conditions, so that both P and C are changed to $\tilde{P}$ and $\tilde{C}$ respectively. When can we say that $\text{argmax}_{x\geq 0}\tilde{\Pi}(x) \geq \text{argmax}_{x\geq 0}\Pi(x)$? By Theorem 1, this holds if $\tilde{\Pi}$ I-dominates Π. Intuitively, we expect this to hold if the increase in the inverse demand is greater than any increase in costs. This idea can be formalized in the following manner.

Assume that all the functions are differentiable, that $\tilde{P}$ and P take strictly positive values, and that the cost functions are strictly increasing. Define $a(x) = \tilde{P}(x)/P(x)$. Differentiating $\tilde{\Pi}(x) = a(x)xP(x) - \tilde{C}(x)$

$$\begin{aligned}\tilde{\Pi}'(x) &= a'(x)xP(x) + a(x)(xP(x))' - \tilde{C}'(x) \\ &\geq a(x)(xP(x))' - \left[\frac{\tilde{C}'(x)}{C'(x)}\right] C'(x),\end{aligned}$$

where the inequality holds because $xP(x) > 0$. Now suppose that

$$a(x) = \frac{\tilde{P}(x)}{P(x)} \geq \frac{\tilde{C}'(x)}{C'(x)}, \tag{11}$$

then we obtain $\tilde{\Pi}'(x) \geq a(x)\Pi'(x)$. By Proposition 1, $\tilde{\Pi}$ I-dominates Π if a is increasing and (11) holds; in other words, the ratio of the inverse demand functions is increasing in x and greater than the ratio of the marginal costs.[13]

Example 4. We wish to show, in the context of a standard optimal growth model, that lowering the discount rate of the representative agent leads to capital deepening: specifically, a higher capital stock at all times. Formally, the agent solves

$$\max\ U(c,k) = \int_0^\infty e^{-\delta s} u(c(s), k(s), s) ds \ \text{ subject to}$$

[13]Note that our argument does not require that P be decreasing in x.

(a) $\dot{k}(t) = H(c(t), k(t), t)$; (b) $k(t) \geq 0$ and $0 \leq c(t) \leq Q(k(t), t)$; and (c) $k(0) = k_0$.

The scalars $c(t)$ and $k(t)$ are the consumption and capital stock at time t respectively. Q is the production function, u the felicity function, and δ the discount rate. It is standard to have $H(c(t), k(t), t) = Q(k(t), t) - c(t) - \eta k(t)$ where $\eta \in (0, 1)$ is the (constant) rate of depreciation, but our result does not rely on this functional form. We also allow felicity to depend on consumption and capital (rather than just the former) and both the felicity and production functions may vary with t directly. In these respects, our model specification is more general than what is often assumed.

We say that capital is *beneficial* if the optimal value of $\int_{t_1}^{t_2} e^{-\delta s} u(c(s), k(s), s) ds$ subject to (a), (b), and the boundary conditions $k(t_1) = k_1$ and $k(t_2) = k_2$ is strictly increasing in k_1. In other words, raising the capital stock at t_1 strictly increases the utility achieved in the period $[t_1, t_2]$. Clearly, this is a mild condition which, in essence, is guaranteed if felicity strictly increases with consumption and production strictly increases with capital. This condition is all that is needed to guarantee that a lower discount rate leads to capital deepening.[14]

PROPOSITION 3: *Suppose that capital is beneficial and $(\bar{c}, \bar{k})$ and $(\hat{c}, \hat{k})$ are solutions to the optimal growth problem at discount rates $\bar{\delta}$ and $\hat{\delta}$ respectively. If $\hat{\delta} < \bar{\delta}$, then $\hat{k}(t) \geq \bar{k}(t)$ for all $t \geq 0$.*

Proof: First, observe that the function $F(\cdot, T, \hat{\delta})$ I-dominates $F(\cdot, T, \bar{\delta})$, where

$$F(t, T, \delta) \equiv \int_t^T e^{-\delta s} \left[u(\hat{c}(s), \hat{k}(s), s) - u(\bar{c}(s), \bar{k}(s), s) \right] ds \text{ with } \delta = \bar{\delta},\ \hat{\delta}. \qquad (12)$$

This follows from Proposition 1 since $F_t(t, T, \hat{\delta}) = e^{(\bar{\delta} - \hat{\delta})t} F_t(t, T, \bar{\delta})$.[15] In particular,

[14]The reader may consult Boyd and Becker (1997) for a discussion of other results (in particular, Amir (1996)) on the response of capital to the discount rate. Typically, these results are valid for multiple capital goods and require, amongst other things, the supermodularity of the value function. When specialized to the case of one capital good, those assumptions are still stronger than the ones made in Proposition 3.

[15]Note that we *cannot* conclude that $F(\cdot, T, \hat{\delta})$ SCP-dominates $F(\cdot, T, \bar{\delta})$ without more information on the integrand in (12). See the discussion at the end of Example 2.

suppose that for all t in some interval $[\underline{t}, T]$ we have $F(t, T, \bar{\delta}) \leq 0$. Since $F(T, T, \bar{\delta}) = 0$, we obtain by the I-dominance property that $F(\underline{t}, T, \hat{\delta}) \leq F(T, T, \hat{\delta}) = 0$. Indeed, we can say more. If $F(t, T, \bar{\delta}) < 0$ for t in a set of non-zero measure in $[\underline{t}, T]$, then $F(\underline{t}, T, \hat{\delta}) < 0$.[16]

Suppose that, contrary to our claim, there is some time t' at which $\bar{k}(t') > \hat{k}(t')$. Let $\underline{t}$ be the largest t below t' such that $\bar{k}(\underline{t}) = \hat{k}(\underline{t})$; the existence of $\underline{t}$ is guaranteed by the continuity of $\bar{k}$ and $\hat{k}$ and the fact that $\bar{k}(0) = \hat{k}(0)$. Let T be the *earliest* time after t' at which $\bar{k}(T) = \hat{k}(T)$. Set $T = \infty$ if no such time exists. So for t in the interval $[\underline{t}, T]$, we have $\bar{k}(t) \geq \hat{k}(t)$, with a strict inequality for t in $(\underline{t}, T)$. For a given t, denote by $(\tilde{c}, \tilde{k})$ the path that maximizes $\int_t^T e^{-\bar{\delta}s} u(c(s), k(s), s) ds$ subject to (a), (b), and the boundary conditions $k(t) = \hat{k}(t)$ and $k(T) = \bar{k}(T)$. We have

$$\int_t^T e^{-\bar{\delta}s} u(\bar{c}(s), \bar{k}(s), s) ds \geq \int_t^T e^{-\bar{\delta}s} u(\tilde{c}(s), \tilde{k}(s), s) ds \geq \int_t^T e^{-\bar{\delta}s} u(\hat{c}(s), \hat{k}(s), s) ds.$$

The second inequality follows from the optimality of $(\tilde{c}, \tilde{k})$. The first inequality is an equality if $t = \underline{t}$ and (by the fact that capital is beneficial) a strict inequality if t is in $(\underline{t}, T)$. So $F(t, T, \bar{\delta}) \leq 0$ for t in $[\underline{t}, T]$, with a strict inequality for t in $(\underline{t}, T)$. We have shown in the preceding paragraph that this implies that $F(\underline{t}, T, \hat{\delta}) < 0$. This is a contradiction because the optimality of $(\hat{c}, \hat{k})$ at the discount rate $\hat{\delta}$ means that it cannot accumulate strictly less utility over the interval $[\underline{t}, T]$ than $(\bar{c}, \bar{k})$. QED

4. The interval dominance order when the state is uncertain

Consider the following problem. Let $\{f(\cdot, s)\}_{s \in S}$ be a family of functions parameterized by s in S, an interval of R, with each function $f(\cdot, s)$ mapping Y, an interval of R, to R. Assume that all the functions are quasiconcave, with their peaks increasing in s; by this we mean that $\text{argmax}_{x \in Y} f(x, s'') \geq \text{argmax}_{x \in Y} f(x, s')$ whenever $s'' > s'$.

[16]We can see this by re-tracing the proof of Lemma 1. In Lemma 1, the inequality (3) is strict if (a) $\int_x^{x''} h(t) dt > 0$ for x in a set with positive measure in (x', x'') and (b) α is a strictly increasing function. Note that in this application, $\alpha(t) = e^{(\bar{\delta} - \hat{\delta})t}$, which is strictly increasing in t.

(Since $f(\cdot, s)$ is quasiconcave, its set of maximizers is either a singleton or an interval.) We shall refer to such a family of functions as a QCIP family, where QCIP stands for *quasiconcave with increasing peaks*. Interpreting s to be the state of the world, an agent has to choose x under uncertainty, i.e., before s is realized. We assume the agent maximizes the expected value of his objective; formally, he maximizes

$$F(x, \lambda) = \int_{s \in S} f(x, s)\lambda(s) ds,$$

where $\lambda : S \to R$ is the density function defined over the states of the world. It is natural to think that if the agent considers the higher states to be more likely, then his optimal value of x will increase. Is this true? More generally, we can ask the same question if the functions $\{f(\cdot, s)\}_{s \in S}$ form an IDO family, i.e., a family of regular functions $f(\cdot, s) : X \to R$, with $X \subseteq R$, such that $f(\cdot, s'')$ I-dominates $f(\cdot, s')$ whenever $s'' > s'$.

One way of formalizing the notion that higher states are more likely is via the monotone likelihood ratio (MLR) property. Let λ and γ be two density functions defined on the interval S of R and assume that $\lambda(s) > 0$ for s in S. We call γ an *MLR shift* of λ if $\gamma(s)/\lambda(s)$ is increasing in s. For density changes of this kind, there are two results that come close, though not quite, to addressing the problem we posed.

Ormiston and Schlee (1993) identify some conditions under which an MLR shift in the density function will raise the agent's optimal choice. Amongst other conditions, they assume that $F(\cdot; \lambda)$ is quasiconcave. This will hold if all the functions in the family $\{f(\cdot, s)\}_{s \in S}$ are concave but will not generally hold if the functions are just quasiconcave. Athey (2002) has a related result which says that an MLR shift will lead to a higher optimal choice of x provided $\{f(\cdot, s)\}_{s \in S}$ is an SCP family. As we had already pointed out in Example 1, a QCIP family need not be an SCP family.

The next result gives the solution to the problem we posed.

THEOREM 2: *Let S be an interval of R and $\{f(\cdot, s)\}_{s \in S}$ be an IDO family. Then $F(\cdot, \gamma) \succeq_I F(\cdot, \lambda)$ if γ is an MLR shift of λ. Consequently,* $\text{argmax}_{x \in X} F(x, \gamma) \geq$

$\mathrm{argmax}_{x\in X} F(x,\lambda)$.

Notice that since $\{f(\cdot,s)\}_{s\in S}$ in Theorem 2 is assumed to be an IDO family, we know (from Theorem 1) that $\mathrm{argmax}_{x\in X} f(x,s'') \geq \mathrm{argmax}_{x\in X} f(x,s')$. Thus Theorem 2 guarantees that the comparative statics which holds when s is known also holds when s is unknown but experiences an MLR shift.[17,18]

The proof of Theorem 2 requires a lemma (stated below). Its motivation arises from the observation that if $g \succeq_{SC} f$, then for any $x'' > x'$ such that $g(x') - g(x'') \geq (>)\, 0$, we must also have $f(x') - f(x'') \geq (>)\, 0$. Lemma 2 is the (less trivial) analog of this observation in the case when $g \succeq_I f$.

Lemma 2: *Let X be a subset of R and f and g two regular functions defined on X. Then $g \succeq_I f$ if and only if the following property holds:*

(M) if $g(x') \geq g(x)$ for x in $[x', x'']$ then

$$g(x') - g(x'') \geq (>)\, 0 \implies f(x') - f(x'') \geq (>)\, 0.$$

Proof: Suppose $x' < x''$ and $g(x') \geq g(x)$ for x in $[x', x'']$. There are two possible ways for property (M) to be violated. One possibility is that $f(x'') > f(x')$. By regularity, we know that $\mathrm{argmax}_{x\in[x',x'']} f(x)$ is nonempty; choosing x^* in this set, we have $f(x^*) \geq f(x)$ for all x in $[x', x^*]$, with $f(x^*) \geq f(x'') > f(x')$. Since $g \succeq_I f$, we must have $g(x^*) > g(x')$, which is a contradiction.

The other possible violation of (M) occurs if $g(x') > g(x'')$ but $f(x') = f(x'')$. By regularity, we know that $\mathrm{argmax}_{x\in[x',x'']} f(x)$ is nonempty, and if f is maximized at x^* with $f(x^*) > f(x')$, then we are back to the case considered above. So assume that x' and x'' are both in $\mathrm{argmax}_{x\in[x',x'']} f(x)$. Since $f \succeq_I g$, we must have $g(x'') \geq g(x')$, contradicting our initial assumption.

[17] We are echoing an observation that was made by Athey (2002) in a similar context.

[18] Apart from being interesting in itself, the monotonicity property guaranteed by Theorem 2 can play a crucial role in establishing the existence of a pure strategy equilibrium in certain Bayesian games (see Athey (2001)).

So we have shown that (M) holds if $g \succeq_I f$. The proof that (M) implies $g \succeq_I f$ is similar. QED

Proof of Theorem 2: This consists of two parts. Firstly, we prove that if $F(x'', \lambda) \geq F(x, \lambda)$ for all x in $[x', x'']$, then, for any $\tilde{s}$ in S,

$$\int_{\tilde{s}}^{s^*} \left(f(x'', s) - f(x', s)\right) \lambda(s) ds \geq 0 \tag{13}$$

(where s^* denotes the supremum of S). Assume instead that there is $\bar{s}$ such that

$$\int_{\bar{s}}^{s^*} \left(f(x'', s) - f(x', s)\right) \lambda(s) ds < 0. \tag{14}$$

By the regularity of $f(\cdot, \bar{s})$, there is $\bar{x}$ that maximizes f in $[x', x'']$. In particular, $f(\bar{x}, \bar{s}) \geq f(x, \bar{s})$ for all x in $[\bar{x}, x'']$. Since $\{f(\cdot, s)\}_{s \in S}$ is an IDO family of regular functions, we also have $f(\bar{x}, s) \geq f(x'', s)$ for all $s \leq \bar{s}$ (using Lemma 2). Thus

$$\int_{s_*}^{\bar{s}} \left(f(\bar{x}, s) - f(x'', s)\right) \lambda(s) ds \geq 0, \tag{15}$$

where s_* is the infimum of S. Notice also that $f(\bar{x}, \bar{s}) \geq f(x, \bar{s})$ for all x in $[x', \bar{x}]$, which implies that $f(\bar{x}, s) \geq f(x', s)$ for all $s \geq \bar{s}$. Aggregating across s we obtain

$$\int_{\bar{s}}^{s^*} \left(f(\bar{x}, s) - f(x', s)\right) \lambda(s) ds \geq 0. \tag{16}$$

It follows from (14) and (16) that

$$\begin{aligned} \int_{\bar{s}}^{s^*} \left(f(\bar{x}, s) - f(x'', s)\right) \lambda(s) ds &= \int_{\bar{s}}^{s^*} \left(f(\bar{x}, s) - f(x', s)\right) \lambda(s) ds \\ &\quad + \int_{\bar{s}}^{s^*} \left(f(x', s) - f(x'', s)\right) \lambda(s) ds \\ &> 0. \end{aligned}$$

Combining this with (15), we obtain $\int_{s_*}^{s^*} \left(f(\bar{x}, s) - f(x'', s)\right) \lambda(s) ds > 0$; in other words, $F(\bar{x}, \lambda) > F(x'', \lambda)$, which is a contradiction.

Given (13), the function $H(\cdot, \lambda) : [s_*, s^*] \to R$ defined by

$$H(\tilde{s}, \lambda) = \int_{s_*}^{\tilde{s}} \left(f(x'', s) - f(x', s)\right) \gamma(s) ds$$

satisfies $H(s^*, \lambda) \geq H(\tilde{s}, \lambda)$ for all $\tilde{s}$ in $[s_*, s^*]$. Defining $H(\cdot, \gamma)$ in an analogous fashion, we also have $H'(\tilde{s}, \gamma) = [\gamma(s)/\lambda(s)]H'(\tilde{s}, \lambda)$ for $\tilde{s}$ in S. Since γ is an upward MLR shift of λ, the ratio $\gamma(s)/\lambda(s)$ is increasing in s. By Proposition 1, $H(\cdot, \gamma) \succeq_I H(\cdot, \lambda)$. In particular, we have $H(s^*, \gamma) \geq (>) H(s_*, \gamma) = 0$ if $H(s^*, \lambda) \geq (>) H(s_*, \lambda) = 0$. Re-writing this, we have $F(x'', \gamma) \geq (>) F(x', \gamma)$ if $F(x'', \lambda) \geq (>) F(x', \lambda)$. QED

Note that Theorem 2 remains true if S is not an interval; in Appendix A, we prove Theorem 2 in the case where S is a finite set of states. We also show in that appendix that the MLR condition in Theorem 2 cannot be weakened in the following sense: if γ is *not* an MLR shift of λ, then there is an SCP (hence IDO) family $\{f(\cdot, s)\}_{s \in S}$ such that $\text{argmax}_{x \in X} F(x, \gamma) < \text{argmax}_{x \in X} F(x, \lambda)$.

We turn now to two simple applications of Theorem 2.

Example 1 continued. Recall that in state s, the firm's profit is $\Pi(x, s)$. It achieves its maximum at $x^*(s) = s$, with $\Pi(s, s) = (1 - c(s) - D)s$, which is strictly positive by assumption. The firm has to choose its capacity before the state of the world is realized; we assume that s is drawn from S, an interval in R, and has a distribution given by the density function $\lambda : S \to R$. We can think of the firm as maximizing its expected profit, which is $\int_S \Pi(x, s)\lambda(s)ds$, or more generally, its expected utility from profit which is $U(x, \lambda) = \int_S u(\Pi(x, s), s)\lambda(s)ds$, where, for each s, the function $u(\cdot, s) : R \to R$ is strictly increasing. The family $\{u(\Pi(\cdot, s), s)\}_{s \in S}$ consists of quasiconcave functions, the peaks of which are increasing in s. This is an IDO family. By Theorem 2, we know that an upward MLR shift of the density function will lead the firm to choose a greater capacity.

Example 5. Consider a firm that has to decide on when to launch a new product. The more time the firm gives itself, the more it can improve the quality of the product and its manufacturing process, but it also knows that there is a rival about to launch a similar product. In formal terms, we assume that the firm's profit (if it is not anticipated by its rival) is an increasing function of time $\phi : R_+ \to R_+$. If the rival launches its product at time s, then the firm's profit falls to $w(s)$ (in R). In other words, the firm's profit function in state s, denoted by $\pi(\cdot, s)$ satisfies $\pi(t,s) = \phi(t)$ for $t \leq s$ and $\pi(t,s) = w(s)$ for $t > s$, where $w(s) < \phi(s)$. Clearly, each $\pi(\cdot, s)$ is a quasiconcave function and $\{\pi(\cdot, s)\}_{s\in S}$ is an IDO family. The firm decides on the launch date t by maximizing $F(t,\lambda) = \int_{s\in S} \pi(t,s)\lambda(s)ds$, where $\lambda : R_+ \to R$ is the density function over s. By Theorem 2, if the firm thinks that it is less likely that the rival will launch early, in the sense that there is an MLR shift in the density function, then it will decide on a later launch date.

Note that IDO imposes no restrictions on the function $w : R_+ \to R$, which gives the firm's profit should it be anticipated by its rival at time s. One can check that $\{\pi(\cdot, s)\}_{s\in S}$ is an SCP family if (and only if) w is an increasing function of s, but Theorem 2 gives us the desired conclusion without making this stronger assumption.

5. Comparing Information Structures[19]

Consider an agent who, as in the previous section, has to make a decision before the state of the world (s) is realized, where the set of possible states S is a subset of R. Suppose that, before he makes his decision, the agent observes a signal z. This signal is potentially informative of the true state of the world; we refer to the collection $\{H(\cdot|s)\}_{s\in S}$, where $H(\cdot|s)$ is the distribution of the signal z conditional on s, as the *information structure* of the decision maker's problem. (Whenever convenient, we shall simply call this information structure H.) We assume that, for every s, $H(\cdot|s)$ admits a density function and has the compact interval $Z \subset R$ as its support. We say that H is *MLR-ordered* if the density function of $H(\cdot|s'')$ is an MLR shift of the density function of $H(\cdot|s')$ whenever $s'' > s'$.

We assume that the agent has a prior distribution Λ on S. We allow either of the following: (i) S is a compact interval and Λ admits a density function with S as its support or (ii) S is finite and Λ has S as its support. The agent's *decision rule* (under

[19]We are very grateful to Ian Jewitt for introducing us to the literature in this section and the next and for extensive discussions.

H) is a measurable map from Z to the set of actions X (contained in R). Denoting the utility of action x in state s by $u(x,s)$, the decision rule $\phi : Z \to X$ gives the ex ante expected utility $\mathcal{U}(\phi, H, \Lambda) = \int_S \int_Z u(\phi(z), s) dH(z|s)\, d\Lambda(s)$.

We denote the agent's posterior distribution (on S) upon observing z by Λ_H^z. The ex ante utility of an agent with the decision rule $\phi : Z \to X$ may be equivalently written as

$$\mathcal{U}(\phi, H, \Lambda) = \int_{z \in Z} \left[\int_{s \in S} u(\phi(z), s) d\Lambda_H^z \right] dM_{H,\Lambda}$$

where $M_{H,\Lambda}$ is the marginal distribution of z. A decision rule $\hat{\phi} : Z \to X$ that maximizes the agent's (posterior) expected utility at each realized signal is called an *H-optimal* decision rule. We assume that X is compact and that $\{u(\cdot, s)\}_{s \in S}$ is an equicontinuous family. This guarantees that the map from x to $\int_S u(x,s) d\Lambda_H^z(s)$ is continuous, so the problem $\max_{x \in X} \int_S u(x,s) d\Lambda_H^z(s)$ has a solution at every value of z; in other words, $\hat{\phi}$ exists. The agent's ex ante utility using such a rule is denoted by $\mathcal{V}(H, \Lambda, u)$.

Consider now an alternative information structure given by $G = \{G(\cdot|s)\}_{s \in S}$; we assume that $G(\cdot|s)$ admits a density function and has the compact interval Z as its support. What conditions will guarantee that the information structure H is more favorable than G in the sense of offering the agent a higher ex ante utility; in other words, how can we guarantee that $\mathcal{V}(H, \Lambda, u) \geq \mathcal{V}(G, \Lambda, u)$?

It is well known that this holds if H is *more informative* than G according to the criterion developed by Blackwell (1953); furthermore, this criterion is also necessary if one does not impose significant restrictions on u (see Blackwell (1953) or, for a recent textbook treatment, Gollier (2001)). We wish instead to consider the case where a significant restriction *is* imposed on u; specifically, we assume that $\{u(\cdot, s)\}_{s \in S}$ is an IDO family. We show that, in this context, a different notion of informativeness due to Lehmann (1988) is the appropriate concept.[20]

[20] Jewitt (2006) gives the precise sense in which Lehmann's concept is weaker than Blackwell's (see also Lehmann (1988) and Persico (2000)) and also discusses its relationship with the concept

Our assumptions on H guarantee that, for any s, $H(\cdot|s)$ admits a density function with support Z; therefore, for any (z, s) in $Z \times S$, there exists a unique element in Z, which we denote by $T(z, s)$, such that $H(T(z, s)|s) = G(z|s)$. We say that H is *more accurate* than G if T is an increasing function of s.[21] Our goal in this section is to prove the following result.

THEOREM 3: *Suppose $\{u(\cdot, s)\}_{s \in S}$ is an IDO family, G is MLR-ordered, and Λ is the agent's prior distribution on S. If H is more accurate than G, then*

$$\mathcal{V}(H, \Lambda, u) \geq \mathcal{V}(G, \Lambda, u). \tag{17}$$

This theorem generalizes a number of earlier results. Lehmann (1988) establishes a special case of Theorem 3 in which $\{u(\cdot, s)\}_{s \in S}$ is a QCIP family. Persico (1996) has a version of Theorem 3 in which $\{u(\cdot, s)\}_{s \in S}$ is an SCP family, but he requires the optimal decision rule to vary smoothly with the signal, a property that is not generally true without the *sufficiency* of the first order conditions for optimality. Jewitt (2006) proves Theorem 3 for the general SCP case.[22]

To prove Theorem 3, we first note that if G is MLR-ordered, then the family of posterior distributions $\{\Lambda_G^z\}_{z \in Z}$ is also MLR-ordered, i.e., if $z'' > z'$ then $\Lambda_G^{z''}$ is an MLR shift of $\Lambda_G^{z'}$.[23] Since $\{u(\cdot, s)\}_{s \in S}$ is an IDO family, Theorem 2 guarantees that

of *concordance*. Some papers with economic applications of Lehmann's concept of informativeness are Persico (2000), Athey and Levin (2001), Levin (2001), Bergemann and Valimaki (2002), and Jewitt (2006). Athey and Levin (2001) explore other related concepts of informativeness and their relationship with the payoff functions. The manner in which these papers are related to Lehmann's (1988) result and to each other is not straightforward; Jewitt (2006) provides an overview.

[21]The concept is Lehmann's; the term *accuracy* follows Persico (2000).

[22]However, there is a sense in which it is incorrect to say that Theorem 3 generalizes Lehmann's result. The criterion employed by us here (and indeed by Persico (1996) and Jewitt (2006) as well) - comparing information structures with the ex ante utility - is less stringent than the criterion Lehmann used. In the next section we shall compare information structures using precisely the same criterion as Lehmann and prove a result (Corollary 1) that is stronger than Theorem 3.

[23]To be precise, in the case when S is an interval, we mean that their associated density functions

there exists a G-optimal decision rule that is increasing in z. Therefore, Theorem 3 is valid if we can show that for *any* increasing decision rule $\psi : Z \to X$ under G there is a rule $\phi : Z \to X$ under H that gives a higher ex ante utility, i.e.,

$$\int_S \int_Z u(\phi(z), s) dH(z|s)\, d\Lambda(s) \geq \int_S \int_Z u(\psi(z), s) dG(z|s)\, d\Lambda(s).$$

This inequality in turn follows from aggregating (across s) the inequality (18) below.

PROPOSITION 4: *Suppose $\{u(\cdot, s)\}_{s\in S}$ is an IDO family and H is more accurate than G. Then for any increasing decision rule $\psi : Z \to X$ under G, there is an increasing decision rule $\phi : Z \to X$ under H such that, at each state s, the distribution of utility induced by ϕ and $H(\cdot|s)$ first order stochastically dominates the distribution of utility induced by ψ and $G(\cdot|s)$. Consequently, at each state s,*

$$\int_Z u(\phi(z), s) dH(z|s) \geq \int_Z u(\psi(z), s) dG(z|s). \tag{18}$$

(At a given state s, a decision rule ρ and a distribution on z induces a distribution of utility in the following sense: for any measurable set U of R, the probability of $\{u \in U\}$ equals the probability of $\{z \in Z : u(\rho(z), s) \in U\}$. So it is meaningful to refer, as this proposition does, to the distribution of utility at each s.)

Our proof of Proposition 4 requires the following lemma.

LEMMA 3: *Suppose $\{u(\cdot, s)\}_{s\in S}$ is an IDO family and H is more accurate than G. Then for any increasing decision rule $\psi : Z \to X$ under G, there is an increasing decision rule $\phi : Z \to X$ under H such that, for all (z, s),*

$$u(\phi(T(z, s)), s) \geq u(\psi(z), s). \tag{19}$$

Proof: We shall only demonstrate here how we construct ϕ from ψ in the case where ψ takes only finitely many values. This is true, in particular, when the set of

are MLR-ordered. In the case when S consists of finitely many elements, the definition of MLR-ordering is given in Appendix A.

actions X is finite. The extension to the case where the range of ψ is infinite is shown in Appendix B; the proof that ϕ is increasing is also postponed to that appendix. The proof below assumes that S is a compact interval, but it can be modified in an obvious way to deal with the case where S is finite.

For every $\bar{t}$ in Z and s in S, there is a unique $\bar{z}$ in Z such that $\bar{t} = T(\bar{z}, s)$. This follows from the fact that $G(\cdot|s)$ is a strictly increasing continuous function (since it admits a density function with support Z). We write $\bar{z} = \tau(\bar{t}, s)$; note that because T is increasing in both its arguments, the function $\tau : Z \times S \to Z$ is decreasing in s. Note also that (19) is equivalent to

$$u(\phi(t), s) \geq u(\psi(\tau(t, s)), s). \tag{20}$$

We will now show how $\phi(t)$ may be chosen to satisfy (20).

Note that because τ is decreasing and ψ is increasing the function $\psi(\tau(t, \cdot))$ is decreasing in s. This, together with our assumption that ψ takes finitely many values, allow us to partition $S = [s^*, s^{**}]$ into the sets S_1, S_2,..., S_M, where M is odd, with the following properties: (i) if $m > n$, then any element in S_m is greater than any element in S_n; (ii) whenever m is odd, S_m is a singleton, with $S_1 = \{s^*\}$ and $S_M = \{s^{**}\}$; (iii) when m is even, S_m is an open interval; (iv) for any s' and s'' in S_m, we have $\psi(\tau(t, s')) = \psi(\tau(t, s''))$; and (v) for s'' in S_m and s' in S_n such that $m > n$, $\psi(\tau(t, s')) \geq \psi(\tau(t, s''))$.

In other words, we have partitioned S into finitely many sets, so that within each set, $\psi(\tau(t, \cdot))$ takes the same value. Denoting $\psi(\tau(t, s))$ for s in S_m by ψ_m, (v) says that $\psi_1 \geq \psi_2 \geq \psi_3 \geq ... \geq \psi_M$. Establishing (20) involves finding $\phi(t)$ such that

$$u(\phi(t), s_m) \geq u(\psi_m, s_m) \text{ for any } s_m \in S_m;\ m = 1, 2, ..., M. \tag{21}$$

In the interval $[\psi_2, \psi_1]$, we pick the largest action $\hat{\phi}_2$ that maximizes $u(\cdot, s^*)$ in that interval. This exists because $u(\cdot, s^*)$ is continuous and $X \cap [\psi_2, \psi_1]$ is compact. By the IDO property,

$$u(\hat{\phi}_2, s_m) \geq u(\psi_m, s_m) \text{ for any } s_m \in S_m;\ m = 1, 2. \tag{22}$$

Recall that S_3 is a singleton; we call that element s_3. The action $\hat{\phi}_4$ is chosen to be the largest action in the interval $[\psi_4, \hat{\phi}_2]$ that maximizes $u(\cdot, s_3)$. Since ψ_3 is in that interval, we have $u(\hat{\phi}_4, s_3) \geq u(\psi_3, s_3)$. Since $u(\hat{\phi}_4, s_3) \geq u(\psi_4, s_3)$, the IDO property guarantees that $u(\hat{\phi}_4, s_4) \geq u(\psi_4, s_4)$ for any s_4 in S_4. Using the IDO property again (specifically, Lemma 2), we have $u(\hat{\phi}_4, s_m) \geq u(\hat{\phi}_2, s_m)$ for s_m in S_m ($m = 1, 2$) since $u(\hat{\phi}_4, s_3) \geq u(\hat{\phi}_2, s_3)$. Combining this with (22), we have found $\hat{\phi}_4$ in $[\psi_4, \hat{\phi}_2]$ such that

$$u(\hat{\phi}_4, s_m) \geq u(\psi_m, s_m) \text{ for any } s_m \in S_m;\ m = 1, 2, 3, 4. \tag{23}$$

We can repeat the procedure finitely many times, at each stage choosing $\hat{\phi}_{m+1}$ (for m odd) as the largest element maximizing $u(\cdot, s_m)$ in the interval $[\psi_{m+1}, \hat{\phi}_{m-1}]$, and finally, choosing $\hat{\phi}_{M+1}$ as the largest element maximizing $u(\cdot, s^{**})$ in the interval $[\psi_M, \hat{\phi}_{M-1}]$. It is clear that $\phi(t) = \hat{\phi}_M$ will satisfy (21). QED

Proof of Proposition 4: Let $\tilde{z}$ denote the random signal received under information structure G and let $\tilde{u}^G$ denote the (random) utility achieved when the decision rule ψ is used. Correspondingly, we denote the random signal received under H by $\tilde{t}$, with $\tilde{u}^H$ denoting the utility achieved by the rule ϕ, as constructed in Lemma 3. Observe that for any fixed utility level u' and at a given state s',

$$\begin{aligned}
\Pr[\tilde{u}^H \leq u' | s = s'] &= \Pr[u(\phi(\tilde{t}), s') \leq u' | s = s'] \\
&= \Pr[u(\phi(T(\tilde{z}, s')), s') \leq u' | s = s'] \\
&\leq \Pr[u(\psi(\tilde{z}), s') \leq u' | s = s'] \\
&= \Pr[\tilde{u}^G \leq u' | s = s']
\end{aligned}$$

where the second equality comes from the fact that, conditional on $s = s'$, the distribution of $\tilde{t}$ coincides with that of $T(\tilde{z}, s')$, and the inequality comes from the fact that $u(\phi(T(z, s')), s') \geq u(\psi(z), s')$ for all z (by Lemma 3).

Finally, the fact that, given the state, the conditional distribution of $\tilde{u}^H$ first order stochastically dominates $\tilde{u}^G$ means that the conditional mean of $\tilde{u}^H$ must also be higher than that of $\tilde{u}^G$. QED

Example 1 continued. As a simple application of Theorem 3, we return again to this example (previously discussed in Sections 2 and 4), where a firm has to decide on its production capacity before the state of the world is realized. Recall that the profit functions $\{\Pi(\cdot, s)\}_{s \in S}$ form an IDO (though not necessarily SCP) family. Suppose that before it makes its decision, the firm receives a signal z from the information structure G. Provided G is MLR-ordered, we know that the posterior distributions (on S) will also be MLR-ordered (in z). It follows from Theorem 2 that a higher signal will cause the firm to decide on a higher capacity. Assuming the firm is risk neutral, its ex ante expected profit is $\mathcal{V}(G, \Lambda, \Pi)$, where Λ is the firm's prior on S. Theorem 3 tells us that a more accurate information structure H will lead to a higher ex ante expected profit; the difference $\mathcal{V}(H, \Lambda, \Pi) - \mathcal{V}(G, \Lambda, \Pi)$ represents what the firm is willing to spend for the more accurate information structure.

We can clearly extend our analysis of Example 5 (in Section 4) in a similar way, i.e., we can introduce and compare information structures available to the firm for deciding the date of its product launch.

It is worth pointing out that our use of Proposition 4 to prove Theorem 3 (via (18)) does not fully exploit the property of first order stochastic dominance that Proposition 4 obtains. Our next application is one where this stronger conclusion is crucial.

Example 6. There are N investors, with investor i having wealth $w_i > 0$ and the strictly increasing Bernoulli utility function v_i. These investors place their wealth with a manager who decides on an investment policy; specifically, the manager allocates the total pool of funds $W = \sum_{i=1}^{N} w_i$ between a risky asset, with return s in state s, and a safe asset with return $r > 0$. Denoting the fraction invested in the risky asset by x, investor i's utility (as a function of x and s) is given by $u_i(x, s) = v_i((xs + (1-x)r)w_i)$. It is easy to see that $\{u_i(\cdot, s)\}_{s \in S}$ is an IDO family. Indeed, it is also an SCP and a QCIP family: for $s > r$, $u_i(\cdot, s)$ is strictly increasing in x; for $s = r$, $u_i(\cdot, r)$ is the constant $v_i(rw_i)$; and for $s < r$, $u_i(\cdot, s)$ is strictly decreasing in x.

Before she makes her portfolio decision, the manager receives a signal z from some information structure G. She employs the decision rule ψ, where $\psi(z)$ (in $[0,1]$) is the fraction of W invested in the risky asset. We assume that ψ is increasing in the signal; in the next section we shall give the precise sense in which this restriction involves no loss of generality. Suppose that the manager now has access to a superior information structure H. By Proposition 4, there is an increasing decision rule ϕ under H such that, at any state s, the distribution of investor k's utility under H and ϕ first order stochastically dominates the distribution of k's utility under G and ψ. In particular, (18) holds for $u = u_k$. Aggregating across states we obtain $\mathcal{U}_k(\phi, H, \Lambda_k) \geq \mathcal{U}_k(\psi, G, \Lambda_k)$, where Λ_k is investor k's (subjective) prior; in other words, k's ex ante utility is higher with the new information structure and the new decision rule.

But even more can be said because, for any other investor i, $u_i(\cdot, s)$ is a strictly increasing transformation of $u_k(\cdot, s)$, i.e., there is a strictly increasing function f such that $u_i = f \circ u_k$. It follows from Proposition 4 that (18) is true, not just for $u = u_k$ but for $u = u_i$. Aggregating across states, we obtain $\mathcal{U}_i(\phi, H, \Lambda_i) \geq \mathcal{U}_i(\psi, G, \Lambda_i)$, where Λ_i is investor i's prior.

To summarize, we have shown the following: though different investors may have different attitudes towards risk aversion and different priors, the greater accuracy of H compared to G allows the manager to implement a new decision rule that gives greater ex ante utility to *every* investor.

Finally, we turn to the following question: how important is the accuracy criterion to the results of this section? For example, Theorem 3 tells us that when H is more accurate than G, it gives the agent a higher ex ante utility for any prior that he may have on S. This raises the possibility that the accuracy criterion may be weakened if we only wish H to give a higher ex ante utility than G for a *particular* prior. However, this is not the case, as the next result shows.

Proposition 5: *Let S be finite, and H and G two information structures on S. If (17) holds at a given prior Λ^* which has S as its support and for any SCP family $\{u(\cdot, s)\}_{s\in S}$, then (17) holds at any prior Λ which has S as its support and for any SCP family.*

Proof: For the distribution Λ^* (Λ) we denote the probability of state s by $\lambda^*(s)$ ($\lambda(s)$). Given the SCP family $\{u(\cdot, s)\}_{s\in S}$, define the family $\{\tilde{u}(\cdot, s)\}_{s\in S}$ by $\tilde{u}(x, s) = [\lambda(s)/\lambda^*(s)]u(x, s)$. The ex ante utility of the decision rule ϕ under H, when the agent's utility is $\tilde{u}$, may be written as

$$\tilde{\mathcal{U}}(\phi, H, \Lambda^*) = \sum_{s\in S} \lambda^*(s) \int \tilde{u}(\phi(z), s) dH(z|s).$$

Clearly,

$$\mathcal{U}(\phi, H, \Lambda) \equiv \sum_{s\in S} \lambda(s) \int u(\phi(z), s) dH(z|s) = \tilde{\mathcal{U}}(\phi, H, \Lambda^*).$$

From this, we conclude that

$$\mathcal{V}(H, \Lambda, u) = \mathcal{V}(H, \Lambda^*, \tilde{u}). \tag{24}$$

Crucially, the fact that $\{u(\cdot, s)\}_{s\in S}$ is an SCP family, guarantees that $\{\tilde{u}(\cdot, s)\}_{s\in S}$ is also an SCP family. By assumption, $\mathcal{V}(H, \Lambda^*, \tilde{u}) \geq \mathcal{V}(G, \Lambda^*, \tilde{u})$. Applying (24) to both sides of this inequality, we obtain $\mathcal{V}(H, \Lambda, u) \geq \mathcal{V}(G, \Lambda, u)$. QED

Loosely speaking, this result says that if we wish to have ex ante utility comparability for *any* SCP family (or, even more strongly, any IDO family), then fixing the prior does not lead to a weaker criterion of informativeness. A weaker criterion can only be obtained if we fix the prior *and* require ex ante utility comparability for a smaller class of utility families.[24]

To construct a converse to Theorem 3, we assume that there are two states and two actions and that the actions are *non-ordered with respect to u* in the sense that x_1 is the better action in state s_1 and x_2 the better action in s_2, i.e., $u(x_1, s_1) > u(x_2, s_1)$

[24] This possibility is explored in Athey and Levin (2001).

and $u(x_1, s_2) < u(x_2, s_2)$. This condition guarantees that information on the state is potentially useful; if it does not hold, the decision problem is clearly trivial since either x_1 or x_2 will be unambiguously superior to the other action. Note also that the family $\{u(\cdot, s_1), u(\cdot, s_2)\}$ is an IDO family. We have the following result.

PROPOSITION 6: *Suppose that $S = \{s_1, s_2\}$, $X = \{x_1, x_2\}$, and that the actions are non-ordered with respect to u. If H is MLR-ordered and* not *more accurate than G, then there is a prior $\bar{\Lambda}$ on S such that $\mathcal{V}(H, \bar{\Lambda}, u) < \mathcal{V}(G, \bar{\Lambda}, u)$.*

Proof: Since H is not more accurate than G, there is $\bar{z}$ and $\bar{t}$ such that

$$G(\bar{z}|s_1) = H(\bar{t}|s_1) \text{ and } G(\bar{z}|s_2) < H(\bar{t}|s_2). \tag{25}$$

Given any prior Λ, and with the information structure H, we may work out the posterior distribution and the posterior expected utility of any action after receipt of a signal. We claim that there is a prior $\bar{\Lambda}$ such that, action x_1 maximizes the agent's posterior expected utility after he receives the signal $z < \bar{t}$ (under H), and action x_2 maximizes the agent's posterior expected utility after he receives the signal $z \geq \bar{t}$. This result follows from the assumption that H is MLR-ordered and is proved in Appendix B.

Therefore, the decision rule ϕ such that $\phi(z) = x_1$ for $z < \bar{t}$ and $\phi(z) = x_2$ for $z \geq \bar{t}$ maximizes the agent's ex ante utility, i.e.

$$\begin{aligned}
\mathcal{V}(H, \bar{\Lambda}, u) &= \mathcal{U}(\phi, H, \bar{\Lambda}) \\
&= \bar{\lambda}(s_1)\ \{ u(x_1|s_1)H(\bar{t}|s_1) + u(x_2|s_1)[1 - H(\bar{t}|s_1)] \} \\
&\quad + \bar{\lambda}(s_2)\ \{ u(x_1|s_2)H(\bar{t}|s_2) + u(x_2|s_2)[1 - H(\bar{t}|s_2)] \}.
\end{aligned}$$

Now consider the decision rule ψ under G given by $\psi(z) = x_1$ for $z < \bar{z}$ and $\psi(z) = x_2$ for $z \geq \bar{z}$. We have

$$\begin{aligned}
\mathcal{U}(\psi, G, \bar{\Lambda}) &= \bar{\lambda}(s_1)\ \{ u(x_1|s_1)G(\bar{z}|s_1) + u(x_2|s_1)[1 - G(\bar{z}|s_1)] \} \\
&\quad + \bar{\lambda}(s_2)\ \{ u(x_1|s_2)G(\bar{z}|s_2) + u(x_2|s_2)[1 - G(\bar{z}|s_2)] \}.
\end{aligned}$$

Comparing the expressions for $\mathcal{U}(\psi, G, \bar{\Lambda})$ and $\mathcal{U}(\phi, H, \bar{\Lambda})$, bearing in mind (25), and the fact that x_2 is the optimal action in state s_2, we see that

$$\mathcal{U}(\psi, G, \bar{\Lambda}) > \mathcal{U}(\phi, H, \bar{\Lambda}) = \mathcal{V}(H, \bar{\Lambda}, u).$$

Therefore, $\mathcal{V}(G, \bar{\Lambda}, u) > \mathcal{V}(H, \bar{\Lambda}, u)$. QED

6. The Completeness of Increasing Decision Rules

The model of information we constructed in the last section is drawn from statistical decision theory.[25] In the context of statistical decisions, the agent is a statistician conducting an experiment, the signal is the random outcome of the experiment, and the state of the world should be interpreted as a parameter of the model being considered by the statistician. In the previous section we identified conditions under which a Bayesian statistician, i.e., a statistician who has a prior on the states of the world, will pick an increasing decision rule. Our objective in this section is to strengthen that conclusion: we show that, under the same conditions, statisticians who use other criteria for choosing their decision rule may also confine their attention to increasing decision rules. This conclusion follows from a natural application of the informativeness results of the previous section (in particular, Proposition 4).

Consider an information structure $G = \{G(\cdot|s)\}_{s\in S}$; as in the previous section, we assume that the signal z is drawn from a compact interval Z and that its distribution $G(\cdot|s)$ admits a density function with Z as its support. The set of possible states, S, may either be a compact interval or a finite set of points. Unlike the previous section, we do not assume that the statistician has a prior on S. The utility associated to each action and state is given by the function $u : X \times S \to R$. We assume that the set of actions X is a compact subset of R and that u is continuous in x.

[25] For an introduction to statistical decision theory, see Blackwell and Girshik (1954) or Berger (1985). Statistical decision theory has been used in econometrics to study portfolio allocation, treatment choice, and model uncertainty in macroeconomic policy, amongst other areas. See the survey of Hirano (2008).

Let D be the set of all decision rules (which are measurable maps from Z to X). The *expected utility of the rule* ψ *in state* s (denoted by $\mathcal{E}_G(\psi, s)$) is defined as

$$\mathcal{E}_G(\psi, s) = \int_{z \in Z} u(\psi(z), s) dG(z|s).$$

Note that our assumptions guarantee that this is well defined.

The decision rule $\tilde{\psi}$ is said to be *at least as good as* another decision ψ, if it gives higher expected utility in all states, i.e., $\mathcal{E}_G(\tilde{\psi}, s) \geq \mathcal{E}_G(\psi, s)$ for all s in S. A subset D' of decision rules forms an *essentially complete class* if for any decision rule ψ, there is a rule $\tilde{\psi}$ in D' that is at least as good as ψ. Results that identify some subset of decision rules as an essentially complete class are called *complete class theorems.*[26]

It is useful to identify such a class of decision rules because, while statisticians may differ on the criterion they adopt in choosing amongst decision rules, it is typically the case that a rule satisfying their preferred criterion can be found in an essentially complete class. Consider the Bayesian statistician with prior Λ; her ex ante utility from the rule ψ (recall that this is denoted by $\mathcal{U}(\psi, G, \Lambda)$), can be written as $\int_{s \in S} \mathcal{E}_G(\psi, s)\, d\Lambda(s)$. It is clear that if $\mathcal{U}(\psi, G, \Lambda)$ is maximized at $\psi = \hat{\psi}$, then a rule $\tilde{\psi}$ in the complete class that is at least as good as $\hat{\psi}$ will also maximize $\mathcal{U}(\psi, G, \Lambda)$.

A non-Bayesian statistician will choose a rule using a different criterion; the two most commonly used are the *maxmin* and *minimax regret* criteria. The maxmin criterion evaluates a decision rule according to the lowest utility the rule could bring; formally, a rule $\hat{\psi}$ satisfies this criterion if it solves $\max_{\psi \in D} \{\min_{s \in S} \mathcal{E}_G(\psi, s)\}$. The *regret* of a decision rule ψ, which we denote by $r(\psi)$, is defined as $\max_{s \in S}[\max_{\psi' \in D} \mathcal{E}_G(\psi', s) - \mathcal{E}_G(\psi, s)]$. In other words, the regret of a rule ψ is the utility gap between the ideal rule (if s is known) and the rule ψ. A rule $\hat{\psi}$ satisfies the minimax regret criterion if it solves $\min_{\psi \in D} r(\psi)$. It is not hard to check that if a decision rule for either

[26]Some readers may wonder why, in our definition of a decision rule, we did not allow the statistician to mix actions at any given signal. The answer is that, when the signal space is atomless (as we have assumed), allowing her to do so will not make a difference since the set of decision rules involving only pure actions form an essentially complete class (see Blackwell (1951)).

the maxmin or minimax regret criterion exists, then such a rule can be found in an essentially complete class.[27]

The following complete class theorem is the main result of this section. It provides a justification for the statistician to restrict her attention to increasing decision rules.

THEOREM 4: *Suppose $\{u(\cdot, s)\}_{s\in S}$ is an IDO family and G is MLR-ordered. Then the increasing decision rules form an essentially complete class.*

Theorem 4 generalizes the complete class theorem of Karlin and Rubin (1956), which in turn generalizes Blackwell and Girshik (1954, Theorem 7.4.3). Karlin and Rubin (1956) establishes the essential completeness of increasing decision rules under the assumption that $\{u(\cdot, s)\}_{s\in S}$ is a QCIP family, which is a special case of our assumption that $\{u(\cdot, s)\}_{s\in S}$ forms an IDO family. Note that Theorem 4 is not known even for the case where $\{u(\cdot, s)\}_{s\in S}$ forms an SCP family.

An immediate application of Theorem 4 is that it allows us to generalize Theorem 3. Recall that Theorem 3 tells us that, for the Bayesian statistician, if H is more accurate than G, then H achieves a higher ex ante utility. Combining Theorem 4 and Proposition 4 allows us to strengthen that conclusion by establishing the superiority of H over G under a more stringent criterion; specifically, Corollary 1 below tells us that when an experiment H is more accurate than G, then H is capable of achieving higher expected utility in all states.

COROLLARY 1: *Suppose $\{u(\cdot, s)\}_{s\in S}$ is an IDO family, H is more accurate than G, and G is MLR-ordered. Then for any decision rule $\psi : Z \to X$ under G, there is an increasing decision rule $\phi : Z \to X$ under H such that*

$$\mathcal{E}_H(\phi, s) \geq \mathcal{E}_G(\psi, s) \quad \text{at each } s \in S. \tag{26}$$

[27]In the context of statistical decisions, the maxmin criterion was first studied by Wald; the minimax regret criterion is due to Savage. Discussion and motivation for the maxmin and minimax regret criteria can be found in Blackwell and Girshik (1954), Berger (1985), and Manski (2005). For some recent applications of the minimax regret criterion, see Manski (2004, 2005) and Manski and Tetenov (2007); a closely related criterion is employed in Chamberlain (2000).

Proof: If ψ is not increasing, then by Theorem 4, there is an increasing rule $\bar{\psi}$ that is at least as good as ψ. Proposition 4 in turn guarantees that there is an increasing decision rule ϕ under H that is at least as good as $\bar{\psi}$ since (18) says that $\mathcal{E}_H(\phi, s) \geq \mathcal{E}_G(\bar{\psi}, s)$ at each s. QED

Corollary 1 is a generalization of Lehmann (1988) which establishes a version of this result in the case where $\{u(\cdot, s)\}_{s\in S}$ is a QCIP family. This result is not known even for the case where $\{u(\cdot, s)\}_{s\in S}$ forms an SCP family.

Proof of Theorem 4: The idea of the proof is to show that the statistician who uses a strategy that is not increasing is in some sense debasing the information made available to her by G.

Having assumed that Z is a compact interval, we can assume, without further loss of generality, that $Z = [0, 1]$. Suppose that ψ is a decision rule (not necessarily increasing) under G. Here we confine ourselves to the case where ψ takes only finitely many values, an assumption which certainly holds if the set of actions X is finite. The case where ψ has an infinite range is covered in Appendix B.

Suppose that the actions taken under ψ are exactly x_1, x_2,...x_n (arranged in increasing order). We construct a new information structure $\bar{G} = \{\bar{G}(\cdot|s)\}_{s\in S}$ along the following lines. For each s, $\bar{G}(0|s) = 0$ and $\bar{G}(k/n|s) = \Pr_G[\psi(z) \leq x_k|s]$ for $k = 1, 2, ..., n$, where the right side of the second equation refers to the probability of $\{z \in Z : \psi(z) \leq x_k\}$ under the distribution $G(\cdot|s)$. We define $t_k(s)$ as the unique element in Z that obeys $G(t_k(s)|s) = \bar{G}(k/n|s)$. (Note that $t_0(s) = 0$ for all s.) Any z in $(\,(k-1)/n, k/n\,)$ may be written as $z = \theta\,[(k-1)/n] + (1-\theta)\,[k/n]$ for some θ in $(0, 1)$; we define

$$\bar{G}(z|s) = G(\theta t_{k-1}(s) + (1-\theta)t_k(s)|s). \tag{27}$$

This completely specifies $\bar{G}$.

Define a new decision rule $\bar{\psi}$ by $\bar{\psi}(z) = x_1$ for z in $[0, 1/n]$; for $k \geq 2$, we have $\bar{\psi}(z) = x_k$ for z in $(\,(k-1)/n, k/n\,]$. This is an increasing decision rule since we have arranged x_k to be increasing with k. It is also clear from our construction of $\bar{G}$ and $\bar{\psi}$ that, at each state s, the distribution of utility induced by $\bar{G}$ and $\bar{\psi}$ equals the distribution of utility induced by G and ψ.

We claim that G is more accurate than $\bar{G}$. Provided this is true, Proposition 4 says that there is an increasing decision rule ϕ under G that is at least as good as $\bar{\psi}$ under $\bar{G}$, i.e., at each s, the distribution of utility induced by G and ϕ first order stochastically dominates that induced by $\bar{G}$ and $\bar{\psi}$. Since the latter coincides with the distribution of utility induced by G and ψ, the proof is complete.

That G is more accurate than $\bar{G}$ follows from the assumption that G is MLR-ordered. Denote the density function associated to the distribution $G(\cdot|s)$ by $g(\cdot|s)$. The probability of $Z_k = \{z \in Z : \psi(z) \leq x_k\}$ is given by $\int \mathbf{1}_{Z_k}(z)g(z|s)dz$, where $\mathbf{1}_{Z_k}$ is the indicator function of Z_k. By the definition of $\bar{G}$, we have

$$\bar{G}(k/n|s) = \int \mathbf{1}_{Z_k}(z)g(z|s)dz.$$

Recall that $t_k(s)$ is defined as the unique element that obeys $G(t_k(s)|s) = \bar{G}(k/n|s)$; equivalently,

$$\bar{G}(k/n|s) - G(t_k(s)|s) = \int \left[\mathbf{1}_{Z_k}(z) - \mathbf{1}_{[0,t_k(s)]}(z)\right] g(z|s)dz = 0. \tag{28}$$

The function W given by $W(z) = \mathbf{1}_{Z_k}(z) - \mathbf{1}_{[0,t_k(s)]}(z)$ has the following single-crossing type condition: $z > t_k(s)$, we have $W(z) \geq 0$ and for $z \leq t_k(s)$, we have $W(z) \leq 0$.[28] Let $s' > s$; since $G(\cdot|s)$ is MLR-ordered, $g(z|s')/g(z|s)$ is an increasing function of z. By a standard result (see, for example, Athey (2002, Lemma 5)), we have

$$\bar{G}(k/n|s') - G(t_k(s)|s') = \int \left[\mathbf{1}_{Z_k}(z) - \mathbf{1}_{[0,t_k(s)]}(z)\right] g(z|s')dz \geq 0. \tag{29}$$

This implies that $t_k(s') \geq t_k(s)$ since $G(t_k(s')|s') = \bar{G}(k/n|s')$.

To show that G is more accurate than $\bar{G}$, we require $T(z,s)$ to be increasing in s, where T is defined by $G(T(z,s)|s) = \bar{G}(z|s)$. For $z = k/n$, $T(z,s) = t_k(s)$, which

[28] This property is related to but not the same as the single crossing property we have defined in this paper; Athey (2002) refers to this property as SC1 and the one we use as SC2.

we have shown is increasing in s. For z in the interval $(\,(k-1)/n, k/n\,)$, recall (see (27)) that $\bar{G}(z)$ was defined such that $T(z,s) = \theta t_{k-1}(s) + (1-\theta)t_k(s)$. Since both t_{k-1} and t_k are increasing in s, $T(z,s)$ is also increasing in s. QED

It is clear from our proof of Theorem 4 that we can in fact give a sharper statement of that result; we do so below.

THEOREM 4*: *Suppose $\{u(\cdot,s)\}_{s\in S}$ is an IDO family and G is MLR-ordered. Then for any decision rule $\psi : Z \to X$, there is an increasing decision rule $\phi : Z \to X$ such that, at each s, the distribution of utility induced by G and ϕ first order stochastically dominates the distribution of utility induced by G and ψ.*

Complete class theorems obviously have a role in statistics and econometrics, but they are also relevant in theoretical economic modeling. The next example gives an instance of such an application.

Example 6 continued. Recall that we assume in Section 5 that the manager's decision rule under G is increasing in the signal. This restriction on the decision rule needs to be justified; provided G is MLR-ordered, this justification is provided by Theorem 4*.

Let $\tilde{\psi}$ be a (not necessarily increasing) decision rule. Theorem 4* tells us that for some investor k, there is an increasing decision rule $\psi : Z \to X$ such that, at each s, the distribution of u_k induced by G and ψ first order stochastically dominates the distribution of u_k induced by G and $\tilde{\psi}$. This implies that

$$\int_{z\in Z} u_k(\phi(z),s)dG(z|s) \geq \int_{z\in Z} u_k(\tilde{\psi}(z),s)dG(z|s). \tag{30}$$

Aggregating this inequality across states, we obtain $\mathcal{U}_k(\psi, G, \Lambda_k) \geq \mathcal{U}_k(\tilde{\psi}, G, \Lambda_k)$, i.e., the increasing rule gives investor k a higher ex ante utility.

However, we can say more because, for any other investor i, $u_i(\cdot,s)$ is just an increasing transformation of $u_k(\cdot,s)$, i.e., there is a strictly increasing function f such that $u_i = f \circ u_k$. Appealing to Theorem 4* again, we see that (30) is true if u_k is replaced with u_i. Aggregating this inequality across states gives us $\mathcal{U}_i(\psi, G, \Lambda_i) \geq$

$\mathcal{U}_i(\tilde{\psi}, G, \Lambda_i)$. In short, we have shown the following: any decision rule admits an increasing decision rule that (weakly) raises the ex ante utility of *every* investor. This justifies our assumption that the manager uses an increasing decision rule.

Our final application generalizes an example found in Manski (2005, Proposition 3.1) on monotone treatment rules.

Example 7. Suppose that there are two ways of treating patients with a particular medical condition. Treatment A is the status quo; it is known that a patient who receives this treatment will recover with probability $\bar{p}^A$. Treatment B is a new treatment whose effectiveness is unknown. The probability of recovery with this treatment, p^B, corresponds to the unknown state of the world and takes values in some set P. We assume that the planner receives a signal z of p^B that is MLR-ordered with respect to p^B. (Manski (2005) considers the case of N subjects who are randomly selected to receive Treatment B, with z being the the number who are cured. Clearly, the distribution of z is binomial; it is also not hard to check that it is MLR-ordered with respect to p^B.)

Normalizing the utility of a cure at 1 and that of no cure at 0, the planner's expected utility when a member of the population receives treatment A is $\bar{p}^A$. Similarly, the expected utility of treatment B is p^B. Therefore, the planner's utility if she subjects fraction x of the population to B (and the rest to A) is

$$u(x, p^B) = (1-x)\,\bar{p}^A + x\,p^B. \tag{31}$$

The planner's decision (treatment) rule maps z to the proportion x of the (patient) population who will receive treatment B. As pointed out in Manski (2005), $\{u(x,\cdot)\}_{p^B \in P}$ is a QCIP family, and so Karlin and Rubin (1956) guarantee that decision rules where x increases with z form an essentially complete class.

Suppose now that the planner has a different payoff function, that takes into account the cost of the treatment. We denote the cost of having fraction x treated with B and the rest with A by $C(x)$. Then the payoff function is

$$u(x, p^B) = (1-x)\,\bar{p}^A + x\,p^B - C(x).$$

If the cost of treatments A and B are both linear, or more generally if C is convex, then one can check that $\{u(x,\cdot)\}_{p^B \in P}$ will still be a QCIP family. We can then appeal to Karlin and Rubin (1956) to obtain the essential completeness of the increasing decision rules. But there is no particular reason to believe that C is convex; indeed C will never be convex if the presence of scale economies leads to the total cost of having both treatments in use being more expensive than subjecting the entire population to one treatment or the other. (Formally, there is $x^* \in (0,1)$ such that $C(0) < C(x^*) > C(1)$.) However, u is supermodular in (x, p^B) whatever the shape of C, so $\{u(x,\cdot)\}_{p^B \in P}$ *is* an IDO family; Theorem 4 tells us that the planner may confine herself to increasing decision rules since they form an essentially complete class.[29]

Appendix A

Our objective in this section is to prove Theorem 2 and its converse in the case where S is a finite set. Suppose that $S = \{s_1, s_2, ..., s_N\}$ and the agent's objective function is

$$F(x, \lambda) = \sum_{i=1}^{N} f(x, s_i)\lambda(s_i)$$

[29]Manski and Tetenov (2007) prove a complete class theorem with a different variation on the payoff function (31). For the payoff function (31), Manski (2005) in fact showed a sharper result: the planner will choose a rule where the *whole* population is subject to B (A) if the number of treatment successes in the sample goes above (below) a particular threshold. The modification of the payoff function in Manski and Tetenov (2007) is motivated in part by the desire to obtain fractional treatment rules, in which, for a non-negligible set of sample outcomes, *both* treatments will be in use. In our variation on (31), it is clear that fractional treatment is plausible if large values of x involve very high costs.

where $\lambda(s_i)$ is the probability of state s_i. We assume that $\lambda(s_i) > 0$ for all s_i. Let γ be another distribution with support S. We say that γ *is an MLR-shift of* λ if $\gamma(s_i)/\lambda(s_i)$ is increasing in i.

Proof of Theorem 2 for the case of finite S: Suppose $F(x'', \lambda) \geq F(x, \lambda)$ for x in $[x', x'']$. We denote $(f(x'', s_i) - f(x', s_i))\lambda(s_i)$ by a_i and define $A_k = \sum_{i=k}^{N} a_i$. By assumption, $A_1 = F(x'', \lambda) - F(x', \lambda) \geq 0$; we claim that $A_k \geq 0$ for any k (this claim is analogous to (13) in the proof of Theorem 2).

Suppose instead that there is $M \geq 2$ such that

$$A_M = \sum_{i=M}^{N} \left(f(x'', s_i) - f(x', s_i)\right) \lambda(s_i) < 0.$$

As in the proof of Theorem 2 in the main part of the paper, we choose $\bar{x}$ that maximizes $f(\cdot, s_M)$ in $[x', x'']$. By the IDO property and Lemma 2, we have

$$f(\bar{x}, s_i) - f(x'', s_i) \geq 0 \text{ for } i \leq M \text{ and} \tag{32}$$

$$f(\bar{x}, s_i) - f(x', s_i) \geq 0 \text{ for } i \geq M. \tag{33}$$

(These inequalities (32) and (33) are analogous to (15) and (16) respectively.) Following the argument used in Theorem 2, these inequalities lead to

$$A_1 = \sum_{i=1}^{N} (f(x'', s_i) - f(x', s_i))\lambda(s_i) < 0, \tag{34}$$

which is a contradiction. Therefore $A_M \geq 0$.

Denoting $\gamma(s_i)/\lambda(s_i)$ by b_i, we have

$$F(x'', \gamma) - F(x', \gamma) = \sum_{i=1}^{N} a_i b_i.$$

It is not hard to check that[30]

$$A_1 b_1 + \sum_{i=2}^{N} A_i(b_i - b_{i-1}) = \sum_{1=1}^{N} a_i b_i. \tag{35}$$

[30] This is just a discrete version of integration by parts.

Since γ is an MLR shift of λ, $b_i - b_{i-1} \geq 0$ for all i and so $\sum_{i=2}^{N} A_i(b_i - b_{i-1}) \geq 0$. Thus (35) guarantees that $\sum_{1=1}^{N} a_i b_i \geq A_1 b_1$; in other words,

$$F(x'', \gamma) - F(x', \gamma) \geq \frac{\gamma(s_1)}{\lambda(s_1)} \left[F(x'', \lambda) - F(x', \lambda) \right]. \qquad (36)$$

It follows that the left hand side is nonnegative (positive) if the right side is nonnegative (positive), as required by the IDO property. QED

It is natural to ask whether the MLR condition in Theorem 2 can be weakened, i.e., can we obtain $F(\cdot, \gamma) \succeq_I F(\cdot, \lambda)$ for every IDO family without requiring an MLR shift from λ to γ? The following proposition says that the answer is 'No'.

Proposition A1: *If γ is* not *an MLR-shift of λ, there exists an SCP family $\{f(\cdot, s)\}_{s \in S}$ defined on $X \subset R$ such that $argmax_{x \in X} F(x, \gamma) < argmax_{x \in X} F(x, \lambda)$.*

Proof: Suppose that γ is not an MLR shift of λ, so there are states k and $k+1$ such that $\gamma(s_k)/\lambda(s_k) > \gamma(s_{k+1})/\lambda(s_{k+1})$. It follows that there is positive scalar a such that

$$\frac{\lambda(s_{k+1})a}{\lambda(s_k)} > 1 > \frac{\gamma(s_{k+1})a}{\gamma(s_k)}. \qquad (37)$$

Let $X = \{x_1, x_2\}$, with $x_1 < x_2$, and define $\{f(\cdot, s)\}_{s \in S}$ in the following way: for $i < k$, $f(x_1, s_i) = \epsilon$ and $f(x_2, s_i) = 0$; for $i = k$, $f(x_1, s_k) = 1$ and $f(x_2, s_k) = 0$; for $i = k+1$, $f(x_1, s_{k+1}) = 0$ and $f(x_2, s_{k+1}) = a$; finally, for $i > k+1$, $f(x_1, s_i) = 0$ and $f(x_2, s_i) = \epsilon$. This is an SCP family provided $\epsilon > 0$ since $f(x_2, s_i) - f(x_1, s_i)$ takes values $-\epsilon$, -1, a, and finally ϵ as i increases.

Note that the first inequality in (37) says that $\lambda(s_{k+1})a > \lambda(s_k)$, so if ϵ is sufficiently small,

$$F(x_2, \lambda) = \lambda(s_{k+1})a + \sum_{i=k+2}^{N} \lambda(s_i)\epsilon > \sum_{i=1}^{k-1} \lambda(s_i)\epsilon + \lambda(s_k) = F(x_1, \lambda).$$

Now the second inequality in (37) says that $\gamma(s_k) > \gamma(s_{k+1})a$, so if ϵ is sufficiently small,

$$F(x_1, \gamma) = \sum_{i=1}^{k-1} \gamma(s_i)\epsilon + \gamma(s_k) > \gamma(s_{k+1})a + \sum_{i=k+2}^{N} \gamma(s_i)\epsilon = F(x_2, \gamma).$$

So the shift from λ to γ has *reduced* the optimal value of x from x_2 to x_1. QED

Appendix B

Proof of Lemma 3 continued: We first show that ϕ (as constructed in the proof in Section 5) is an increasing rule. We wish to compare $\phi(t')$ against $\phi(t)$ where $t' > t$. Note that the construction of $\phi(t)$ first involves partitioning S into subsets S_1, $S_2, ...S_M$ obeying properties (i) to (v). In particular, (v) says that for any s in S_m, $\psi(\tau(t,s))$ takes the same value, which we denote by ψ_m. To obtain $\phi(t')$, we first partition S into disjoints sets S'_1, S'_2,..., S'_L, where L is odd, with the partition satisfying properties (i) to (v). The important thing to note is that, for any s, we have

$$\psi(\tau(t', s)) \geq \psi(\tau(t, s)). \tag{38}$$

This is clear: both ψ and $\tau(\cdot, s)$ are increasing functions and $t' > t$. We denote $\psi(\tau(t', s))$ for s in S'_k by ψ'_k. Any s in S belongs to some S_m and some S'_k, in which case (38) may be re-written as

$$\psi'_k \geq \psi_m. \tag{39}$$

The construction of $\phi(t')$ involves the construction of $\hat{\phi}'_2$, $\hat{\phi}'_4$, etc. The action $\hat{\phi}'_2$ is the largest action maximizing $u(\cdot, s^*)$ in the interval $[\psi'_2, \psi'_1]$. Comparing this with $\hat{\phi}_2$, which is the largest action maximizing $u(\cdot, s^*)$ in the interval $[\psi_2, \psi_1]$, we know that $\hat{\phi}'_2 \geq \hat{\phi}_2$ since (following from (39)) $\psi'_2 \geq \psi_2$ and $\psi'_1 \geq \psi_1$.

By definition, $\hat{\phi}'_4$ is the largest action maximizing $u(\cdot, s'_3)$ in $[\psi'_4, \hat{\phi}'_2]$, where s'_3 refers to the unique element in S'_3. Let $\bar{m}$ be the largest odd number such that $s_{\bar{m}} \leq s'_3$. (Recall that $s_{\bar{m}}$ is the unique element in $S_{\bar{m}}$.) By definition, $\hat{\phi}_{\bar{m}+1}$ is the largest element maximizing $u(\cdot, s_{\bar{m}})$ in $[\psi_{\bar{m}+1}, \hat{\phi}_{\bar{m}-1}]$. We claim that $\hat{\phi}'_4 \geq \hat{\phi}_{\bar{m}+1}$. This is an application of Theorem 1. If follows from the following: (i) $s'_3 \geq s_{\bar{m}}$, so $u(\cdot, s'_3) \succeq_I u(\cdot, s_{\bar{m}})$; (ii) the manner in which $\bar{m}$ is defined, along with (39), guarantees that $\psi'_4 \geq \psi_{\bar{m}+1}$; and (iii) we know (from the previous paragraph) that $\hat{\phi}'_2 \geq \hat{\phi}_2 \geq \hat{\phi}_{\bar{m}-1}$.

So we obtain $\hat{\phi}'_4 \geq \hat{\phi}_{\bar{m}+1} \geq \phi(t)$. Repeating the argument finitely many times (on $\hat{\phi}'_6$ and so on), we obtain $\phi(t') \geq \phi(t)$. This completes the proof of Lemma 3 in the special case where ψ takes finitely many values.

Extension to the case where the range of ψ is infinite.

The strategy is to approximate ψ with a sequence of simpler decision rules. Let $\{A_n\}_{n\geq 1}$ be a sequence of finite subsets of X such that $A_n \subset A_{n+1}$ and $\cup_{n\geq 1} A_n$ is dense in X. (This sequence exists because X is compact.) The function $\psi_n : Z \to X$ is defined as follows: $\psi_n(z)$ is the largest element in A_n that is less than or equal to $\psi(z)$. The sequence of decision rules ψ_n has the following properties:

(i) ψ_n is increasing in z;

(ii) $\psi_{n+1}(z) \geq \psi_n(z)$ for all z;

(iii) the range of ψ_n is finite; and

(iv) the increasing sequence ψ_n converges to ψ pointwise.

Since ψ_n takes only finitely many values, we know there is an increasing decision rule ϕ_n (as defined in the proof of Lemma 3 in Section 5) such that for all (z, s),

$$u(\phi_n(T(z, s)), s) \geq u(\psi_n(z), s). \tag{40}$$

We claim that ϕ_n is also an increasing sequence. This follows from the fact that, for all (t, s),

$$\psi_{n+1}(\tau(t, s)) \geq \psi_n(\tau(t, s)). \tag{41}$$

This inequality plays the same role as (38); the latter was used to show that $\phi(t') \geq \phi(t)$. Mimicking the argument there, (41) tells us that $\phi_{n+1}(t) \geq \phi_n(t)$.

Since ϕ_n is an increasing sequence and X is compact, it has a limit, which we denote as ϕ. Since, for each n, ϕ_n is an increasing decision rule, ϕ is also an increasing decision rule. For each n, (40) holds; taking limits, and using the continuity of u with respect to x, we obtain $u(\phi(T(z, s)), s) \geq u(\psi(z), s)$. QED

Proof of Proposition 6 continued: It remains for us to show how $\bar{\Lambda}$ is constructed. We denote the density function of $H(\cdot|s)$ by $h(\cdot|s)$. It is clear that since the actions are non-ordered, we may choose $\bar{\lambda}(s_1)$ and $\bar{\lambda}(s_2)$ (the probabilities of s_1 and s_2

respectively) such that

$$\bar{\lambda}(s_1)h(\bar{t}|s_1)\,[u(x_1,s_1)-u(x_2,s_1)] = \bar{\lambda}(s_2)h(\bar{t}|s_2)\,[u(x_2|s_2)-u(x_1,s_2)]. \tag{42}$$

Re-arranging this equation, we obtain

$$\bar{\lambda}(s_1)h(\bar{t}|s_1)u(x_1|s_1)+\bar{\lambda}(s_2)h(\bar{t}|s_2)u(x_1|s_2) = \bar{\lambda}(s_1)h(\bar{t}|s_1)u(x_2|s_1)+\bar{\lambda}(s_2)h(\bar{t}|s_2)u(x_2|s_2).$$

Therefore, given the prior $\bar{\Lambda}$, the posterior distribution after observing $\bar{t}$ is such that the agent is indifferent between actions x_1 and x_2.

Suppose the agent receives the signal $z < \bar{t}$. Since H is MLR-ordered, we have

$$\frac{h(z|s_1)}{h(\bar{t}|s_1)} \geq \frac{h(z|s_2)}{h(\bar{t}|s_2)}.$$

This fact, together with (42) guarantee that

$$\bar{\lambda}(s_1)h(z|s_1)\,[u(x_1,s_1)-u(x_2,s_1)] \geq \bar{\lambda}(s_2)h(z|s_2)\,[u(x_2|s_2)-u(x_1,s_2)].$$

Re-arranging this equation, we obtain

$$\bar{\lambda}(s_1)h(z|s_1)u(x_1|s_1)+\bar{\lambda}(s_2)h(z|s_2)u(x_1|s_2) \;\geq$$

$$\bar{\lambda}(s_1)h(z|s_1)u(x_2|s_1)+\bar{\lambda}(s_2)h(z|s_2)u(x_2|s_2).$$

So, after observing $z < \bar{t}$, the (posterior) expected utility of action x_1 is greater than that of x_2. In a similar way, we can show that x_2 is the optimal action after observing a signal $z \geq \bar{t}$. QED

Proof of Theorem 4 for the case where ψ has an infinite range: We construct an alternative experiment $\bar{G}$ and increasing decision rule $\bar{\psi}$ with the following two properties:

(P1) at each state s, the distribution of losses induced by $\bar{G}$ and $\bar{\psi}$ equals that induced by G and ψ;

(P2) G is more accurate than $\bar{G}$.

An application of Proposition 4 then guarantees that there is an increasing decision rule under G that is at least as good as ψ. Thus our proof is essentially the same as the one we gave for the finite case in Section 6, except that construction of $\bar{G}$ and $\bar{\psi}$ is somewhat more complicated.

Since X is compact, there is a smallest compact interval M containing X. At a given state s, we denote the distribution on M induced by G and ψ by $F(\cdot|s)$, i.e., for any x in M, we have $F(x|s) = \Pr_G[\psi(z) \leq x|s]$. There are two noteworthy features of $\{F(\cdot|s)\}_{s\in S}$:

(i) For a fixed $\bar{s}$, we may partition M into (disjoint) contour sets, $U_{\bar{s}}(r)$, i.e., $U_{\bar{s}}(r) = \{x \in M : F(x|\bar{s}) = r\}$. It is possible that for some r, $U_{\bar{s}}(r)$ is empty, but if it is nonempty then it has a minimum and the minimum is in X (and not just in M). Crucially, this partition is *common across all states s*. In other words, for any other state s, there is some r' such that $U_s(r') = U_{\bar{s}}(r)$.

(ii) The atoms of $F(\cdot|s)$ also do not vary with s; i.e., if x is an atom for $F(\cdot|\bar{s})$, then it is an atom for $F(\cdot|s)$ for every other state s.

These two features follow easily from the definition of F, the compactness of X, and the fact that $G(\cdot|s)$ is atomless and has support Z at every state s.

To each element x in M we associate a number $\epsilon(x)$, where $\epsilon(x) > 0$ if and only if x is an atom and $\sum_{x\in X} \epsilon(x) < \infty$. (Note that there are at most countably many atoms, so the infinite summation makes sense.) We define the map $Y : M \to R$ where $Y(x) = x + \sum_{\{x'\in M: x'\leq x\}} \epsilon(x')$. It is clear that this map is a strictly increasing and hence 1-1 map. Let $Y^* = \cup_{x\in M}[Y(x) - \epsilon(x), Y(x)]$. The difference between Y^* and the range of Y, i.e., set $Y^* \setminus Y(M)$, may be written in the form $\cup_{n=1}^{\infty} I_n$, where $I_n = [\,Y(a_n) - \epsilon(a_n)\,, Y(a_n)\,)$ and $\{a_n\}_{n\in N}$ is the set of atoms. (Loosely speaking, the 'gaps' I_n arise at every atom.)

We define the distribution $\tilde{G}(\cdot|s)$ on Y^* in the following way. For y in $Y(M)$, $\tilde{G}(y|s) = F(Y^{-1}(y)|s)$. For $y = Y(a_n) - \epsilon(a_n)$, define $\tilde{G}(y|s)$ as the limit of $\tilde{G}(y_n|s)$ where y_n is some sequence in $Y(M)$ tending to y from the left; if no such sequence exists (which occurs if and only if there is an atom at the smallest element of X), let

$\tilde{G}(y|s) = 0$. (One can easily check that this definition is unambiguous.) It remains for us to define $\tilde{G}(y|s)$ for y in the open interval $(\,Y(a_n) - \epsilon(a_n)\,, Y(a_n)\,)$. For $y = C(a_n)$ or $y = C(a_n) - \epsilon(a_n)$, define $t(y)$ by $G(t(y)|s) = \tilde{G}(y|s)$. Any element y in $(\,Y(a_n) - \epsilon(a_n)\,, Y(a_n)\,)$ may be written as $\theta\,[Y(a_n) - \epsilon(a_n)] + (1-\theta)\,[Y(a_n)]$. We define

$$\tilde{G}(y|s) = G\left(\theta t\left(C(a_n) - \epsilon(a_n)\right) + (1-\theta)t\left(C(a_n)\right)\,|s\right).$$

We have now completely specified the distribution $\tilde{G}(\cdot|s)$. Note that we have constructed this distribution to be atomless, so for every number r in $[0,1]$, the set $\{y \in Y^* : \tilde{G}(y|s) = r\}$ is nonempty. Indeed, following from observation (i) above, this set has a smallest element, which we denote by $\hat{y}(r)$. We define $Y^{**} = \{\hat{y}(r) : r \in [0,1]\}$. Observation (i) also tells us that Y^{**} does not vary with s. We denote the restriction of $\tilde{G}(\cdot|s)$ to Y^{**} by $\bar{G}(\cdot|s)$. Therefore, for any r in $[0,1]$ and any state s, there is a unique y in Y^{**} such that $\bar{G}(y|s) = r$.

One can check that property (P2) (stated above) holds: G is more accurate than $\bar{G}$. Formally, the map T defined by $G(T(y,s)|s) = \bar{G}(y|s)$ exists and has the property that $T(y,s)$ is increasing in s; the proof is substantially the same as that for the finite case. Furthermore, the map $T(\cdot|s)$ has a unique inverse (in Y^{**}). So we have identified precisely the properties of T needed for the application of Proposition 4.

Consider the decision rule $\bar{\psi} : Y^{**} \to X$ defined as follows: if y is in $Y(M)$, define $\bar{\psi}(y) = Y^{-1}(y)$; if y is in $[\,Y(a_n) - \epsilon(a_n)\,, Y(a_n)\,)$, define $\bar{\psi}(y) = a_n$. It is not hard to verify that $\bar{G}$ and $\bar{\psi}$ generate the same distribution of losses as G and ψ, as required by (P1). QED

REFERENCES

AMIR, R. (1996): "Sensitivity Analysis in Multisector Optimal Economic Dynamics" *Journal of Mathematical Economics*, 25, 123-141.

ARROW, K. AND D. LEVHARI (1969): "Uniqueness of the internal rate of return with variable life of investment," *Economic Journal*, 79(315), 560-566.

ASHWORTH, S. AND E. BUENO DE MESQUITA (2006): "Monotone Comparative Statics in Models of Politics," *American Journal of Political Science* 50(1), 214-231.

ATHEY, S. (2001): "Single Crossing Properties and the Existence of Pure Strategy Equilibria in Games of Incomplete Information," *Econometrica*, 69(4), 861-890.

ATHEY, S. (2002): "Monotone Comparative Statics under Uncertainty," *Quarterly Journal of Economics*, 117(1), 187-223.

ATHEY, S. AND J. LEVIN (2001): "The Value of Information in Monotone Decision Problems," Stanford Working Paper 01-003. BOYD (1997): *Capital Theory, Equilibrium Analysis, and Recursive Utility.* Oxford: Blackwell.

BERGEMANN, D. and J. VALIMAKI (2002): "Information Acquisition and Efficient Mechanism Design," *Econometrica*, 70, 1007-1033.

BERGER, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis* New York: Springer-Verlag

BLACKWELL, D. (1951): "On a Theorem of Lyapunov," *The Annals of Mathematical Statistics,* 22(1), 112-114.

BLACKWELL, D. (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24(2), 265-272.

BLACKWELL, D. AND M. A. GIRSHIK (1954): *Theory of Games and Statistical Decisions.* New York: Dover Publications

CHAMBERLAIN, G. (2000): "Econometrics and Decision Theory," *Journal of Econometrics*, 95, 255-283.

GOLLIER, C. (2001): *The Economics of Risk and Time.* Cambridge: MIT Press.

HIRANO, K. (2008): "Decision theory in econometrics," in *The New Palgrave Dictionary of Economics, 2nd Edition.* Eds. S. DURLAUF and L. E. BLUME. Palgrave Macmillan.

JEWITT, I. (2006): "Information Order in Decision and Agency Problems," Personal Manuscript.

KARLIN, S. AND H. RUBIN (1956): "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *The Annals of Mathematical Statistics,* 27(2), 272-299.

LEHMANN, E. L. (1988): "Comparing Location Experiments," *The Annals of Statistics,* 16(2), 521-533.

LEVIN, J. (2001): "Information and the Market for Lemons," *Rand Journal of Economics,* 32(4), 657-666.

MANSKI, C. (2004): "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72(4), 1221-1246.

MANSKI, C. (2005): *Social Choice with Partial Knowledge of Treatment Response,* Princeton: Princeton University Press.

MANSKI, C. AND A. TETENOV (2007): "Admissible treatment rules for a risk-averse planner with experimental data on an innovation," *Journal of Statistical Planning and Inference*, 137(6), 1998-2010.

MILGROM, P. AND J. ROBERTS (1990): "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica,* 58(6), 1255-1277.

MILGROM, P. AND C. SHANNON (1994): "Monotone Comparative Statics," *Econometrica*, 62(1), 157-180.

ORMISTON, M. B. AND E. SCHLEE (1993): "Comparative Statics under uncertainty for a class of economic agents," *Journal of Economic Theory*, 61, 412-422.

PERSICO, N. (1996): "Information Acquisition in Affiliated Decision Problems." Discussion Paper No. 1149, Department of Economics, Northwestern University.

PERSICO, N. (2000): "Information Acquisition in Auctions," *Econometrica*, 68(1), 135-148.

QUAH, J. (2007): "The comparative statics of constrained optimization problems," *Econometrica*, 75(2), 401-431.

QUAH, J. K.-H. AND B. STRULOVICI (2007): "Comparative Statics with the Interval Dominance Order II." Incomplete Manuscript; available at http://www.economics.ox.ac.uk/Members/john.quah/hlp30.pdf

TOPKIS, D. M. (1978): "Minimizing a Submodular Function on a lattice," *Operations Research*, 26, 305-321.

TOPKIS, D. M. (1998): *Supermodularity and Complementarity.* Princeton: Princeton University Press.

VIVES, X. (1990): "Nash Equilibrium with Strategic Complementarities," *Journal of Mathematical Economics*, 19, 305-21.

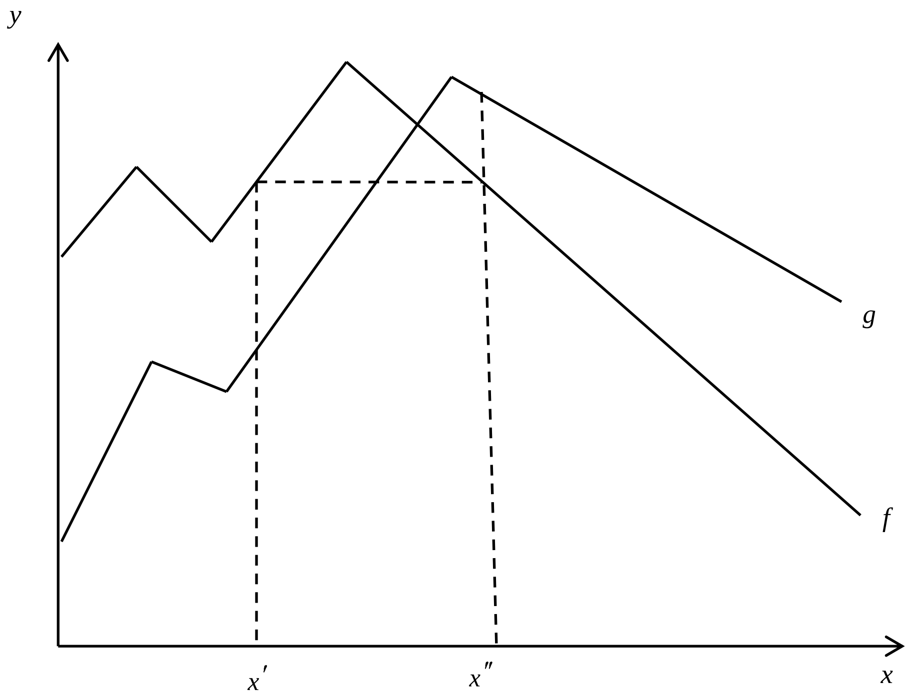

Figure 1

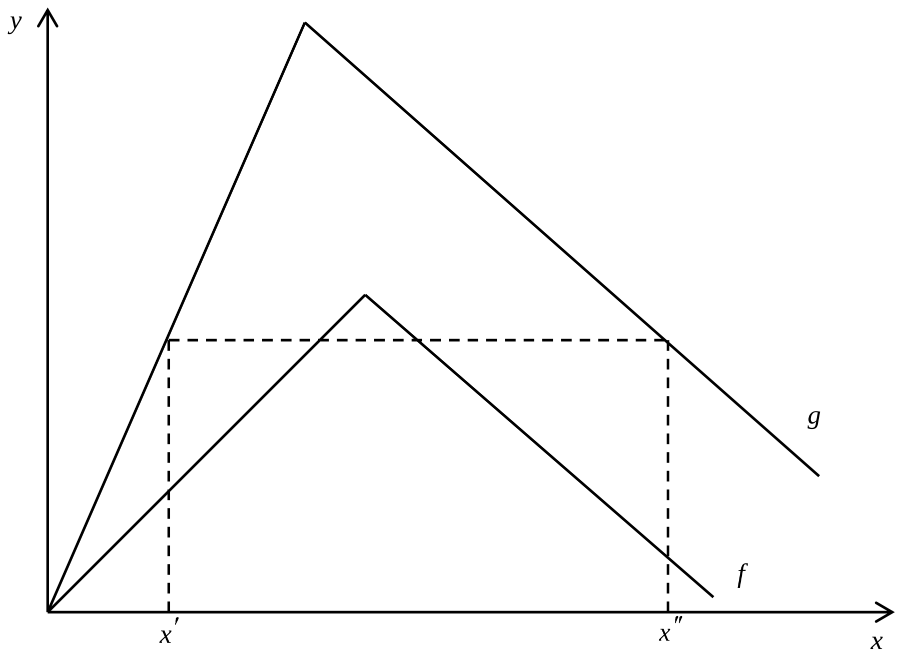

Figure 2