

The Value of Information in Monotone Decision Problems*

Susan Athey
M.I.T. and NBER

Jonathan Levin
M.I.T.

First Draft: September 1997

This Draft: November 1998

Abstract

A seminal theorem due to Blackwell (1951) shows that every Bayesian decision-maker prefers an informative signal Y to another signal X if and only if Y is statistically sufficient for X . Sufficiency is an unduly strong requirement in most economic problems because it does not incorporate any structure the model might impose. In this paper, we develop a general theory of information that allows us to characterize the information preferences of decision-makers based on how their marginal returns to acting vary with the underlying (unknown) state of the world. Our analysis imposes one central restriction: we consider “monotone decision problems,” whereby all decision-makers in the relevant class choose higher actions when higher values of the signal are realized. We show how this restriction can be exploited to characterize information preferences using stochastic dominance orders over the distributions of posterior beliefs generated by different signals. Of particular interest for applied modeling, we identify conditions under which one decision-maker has a higher marginal value of information than another decision-maker, and thus will acquire more information. The results are applied to oligopoly models, labor markets with adverse selection, hiring problems, and a coordination game.

JEL Classification Numbers: C44, C60, D81.

Keywords: Bayesian decision problem, value of information, stochastic dominance, stochastic orderings, decision-making under uncertainty.

*We would like to thank Scott Ashworth, Kyle Bagwell, Glenn Ellison, John Geanakoplos, Bengt Holmström, Eric Lehmann, as well as seminar participants at MIT, Yale, the 1998 Summer Meetings of the Econometric Society, and the Pacific Institute of Mathematical Sciences for helpful discussions. Athey thanks the Cowles Foundation at Yale University and NSF Grant SBR-9631760 for support.

1 Introduction

In a Bayesian decision problem, an agent who is uncertain about the true state of the world must choose an action after observing an imperfectly informative signal. Blackwell (1951, 1953) proved the seminal result that *every* agent faced with such a decision problem will prefer (ex ante) an informative signal \tilde{y} to another signal \tilde{x} if and only if \tilde{y} is statistically sufficient for \tilde{x} . This notion of “better information” is useful and intuitive, but it is also quite demanding (as noted by Blackwell himself).¹ For economic modeling, we might expect that sufficiency is far stronger than needed to compare information structures— in sharp contrast to most economic models, Blackwell’s theorem places no restrictions on the decision-maker’s payoff function.

In this paper, we show that by exploiting properties of the decision-maker’s preferences, it is possible to relax sufficiency and derive comparisons among a richer set of information structures. For many different classes of decision-makers, we derive conditions under which a signal \tilde{y} is preferred to another signal \tilde{x} by all members of the class. In most cases, the informativeness order can be represented as a stochastic dominance ordering over the distribution of posteriors that might arise from each information source. We also provide conditions under which one decision-maker will have a higher marginal value of information than another decision-maker, and thus will acquire more information. The results are applied to oligopoly models, labor markets with adverse selection, hiring problems, and a coordination game.

The starting point of our analysis is to focus attention on *monotone decision problems*, each consisting of a prior belief on the state of the world $H(\omega)$, a class of payoff functions U , and a set of admissible signals $\{\tilde{x}\}$. The class of payoff functions is defined by how the decision-maker’s incremental returns to taking an action (a) vary with the state of the world (ω). For example, in many economic problems, it is assumed that the incremental returns to an action are nondecreasing in the state of the world (i.e., the payoff function $u(\omega, a)$ is supermodular). Given such a class, it is possible to (partially) order posterior beliefs about the state of the world so that “higher” beliefs induce higher actions for all decision-makers in the class. For a signal to be admissible to a monotone decision problem, we require that all of the posteriors that might be generated from different realizations of the signal can be ordered in this way. To illustrate, for the class of decision-makers with supermodular payoffs, higher posteriors in the sense of First Order Stochastic

¹See Blackwell and Girshik (1954). For \tilde{y} to be sufficient for \tilde{x} , all of the posteriors, no matter how unlikely, generated by \tilde{x} must be in the convex hull of the set of posteriors generated by \tilde{y} . There are many unsatisfying examples of distributions which cannot be ranked according to sufficiency. Unless a signal \tilde{x} is normally distributed, \tilde{x} cannot be more informative than a normally distributed signal \tilde{y} in the Blackwell order. See Lehmann (1988) for further examples.

Dominance (FOSD) induce higher actions, and the signal \tilde{x} is admissible only if the corresponding set of posteriors can be totally ordered by FOSD. In general, different classes of decision-makers will induce different “stochastic dominance” orderings over posteriors (for instance, if the marginal returns to acting are concave in the state of the world, the relevant order is Second Order Stochastic Dominance (SOSD)).

We show that a natural information ordering over signals is available for monotone decision problems. We find that a signal \tilde{y} is preferred to \tilde{x} for a class of decision-makers (and thus \tilde{y} is “more informative”), if the “high” posteriors induced by \tilde{y} are (on average) higher than the “high” posteriors induced by \tilde{x} , and the “low” posteriors induced by \tilde{y} are (on average) lower than the “low” posteriors induced by \tilde{x} . The terms “high” and “low” refer to the stochastic dominance order induced by the restriction on the decision-maker’s payoff function.

To fix ideas, return again to the example of supermodular payoff functions. Recalling that the posteriors of the signal \tilde{x} are totally ordered, we can index them with a parameter α_x , so that higher values of the index correspond to higher ordered (by FOSD) posteriors. So that this index will be comparable across signals, it is normalized to have a uniform distribution ex ante: with probability α_x , a posterior lower (in the FOSD order) than $F(\omega|\alpha_x)$ is realized after observing \tilde{x} . We similarly index the posteriors $G(\omega|\alpha_y)$ generated by \tilde{y} . Then a signal \tilde{y} more informative than \tilde{x} for all supermodular payoff functions if, for all $\alpha \in [0, 1]$,

$$G(\omega|\alpha_y \geq \alpha) \succeq_{FOSD} F(\omega|\alpha_x \geq \alpha).$$

The expression $G(\omega|\alpha_y \geq \alpha)$ represents the average over the highest $1 - \alpha$ fraction of the posteriors generated by \tilde{y} . Thus, a signal \tilde{y} is more informative than \tilde{x} if high realizations of ω are more likely when highly-ranked posteriors are realized. For other classes of payoff functions, other stochastic dominance orders (such as SOSD) will be relevant for the comparison of average posteriors.

The conditions we derive are *sufficient* for all decision-makers in a given class to prefer one signal to another; they are *necessary* when we compare small (differential) changes in the signal structure. We also show that the theory can be generalized by considering orders based on single crossing properties rather than stochastic dominance.

Our second objective is to derive conditions under which one decision-maker will have a higher marginal value for information than another, and thus will acquire more information when information is costly. We find that if u and v are in the same class of payoff functions, and the two decision-makers consider purchasing signals ranked according to our criteria, then decision-maker u buys more information than v if u ’s preferences over the distribution of posteriors when using an optimal decision rule are “more sensitive” than v ’s. The meaning of “sensitive” is determined by

the class of payoff functions under consideration. Although our conditions depend on the optimal policies chosen by the agents, and thus are not primitive, they can be verified in many applications.

Our results are related to work in statistics by Lehmann (1988). He considered one specific class of monotone decision problems — those where the decision-maker’s payoffs satisfy a single crossing property, and where the signals satisfy the monotone likelihood ratio property. For such problems, Lehmann (1988) derived a new information ordering that relaxes Blackwell’s sufficiency criterion. Lehmann’s *effectiveness* ordering has already entered economics in the context of auctions (Persico, 1997), principal-agent problems (Jewitt, 1997), and implicit incentive models (Dewatripont, Jewitt and Tirole, 1997). While our methods are quite different – Lehmann uses an approach based on statistical hypothesis testing – we obtain the effectiveness ordering as a special case. Closely related to Lehmann’s work is that of Persico (1996), who studied the same specific class of decision problems. His paper develops an approach to ranking decision-makers in terms of their incentives to acquire information.² Our approach to information acquisition builds directly on his, and provides a significant generalization to other classes of monotone decision problems.

The paper develops as follows. In the next section we describe the model and briefly review the standard approach to information and some preliminary results on stochastic orderings. In Section 3, we introduce the idea of monotone decision problems (MDPs), and discuss some important classes of MDPs. Section 4 includes our main theorems on ordering information structures, and characterizes the monotone information order for several classes of MDPs. Section 5 presents results on ordering payoff functions in terms of their marginal value for information. Section 6 gives some economic applications — to information gathering by firms, adverse selection in labor markets, a coordination game under uncertainty, and a hiring problem. The last section discusses some possible extensions and concludes.

2 The Model

2.1 The Bayesian Decision Problem

A decision-maker (DM) who is uncertain about the true state of the world must take an action after observing an informative signal. The state of the world is denoted $\omega \in \Omega$, where $\Omega \subseteq \mathbb{R}$ is an interval. Let \mathcal{P} denote the set of all probability distributions on Ω . The DM must choose an action $a \in A \subset \mathbb{R}$. Her payoff $u(\omega, a)$ depends on both her action and the true state; we assume u

²Persico (1997) further established that similar techniques can be used to rank the revenue to the auctioneer under different auction formats.

is a bounded measurable function taking $\Omega \times \mathbb{R} \rightarrow \mathbb{R}$. Throughout, we will maintain the following assumption about the set of available actions.

(A) Either A is finite, or A is a compact interval of \mathbb{R} and $u(\omega, a)$ is continuous in a .

The DM has prior distribution $H(\omega)$. Before acting, the DM observes some informative random variable \tilde{x} , with support $\mathcal{X} \subseteq \mathbb{R}$, and forms a posterior distribution $F(\omega|x)$. The joint distribution of (ω, x) is then written $F(\omega, x)$, while the marginal distributions are denoted $F(x)$ and $F(\omega) \equiv H(\omega)$.³ We will refer to F as an “information structure.”

Observe that many different information structures can be equivalent from the perspective of decision-making, since only the posterior generated by a signal realization affects behavior (the value of the signal does not). In particular, it does not matter if the DM observes \tilde{x} or some $T(\tilde{x})$, so long as $F(\omega|T(\tilde{x})=T(x)) = F(\omega|\tilde{x}=x)$ for all x . The payoff-relevant features of an information structure, F , can be *uniquely* characterized in terms of the probability measure induced on the space of posteriors \mathcal{P} , which we write as μ_F .⁴ We will begin by working with this abstract characterization of an information structure before mapping back to the first formulation in terms of the joint distribution $F(\omega, x)$.

The decision-maker’s problem, given a posterior distribution $P \in \mathcal{P}$, is to solve

$$\max_{a \in A} \int_{\Omega} u(\omega, a) dP(\omega) \tag{1}$$

to obtain an optimal action $a^*(P)$, and a realized payoff $u(\omega, a^*(P))$. We define the (ex ante) *value* of the decision problem as

$$V^*(F, u) = \int_{\mathcal{P}} \int_{\Omega} u(\omega, a^*(P)) dP(\omega) d\mu_F(P). \tag{2}$$

2.2 The Classical Approach to Information

The classical approach to information, due to Blackwell (1951, 1953), begins by writing $V^*(F, u)$ as

$$V^*(F, u) = \int_{\mathcal{P}} \left\{ \max_{a(P)} \int_{\Omega} u(\omega, a(P)) dP(\omega) \right\} d\mu_F(P) \equiv \int_{\mathcal{P}} u^*(P) d\mu_F(P). \tag{3}$$

³Note that $F(\omega) = H(\omega)$ is needed to ensure that the expectation of the posterior is the prior. Taking the joint distribution of (ω, x) as primitive is a departure from the analyses of Blackwell (1951) and Lehmann (1988), which take as primitive the distribution of $x|\omega$. This means that their information orders are the same *for all prior distributions* H . Our information rankings will be given *relative to a fixed prior* H . Thus, we may sensibly consider averages over subsets of the posterior distributions.

⁴This construction is introduced in Blackwell’s original article (Blackwell, 1951). Endow \mathcal{P} with the weak topology. Then $F(\omega|x)$ is measurable with respect to the Borel σ -algebra and $\mu_F(E) = \int_{x:F(\omega|x) \in E} dF(x)$.

Using revealed preference, it is straightforward to show that $u^*(P)$ will be convex in P . Simply observe that for $\lambda \in [0, 1]$, $P^1, P^2 \in \mathcal{P}$,

$$\max_a \left\{ \lambda \int_{\Omega} u dP^1 + (1 - \lambda) \int_{\Omega} u dP^2 \right\} \leq \max_a \lambda \int_{\Omega} u dP^1 + \max_a (1 - \lambda) \int_{\Omega} u dP^2. \quad (4)$$

Now suppose G is an alternative information structure which induces a measure μ_G over posteriors distributions, and that μ_G *stochastically dominates* μ_F for *convex functions*, i.e. $\int_{\mathcal{P}} \varphi d\mu_G \geq \int_{\mathcal{P}} \varphi d\mu_F$ for any convex function $\varphi : \mathcal{P} \rightarrow \mathbb{R}$. Clearly, if μ_G stochastically dominates μ_F for convex functions, then $V^*(G, u) \geq V^*(F, u)$ for any payoff function u and action set A , i.e. every decision-maker will find G more valuable than F ; Blackwell (1951) showed the converse via a separation argument. Thus, Blackwell's order can be thought of as a multivariate generalization of the (perhaps more familiar) mean-preserving spread of Rothschild and Stiglitz (1970). The geometric representation of a mean-preserving spread in multiple dimensions can be formalized using what is known as a "dilation," so that μ_G is a dilation of μ_F .⁵ It can further be shown that for a signal \tilde{y} to be sufficient for \tilde{x} , the set of posteriors generated by \tilde{x} must lie in the convex hull of those generated by \tilde{y} .

Sufficiency can be understood in the context of the following three state, two signal example. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$, where $\omega_1 < \omega_2 < \omega_3$, and suppose the prior on ω is uniform $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Consider a signal \tilde{y} that is equally likely to induce posterior beliefs $(\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$ and $(0, \frac{1}{2}, \frac{1}{2})$, and another signal \tilde{x} that is equally likely to induce posterior beliefs $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ and $(\frac{1}{6}, \frac{5}{12}, \frac{5}{12})$. As illustrated in Figure 1, the posteriors generated by \tilde{y} are a mean-preserving spread of the posteriors generated by \tilde{x} . So \tilde{y} will be sufficient for \tilde{x} . But this is clearly very special, since \tilde{y} can be sufficiency ranked relative to another signal \tilde{z} only if the posteriors generated by \tilde{z} lie on the line between $(\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$ and $(0, \frac{1}{2}, \frac{1}{2})$. The ranking will be upset by even the slightest perturbation of \tilde{z} leading to a potential posterior belief off of the line. This lack of robustness is unappealing.⁶ Figure 1 also highlights the fact that *no* two-point information structure can be sufficient for \tilde{y} (the prior is fixed, and so no further spreading is possible). In spite of this, \tilde{y} is clearly not uniquely suited to making all types of inferences. For instance, \tilde{y} provides no information at all as to the relative likelihood of ω_2 versus ω_3 . And while \tilde{y} provides significant information as to whether ω is low or high, many other two-point information structures might be more informative on this question. Since it seems reasonable

⁵ A dilation $D : P \rightarrow D_P$ is a mapping from \mathcal{P} into the set of probability measures on \mathcal{P} such that $\int_{\mathcal{P}} Q dD_P(Q) = P$. That is, D associates with each $P \in \mathcal{P}$ a non-degenerate probability measure on \mathcal{P} with mean P . For our problem, $\mu_G = \int_{\mathcal{P}} D_P d\mu_F(P)$. Strassen (1965) has shown that other stochastic dominance relations also have this sort of representation.

⁶ Indeed, it has motivated work in statistics on "approximate sufficiency"; see Le Cam (1964).

to suppose that particular classes of decision-makers — especially the classes of decision-makers that appear in economic models — care only about making certain types of inferences, the example suggests that it should be possible to derive other informativeness criteria.

Unfortunately, the classical approach is ill-suited to exploiting any structure one might impose on the decision-maker’s preferences. Most important from our perspective, and from that of many economic models, are assumptions about how the returns to taking a higher action, $r(\omega) = u(\omega, a^H) - u(\omega, a^L)$, depend on the state — for instance $r(\omega)$ might be increasing, or concave, or polynomial in the state of the world, or positive if and only if the state variable is greater than some critical level ω_0 . Using the above approach it is hard to see how to incorporate such a restriction. For example, suppose one restricted attention to the class of supermodular payoff functions ($r(\omega)$ increasing in ω). First, it is not clear that this implies anything about $u^*(\omega, P) = u(\omega, a^*(P))$; and second, even if it did, we would still need to connect this property with a meaningful characterization of “better” information. In the next few sections, we will show that by placing a total order over the posteriors (such as a stochastic dominance order), it is possible to overcome these difficulties.

3 Monotone Decision Problems and Stochastic Orders

3.1 Preliminaries: Stochastic Dominance and Single Crossing

Since we will make extensive use of stochastic orderings, we first review a few definitions and known facts.⁷ Assume that U is some set of measurable functions taking $\mathbb{R}^n \rightarrow \mathbb{R}$, and let P^H, P^L be two probability distributions on \mathbb{R}^n . We say that P^H *stochastically dominates* P^L with respect to U , written $P^H \succ_{SD-U} P^L$ if

$$\int u(\mathbf{z}) dP^H(\mathbf{z}) \geq \int u(\mathbf{z}) dP^L(\mathbf{z}) \quad \text{for all } u \in U. \quad (SD)$$

If U is the set of nondecreasing univariate functions, then \succ_{SD-U} is the standard First Order Stochastic Dominance (FOSD) relationship. If U is the set of concave univariate functions, then \succ_{SD-U} corresponds to Second Order Stochastic Dominance (SOSD).

Any set of functions U induces a stochastic dominance order. Some critical features of stochastic dominance orders for our purposes can be understood using the notion of a closed convex cone, as follows.

⁷For further treatments of this material, see *inter alia*, Karlin and Studden (1966), Karlin (1968), Jewitt (1986), Border (1991) and Athey (1998a,1998b).

Definition 1 A set U is a closed convex cone (ccc) if (a) $u, v \in U$ implies that $\alpha u + \beta v \in U$ for any $\alpha, \beta > 0$, and (b) U is closed under the weak topology.

If $\int u dP^H \geq \int u dP^L$ for all $u \in U$, then the same inequality will hold for any function v in the closed convex cone generated by U (denoted $ccc(U)$). Further, since P^H and P^L are probability distributions, then (SD) holds for the functions $u(\mathbf{z}) \equiv 1$ and $u(\mathbf{z}) \equiv -1$ (denoted $\{\mathbf{1}, -\mathbf{1}\}$). Thus, the set U generates the same stochastic dominance order as $ccc(U \cup \{\mathbf{1}, -\mathbf{1}\})$.⁸

A weaker notion than stochastic dominance is that of *stochastic single crossing*.⁹ We write $P^H \succ_{SC-U} P^L$ if

$$\int u(\mathbf{z}) dP^L(\mathbf{z}) \geq 0 \quad \Rightarrow \quad \int u(\mathbf{z}) dP^H(\mathbf{z}) \geq 0 \quad \text{for all } u \in U. \quad (SC)$$

If $\{\mathbf{1}, -\mathbf{1}\} \in U$, then one can show (Athey, 1998a) that \succ_{SC-U} is equivalent to \succ_{SD-U} .¹⁰ If U does not contain constant functions, however, \succ_{SC-U} is weaker than \succ_{SD-U} . The distinction between stochastic dominance and stochastic single crossing will become relevant in our analysis of information orders, since some sets of payoff functions we wish to consider (most notably, single crossing payoff functions, which arise in auction games and portfolio problems) do not contain the constant functions.

The stochastic single crossing order is used in our analysis because it implies a comparative statics prediction about monotonicity of the optimal policy.¹¹

Lemma 1 Let U_1 be some set of functions taking $\Omega \rightarrow \mathbb{R}$. Suppose that $u(\omega, a) : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$ and that for all $a^H > a^L$, $u(\omega, a^H) - u(\omega, a^L) \in U_1$. Let $a^*(P) = \arg \max_{a \in A} \int_{\Omega} u(\omega, a) dP(\omega)$. Then $P^H \succ_{SC-U_1} P^L$ implies that there exists a selection $\hat{a}^*(P)$ from $a^*(P)$ such that $\hat{a}^*(P^H) \geq \hat{a}^*(P^L)$.

⁸In fact, just this insight allows us to characterize stochastic dominance orderings, since we can also consider a the order induced by a much smaller set E_U (for the case of nondecreasing functions, a set of indicator functions), loosely referred to as extreme points, so long as $U = ccc(E_U \cup \{\mathbf{1}, -\mathbf{1}\})$. See Border (1991) or Athey (1998a) for more discussion.

⁹Various definitions of single crossing properties (using different combinations of weak and strict inequalities) have been proposed by different authors in economics (Milgrom and Shannon, 1994; Shannon, 1995) and statistics (Karlin, 1968). The definition in (SC) is not the most natural variation for comparative statics theorems. However, as it involves only weak inequalities, (SC) is especially appropriate for working with closed convex cones and stochastic orders.

¹⁰This result hinges on the assumption that P^H and P^L are probability distributions.

¹¹To simplify the notation and the statement of the result we consider only sufficiency here, but with minor qualifications, the single crossing condition is in fact necessary for comparative statics as well. See Milgrom and Shannon (1994) and Shannon (1995).

Proof. Define a function $v(\tau, a) : \{0, 1\} \times \mathbb{R} \rightarrow \mathbb{R}$, with $v(1, a) = \int_{\Omega} u(\omega, a) dP^H(\omega)$ and $v(0, a) = \int_{\Omega} u(\omega, a) dP^L(\omega)$. Then $v(\tau, a)$ is bivariate weak single crossing in (τ, a) and by Shannon (1995), there exists a selection from $a^*(\tau) = \arg \max_{a \in A} v(\tau, a)$ that is nondecreasing. \square

3.2 Monotone Decision Problems

Our approach is to restrict attention to particular classes of decision problems. Each class will be characterized by a restriction on the DM's payoff function and a corresponding restriction on the set of admissible information structures.

Definition 2 *The pair (U_2, \mathcal{F}) constitutes a class of monotone decision problems if there exists a prior $H(\omega)$ and some set U_1 of bounded measurable functions taking Ω into \mathbb{R} , such that:*

(MDP-U) *For all $u \in U_2$, if $a^H, a^L \in \mathbb{R}$ and $a^H \geq a^L$, then $u(\omega, a^H) - u(\omega, a^L) \in U_1$. And moreover, if $u : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$ is some bounded measurable function with incremental returns in U_1 , then $u \in U_2$.*

(MDP-F) *For all $F \in \mathcal{F}$, $F(\omega) \equiv H(\omega)$, and further if $x^H, x^L \in \text{support}(F)$ and $x^H \geq x^L$, then $F(\omega|x^H) \succ_{SC-U_1} F(\omega|x^L)$. And moreover, if F is an information structure with prior $H(\omega)$ and posteriors completely ordered in x by \succ_{SC-U_1} , then $F \in \mathcal{F}$.*

If (U_2, \mathcal{F}) satisfy this definition, we refer to them as an *MDP pair*. The first condition, (*MDP-U*), implies that every DM under consideration has incremental returns to acting that rely on the state in some pre-specified way (for example, they might be nondecreasing or concave); the second, (*MDP-F*), implies that for every admissible information structure, the induced posterior beliefs are completely ordered in the sense of stochastic single crossing.

What do these restrictions buy us? Let (U_2, \mathcal{F}) be an *MDP pair*. By (*MDP-U*) and Lemma 1, each DM $u \in U_2$ has an optimal policy $a^*(P)$ that is monotone in P when posteriors are ordered by \succ_{U_1-SC} . And (*MDP-F*) implies that if $F \in \mathcal{F}$, then *every* posterior associated with F can be ordered in this way and indexed by x . Thus, each DM $u \in U_2$ takes higher actions when she receives higher realizations of the signal.

In this way, the *MDP* restrictions impose an *ordinal* structure, so that one can heuristically think of the posteriors generated by \tilde{x} going from low to high along a single dimension. However, the goal in this paper is to compare different information structures. While (*MDP-F*) implies that the posteriors for a given signal can be totally ordered, it provides no guidance as to how to compare the posteriors arising from different signals. Thus, we introduce a *cardinal* index for the

posteriors generated by each signal, $\alpha \in [0, 1]$, so that we may “match” for comparison posteriors from different signals. This is accomplished using a strictly increasing function $T_F : \mathcal{X} \rightarrow [0, 1]$, so that $T_F(x) = \alpha$.

For an important set of cases, we will show that it is appropriate to match posteriors according to their percentile in the ex ante signal distribution. Formally, for an information structure $F(\omega, x)$ with marginal distribution $F(x)$, we let $T_F(x) = F(x)$. Using this index, we can let $\tilde{F}(\omega, \alpha) = F(\omega, F^{-1}(\alpha))$ on $\Omega \times [0, 1]$,¹² so that $\tilde{F}(\omega|\alpha)$ is the “ α -percentile posterior.” The probability of observing a posterior below $\tilde{F}(\omega|\alpha)$ in the \succ_{U_1-SC} order is given by α , so that $\tilde{F}(\alpha)$ is the uniform distribution on $[0, 1]$. As illustrated in Figure 2, we can use such an index to compare the realizations of two signals, \tilde{x} and \tilde{y} , according their α -percentile.

Using this construction, we can represent a decision-maker’s policy function $a : [0, 1] \rightarrow A$ in terms of the action it prescribes for an α -indexed posterior. Policy functions $a(\alpha)$ can then be analyzed without reference to a particular information structure. Using this representation, for any probability distribution \tilde{F} on $\Omega \times [0, 1]$, we define the ex ante expected value from using the policy a by

$$V(\tilde{F}, u, a) = \int_{[0,1]} \int_{\Omega} u(\omega, a(\alpha)) d\tilde{F}(\omega, \alpha). \quad (5)$$

An important consequence of $(MDP-U)$ follows: for any nondecreasing $a(\alpha)$,

$$u(\omega, a(\alpha^H)) - u(\omega, a(\alpha^L)) \in U_1 \text{ for } \alpha^H > \alpha^L. \quad (6)$$

This in turn implies that $u(\omega, a(\alpha))$ viewed as a function of (ω, α) can be extended to be in U_2 . In words, whenever the DM uses a monotone policy, the incremental returns to having a higher α -indexed posterior retain the properties we assumed for the incremental returns of the primitive payoff function, and thus $u(\omega, a(\alpha))$ inherits the properties specified by U_2 .

The discussion of this subsection can be formally summarized as follows.

Theorem 2 *Suppose (U_2, \mathcal{F}) are an MDP pair and that (A) holds. Consider any $(u, F) \in (U_2, \mathcal{F})$. Then for any $T_F : \mathcal{X} \rightarrow [0, 1]$ continuous and strictly increasing, $\tilde{F}(\omega, \alpha) \equiv F(\omega, T_F^{-1}(\alpha)) \in \mathcal{F}$. Further, there exists some nondecreasing $a^F : [0, 1] \rightarrow A$ such that:*

¹²We can, without loss of generality, take $F(x)$ to be continuous and strictly increasing. Lehmann (1988, p. 527) provided a construction which shows that one can always take $F(x)$ to be continuous. If $F(x)$ is constant over some interval $[x_0, x_1]$, notice that the DM would experience no payoff loss from observing $x^* = x$ on $x < x_0$, $x^* = x_0$ on $[x_0, x_1]$ and $x^* = x - (x_1 - x_0)$. And $F^*(x^*)$ will be strictly increasing (one needs a more involved construction if \mathcal{X} is not compact). Then $F(x) : \mathcal{X} \rightarrow [0, 1]$ is a bijection (F is continuous and strictly monotone), so \tilde{F} is well defined.

(i) For all $\alpha \in [0, 1]$, $a^F(\alpha) \in \arg \max_{a \in A} \int_{\Omega} u(\omega, a) d\tilde{F}(\omega|\alpha)$.

(ii) $V(\tilde{F}, u, a^F(\cdot)) = V^*(F, u)$.

(iii) For any $\alpha^H > \alpha^L$, $u(\omega, a^F(\alpha^H)) - u(\omega, a^F(\alpha^L)) \in U_1$.

(iv) There exists $u^F : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$, $u^F \in U_2$, with $u^F(\omega, \alpha) = u(\omega, a^F(\alpha))$.

Proof. (i) Because A is compact, for all $P \in \mathcal{P}$, $\int_{\Omega} u(\omega, a) dP(\omega)$ attains a maximum on A . For any $\alpha^H \geq \alpha^L$, monotonicity of T and $(MDP-\mathcal{F})$ implies that $\tilde{F}(\omega|\alpha^H) \succ_{SC-U_1} \tilde{F}(\omega|\alpha^L)$. So from Lemma 1, there will exist a nondecreasing selection from the set of optimizers, denoted $a^F(\alpha)$. (ii) By (i), $\int_{\Omega} u(\omega, a^F(\alpha)) d\tilde{F}(\omega|\alpha) = \max_{a \in A} \int_{\Omega} u(\omega, a) dF(\omega|T_F^{-1}(\alpha))$. The result then follows immediately from the definitions of the distributions. (iii) Let $a^H = a^F(\alpha^H)$, $a^L = a^F(\alpha^L)$ and note that $a^H \geq a^L$. But then $u(\omega, a^F(\alpha^H)) - u(\omega, a^F(\alpha^L)) = u(\omega, a^H) - u(\omega, a^L) \in U_1$ by $(MDP-U)$. (iv) The function $u^F(\omega, \alpha)$ can be defined to equal $u(\omega, a^F(\alpha))$ for $\alpha \in [0, 1]$, to equal $u(\omega, a^F(0))$ for $\alpha < 0$, and to equal $u(\omega, a^F(1))$ for $\alpha > 1$. It follows from $(MDP-U)$ that $u^F \in U_2$. \square

3.3 Examples

Our framework covers many problems of economic interest. In this section, we discuss four common classes of problems and then show briefly how the theory can be applied to other cases.

Example 1. *Supermodular (Incremental returns increasing in ω).* Suppose U_1 is the set of nondecreasing functions. Then U_2 is the set of functions $u(\omega, a)$ that are supermodular in (ω, a) . In words, the incremental return to a higher action is nondecreasing in the state of the world. Numerous economic applications take this form (see Milgrom and Roberts (1990)); for example, a might represent the level of investment for a firm, where the marginal returns are indexed by ω . Since U_1 contains constant functions, we have $P^H \succ_{SC-U_1} P^L$ if and only if P^H is higher than P^L according to FOSD, denoted $P^H \succ_{FOSD} P^L$. So $(MDP-\mathcal{F})$ implies that if $F \in \mathcal{F}$, the induced posteriors must be ranked by FOSD, which requires that $F(\omega|x)$ is decreasing in x : a higher signal corresponds to a higher probability that the state of the world is high. Figure 3 illustrates the FOSD ordering over posteriors for a 3-state example.

Example 2. *Concave returns (Incremental returns concave in ω).* Suppose U_1 is the set of concave functions. Then U_2 is the set of functions $u(\omega, a)$ that have concave incremental returns. Such a

payoff function can arise naturally in binary decision problems ($a \in \{0, 1\}$) where the DM must decide whether or not to undertake some risky venture with concave payoff $r(\omega)$ (in Section 5 we present a hiring problem of this type). More generally, if $u \in U_2$, then the DM becomes strongly more risk averse as she raises her action.¹³ For this case $P^H \succ_{SC-U_1} P^L$ if and only if P^H dominates P^L according to SOSD ($P^H \succ_{SOSD} P^L$): the DM takes a higher action when the posterior about ω is “less risky.” So if $F \in \mathcal{F}$, then $F(\omega|x^H) \succ_{SOSD} F(\omega|x^L)$. This requires that $E[\omega|x]$ is constant in x and for any ω , $\int_{-\infty}^{\omega} F(\tilde{\omega}|x)d\tilde{\omega}$ is nonincreasing in x . Alternatively, for any $x^L < x^H$, $F(\omega|x^L)$ can be attained from $F(\omega|x^H)$ by a sequence of mean-preserving spreads (Rothschild and Stiglitz, 1970).

Example 3. *WSC(ω_0) (Incremental returns weak single crossing at ω_0).* Suppose U_1 is the set of functions $r(\omega)$ such that $r(\omega) \leq 0$ for $\omega < \omega_0$ and $r(\omega) \geq 0$ for $\omega > \omega_0$, i.e. the functions that cross zero from below at ω_0 . We say such a function satisfies $WSC(\omega_0)$. Payoff functions in this class arise in the context of investment under uncertainty problems, where ω might represent the return on a risky asset, ω_0 is the return on a risk-free asset, a is the portfolio weight on the risky asset, and investor payoffs are given by $v(a\omega + (1-a)\omega_0)$. In this case, if P^H, P^L have densities p^H, p^L with respect to Lebesgue measure, then $P^H \succ_{SC-U_1} P^L$ if and only if $p^H(\omega) - \frac{p^H(\omega_0)}{p^L(\omega_0)}p^L(\omega)$ satisfies $WSC(\omega_0)$ in ω (Athey, 1998b). Thus if $F \in \mathcal{F}$, this reduces to $\frac{f(\omega|x^H)}{f(\omega|x^L)} \leq (\geq) \frac{f(\omega_0|x^H)}{f(\omega_0|x^L)}$ as $\omega < (>)\omega_0$. A high signal means that it is more likely that the true state ω is greater than ω_0 . Since the class of $WSC(\omega_0)$ functions does not contain constant functions, a stronger condition is required to order posteriors by stochastic dominance. For convenience, define the set of functions $\overline{WSC}(\omega_0)$ as the set of functions $r(\omega)$ satisfying $WSC(\omega_0)$ and $r(\omega_0) = 0$. We have $P^H \succ_{SD-U_1} P^L$ if and only if $p^H(\omega) - p^L(\omega)$ is $\overline{WSC}(\omega_0)$. In particular, the densities must cross at ω_0 , a restriction not imposed by \succ_{SC-U_1} .

Example 4. *Weak Single Crossing (Incremental returns single crossing in ω).* Suppose U_1 is the set of functions $r(\omega)$ that cross zero from below at some point ω_0 . In other words, U_1 is the union of all the $WSC(\omega_0)$ sets. Payoff functions $u(\omega, a)$ with incremental returns that are single crossing also arise throughout economics — for instance in bidding and pricing problems.¹⁴ For the class of single crossing functions, we have $P^H \succ_{SC-WSC} P^L$ if and only if $P^H \succ_{SC-WSC(\omega_0)} P^L$ for all ω_0 , which implies that (where the densities exist) $p^H(\omega)/p^L(\omega)$ is nondecreasing in ω . This order

¹³Recall that $r^H(\omega)$ is strongly more risk averse than $r^L(\omega)$ if there is some $\lambda \geq 0$ such that $r^H(\omega) - \lambda r^L(\omega)$ is concave. See Ross (1981) or Jewitt (1986).

¹⁴See Milgrom and Shannon (1994) for an extensive discussion of the single crossing property, and Athey (1998b) for applications in stochastic problems.

has received a great deal of attention in economics and statistics, and it is known as the Monotone Likelihood Ratio (MLR) order ($P^H \succ_{MLR} P^L$). Then if $F \in \mathcal{F}$, $F(\omega|x)$ must be ordered by MLR, and if a density exists, $f(\omega|x^H)/f(\omega|x^L)$ is nondecreasing in ω .¹⁵

Note that the class of payoff functions with single crossing incremental returns includes the class of supermodular payoff functions and the class of payoff functions with incremental returns that are $WSC(\omega_0)$. But expanding the set of payoff functions comes at a cost — we have to limit the set of information structures that we can attempt to compare. For instance, requiring the posteriors to be ordered by MLR is significantly stronger than requiring the posteriors to be ordered by FOSD.¹⁶ This is illustrated in Figure 3.

More generally, our theory applies if U_1 is some arbitrary closed convex cone of functions, such as linear functions or quadratic functions or functions with positive n^{th} derivatives. For any such case, there is a rich theory of stochastic dominance that allows one to characterize the relation \succ_{SD-U_1} , and hence the relevant restriction on posterior beliefs, in terms of the “extreme points” of the cone U_1 .¹⁷ We leave this pursuit to the reader.

4 Ordering Information Structures

This section contains our main results on ordering information structures. We begin by identifying a *sufficient* condition for all DMs with payoffs $u \in U_2$ to prefer an information structure $G \in \mathcal{F}$ to another information structure $F \in \mathcal{F}$. We call this condition the “monotone information order” (*MIO*) for (U_2, \mathcal{F}) ; it can be characterized using the stochastic dominance order induced by U_1 . We show that (*MIO*) is not just sufficient, but also *necessary*, for small (differential) changes in the information structure when U_1 is a closed convex cone that contains the constant functions. Section 4.2 provides a general analysis of informativeness for classes of MDPs based on stochastic

¹⁵Athey (1998b) also shows that the set of log-supermodular payoff functions (where u is positive and $u(\omega, a^H)/u(\omega, a^L)$ is nondecreasing in ω) induce the same stochastic single crossing order, \succ_{SC-U_1} , despite the fact that the set of log-supermodular functions is smaller than the set of payoff functions with incremental returns that are single crossing. However, the set of log-supermodular payoffs does not satisfy (*MDP-U*), and thus our analysis below does not apply directly to this class of payoffs. To see an example where log-supermodular payoffs arise, suppose that the action a is price, firms maximize $(a - c)D(a, \omega)$, and higher states of the world correspond to more inelastic demand.

¹⁶In particular, assuming F is indexed by the MLR is equivalent to assuming the $F(\omega|x)$ is ordered by FOSD for all prior distributions $H(\omega)$ (Milgrom, 1981).

¹⁷That is, we check (SD) for some set of functions E_U such that $ccc(E_U \cup \{\mathbf{1}, -\mathbf{1}\}) = U$. See Karlin and Studden, 1966; Border, 1991; Jewitt, 1986; Athey 1998a.

single crossing. Recall from Section 3.1 that when U_1 does not contain the constant functions, \succ_{SC-U_1} is weaker than \succ_{SD-U_1} . Thus, the ordering based on stochastic single crossing can be used to compare a greater variety of information structures when U_1 does not contain constant functions. We derive Lehmann's (1988) effectiveness order as a corollary, and relate our work to his. Examples are provided.

4.1 Monotone Information Orders Using Stochastic Dominance

We begin by stating a sufficient condition for G to be more informative than F to a given class of decision-makers.

Theorem 3 *Let (U_2, \mathcal{F}) be an MDP pair and suppose (A) holds. Consider any $F, G \in \mathcal{F}$. Then $V^*(G, u) \geq V^*(F, u)$ for all $u \in U_2$ if for all $\alpha \in (0, 1)$*

$$G(\omega|G(y) \geq \alpha) \succ_{SD-U_1} F(\omega|F(x) \geq \alpha). \quad (MIO)$$

If the assumptions of Theorem 3 hold and (MIO) obtains, we write $G \succ_{MIO-U_1} F$, i.e. G is greater than F in the monotone information order induced by payoff functions with incremental returns in U_1 . The condition (MIO) is easy to interpret. It says that the high posteriors induced by G (where "high" means $G(\tilde{y}) \geq \alpha$) are on average higher (according to stochastic dominance) than the corresponding high posteriors induced by F .¹⁸ This condition is also equivalent to saying that the low posteriors induced by G are lower: observe that

$$\alpha F(\omega|F(x) \leq \alpha) + (1 - \alpha)F(\omega|F(x) > \alpha) = H(\omega), \quad (7)$$

where $H(\omega)$ is the prior. So an equivalent expression to (MIO) is that for all $\alpha \in [0, 1]$,

$$F(\omega|F(y) \leq \alpha) \succ_{SD-U_1} G(\omega|G(x) \leq \alpha). \quad (8)$$

Notice that in this condition, signals are implicitly indexed by their *ex ante* percentile. This motivates us to use the index $\alpha_x = F(x)$, as described in Section 3.2 and illustrated in Figure 2. Thus, throughout this subsection, we let $\tilde{F}(\omega, \alpha) = F(\omega, F^{-1}(\alpha))$ and let $\tilde{G}(\omega, \alpha) = G(\omega, G^{-1}(\alpha))$.

¹⁸Analogous to Blackwell, one can interpret (MIO) as saying that the G posteriors are "more spread out" than the F posteriors. To see this, note that (MIO) is equivalent to saying that for all $\alpha \in [0, 1]$,

$$\int_0^\alpha G(\omega|G(y) = \tilde{\alpha})d\tilde{\alpha} \prec_{SD-U_1} \int_0^\alpha F(\omega|F(x) = \tilde{\alpha})d\tilde{\alpha}$$

which can be interpreted in a manner similar to the second order stochastic dominance condition for comparing two distributions, F^H and F^L , on $[0, 1]$, which requires that $\int_0^z F^H(\tilde{z})d\tilde{z} \leq \int_0^z F^L(\tilde{z})d\tilde{z}$ with equality when $z = 1$.

Proof of Theorem 3. We show that (MIO) actually implies a stronger result, namely that under (MIO), for any $u \in U_2$, $V(\tilde{G}, u, a) \geq V(\tilde{F}, u, a)$ for any $a(\alpha)$ nondecreasing. This will imply that G is more informative than F under the conditions of Theorem 3. To see this, observe that $a^F(\alpha) \in \arg \max_a \int_{\Omega} u(\omega, a) dF(\omega|F(x)=\alpha)$ is nondecreasing, and thus by revealed preference

$$V^*(G, u) \geq V(\tilde{G}, u, a^F) \geq V(\tilde{F}, u, a^F) \equiv V^*(F, u).$$

Suppose that $A = \{a_1, \dots, a_n\}$ is finite, with $a_{i+1} > a_i$. Define

$$r_i(\omega) \equiv u(\omega, a_{i+1}) - u(\omega, a_i).$$

Consider some arbitrary monotone increasing policy $a : [0, 1] \rightarrow A$. We can find $\alpha_0 = 0 \leq \alpha_1 \leq \dots \leq \alpha_{n-1} \leq \alpha_n = 1$ such that $a(\alpha) = a_i$ on $[\alpha_{i-1}, \alpha_i]$. Then we have

$$\begin{aligned} V(\tilde{F}, u, a) &\equiv \int_{\Omega} \int_{[0,1]} u(\omega, a(\alpha)) d\tilde{F}(\omega, \alpha) \\ &= \int_{\Omega} \sum_{i=1}^n u(\omega, a_i) \left[\tilde{F}(\alpha_i|\omega) - \tilde{F}(\alpha_{i-1}|\omega) \right] d\tilde{F}(\omega) \\ &= E[u(\omega, a_1)] + \int_{\Omega} \left\{ \sum_{i=1}^{n-1} [u(\omega, a_{i+1}) - u(\omega, a_i)] \left[1 - \tilde{F}(\alpha_i|\omega) \right] \right\} d\tilde{F}(\omega) \\ &= E[u(\omega, a_1)] + \sum_{i=1}^{n-1} (1 - \alpha_i) \int_{\Omega} r_i(\omega) d\tilde{F}(\omega|\tilde{\alpha}_x \geq \alpha_i) \\ &\leq E[u(\omega, a_1)] + \sum_{i=1}^{n-1} (1 - \alpha_i) \int_{\Omega} r_i(\omega) d\tilde{G}(\omega|\tilde{\alpha}_y \geq \alpha_i) = V(\tilde{G}, u, a). \end{aligned}$$

The equalities follow by algebraic manipulation and Bayes' rule. The inequality follows directly from (MIO) since, by (MDP-U), we know that for each i , $r_i(\omega) \in U_1$. The case of A compact follows via a limiting argument and is deferred until the next section. \square

The idea in the proof is very simple. Starting from action a_1 , every new increment to the choice of action entails a new (interim) gamble on ω , described by $r_i(\omega) \equiv u(\omega, a_{i+1}) - u(\omega, a_i) \in U_1$. Formally, payoffs for a given action can be written as the sum over these incremental gambles:

$$u(\omega, a_i) = u(\omega, a_1) + \sum_{j=1}^{i-1} r_j(\omega). \quad (9)$$

Jumping up to a higher action (say, from a_i to a_{i+1}) at some posterior indexed by $F(x) = \alpha_i$ means taking on the gamble for all higher-ranked posteriors as well, since the DM uses a monotone policy.

The effect on *ex ante* payoffs is then given by $E[r_i(\omega)|F(\tilde{x}) \geq \alpha_i]$. Preferences over the gambles $r_i(\omega)$ are described by \succ_{SD-U_1} , and (MIO) requires that *every* such gamble is more favorable under G than under F . As this argument applies when the DM uses the optimal policy for signal \tilde{x} , there is no possible way for a decision-maker in the class to do better, from an *ex ante* standpoint, using F instead of G .

In the proof, we showed that (MIO) implies $V(\tilde{G}, u, a) \geq V(\tilde{F}, u, a)$ for *any* $u \in U_2$ and $a(\alpha)$ nondecreasing. Since (MIO) has such powerful consequences, it might seem to be “too strong” as an informativeness order. At a minimum, it seems reasonable to compare $V(\tilde{G}, u, a^F) \geq V(\tilde{F}, u, a^F)$ only for policies a^F which are optimal for information structure F . However, for important classes of monotone decision problems, we show that restricting the policy function in this way does not permit a weakening of (MIO).

To see why this might be true, consider the simplest case of $A = \{0, 1\}$ and some $u \in U_2$. The optimal policy for u (denoted $a^{F,u}$) entails jumping from $a = 0$ to $a = 1$ at a posterior indexed by $F(\tilde{x}) = \alpha^{F,u}$. Now consider utility functions of the form $w(\omega, a) = u(\omega, a) + aK$, where $K \neq 0$ is an arbitrary constant. There is no particular reason for the policy of jumping to $a = 1$ at $\alpha^{F,u}$ to be optimal for w under F ; however, a DM with payoff w who does use $a^{F,u}$ (suboptimally), will prefer G to F if and only if the DM with payoff u does as well:

$$V(\tilde{G}, u, a^{F,u}) - V(\tilde{F}, u, a^{F,u}) - [V(\tilde{G}, w, a^{F,u}) - V(\tilde{F}, w, a^{F,u})] \quad (10)$$

$$= (1 - \alpha) \int_{\Omega} K dG(\omega|G(\tilde{y}) \geq \alpha) - (1 - \alpha) \int_{\Omega} K dF(\omega|F(\tilde{x}) \geq \alpha) = 0. \quad (11)$$

Similarly, a DM with payoff u who uses $a^{F,w}$ (suboptimally) prefers G to F if and only if $V(\tilde{G}, w, a^{F,w}) - V(\tilde{F}, w, a^{F,w})$.¹⁹ To extend this idea, suppose that for every K , $w_K = u + aK \in U_2$. Then by varying K , we potentially can recover *any* policy of the form $a = 0$ for $F(\tilde{x}) < \alpha$, $a = 1$ for $F(\tilde{x}) \geq \alpha$ as the optimal policy of some w_K . And thus for *all* payoff functions $\{w_K \in U_2\}$ to prefer G to F using their optimal response to F , it will be *necessary* that $V(\tilde{G}, u, a) \geq V(\tilde{F}, u, a)$ for all monotone policies $a(\alpha)$.

This heuristic argument can be made exact under the following condition:

$$\underline{(\mathbf{U}_1\text{-C})} \quad U_1 = ccc(U_1 \cup \{\mathbf{1}, -\mathbf{1}\}).$$

¹⁹Note that this argument turns critically on our decision (implicit in (MIO)) to compare G to F using (fixed) policy functions that depend on a posterior’s rank in the *ex ante* distribution, $\alpha_y = G(y)$. Had we used policy functions $a : [0, 1] \rightarrow A$ based on some other index (not the marginal), the policy of increasing the action at $\alpha_y = \alpha$ for information structure G would no longer correspond to choosing a higher action when $G(\tilde{y}) \geq \alpha_i$ and (11) would fail.

That is, U_1 is a closed convex cone, and it contains the constant functions. Under this condition, $r \in U_1$ implies that $r + K \in U_1$ for any K . Thus – crucially for our purposes – if $u \in U_2$, then $u + aK \in U_2$.

The following theorem establishes that, when (U_1-C) holds, then for small (differential) changes in the information structure, (MIO) is both *necessary and sufficient* for a ranking of information structures.

Theorem 4 *Let (U_2, \mathcal{F}) be an MDP pair such that (U_1-C) and (A) holds. Let $F^\theta(\omega, x)$ be smoothly parametrized by θ , with $F^\theta \in \mathcal{F}$ for all θ . Then $\frac{d}{d\theta} V^*(F^\theta, u) \geq 0$ for all $u \in U_2$, if and only if $F^{\theta+d\theta} \succ_{MIO-U_1} F^\theta$.*

Proof. Sufficiency follows from above. We prove a stronger result, that (MIO) is necessary for $V(\tilde{G}, u, a^F) \geq V(\tilde{F}, u, a^F)$, where a^F is the optimal policy for information structure F . We proceed by assuming that (MIO) fails and constructing a contradiction. Suppose $G \not\succeq_{MIO-U_1} F$. Then there is some $\bar{\alpha}$, and $\bar{r} \in U_1$ such that

$$\int_{\Omega} \bar{r}(\omega) dG(\omega|G(\tilde{y}) \leq \bar{\alpha}) > \int_{\Omega} \bar{r}(\omega) dF(\omega|F(\tilde{x}) \leq \bar{\alpha}). \quad (12)$$

Let $A = \{0, 1\}$. Then compute $K_{\bar{\alpha}} \equiv E[\bar{r}(\omega)|F(\tilde{x}) = \bar{\alpha}]$, and define

$$\bar{u}(\omega, a) = a(\bar{r}(\omega) - K_{\bar{\alpha}}).$$

By (U_1-C) , since $\bar{r} \in U_1$, then $\bar{u} \in U_2$. An optimal policy for this DM is to set $a^F(\alpha) = 1$ if $\alpha > \bar{\alpha}$ and zero otherwise. Then $V(\tilde{G}, u, a^F) \geq V(\tilde{F}, u, a^F)$, if and only if

$$\int_{[0,1]} \int_{\Omega} \bar{u}(\omega, a^F(\alpha)) dG(\omega, G^{-1}(\alpha)) \geq \int_{[0,1]} \int_{\Omega} \bar{u}(\omega, a^F(\alpha)) dF(\omega, F^{-1}(\alpha)) \quad (13)$$

$$\iff \int_{\Omega} [\bar{r}(\omega) - K_{\bar{\alpha}}] dG(\omega|G(\tilde{y}) \leq \bar{\alpha}) \leq \int_{\Omega} [\bar{r}(\omega) - K_{\bar{\alpha}}] dF(\omega|F(\tilde{x}) \geq \bar{\alpha}) \quad (14)$$

But, the expected value of a constant function is simply that constant. So (13) and (14) contradict (12).

For the case of small changes, consider F^θ , a smoothly parameterized family, and let $a^\theta(\alpha)$ be the optimal policy. Then, by the envelope theorem, $\frac{d}{d\theta} V(F^\theta, u, a^\theta)|_{\theta=\bar{\theta}} = \frac{\partial}{\partial \theta} V(F^\theta, u, a^{\bar{\theta}})|_{\theta=\bar{\theta}}$. Then apply the above argument to compare $F^{\bar{\theta}+d\theta}$ and $F^{\bar{\theta}}$ at the policy $a^{\bar{\theta}}$. \square

It is also possible to view the necessity result through the lens of statistical hypothesis testing. Consider testing the null hypothesis $H_0 : r(\omega) \geq 0$ against the alternative $H_1 : r(\omega) < 0$, for some

$r(\omega) \in U_1$. However, we impose a constraint on the decision-maker: the test has an *average size* of $1 - \alpha$. That is, the *ex ante* probability (over all possible signals) of accepting the null hypothesis is $1 - \alpha$. A decision-maker facing this problem would then solve:

$$\begin{aligned} \max_{a(x) \in \{0,1\}} \quad & \int_{\mathcal{X}} \int_{\Omega} a(x)r(\omega)dF(\omega, x) \\ \text{s.t.} \quad & \int_{\mathcal{X}} a(x)dF(x) = 1 - \alpha \end{aligned}$$

Assuming the conditions of Theorem 4 apply, an optimal policy for this testing problem will set $a(x) = 1$ if and only if $F(\tilde{x}) \geq \alpha$. The *ex ante* expected payoff will be $(1 - \alpha) \cdot E[r(\omega)|F(\tilde{x}) \geq \alpha]$, which, by (MIO), is larger than the payoff from solving the inference problem associated with G , $(1 - \alpha) \cdot E[r(\omega)|G(\tilde{y}) \geq \alpha]$. Thus, (MIO) can be interpreted as requiring that better information allows (on average) better inference about the returns to taking a higher action.²⁰

4.1.1 Examples

We now revisit two of our examples from above.

Example 1. *Supermodular (Incremental returns increasing in ω).* Since U_1 is the set of non-decreasing functions, the relation \succ_{SD-U_1} corresponds to FOSD. Thus G is higher than F in the supermodular monotone information order (MIO-SPM) if and only if for all $\alpha \in (0, 1)$, $\tilde{G}(\omega|\tilde{\alpha}_y \geq \alpha) \succ_{FOSD} \tilde{F}(\omega|\tilde{\alpha}_x \geq \alpha)$. That is, high signals from G lead on average to higher posterior beliefs than high signals from F . Recall from above that if $F, G \in \mathcal{F}$, then $\tilde{F}(\omega|\alpha)$, $\tilde{G}(\omega|\alpha)$ are increasing in the sense of FOSD as α increases. So a higher signal is “good news” about the state of the world. If $G \succ_{MIO-U_1} F$ then high signals are (on average) “better news” under G than under F . Since high signals lead to high actions, one can think intuitively about G bringing about a better match between actions and the true state of the world.

It is interesting to relate (MIO-SPM) to sufficiency. To do this, we return to our earlier three-state, two-signal example. Recall the DM has uniform prior over the three states $\omega_1 < \omega_2 < \omega_3$. As illustrated in Figure 4, let \tilde{y} be a signal that put equal likelihood on posteriors $(\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$ and $(0, \frac{1}{2}, \frac{1}{2})$. Note that \tilde{y} is admissible for supermodular MDPs since $(0, \frac{1}{2}, \frac{1}{2}) \succ_{FOSD} (\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$. Consider an information structure \tilde{z} that puts equal likelihood on posteriors $(\frac{2}{3}, \frac{1}{3}, 0)$ and $(0, \frac{1}{3}, \frac{2}{3})$. Clearly \tilde{z} is also admissible for supermodular MDPs. Further, from the discussion in Section 3.3, \tilde{y} and \tilde{z}

²⁰To see why this alternative approach is essentially equivalent to our necessity proof, let $\bar{\lambda}$ be the Lagrange multiplier on the constrained optimization problem for \bar{r} , $\bar{\alpha}$ (defined as in the proof). Then \bar{u} in the proof equals $a(\bar{r}(\omega) + \bar{\lambda})$.

cannot be compared by Blackwell’s sufficiency criteria. However, all supermodular decision-makers prefer \tilde{z} to \tilde{y} , i.e. $\tilde{z} \succ_{MIO-SPM} \tilde{y}$. The reason is simple: \tilde{z} ’s posteriors are more spread out according to FOSD. When the DM sees a low signal realization from \tilde{z} , her beliefs about the state are truly pessimistic, while the converse holds for high signal realizations.

Figure 5 illustrates this shift in a different way. For each signal, we label the lower (by FOSD) posterior L , and the higher posterior H . We see that signals that are higher according to (*MIO-SPM*) place more probability weight on (ω_2, L) , and (ω_3, H) . In fact, using ideas from bivariate stochastic dominance, it is possible to show that (*MIO-SPM*) can in general be characterized this way. Results from Meyer (1991) can be used to show that $G \succ_{MIO-SPM} F$ if and only if G is obtained from F using a “marginal-preserving spread” of the form illustrated in Figure 3.²¹

As a further comparison, consider the signal structure \tilde{x} that puts equal likelihood on posteriors $(\frac{2}{3}, 0, \frac{1}{3})$ and $(0, \frac{2}{3}, \frac{1}{3})$. It follows easily that \tilde{x} is admissible for supermodular decision problems and that $\tilde{y}, \tilde{z} \succ_{MIO-SPM} \tilde{x}$. Yet no two-point information structure is sufficient for \tilde{x} . Moreover, if we perturb \tilde{x} so that it generates posteriors $(\frac{2}{3} - \varepsilon, \varepsilon + \eta, \frac{1}{3} - \eta)$ and $(\varepsilon, \frac{2}{3} - \varepsilon - \eta, \frac{1}{3} + \eta)$ with equal probability (where $\varepsilon > 0, \eta > 0$, and $\varepsilon + 2\eta < \frac{1}{3}$), then we still have $\tilde{y} \succ_{MIO-SPM} \tilde{x}$, even though the two signals are not Blackwell comparable (illustrated in Figure 4). So there is a sense in which our information order is *robust*.

Our examples highlight the fact that when there are three or more states of the world, there is a significant restriction in Blackwell’s requirement that the posteriors of the “bad” signal must lie in the convex hull of the “good” signal posteriors. If there are only two states of the world, $|\Omega| = 2$ (the state of the world is a “dichotomy”), then this restriction is no longer severe, since all posteriors can be summarized in a single dimension. It can be shown that on dichotomies (*MIO-SPM*) is equivalent to sufficiency.

Example 2. *Concave (Incremental returns concave in ω).* When U_1 is the set of concave functions, the relation \succ_{SD-U_1} corresponds to SOSD. Recall from above that (*MDP-F*) implies that $\tilde{F}(\omega|\alpha)$ becomes “less risky” (by SOSD) as α increases. So $G \succ_{MIO-CV} F$ if, for any α , $\tilde{G}(\omega|\tilde{\alpha}_y \geq \alpha) \succ_{SOSD} \tilde{F}(\omega|\tilde{\alpha}_x \geq \alpha)$: high signals lead on average to less risky posteriors under G than under F . Since high signals lead to high actions and high actions lead to interim preferences that are more risk-averse, one can see immediately that having a better match between high signals and low risk will be valuable to the DM. We give an example in Section 6 where such preferences arise. Interestingly (*MIO-CV*) implies sufficiency when there are only two *or* three possible states of the

²¹See also Levy and Paroush (1974) and Shaked and Shantikumar (1997) for more discussion of the stochastic dominance order associated with supermodular functions.

world.

Just as we could find an *MDP* pair (U_2, \mathcal{F}) corresponding to *any* set of incremental return functions U_1 that forms a closed convex cone, we can describe the relevant monotone information order condition. Once again, \succ_{MIO-U_1} says that if the average posteriors under G stochastically dominate the average posteriors under F (with respect to functions in U_1), then G is more informative than F for the class of problems in (U_2, \mathcal{F}) . And we know that any stochastic dominance relationship induced by a closed convex cone can be described in terms of the extreme points of the cone.²²

4.2 Monotone Information Orders Using Stochastic Single Crossing

In the previous section we indexed posteriors by their *ex ante* percentile and then compared information structures, finding that informativeness rankings could be derived using familiar notions of stochastic dominance. For this approach to be “tight,” we required U_1 to contain the positive and negative constant functions. When U_1 does not contain the constants, but does contain some other function \hat{r} and its additive inverse $-\hat{r}$, we can extend our results using stochastic single crossing in place of stochastic dominance.

We introduce an additional pair of restrictions (jointly denoted (R)) in order for (U_2, \mathcal{F}) to be admissible:

(R-U) U_1 is a closed convex cone, and there some $\hat{r}(\omega) \geq 0$ such that $\hat{r}, -\hat{r} \in U_1$.

(R-F) For an \hat{r} satisfying (R-U), for all $F \in \mathcal{F}$ with associated signal \tilde{x} , there exists $b > 0$ such that $E[\hat{r}(\omega)|x] \geq b$ for all x in support(\tilde{x}).

If U_1 contains the constant functions, then (R) holds with $\hat{r}(\omega) = 1$. If U_1 is the class of $WSC(\omega_0)$ functions, then (R) holds with $\hat{r} = 1_{\{\omega \approx \omega_0\}}$ (that is, \hat{r} is an indicator function for a “small” interval containing ω_0).

We begin with a Lemma that generalizes a key step in the proof of sufficiency above.

Lemma 5 *Let U_2 satisfy (MDP-U) and assume (A) holds. Consider two information structures F and G , where $F(\omega) = G(\omega)$, and continuous strictly increasing functions $T_F : \mathcal{X} \rightarrow [0, 1]$,*

²²It is interesting to observe in passing that if the DM’s marginal returns are *both* nondecreasing and concave, we can concatenate our conditions — as is often done for straightforward preferences over gambles. For example, such a DM would be better off after a sequence of two changes to $\tilde{F}(\omega|\tilde{\alpha}_x \geq \alpha)$: one that makes higher states more likely (according to FOSD) and one that reduces risk (according to SOSD).

$T_G : \mathcal{Y} \rightarrow [0, 1]$. Define $m(\omega, \alpha) = G(\omega, T_G^{-1}(\alpha)) - F(\omega, T_F^{-1}(\alpha))$. Then

$$\int_{\Omega} \int_{[0,1]} u(\omega, a(\alpha)) dm(\omega, \alpha) \geq 0 \quad (15)$$

for all $u \in U_2$ and all $a : [0, 1] \rightarrow A$ nondecreasing, if and only if for all $r \in U_1$ and all $\alpha \in (0, 1)$,

$$\int_{\Omega} r(\omega) d_{\omega} m(\omega, \alpha) \leq 0. \quad (16)$$

An important thing to note is that $m(\cdot, \alpha)$ is *not necessarily* a probability distribution on ω ; checking (16) when $r(\omega) \equiv 1$ entails checking that $\Pr(T_G(\tilde{y}) \leq \alpha) \leq \Pr(T_F(\tilde{x}) \leq \alpha)$. Thus, our choice of index in the last subsection can be immediately understood: if $\{\mathbf{1}, -\mathbf{1}\} \in U_1$, (16) holds for all $r \in U_1$ only if $\Pr(T_G(\tilde{y}) \leq \alpha) = \Pr(T_F(\tilde{x}) \leq \alpha)$, which is always true when $T_F(x) = F(x)$ and $T_G(y) = G(y)$.

When $\hat{r}, -\hat{r} \in U_1$, then (16) requires that

$$\int_{\Omega} \hat{r} d_{\omega} m(\omega, \alpha) = 0. \quad (17)$$

This motivates our new choice of indexing functions, T_F, T_G :

$$T_F(x) = \frac{\int_{\Omega} \hat{r}(\omega) d_{\omega} F(\omega, x)}{\int_{\Omega} \hat{r}(\omega) dF(\omega)} = F(x) \frac{E[\hat{r}|\tilde{x} \leq x]}{E[\hat{r}]}. \quad (18)$$

This choice of T_F, T_G guarantees that (17) will hold, meaning that it remains only to identify a condition under which (16) will hold for all $r \in U_1$, $r \neq \hat{r}$. If $\hat{r} = 1$, then $T_F(x) = F(x)$ as above. If U_1 is the class of WSC(ω_0) functions, then $T_F(x) = F(x|\omega_0)$. In general, $(R-\mathcal{F})$ implies that $T_F : \mathcal{X} \rightarrow [0, 1]$ is strictly increasing.²³

Theorem 6 *Let (U_2, \mathcal{F}) be an MDP pair, suppose (A) and (R) hold. Let $F, G \in \mathcal{F}$. Then $V^*(G, u) \geq V^*(F, u)$ for all $u \in U_2$ if for all $\alpha \in (0, 1)$,*

$$G(\omega|T_G(\tilde{y}) \geq \alpha) \succ_{SC-U_1} F(\omega|T_F(\tilde{x}) \geq \alpha) \quad (MIO')$$

where T_F and T_G are defined by (18).

(MIO') generalizes (MIO) by requiring that the average posteriors be ordered using stochastic single crossing rather than the (stronger) stochastic dominance order. To interpret this, the stochastic single crossing requirement can be restated as follows: for each α , and for all $r \in U_1$,

$$E[r(\omega)|T_G(\tilde{y}) \geq \alpha] \geq 0 \implies E[r(\omega)|T_F(\tilde{x}) \geq \alpha] \geq 0. \quad (19)$$

²³And continuity can be ensured (without changing the information content) using Lehmann's (1988) construction.

This contrasts with stochastic dominance, which requires $E[r(\omega)|T_G(\tilde{y}) \geq \alpha] \geq E[r(\omega)|T_F(\tilde{x}) \geq \alpha]$. If U_1 contains the constant functions, then stochastic dominance and stochastic single crossing coincide. For this reason, we maintain the notation $G \succ_{MIO-U_1} F$ even when applying (MIO') .

To apply Theorem 6, it is useful to have an alternative condition for (MIO') that requires checking only a single inequality: for all $\alpha \in (0, 1)$ and $r \in U_1$,

$$E[r(\omega)|T_G(\tilde{y}) \geq \alpha] \cdot \Pr(T_G(\tilde{y}) \geq \alpha) \geq E[r(\omega)|T_F(\tilde{x}) \geq \alpha] \cdot \Pr(T_F(\tilde{x}) \geq \alpha). \quad (MIO'')$$

The proof of Theorem 6 proceeds by first establishing the equivalence of (MIO') and (MIO'') , and then applying Lemma 5 to show that the result is implied by (MIO'') .

Proof. It can be shown²⁴ that when U_1 is a closed convex cone, then for each α , (19) holds if and only if there exists a $\lambda(\alpha) \geq 0$ such that

$$E[r(\omega)|T_G(\tilde{y}) \geq \alpha] - \lambda(\alpha)E[r(\omega)|T_F(\tilde{x}) \geq \alpha] \geq 0. \quad (20)$$

for all $r \in U_1$. ($\lambda(\alpha)$ can be interpreted as the Lagrange multiplier in the problem of minimizing $E[r(\omega)|T_G(\tilde{y}) \geq \alpha]$ subject to the constraint that $E[r(\omega)|T_F(\tilde{x}) \geq \alpha] \geq 0$, and the left-hand side of (20) is the minimized value of the objective). But, checking (20) for \hat{r} and $-\hat{r}$, both in U_1 by $(R-U)$, implies that $\lambda(\alpha) = E[\hat{r}(\omega)|T_G(\tilde{y}) \geq \alpha]/E[\hat{r}(\omega)|T_F(\tilde{x}) \geq \alpha]$. But, the latter ratio is in turn equal to $\Pr(T_G(\tilde{y}) \geq \alpha)/\Pr(T_F(\tilde{x}) \geq \alpha)$ (simply substitute in from the definitions of T_F and T_G in (18)). So, (20) is equivalent to (MIO'') . To complete the proof, let $\tilde{F}(\omega, \alpha) = F(\omega, T_F^{-1}(\alpha))$ and likewise for G . By Lemma 5, (MIO'') is equivalent to the following: for all $u \in U_2$ and all monotone decision policies $a : [0, 1] \rightarrow A$, $V(\tilde{G}, u, a(\cdot)) \geq V(\tilde{F}, u, a(\cdot))$. So for all $u \in U_2$, if a^F is an optimal (monotone) policy for u under F , $V^*(G, u) \geq V(\tilde{G}, u, a^F) \geq V^*(F, u)$. \square

We now show that (MIO') is tight for a larger class of decision problems than that identified in Theorem 4 above.

Theorem 7 *Let (U_2, \mathcal{F}) be an MDP pair and suppose (A) and (R) hold. Let $F^\theta(\omega, x)$ be smoothly parametrized by θ , with $F^\theta \in \mathcal{F}$ for all θ . Then $\frac{d}{d\theta}V(F^\theta, u) \geq 0$ for all $u \in U_2$, if and only if $F^{\theta+d\theta} \succ_{MIO-U_1} F^\theta$.*

Proof. As in Theorem 4, we prove a stronger result, that (MIO) is necessary for $V(\tilde{G}, u, a^F) \geq V(\tilde{F}, u, a^F)$, where a^F is the optimal policy for information structure F . Suppose $G \not\succeq_{MIO-U_1} F$.

²⁴See Jewitt (1986), Gollier and Kimball (1995), or Athey (1998a) for alternative proofs.

Using the representation (MIO''), this implies that there is some $\bar{\alpha}$, and $\bar{r} \in U_1$ such that

$$\int_{\Omega} \bar{r}(\omega) dG(\omega, T_G^{-1}(\bar{\alpha})) > \int_{\Omega} \bar{r}(\omega) dF(\omega, T_F^{-1}(\bar{\alpha})). \quad (21)$$

Let $A = \{0, 1\}$. Then define

$$K_{\bar{\alpha}} = E[r(\omega)|T_F(\tilde{x}) = \bar{\alpha}] / E[\hat{r}(\omega)|T_F(\tilde{x}) = \bar{\alpha}].$$

The denominator is non-negative by ($R\text{-}\mathcal{F}$). Then define

$$\bar{u}(\omega, a) = a(\bar{r}(\omega) - K_{\bar{\alpha}}\hat{r}(\omega)).$$

By ($R\text{-}U$), $\bar{u} \in U_2$. By ($MDP\text{-}\mathcal{F}$), $E[\bar{r}(\omega) - K_{\bar{\alpha}}\hat{r}(\omega)|T_F(\tilde{x}) = \alpha]$ is single crossing in α , so that an optimal policy for this DM is to set $a^F(\alpha) = 1$ if $\alpha > \bar{\alpha}$ and zero otherwise. Then $V(\tilde{G}, u, a^F) \geq V(F, u, a^F)$, if and only if

$$\int_{[0,1]} \int_{\Omega} \bar{u}(\omega, a^F(\alpha)) dG(\omega, T_G^{-1}(\alpha)) \geq \int_{[0,1]} \int_{\Omega} \bar{u}(\omega, a^F(\alpha)) dF(\omega, T_F^{-1}(\alpha)) \quad (22)$$

$$\iff \int_{\Omega} [\bar{r}(\omega) - K_{\bar{\alpha}}\hat{r}(\omega)] dG(\omega, T_G^{-1}(\bar{\alpha})) \leq \int_{\Omega} [\bar{r}(\omega) - K_{\bar{\alpha}}\hat{r}(\omega)] dF(\omega, T_F^{-1}(\bar{\alpha})) \quad (23)$$

But, we have defined T_F and T_G so that $\int_{\Omega} \hat{r}(\omega) d_{\omega}G(\omega, T_G^{-1}(\alpha)) = \int_{\Omega} \hat{r}(\omega) d_{\omega}F(\omega, T_F^{-1}(\alpha))$ for all α . So (23) contradicts (21). As in Theorem 4, necessity for the case of a smoothly parameterized distribution follows from an application of the Envelope Theorem. \square

If (U_2, \mathcal{F}) are an MDP pair corresponding to U_1 , and $F, G \in \mathcal{F}$, the posteriors induced by each information structure are not required by ($MDP\text{-}\mathcal{F}$) to be totally ordered by stochastic dominance unless U_1 contains the constant functions. If, however, they do happen to be ordered by stochastic dominance, we can show that (MIO) and (MIO') are equivalent.

(MDP-SD) For all $F \in \mathcal{F}$, if $x^H, x^L \in \text{support}(F)$ and $x^H \geq x^L$, then $F(\omega|x^H) \succ_{SD-U_1} F(\omega|x^L)$.

Proposition 8 Let (U_2, \mathcal{F}) be an MDP pair and suppose (A), (R), and (MDP-SD) hold. Let $F, G \in \mathcal{F}$. Then (MIO') is equivalent to (MIO).

Proof. Note that $\int \hat{r} dF(\omega|x)$ and $-\int \hat{r} dF(\omega|x)$ are both increasing in x , so it must be the case that $\int \hat{r} dF(\omega|x) = \int \hat{r} dF(\omega)$ is constant in x . And likewise for G . Simplifying the expressions

from (18), $T_F(x) = F(x)$, $T_G(x) = G(x)$. It follows immediately that (MIO'') (which we know is equivalent to (MIO')) is equivalent to (MIO) . \square

The result obtains because when posteriors are ordered by stochastic dominance and $(R-U)$ holds, $E[\hat{r}|x]$ must be constant in x . It follows that $T_F(x)$ reduces immediately to $F(x)$, our earlier index. In contrast, when the posteriors are ordered by \succ_{SC-U_1} , $E[\hat{r}|x]$ can vary with x .

4.2.1 Examples

We now interpret the monotone information orders corresponding to Examples 3 and 4 from above. In the process we relate our information orders to Lehmann's efficiency criteria.

Example 3. $WSC(\omega_0)$ (*Incremental returns weak single crossing at ω_0*). Recall that if r is $WSC(\omega_0)$, then $r(\omega) \leq (\geq) 0$ as $\omega < (>) 0$. So $\hat{r} = 1_{\{\omega \approx \omega_0\}}$, and $T_F(x) = F(x|\omega_0)$ using Bayes' Rule. Then (MIO') reduces to

$$\Pr(G(\tilde{y}|\omega_0) > \alpha | \omega) < \Pr(F(\tilde{x}|\omega_0) > \alpha | \omega) \quad \text{for } \omega < \omega_0$$

$$\Pr(G(\tilde{y}|\omega_0) > \alpha | \omega) > \Pr(F(\tilde{x}|\omega_0) > \alpha | \omega) \quad \text{for } \omega > \omega_0$$

Alternatively, $F(F^{-1}(\alpha|\omega_0)|\omega) - G(G^{-1}(\alpha|\omega_0)|\omega)$ is $WSC(\omega_0)$. Recall that the DM is essentially interested in knowing whether ω is above or below ω_0 — and $(MDP-\mathcal{F})$ ensures that high signals are “good news” about ω being above ω_0 . The information condition implies that if the true state of the world $\omega > \omega_0$, then the probability of receiving “good news” under G (defined as $\tilde{y} > G^{-1}(\alpha|\omega_0)$) is higher than under F . If we impose the additional assumption $(MDP-SD)$, it follows that for all \tilde{x} , \tilde{y} , $f(\omega_0|x) = g(\omega_0|y) = h(\omega_0)$ (the prior). Then, conditions (MIO) and (MIO') are equivalent, and we can simply check that $g(\omega|G(\tilde{y}) > \alpha) - f(\omega|F(\tilde{x}) > \alpha)$ is $WSC(\omega_0)$ for all α . That is, conditional on a high posteriors, low states are less likely and high states are more likely under G as opposed to F , while neither signal is informative about the state ω_0 .

The monotone information order for $WSC(\omega_0)$ can also be interpreted in terms of hypothesis testing. Consider testing the null hypothesis $H_0 : \omega = \omega_0$ against the alternative $H_1 : \omega < \omega_0$. However, the test is constrained to have a *size* of $1 - \alpha$: conditional on H_0 , the probability of rejecting the null is $1 - \alpha$. Thus, we reject when $F(x|\omega_0) > \alpha$, and likewise for G . Since $MIO-WSC(\omega_0)$ is equivalent to requiring that $F(F^{-1}(\alpha|\omega_0)|\omega) - G(G^{-1}(\alpha|\omega_0)|\omega)$ if $\omega \geq (\leq) \omega_0$, we have the following interpretation. The probability of rejecting the null when the alternative is true ($\omega \leq \omega_0$) is greater for G than F ; and the probability of rejecting the null when the alternative

is false ($\omega > \omega_0$) is smaller for G than F . Thus, the hypothesis test using G is uniformly more powerful for a given size. This is exactly the interpretation given by Lehmann (1988) in his analysis, which will be discussed in more detail below.

To see a very simple example where this order can be applied, return to the portfolio allocation problem discussed in Section 3.3, where payoffs are given by $u(\omega, a) = v(a\omega + (1 - a)\omega_0)$. For this problem, F is more informative than G if it provides a more powerful inference about whether or not investment in the risky asset is worthwhile.

Example 4. *Single Crossing (Incremental returns single crossing in ω).* The monotone information order for the case of single crossing payoff functions and signal distributions with monotone likelihood has been derived by Lehmann (1988). He showed that G is more informative than F for this class of problems ($G \succ_L F$) if for all $x \in \mathcal{X}$, $G^{-1}(F(x|\omega)|\omega)$ is nondecreasing in ω .²⁵ Jewitt (1997) has given further equivalent characterizations. We now show that Lehmann's order can be obtained as a special case of our result.

Recall that we noted above that the set of all functions that cross zero once from below at some ω_0 was the union of all the $WSC(\omega_0)$ sets. By quantifying over all crossing points ω_0 , we can recover the information preferences of DMs with single crossing payoff functions.

Theorem 9 *The following are equivalent: (i) $G \succ_L F$; (ii) $G \succ_{MIO-WSC(\omega_0)} F$ for all $\omega_0 \in \Omega$; (iii) $G \succ_{MIO-SPM} F$ for all prior distributions $H(\omega)$.*

Proof. First note from above that if $F(\omega|x)$ is increasing in $\succ_{SC-WSC(\omega_0)}$ for all ω_0 as x increases, then $F(\omega|x)$ is ordered by MLR.. And similarly, if $F(\omega|x)$ is increasing by FOSD as x increases for all priors H , then F is ordered by MLR.. We first show that (i) and (ii) are equivalent. Suppose $G \succ_{MIO-WSC(\omega_0)} F$. This means that $F(F^{-1}(\alpha|\omega_0)|\omega) - G(G^{-1}(\alpha|\omega_0)|\omega)$ is $WSC(\omega_0)$. For $WSC(\omega_0)$ to hold for every ω_0 , the expression must be increasing in ω . Letting $x = F^{-1}(\alpha|\omega_0)$, and taking $G^{-1}(\cdot|\omega)$ of each term, we have $G^{-1}(F(x|\omega)|\omega) - G^{-1}(F(x|\omega_0)|\omega_0)$ is $WSC(\omega_0)$. This will hold for every ω_0 if and only if $G^{-1}(F(x|\omega)|\omega)$ is increasing in ω , which is exactly Lehmann's condition. Now consider the equivalence of (ii) and (iii). Suppose $G \succ_{MIO-SPM} F$. Then using the definition of FOSD, we have $F(\omega|F(x) \leq \alpha) \leq G(\omega|G(y) \leq \alpha)$. Applying Bayes' Rule, this is equivalent to $F(F^{-1}(\alpha)|\tilde{\omega} \leq \omega) \leq G(G^{-1}(\alpha)|\tilde{\omega} \leq \omega)$. Letting $x = F^{-1}(\alpha)$ and taking

²⁵Lehmann (1988) also considered necessity, although his theorem formally answers a slightly different question. He finds that the efficiency order is necessary and sufficient for one signal to return higher payoffs than another in every possible state (instead of on average, given the prior).The discussion following Example 3 above provides an outline of Lehmann's approach.

$G^{-1}(\cdot|\tilde{\omega} \leq \omega)$ of both sides, we see this is equivalent to $G^{-1}(F(x|\tilde{\omega} \leq \omega)|\tilde{\omega} \leq \omega) \leq G^{-1}(F(x))$. This is implied by $G \succ_L F$, and quantifying over all two-point priors, it implies \succ_L . \square

To see how Lehmann's order can depart from the supermodular monotone information order in practice, return to Figure 4. In that example, recall that $\tilde{z} \succ_{MIO-SPM} \tilde{y} \succ_{MIO-SPM} \tilde{x}$. Using Lehmann's order, $\tilde{z} \succ_L \tilde{y}$, but \tilde{x} does not satisfy MLR, so Lehmann's order does not apply to it. Of course, *none* of the three are ordered by sufficiency.

The analysis can be extended to other payoff classes. For instance, it follows from results in Athey (1998b) that when U_2 is the set of log-supermodular functions the appropriate information order is again \succ_L . Or in another example, consider the set of payoffs with incremental returns that positive only in some intermediate range between ω_0 and ω'_0 , where $\omega_0 < \omega'_0$. This case works out similarly to single crossing at a point.

5 Ordering Payoff Functions

We now provide conditions under which one decision-maker has a higher incremental return to improving her information than another decision-maker. Persico (1997) has investigated this question for the case of single crossing payoff functions and has shown that if $\frac{\partial}{\partial \alpha} u(\omega, a^F(\alpha)) - \frac{\partial}{\partial \alpha} v(\omega, a^F(\alpha))$ is single-crossing in ω , the first decision-maker (u) benefits more from a small increase in information than the second (v), according to Lehmann's information order for single crossing payoff functions. We generalize this result. To do this, we first limit our attention to marginal changes in the information structure. If two signal distributions are linked by a smoothly parameterized family of distributions which is information-ranked all along the way from F to G , we can then compare the incremental return to increasing the signal strength from F to G .

We need a small amount of notation: let Θ be a convex subset of \mathbb{R} , and suppose $\{F^\theta(\omega, x) : \theta \in \Theta\}$ is a family of information structures smoothly parameterized by θ . Let $a^{\theta, u}(\alpha)$ be a nondecreasing selection from $\int_{\Omega} u(\omega, a) d\tilde{F}^\theta(\omega|\alpha)$, and let $u^\theta(\omega, \alpha) = u(\omega, a^{\theta, u}(\alpha))$.

Theorem 10 *Suppose (U_2, \mathcal{F}) is an MDP pair and that the family $\{F^\theta(\omega, x) : \theta \in \Theta\}$ is smoothly parametrized on Θ , with $F^\theta \in \mathcal{F}$ for all θ . Let u, v be bounded measurable payoff functions. If, for some $\bar{\theta}$, $u^{\bar{\theta}}(\omega, \alpha) - v^{\bar{\theta}}(\omega, \alpha) \in U_2$, then $\frac{\partial}{\partial \theta} V(F^{\bar{\theta}}, u) \geq \frac{\partial}{\partial \theta} V(F^{\bar{\theta}}, v)$ at $\bar{\theta}$.*

Proof. By the envelope theorem, we have

$$\frac{\partial}{\partial \theta} V(F^\theta, u) = \int_{\Omega} \int_{[0,1]} u(\omega, a^{\theta, u}(\alpha)) d \left[\frac{\partial}{\partial \theta} \tilde{F}^\theta(\omega, \alpha) \right]. \quad (24)$$

So letting $w(\omega, \alpha) = u(\omega, a^{\theta, u}(\alpha)) - v(\omega, a^{\theta, v}(\alpha))$ and $m(\omega, \alpha) = \frac{\partial}{\partial \theta} \tilde{F}^\theta(\omega, \alpha)$, we have

$$\frac{\partial}{\partial \theta} V(u; F^\theta) - \frac{\partial}{\partial \theta} V(v; F^\theta) = \int_{\Omega} \int_{[0,1]} w(\omega, \alpha) d_{\omega} m(\omega, \alpha). \quad (25)$$

The assumption that \tilde{F}^θ is increasing in \succ_{MIO-U_1} as θ increases implies that for any $r(\omega) \in U_1$, $\alpha' \in (0, 1)$

$$\int_{\Omega} r(\omega) d_{\omega} m(\omega, \alpha) \geq 0. \quad (26)$$

Lemma 5 then implies that (25) evaluated at $\bar{\theta}$ is nonnegative. \square

Theorem 10 says that if $u^{\bar{\theta}} - v^{\bar{\theta}}$ is in U_2 , i.e. if $u^{\bar{\theta}}$ is “more U_2 ”, than $v^{\bar{\theta}}$, then the decision-maker with payoff function u has a higher marginal value for information than the DM with payoff function v . The theorem *does not* require that $u, v \in U_2$, or that the marginal value of information is nonnegative for each agent, although it is somewhat hard to imagine applying the Theorem when this is not the case (since policies might not be monotone).

If u is more sensitive to information than v in response to every signal in the family, we can compare changes in information that are not marginal. From this, we can derive comparative statics on the amount of amount of information acquired.

Theorem 11 *Suppose the conditions of Theorem 10 are satisfied, and that $u^{F^{\bar{\theta}}}(\omega, \alpha) - v^{F^{\bar{\theta}}}(\omega, \alpha) \in U_2$ for every $\theta \in \Theta$, where Θ is a closed interval. Let $C : \Theta \rightarrow \mathbb{R}$ be the cost of information; and let $\theta^*(u) = \arg \max_{\theta \in \Theta} V(\theta; u) - C(\theta)$. Then $\theta^*(u) \geq \theta^*(v)$ (in the strong set order).*

Proof. Let $V(\gamma^H; \theta) = V(u; F^\theta)$, and $V(\gamma^L; \theta) = V(v; F^\theta)$. Applying Theorem 10, $V(\theta; \gamma)$ is supermodular in (θ, γ) , so $V(\theta, \gamma) - C(\theta)$ is supermodular in (θ, γ) . By Topkis’ (1978) Monotonicity Theorem, $\theta^*(\gamma)$ is nondecreasing in the strong set order. \square

A few words are in order about the results of this section. A main drawback to Theorems 10 and 11 is that the conditions on u and v are not primitive, since they depend on the properties of the objective function evaluated at its optimum. In general, verifying the necessary condition may require knowing something about the shape of the optimizer, or about the curvature of the payoff functions. To take an apparently simple example, suppose $u(\omega, a) = v(\omega, a) + g(a)$, where v is supermodular. One can verify that *sufficient* conditions for u to acquire more information than v , where information is ranked by the supermodular information order, are that $a^F(\alpha; u) \geq a^F(\alpha; v)$, that $a_{\alpha}^F(\alpha; u) \geq a_{\alpha}^F(\alpha; v)$, and $v_{\omega aa} \geq 0$, (i.e. *marginal returns* are supermodular). The first

condition will hold if $g(a)$ is increasing,²⁶ and the last is (reasonably) transparent, but even if g is linear, the second condition requires significant assumptions on the curvature of v . What conclusion should we draw from this? While we think that these results hold promise for applied modeling, it may be necessary to place a fair amount of structure on the model to make them operational.

On the other hand, the existing literature (prior to Persico (1997)) provides minimal guidance in this area. For example, an increase in the Blackwell information order increases the ex ante expected value of any function which is convex in the posterior beliefs. Thus, agent u will buy more Blackwell-ordered information than v if and only if $u^*(P) - v^*(P) = \int_{\Omega} [u(\omega, a^{*,u}(P)) - v(\omega, a^{*,v}(P))] dP(\omega)$ is convex in the posterior $P(\omega)$. While convexity of $u^*(P)$ and $v^*(P)$ follows as a simple consequence of optimality, convexity of the difference is not at all transparent. Checking the condition would almost certainly involve non-primitive assumptions similar to the ones described in this section.

6 Applications

We now present several applications of the results developed above. The first two applications are standard decision problems; the third examines adverse selection in a labor market equilibrium, while the fourth studies a coordination game. The examples demonstrate both strengths and weaknesses of the approach. In particular, while comparisons of information structures may be obtained immediately in a broad range of cases, comparative statics on information acquisition often require additional assumptions.

6.1 Application: Cost Uncertainty for Producers

A growing literature in Industrial Organization considers the value of information to firms in oligopoly models (see for instance Mirman, Samuelson, and Schlee (1994) and references therein). The methods outlined above allow for some new results in this general class of problems.

Consider the problem faced by a producer who must choose quantity, q , to maximize some objective. We will compare the importance of information to firms under different market structures. However, we will restrict attention to covert information gathering, whereby we hold the strategies of other agents fixed when we analyze the effects of gathering information.

Write the firm's gross surplus as a general function, $R(q; \beta)$, where β parameterizes the market structure. In particular, we will consider $\beta \in \{M, D, S\}$, representing monopoly, duopoly, and the social planner's total surplus. The firm faces uncertainty about its cost: the total cost of producing

²⁶To see this, let $u(\omega, a, \tau) = v(\omega, a)$ when $\tau = 0$, and $v(\omega, a) + g(a)$ when $\tau = 1$. Then u is supermodular in (a, τ) for every ω , so $a^*(\alpha, \tau)$ will be increasing in (α, τ) . That is, $a^*(\alpha; u) \geq a^*(\alpha; v)$ for every α .

q is given by $C(q; \omega)$. Thus, the firm's profits are given by $\pi(q, \omega; \beta) = R(q; \beta) - C(q; \omega)$. We consider a smoothly parameterized family of information structures, $\{F^\theta(\omega, \alpha) : \theta \in \Theta\}$, where $F^\theta(\alpha) = \alpha$ for each θ and all $\alpha \in [0, 1]$. Expected profits given an α -percentile posterior are denoted $\bar{\pi}(q, \alpha; \beta) = R(q; \beta) - E[C(q, \omega)|\alpha]$. Assume that each of these functions is differentiable in q . Under appropriate assumptions, the optimal choice of quantity is then determined by $MR(q; \beta) = \frac{\partial}{\partial q} R(q; \beta) = \frac{\partial}{\partial q} E[C(q, \omega)|\alpha] = E[MC(q, \omega)|\alpha]$.

Suppose that θ indexes $F^\theta(\omega, \alpha)$ according to (*MIO-SPM*), and that α indexes $F^\theta(\omega|\alpha)$ according to FOSD for all θ . We first apply our analysis from above to characterize increasing information for this problem. If we assume that $MC(q; \omega)$ is decreasing in ω , then $\pi(q, \omega; \beta)$ is supermodular in (q, ω) . Since (*MDP-F*) is satisfied, the optimal choice, $q^\beta(\alpha; \theta)$, will be nondecreasing in α . Better information (according to (*MIO-SPM*)) leads to higher ex ante expected payoffs.

We now turn to consider how the marginal value of information changes under different market conditions. We begin with the following result.

Proposition 12 *Consider the model described above, where for each θ , α^θ indexes $F^\theta(\omega|\alpha^\theta)$ according to FOSD. Assume that the smoothly parameterized family $\{F^\theta(\omega, \alpha)\}$ is ordered by MIO-SPM. In addition, assume that $MC(q; \omega)$ is decreasing in ω , increasing in q , and submodular in (ω, q) . Finally, assume that $q^H(\alpha) > q^L(\alpha)$ and $q^{H'}(\alpha) > q^{L'}(\alpha)$. Then the marginal value of θ is higher for firm β^H than firm β^L .*

Proof. By Theorem 10, it suffices to show that $\pi_q(q^H(\alpha), \omega; \beta^H)q^{H'}(\alpha) - \pi_q(q^L(\alpha), \omega; \beta^L)q^{L'}(\alpha)$ is nondecreasing in ω . If $q^{H'}(\alpha) > q^{L'}(\alpha)$, then our result obtains if $\pi_q(q^H(\alpha), \omega; \beta^H) - \pi_q(q^L(\alpha), \omega; \beta^L)$ is nondecreasing in ω (since each term is separately nondecreasing). Recall that $\pi_q(q^\beta(\alpha), \omega; \beta) = MR(q^\beta(\alpha); \beta) - MC(q^\beta(\alpha), \omega)$ for each firm. Since the first term does not vary with ω , we can consider whether $MC(q^L(\alpha), \omega) - MC(q^H(\alpha), \omega)$ is nondecreasing in ω , which follows since $MC(q, \omega)$ is submodular. \square

To interpret the conditions on the cost function, note that they are satisfied if $C(q, \omega) = q^2/\omega$. To interpret the conditions on the quantities chosen, which of course are not primitive conditions, it is useful to consider some examples. First consider comparing a monopolist's problem with that of the social planner. Let $R(q; \beta^M) = P(q) \cdot q$ and $R(q; \beta^S) = \int_0^q P(t)dt$.

As usual, it follows directly that $q^S(\alpha) > q^M(\alpha)$, that is, the monopolist underprovides quantity. Now consider the terms $q^{S'}(\alpha)$ and $q^{M'}(\alpha)$. The implicit function theorem yields

$$q^{\beta'}(\alpha) = \frac{\frac{\partial}{\partial \alpha} E[MC(q^\beta(\alpha), \omega)|\alpha]}{\frac{\partial}{\partial q} MR(q^\beta(\alpha); \beta) - \frac{\partial}{\partial q} E[MC(q^\beta(\alpha), \omega)|\alpha]}.$$

Thus, $q^{S'}(\alpha) > q^{M'}(\alpha)$ if $MC(q, \omega)$ is submodular in (q, ω) , if $MC(q, \omega)$ is convex in q , and if $P'(q^S(\alpha)) > 2P'(q^M(\alpha)) + q^M(\alpha)P''(q^M(\alpha))$. If P is linear or convex, a sufficient condition for the latter inequality is the commonly assumed requirement that the marginal revenue curve is steeper than the demand curve.

We can also compare the value of information for a monopolist and a duopolist. We suppose that both duopolists have the same initial signal structure as the monopolist, but only one duopolist has the opportunity to (covertly) purchase additional information (it is also possible to study the information acquisition game, but we will not consider that case here). With these symmetry assumptions, the result $q^M(\alpha) > q^D(\alpha)$ obtains so long as marginal revenue is downward sloping. Further, $q^{M'}(\alpha) > q^{D'}(\alpha)$ if marginal cost is submodular, and if

$$2P'(q^M(\alpha)) + q^M(\alpha)P''(q^M(\alpha)) > 2P'(2q^D(\alpha)) + q^D(\alpha)P''(2q^D(\alpha)).$$

This condition is satisfied for the constant elasticity demand curve.

Summarizing, we see that under several reasonable conditions, the social planner has a larger incentive to acquire information than the monopolist, but the monopolist may have a larger incentive to acquire information than the duopolist.

6.2 Application: Screening for a “Target” Hire

An entrepreneur or manager needs to hire a worker or subcontractor to perform a very specific task. She interviews an interested party, who appears to have roughly the right characteristics. The manager now has to make a single take-it-or-leave-it monetary offer based on information gleaned from the interview. What sort of signal from the interview will cause the manager to make a large offer? And what would it mean for the screening process to be more effective?

To model such a situation, we assume that there is some optimal task outcome ω^* . Hiring the agent would result in outcome ω , and a payoff of $r(\omega)$, which achieves a maximum at ω^* , and decreases as ω moves away in both directions. So $r(\omega)$ is concave (but not necessarily symmetric—it’s not just the distance from perfection that matters, but the direction). A wage offer a will be accepted with probability $p(a)$, which is increasing. The manager has a signal x about ω (arising from an information structure $F(\omega, x)$), and chooses a to maximize $p(a)E[r(\omega) - a|x]$. We can write the payoff function $u(\omega, a)$ as $u(\omega, a) = p(a)[r(\omega) - a]$. In order for signals to be ordered in such a way that the manager will offer more money following a “good” signal, regardless of the exact shape of p and r , it must be the case that x orders the posteriors by SOSD (see Example 2 in Section 3.3 for interpretations). If the potential signals from the interview satisfy this condition, $a^F(x)$ will be monotone regardless of the exact shape of p and r .

Now for the question of what constitutes better information. Applying (*MIO-CV*), G will be a more informative interview than F if $\forall \alpha \in [0, 1]$,

$$G(\omega|G(x) > \alpha) \succeq_{SOSD} F(\omega|F(x) > \alpha).$$

The intuition is simple: after the interview the expected outcome if the agent is hired is always ω^* , but there is some residual uncertainty. The manager does not like this residual risk—in particular, the less risk she believes there is, the more aggressively she will pursue the candidate. A more informative screening process is one where high signals (which are more likely to lead to hires) imply less residual risk.

6.3 Application: Adverse Selection and Labor Markets

Our next example suggests how our information orders can be applied to adverse selection markets. Consider the following stylized situation. Workers live for two periods and spend the first period training (in school). Each worker’s productivity is unknown with prior $H(\omega)$, and schooling generates a noisy signal \tilde{x} about underlying ability. Suppose that the joint distribution of (ω, x) is $F(\omega, x)$. This signal is observed by the school, but not by the general labor market. Instead, only the top $1 - \alpha$ fraction of the class “graduates,” while the rest do not — the labor market observes only if a given worker graduated. Each firm has production function $J(\omega)$, increasing in ω , and the labor market is competitive so that workers receive a wage $E[J(\omega)|\mathfrak{I}]$, where \mathfrak{I} is the information available to the market about productivity.

The market wage for graduates will be $E[J(\omega)|F(\tilde{x}) \geq \alpha]$, and for non-graduates $E[J(\omega)|F(\tilde{x}) < \alpha]$. Now suppose that the schooling process becomes more informative about the worker’s ability, in the sense that F increases to G in the supermodular monotone information order (*MIO-SPM*). It follows immediately that the wage for graduates will increase since

$$E[J(\omega)|G(\tilde{y}) \geq \alpha] \geq E[J(\omega)|F(\tilde{x}) \geq \alpha],$$

while the wage decreases for those who do not graduate. The point is that the labor market interprets a failure to graduate as bad news about ability, and as worse news when the schooling process is more revealing. Note that the average wage (and average production) for the whole economy is constant at $E[J(\omega)]$, but inequality increases with information.

To extend this, suppose now that a fraction $1 - \beta$ of workers, $\beta < \alpha$ can be hired into jobs that are “skill-sensitive” — that is, have production function $S(\omega)$, with $S'(\omega) > J'(\omega)$ for all ω . And suppose that $E[S(\omega)] = E[J(\omega)]$ so that “on average” productivity is the same at the two jobs.

Assuming that each worker gets her expected marginal product as a wage, the equilibrium in this two-job labor market has a fraction β of workers, all of whom graduated, going to “skill-sensitive” jobs at a wage $E[S(\omega)|F(\tilde{x}) \geq \alpha]$, the remaining graduates taking less skill-sensitive jobs at a lower wage (they are rationed), and non-graduates receiving $E[J(\omega)|F(\tilde{x}) < \alpha]$ as before.

Consider an increase in the information generated by school screening in the sense of (*MIO-SPM*). This will increase the wage for graduates and decrease the wage for non-graduates. But it will also increase the total production of the economy by leading to better matching between high-skill workers and skill-sensitive jobs. Moreover, as the fraction of skill-sensitive jobs, β , increases toward α , the social returns to better screening by the schools increase (assuming the planner cares only about gross production and not inequality). Similarly, if $S(\omega, \tau)$ is supermodular in τ , then the social returns to better screening will be increasing in τ .

6.4 Application: Coordination Under Uncertainty

This section considers a game where both players can choose to acquire information. The game involves two players with symmetric payoffs $u(\omega, a_i, a_j)$ where a_i is player i 's action. Player i receives signal $\tilde{\alpha}_i$, with joint distribution $F^{\theta_i}(\omega, \alpha_i)$, where $F^{\theta_i}(\alpha_i) = \alpha_i$ and θ_i reflects signal quality. Assume the $\tilde{\alpha}_i$'s are independent conditional on ω . We let $u(\omega, a_1, a_2) = a_1[\omega + \gamma a_2 + K] - \frac{1}{2}\tau a_1^2$, and restrict $\tau > \gamma > 0$. In this supermodular game, players make unobserved choices of signal quality (θ_1, θ_2) and then choose their strategies (a_1, a_2) . We look for symmetric Nash equilibria. Conditional on the information structure, Player 1 maximizes

$$\int_{\Omega} \left[a_1[\omega + \gamma E[a_2|\omega; \theta] + K] - \frac{1}{2}\tau a_1^2 \right] dF^{\theta_1}(\omega|\alpha_1)$$

to find an optimal action $a_1(\alpha_1) = \frac{1}{\tau} [E[\omega|\alpha_1; \theta] + \gamma E[a_2|\alpha_1; \theta] + K]$. In the unique (essentially) symmetric equilibrium, $a_i(x) = \frac{1}{\tau - \gamma} [E[\omega|\alpha_i; \theta_i] + K]$.

The first question to ask is when the optimal policy will be monotone in α_i . Clearly requiring that the posteriors are ordered by FOSD is sufficient; actually all that is needed is that the posteriors be ordered according to their first moments. Since any two distributions can be ranked according to their means, assuming (*MDP-F*) is simply notational. The linearity of payoffs arises since when j plays her equilibrium strategy, $E[a_j|\omega] = \frac{1}{\tau - \gamma}[\omega + K]$, and so $u_i(a_i, \omega) = a_i \frac{\tau}{\tau - \gamma}[\omega + K]$. Thus marginal returns are of the form $u_a(\omega) = c + d\omega$, where $d > 0$. The set of marginal returns satisfying these restrictions forms a convex cone, with corresponding (*MDP*) and (*MIO*) conditions. The condition for monotonicity of the policy is that $E[\omega|\alpha_i]$ be increasing in α_i . The condition

under which G is more informative than F for all decision-makers with linear marginal returns can be derived immediately from Theorem 3: $\forall \alpha \in [0, 1]$,

$$E[\omega|F(x) < \alpha] \geq E[\omega|G(y) < \alpha]. \quad (\text{MIO-LIN})$$

This is the linear monotone information order.

We now consider the information acquisition choice. Assume that θ_i indexes F_i according to (*MIO-LIN*). Expected payoffs to player i are $\Pi(\theta_i, \theta_j) = E[u(\omega, a_i(\alpha_i), a_j(\alpha_j))|\theta_1, \theta_2]$, so the symmetric equilibrium must also satisfy $\Pi_i(\theta_i, \theta_i) - C'(\theta_i) = 0$, the first order condition for information acquisition. We are also interested in how changes in the parameters alter the amount of information gathered in equilibrium. Straightforward algebra yields

$$u^{\theta_i}(\omega, \alpha_i) = \frac{\tau}{(\tau - \gamma)^2} [E[\omega|\alpha_i; \theta_1] + K] \left[(\omega + K) - \frac{1}{2}(E[\omega|\alpha_i; \theta_i] + K) \right],$$

$$\frac{\partial}{\partial \alpha_i} u^{\theta_i}(\omega, \alpha_i) = \frac{\tau}{(\tau - \gamma)^2} [\omega - E[\omega|\alpha_i; \theta_i]] \frac{\partial}{\partial \alpha_i} E[\omega|\alpha_i; \theta_i].$$

Comparative statics follow immediately from an application of Theorem 10. A change in K , the known marginal benefit to acting, has no effect on the amount of information gathered. An increase in the quadratic cost of action, τ , decreases the amount of information gathered. And as is intuitive, an increase in γ , the returns to coordination, increases information gathering. Similarly, agent's actions are increasing γ and K and decreasing in τ . Endogenizing the information structure in this game *reinforces* known complementarities.

7 Conclusion

In this paper, we have obtained a new set of results for the standard Bayesian decision problem. We have provided a general analysis of when one signal is more valuable than another signal to a given class of decision-makers, and also provided conditions under which decision-makers within a given class will differ systematically in their marginal value for information. We have applied this general framework to several payoff classes of particular economic relevance, and described their corresponding monotone information orders. As we have attempted to point out in the previous section, the results can be applied to a variety of economic settings, allowing characterizations of what sort of information is valuable in a given environment, and which sorts of agents care most about acquiring it.

An extremely interesting, but potentially difficult extension of this research, is to analyze the value of information in strategic settings. In such a setting, information may have both a direct and a strategic value. In standard incentive problems, information has only strategic value, while in Persico's (1997) study of auctions, information is acquired secretly so there is only a direct value. In the industrial organization literature discussed in Section 6.1, results combining both direct and strategic benefits are obtained, but they typically rely on restrictive functional form assumptions. The examples from this literature suggest that unambiguous results about the value of information may not be available in many settings. It remains to be seen whether general characterization results can be obtained in game-theoretic models.

8 Appendix

Proof of Lemma 5. Assume first that A is finite, $A = \{a_1, \dots, a_n\}$ with $a_{i+1} > a_i$. Then for any monotone nondecreasing $a(\alpha)$, there exist $\alpha_0 = 0 \leq \alpha_1 \leq \dots \leq \alpha_{n-1} \leq \alpha_n = 1$ such that $a(\alpha) = a_i$ on $[\alpha_{i-1}, \alpha_i]$. Then

$$\begin{aligned}
\int_{\Omega} \int_{[0,1]} u(\omega, a(\alpha)) dm(\omega, \alpha) &= \int_{\Omega} \sum_{i=1}^n u(\omega, a_i) [m(\alpha_i|\omega) - m(\alpha_{i-1}|\omega)] dm(\omega) \\
&= \int_{\Omega} \left\{ \begin{array}{l} u(\omega, a_n)m(1|\omega) - u(\omega, a_1)m(0|\omega) \\ - \sum_{i=1}^{n-1} [u(\omega, a_{i+1}) - u(\omega, a_i)] m(\alpha_i|\omega) \end{array} \right\} dm(\omega) \\
&= - \int_{\Omega} \left\{ \sum_{i=1}^{n-1} [u(\omega, a_{i+1}) - u(\omega, a_i)] m(\alpha_i|\omega) \right\} dm(\omega) \\
&= - \sum_{i=1}^{n-1} \int_{\Omega} r_i(\omega) dm_{\omega}(\omega, \alpha_i)
\end{aligned}$$

for some $r_1(\omega), \dots, r_{n-1}(\omega)$ with $r_i(\omega) \in U_1$. The third equality uses the fact that $m(1|\omega)$, $m(0|\omega)$ are zero a.e. with respect to $m(\omega)$. The last step follows from $(MDP-U)$. Then clearly (16) implies (15). Moreover, suppose (16) fails for some $\hat{\alpha} \in (0, 1)$, $\hat{r} \in U_1$. Then define $\hat{u}(\omega, a) = a\hat{r}(\omega)$. And let $\hat{a}(\alpha) = a_1$ on $[0, \hat{\alpha})$ and $\hat{a}(\alpha) = a_n$ on $[\hat{\alpha}, 1]$. By $(MDP-U)$, $\hat{u} \in U_2$, and \hat{a} is clearly nondecreasing. Substituting into the derivation above,

$$\int_{\Omega} \int_{[0,1]} \hat{u}(\omega, \hat{a}(\alpha)) dm(\omega, \alpha) = -(a_n - a_1) \int_{\Omega} \hat{r}(\omega) d_{\omega} m(\omega, \hat{\alpha}) < 0.$$

So (15) must imply (16).

Now suppose A is compact. We know from above that (16) is equivalent to (15) holding for all $a(\alpha)$ step functions. We show that (15) holds for all nondecreasing functions $a(\alpha)$ if and only if it

holds for all step functions. Let $a(\alpha)$ be some nondecreasing function, and let $a^1(\alpha), a^2(\alpha), \dots$ be a sequence of step functions converging to $a(\alpha)$. Then since u is continuous in a , then $u(\omega, a^k(\alpha))$ will be converging to $u(\omega, a(\alpha))$. And since $u(\omega, a)$ is bounded, we can apply the Lebesgue Convergence Theorem to show

$$\int_{\Omega \times [0,1]} u(\omega, a^k(\alpha)) dm(\omega, \alpha) \longrightarrow \int_{\Omega \times [0,1]} u(\omega, a(\alpha)) dm(\omega, \alpha).$$

Then clearly, (15) will hold for all a nondecreasing if and only if it holds for all step functions. \square

References

- [1] ATHEY, SUSAN (1998a): “Characterizing Properties of Stochastic Objective Functions,” MIT Working Paper 96-1R.
- [2] ATHEY, SUSAN (1998b): “Comparative Statics under Uncertainty: Single Crossing Properties and Log-Supermodularity,” MIT Working Paper 97-11R.
- [3] BLACKWELL, DAVID (1951): “Comparisons of Experiments,” *Proceedings of the Second Berkeley Symposium on Mathematical Statistics*, 93–102.
- [4] BLACKWELL, DAVID (1953): “Equivalent Comparisons of Experiments,” *Annals of Mathematical Statistics*, 24, 265–272.
- [5] BLACKWELL, DAVID AND M. GIRSHIK (1954): *Theory of Games and Statistical Decisions*. John Wiley & Sons.
- [6] BORDER, KIM (1991): “Functional Analytic Tools for Expected Utility Theory,” in *Positive Operators, Riesz Spaces, and Economics*, ed. C.D. Aliprantis, K. Border and W. Luxemburg. Springer Verlag.
- [7] DEWATRIPONT, M., I. JEWITT, AND J. TIROLE (1997): “The Economics of Career Concerns: Theory and an Application to Mission Setting,” Mimeo, University of Bristol.
- [8] GOLLIER, C. AND M. KIMBALL (1995): “Toward a Systematic Approach to the Economic Effects of Uncertainty I: Comparing Risks,” Mimeo, University of Toulouse.
- [9] JEWITT, IAN (1986): “A Note on Comparative Statics and Stochastic Dominance,” *Journal of Mathematical Economics*, 15, 249–254.

- [10] JEWITT, IAN (1997): “Information and Principal Agent Problems,” mimeo, University of Bristol.
- [11] KARLIN, SAMUEL (1968): *Total Positivity: Volume I*. Stanford University Press.
- [12] KARLIN, SAMUEL AND WILLIAM STUDDEN (1966): *Tchebycheff Systems with Applications in Analysis and Statistics*. New York: Interscience.
- [13] LE CAM, LUC (1964): “Sufficiency and Approximate Sufficiency,” *Annals of Mathematical Statistics*, 35,
- [14] LEHMANN, E.L. (1988): “Comparing Location Experiments,” *Annals of Statistics*, 16 (2): 521-533.
- [15] LEVY, HAIM AND JACOB PAROUSH, “Toward Multivariate Efficiency Criteria,” *Journal of Economic Theory* 7 (2) Feb., pp. 129-42.
- [16] MEYER, MARGARET (1990): “Interdependence of Multivariate Distributions: Stochastic Dominance Theorems and an Application to the Measurement of Ex Post Inequality under Uncertainty,” Nuffield College Discussion Paper, no. 49.
- [17] MILGROM, PAUL (1981): “Good News and Bad News: Representation Theorems and Applications,” *Bell Journal of Economics*, 12, 380–391.
- [18] MILGROM, PAUL AND JOHN ROBERTS (1990): “Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities,” *Econometrica*, 58, 1255–1277.
- [19] MILGROM, PAUL AND CHRISTINA SHANNON (1994): “Monotone Comparative Statics,” *Econometrica*, 62, 157–180.
- [20] MIRMAN, L., L. SAMUELSON, AND E. SCHLEE (1994): “Strategic Manipulation of Information in Duopolies.” *Journal of Economic Theory*, 36, 364-384.
- [21] PERSICO, NICOLA (1996): “Information Acquisition in Affiliated Decision Problems,” mimeo, UCLA.
- [22] PERSICO, NICOLA (1997): “Information Acquisition in Auctions,” mimeo, UCLA.
- [23] ROSS, STEPHEN (1981): “Some Stronger Measures of Risk Aversion in the Small and the Large with Applications,” *Econometrica* 49 (3), pp. 621-38.

- [24] ROTHSCHILD, MICHAEL, AND JOSEPH STIGLITZ (1970): “Increasing Risk: I. A Definition,” *Journal of Economic Theory* 2 (3), pp. 225-43.
- [25] SHAKED, MOSHE AND GEORGE SHANTIKUMAR (1997): “Supermodular Stochastic Order and Positive Dependence of Random Variables,” *Journal of Multivariate Analysis*, 61, 86–101.
- [26] SHANNON, CHRIS (1995), “Weak and Strong Monotone Comparative Statics,” *Economic Theory* 5 (2), March, pp. 209-27.
- [27] STRASSEN, V. (1965): “The Existence of Probability Measures with Given Marginals,” *Annals of Mathematical Statistics*, 36, 423–439.
- [28] TOPKIS, DONALD (1978): “Minimizing a Submodular Function on a Lattice,” *Operations Research*, 26, 305–321.