Comparison of Experiments in Monotone Problems*

Alfredo Di Tillio[†] Marco Ottaviani[‡] Peter Norman Sørensen[§]

April 10, 2022

Abstract

Blackwell (1951) characterized when an experiment is more informative than another, in the sense that no rational decision maker would prefer observing the second experiment rather than the first. This paper provides a novel characterization for a binary-signal experiment A to be more informative than another arbitrary experiment B for all decision makers with preferences in Quah and Strulovici's (2009) interval dominance ordered class, encompassing monotone decision problems (Karlin and Rubin, 1956) and single-crossing preferences (Milgrom and Shannon, 1994). We show that if experiment A satisfies the monotone likelihood ratio property, then A is more informative than B if and only if all posterior beliefs induced by B are dominated by (dominate) the belief induced by the highest (lowest) signal from A in the likelihood ratio order. If instead experiment A fails to satisfy the monotone likelihood ratio property, Blackwell's (1951) characterization applies: A is more informative than B if and only if B is a garbling of A.

Keywords: Accuracy; Comparison of experiments.

JEL codes: D83, C90

*We thank Biao Yang for excellent research assistance. We also thank seminar participants at the University of Macau for helpful comments.

[†]Department of Economics and IGIER, Bocconi University, Via Roberto Sarfatti 25, 20136 Milan, Italy. Phone: +39–02–5836–5422. E-mail: alfredo.ditillio@unibocconi.it.

[‡]Department of Economics and IGIER, Bocconi University, Via Roberto Sarfatti 25, 20136 Milan, Italy. Phone: +39–02–5836–3385. E-mail: marco.ottaviani@unibocconi.it.

[§]Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK–1353 Copenhagen K, Denmark. Phone: +45–3532–3056. E-mail: peter.sorensen@econ.ku.dk.

1 Introduction

With the progress of machine learning techniques, a large number of powerful prediction algorithms are becoming available; see Agrawal, Gans, and Goldfarb (2018). How can the value of information corresponding to different prediction machines be compared? Blackwell's (1951, 1953) traditional approach compares experiments without imposing restrictions on preferences. With this approach, however, rather few experiments are comparable. As noticed by Lehmann (1988), Persico (2000), Jewitt (2007) and Quah and Strulovici (2009), the class of experiments that can be compared grows larger once preferences are restricted to a smaller but natural class with a suitably monotonic structure. This paper aims at characterizing informativeness for the restricted domain of monotone preferences, in analogy with Blackwell's (1951) characterization for the universal domain of preferences.

Consider a decision maker (DM) who faces a problem under uncertainty. The DM chooses an action a to maximize the expectation of utility $u(a,\theta)$, where θ is the unknown state of nature. Before deciding, the DM observes a signal s with state-contingent distribution $f(s|\theta)$, that is, an experiment. How can we quantify the value of the experiment to the DM, and how can we compare two experiments? For given utility function and prior $\pi(\theta)$, the answer is simple: the value of experiment is the difference between the ex-anteexpected utilities of the DM with and without the experiment. Then the question becomes how we can order two experiments for an arbitrary set of utilities and priors. The answer clearly depends on the set considered.

Blackwell's (1951, 1953) foundational work characterizes the comparison of experiments when the set is unrestricted and contains all utilities and priors (universal domain). According to Blackwell, an experiment A is more informative than B when a DM with any utility function and prior prefers to observe experiment A over B. As Blackwell demonstrates, this happens if and only if experiment B is a *garbling* of A, that is, B can be obtained by adding noise (unrelated to the underlying state) to experiment A. By insisting on general utilities and priors, Blackwell's ordering leaves many experiments not comparable with each other—for example, Lehmann (1988) shows that, surprisingly, Blackwell's informativeness fails to compare two location experiments with uniform noise. Furthermore, the notion of garbling does not characterize the "more informative than" relation in economic contexts, where preferences are typically restricted to a class smaller than the universal domain—B being a garbling of A remains of course sufficient, but is generally no longer necessary for A to be preferred to B, once we only consider DMs in the smaller class.

Motivated by these issues, we provide a new characterization of when an experiment is more informative than another in every monotone decision problem. More precisely, we restrict the domain of preferences by assuming that $\{u(\cdot,\theta)\}_{\theta\in\Theta}$ is an interval dominance ordered (IDO) family (Quah and Strulovici, 2009). While comparing more experiments than Blackwell, our ordering is widely applicable, because IDO preferences are natural in economic contexts—for example, they include Karlin and Rubin's (1956) monotone preferences and Milgrom and Shannon's (1994) single-crossing preferences. As a proof of concept, here we develop the approach when the dominating experiment A is binary, with two possible signal realizations in its support.[1] Our characterization hinges on whether experiment A satisfies or violates the monotone likelihood ratio (MLR) property—higher signals indicate higher states. We prove that if A satisfies the MLR property, then experiment B is dominated by A precisely when all the posteriors induced by B are contained in the area defined as the interval under the likelihood ratio order between the posteriors induced by the highest and lowest signals from experiment A. This includes non-trivial cases where experiment B violates the MLR property.[2] If instead A fails to satisfy MLR, then Blackwell's characterization applies: the dominated experiment B must be a garbling of A.

As an illustration for our model with three states of nature, and experiments without the MLR property, consider the following problem naturally faced by firms in a small economy. A firm chooses how much to invest. Its utility is the profit earned on this investment, and this depends on the political climate of the next parliament. The states are three possible climates: left-dominated, centrist, and right-dominated. The ideal climate for the firm is centrist, while the worst climate for its profit is left-dominated when businesses must contribute more to the public. The right-dominated climate leaves profit between the two others, as moderate harm comes when exports suffer for diplomatic reasons. The firm's utility satisfies IDO when the three states are ordered centrist $>$ right $>$ left. Various informative experiments can be available. Some directly measure the outlook for firms, and they satisfy MLR. Other experiments instead measure the left-center-right inclinations of the electorate, and these political polls violate MLR, because they instead satisfy MLR when states are differently ordered with centrist as the middle state.

The closest connections in the recent literature are Athey and Levin (2018) and Kim (2021), who consider combinations of experiments with decision problems such that the decision maker optimally chooses a monotone map from signals to actions. Athey and Levin (2018) impose a monotonicity assumption on comparable experiments which appears to rule out our ability to compare any binary experiment A to any other experiment B. In Kim (2021), the decision problems depend on the experiment pair, and the proposed pairwise comparison of experiments is no longer

[1] We also retrict attention to the simplest interesting case, with three possible states of nature.

[2] Lehmann (1988) restricts preferences to the class of problems defined by Karlin and Rubin (1956) and experiments to the family of experiments satisfy the MLR property. For this restricted class of experiments, Lehmann (1988) shows that a notion of greater accuracy is equivalent to his ordering. Intuitively, MLR experiments with less dispersed noise distributions (i.e., with higher density at every quantile) are more informative. Persico (2000) and Jewitt (2007) prove a version of this result for Milgrom and Shannon's (1994) single-crossing preferences; Quah and Strulovici (2009) and Di Tillio, Ottaviani, and Sørensen (2021) extend the result to the more general case of interval dominance ordered (IDO) preferences, which we also assume.

transitive.[3] We instead stay closer to the original intentions of Blackwell (1951) and provide a novel characterization of an incomplete order of all experiments.

2 Setup

The setup has three main components.

First, the unknown state of nature θ, living in space Θ. We hold constant the space Θ, which is assumed to be a finite, ordered set. For simplicity of exposition, we focus on the case with three states, $\Theta = \{\theta_1, \theta_2, \theta_3\}$, ordered $\theta_3 \succ \theta_2 \succ \theta_1$.

Second, an experiment that produces signal s living in signal set S. The experiment specifies S and the collection of conditional probability distributions for signal s on S given any state $\theta \in \Theta$. We restrict attention to finite signal sets. A conditional distribution is then defined by the discrete density function $f(s|\theta)$. We assume that no signal is redundant, that is, for every signal s there exists a state θ with $f(s|\theta) > 0$. Since f completely describes the experiment, we simply call it experiment f.

Third, a decision problem for a decision maker (DM). A decision problem specifies a prior belief $\pi \in \Delta(\Theta)$ for the DM, i.e., a probability distribution over the state space. It also specifies an ordered action set A and a utility function $u : A \times \Theta \to \mathbb{R}$. For simplicity of exposition, we restrict attention to decision problems with finite action sets, and we only consider full support prior beliefs π.

Bringing together all three components, first nature draws state θ, then the DM observes a signal s from its conditional distribution given θ, and finally the DM takes action $a \in A$. The DM is assumed to choose $a \in A$ in order to maximize the posterior expected utility given observed signal s. By Bayes' rule, at signal s the posterior chance of any state $\theta \in \Theta$ is

$$\pi(\theta|s) = \frac{\pi(\theta) f(s|\theta)}{\sum_{\tilde{\theta}\in\Theta} \pi\left(\tilde{\theta}\right) f\left(s|\tilde{\theta}\right)}.$$

The posterior expected utility from action a is

$$u(a|s) = \sum_{\theta\in\Theta} \pi(\theta|s)\, u(a,\theta).$$

Since A is finite, an optimal rule $a(s)$ exists. For any such optimal rule, we can let

$$U = \sum_{\theta\in\Theta} \pi(\theta) \sum_{s\in S} f(s|\theta)\, u(a(s),\theta)$$

denote the maximal utility that this DM can obtain with this experiment.

[3] We will further explain this below.

The main question is which experiments are better for DMs, in the sense of permitting them greater maximal utility. Experiment f is Blackwell more informative than experiment g when experiment f results in greater maximal utility than g for every decision problem. Blackwell found that this definition is equivalent to an intuitive notion of g containing more noise than experiment f, that g is a garbling of f. Namely, experiment g can be replicated by first drawing a signal from f, and then randomizing over the signal set from experiment g in a fashion that depends only on the drawn signal from f, but otherwise not on state θ.

According to Blackwell's definition, many experiments cannot be compared. For example, Lehmann (1988) exhibits uniform location experiments that are not comparable in terms of Blackwell's informativeness. To address this limitation, and shed further light on informativeness in economic contexts, we follow Quah and Strulovici (2009) by considering DMs whose utility functions satisfy the interval dominance order (IDO): for all $\theta' > \theta$ and $a' > a$,

$$u(a',\theta) \geqslant (>)\, u(a,\theta) \quad \implies \quad u(a',\theta') \geqslant (>)\, u(a,\theta')$$

whenever $u(a',\theta) \geqslant u(a'',\theta)$ for all a'' with $a \leqslant a'' \leqslant a'$. In words, if action a' is the best action in the interval $[a,a'] \cap A$ when the state is θ, then a' remains the best action in $[a,a'] \cap A$ when the state is raised to θ'. Quah and Strulovici (2009) note that this class of preferences includes Milgrom and Shannon (1994) single-crossing preferences and Karlin and Rubin (1956) monotone preferences.

Monotonicity of a decision problem means that there is an increasing solution to the problem of maximizing $u(a,\theta)$ over $a \in A$ for given θ. The literature has provided sufficient conditions on the utility function that guarantees this property.[4] The literature has further focused on the question whether the optimal map from signals to actions is monotone. Quah and Strulovici (2009) demonstrate that, given an IDO decision problem and an experiment, there exists a monotone optimal rule $a(s)$, provided also that the experiment satisfies the monotone likelihood ratio (MLR) property: the signal set can be ordered such that, for any $\theta' > \theta$, the likelihood ratio $f\left(s|\theta'\right)/f\left(s|\theta\right)$ rises in s. In words, higher signals are relatively more likely in higher states. Once the signal set is thus ordered, the cumulative distribution function is $F\left(s|\theta\right) = \sum_{\tilde{s} \preceq s} f\left(\tilde{s}|\theta\right)$.

3 Results

An experiment is binary if its signal space contains two elements, $|S| = 2$.

Proposition 1. *Binary MLR experiment f with signal space $\{s_L, s_H\}$ results in greater maximal utility than g with signal space S for every IDO decision problem if and only if, all signals $s \in S$*

[4] Assumptions on the utility function can help to secure that the decision remains monotone if the action set is constrained to a subset $\tilde{A}$ of A. In the appendix, we prove that IDO has this property in the three-element state space. Single-crossing immediately satisfies the property, as shown by Milgrom and Shannon.

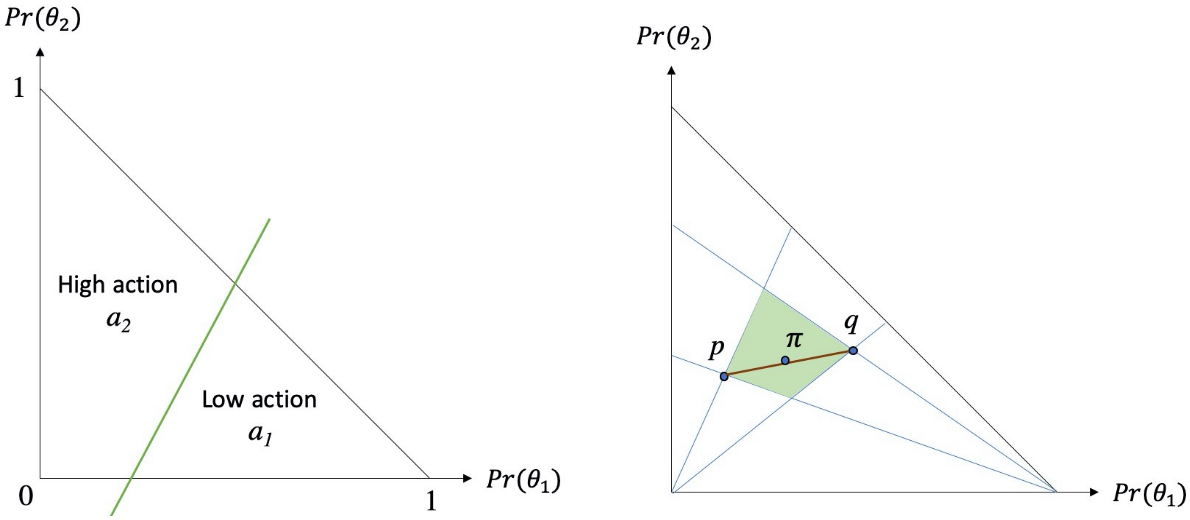

Figure 1: Illustration of (A) Marschak's triangle and (B) the LR box.

satisfy the likelihood ratio ordering

$$\frac{f(s_H|\theta_2)}{f(s_H|\theta_1)} \geqslant \frac{g(s|\theta_2)}{g(s|\theta_1)} \geqslant \frac{f(s_L|\theta_2)}{f(s_L|\theta_1)} \text{ and } \frac{f(s_H|\theta_3)}{f(s_H|\theta_2)} \geqslant \frac{g(s|\theta_3)}{g(s|\theta_2)} \geqslant \frac{f(s_L|\theta_3)}{f(s_L|\theta_2)}. \tag{1}$$

Binary non-MLR experiment f results in greater maximal utility than g for every IDO decision problem if and only if, g is a garbling of f.

We formally prove this proposition in the appendix. In the following, we informally explain two key steps.

Recall that we consider state space $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Any belief over the set of states can be represented by a point in Marschak's triangle, as shown in panel A of Figure 1. Let the points $(1,0)$, $(0,1)$, and $(0,0)$ represent beliefs that assign one-hundred percent weight on state θ_1, θ_2, and θ_3, respectively. Since the signal is binary, experiment f can be represented by the DM's two posteriors (p,q) which are connected by a segment that passes through the prior π, given that the expected posterior must be equal to the prior by the martingale property of beliefs.

For our first key step, suppose that also g is a binary experiment. We will visualize the set of experiments g with posterior pair (r,s) that are less informative than experiment (p,q).

For any decision problem, the expected utility of each action is a linear function of the beliefs, defining a plane above the Marschak triangle. Under binary experiment g, at most two actions from A are needed, those which are optimal at the two possible signals.[5] So, consider a decision problem with binary action set $A = \{a_1, a_2\}$, ordered $a_2 \succ a_1$. The two expected-utility planes

[5]In decision problems where g optimally employs only one action, trivially experiment f can perform at least as well by always taking that same action.

intersect along a line. Projecting this into the Marschak triangle, we obtain the DM's indifference line among a_1 and a_2 (the green line in Panel A of Figure 1). Since the decision problem is monotone and non-trivial, action a_2 must be strictly optimal in state θ_3 and a_1 strictly optimal in state θ_1. Hence the indifference line must cross the base of the triangle and the higher action a_2 must be optimal to the left of this line.

Posterior pair (r,s) defines an experiment which is better than f in case there exists an indifference line that separates (r,s) without separating (p,q). Then (r,s) induces higher expected utility in this problem. Given the possible indifference lines, the set of posterior pairs is therefore restricted to the green box as indicated in the panel B of Figure 1. Specifically, any experiment with a posterior outside the box can be separated by some monotone indifference line.

This box is the set of beliefs that are LR (likelihood ratio) ordered between p and q. That is, for any belief t inside the box, $p \succ_{LR} t \succ_{LR} q$ (weakly).[6] Thus, we call it the LR box. Any experiment with signal posteriors inside the LR box has posteriors that are less LR-variable than (p,q). We have here demonstrated that experiments dominated by (p,q) must be in this box.[7] Our formal proof verifies that having (r,s) in the box is also sufficient for g to be dominated by f.

Note that the box is non-empty only when p and q are LR ordered, i.e., the better experiment f has MLR. When the MLR property is violated, experiment (p,q) only dominates experiments with posteriors inside the $p-q$ segment, which is then a Blackwell garbling of (p,q). To see this, if an experiment has a posterior outside the $p-q$ segment, we can always find a monotone indifference line that separates posteriors of that experiment but does not separate (p,q).

We observe that the green box contains binary experiments which violate the MLR property. Experiments violate the MLR property when for instance some signals are directional (sorting higher from lower states), but other signals are central (sorting middle states from extreme states).[8]

Note also that our construction does not narrow down to our LR box, if we fix some ordering of the signals in g and then only consider decision problems with optimal monotone map from signal to action. For illustration, suppose that s sits to the north-east of π in panel B of Figure 1, outside the LR box, while r sits inside the LR box (on the opposite side of π), in such a fashion that s,r are not LR ordered. If this binary experiment g is considered with the order $s \succ r$, then we

[6] Being to the right of the upward-sloping line through p means that $r_2/r_1 \leqslant p_2/p_1$. Being to the right of the downward-sloping line through p means that $r_3/r_2 \leqslant p_3/p_2$. Thus, $p \succ_{LR} r$. Similarly with the lines through q. Furthermore, when p is the posterior at s_H and r is the posterior at signal s from experiment g, from Bayes' rule, $r_2/r_1 \leqslant p_2/p_1$ is equivalent to the first inequality in (1).

[7] Following standard terminology, we say that experiment f dominates experiment g when f provides greater maximal utility in every IDO decision problem.

[8] Lehmann's (1988) characterization only applies to MLR experiments, as his sufficiency proof crucially rests on the weaker experiment g satisfying the MLR property—it starts from an optimal monotone rule $a(s)$ of the weaker experiment, and then uses accuracy to improve on this rule under the stronger experiment. This, in a sense, that characterization provides only "half" of the green box squeezed between the two diagonals through prior belief π.

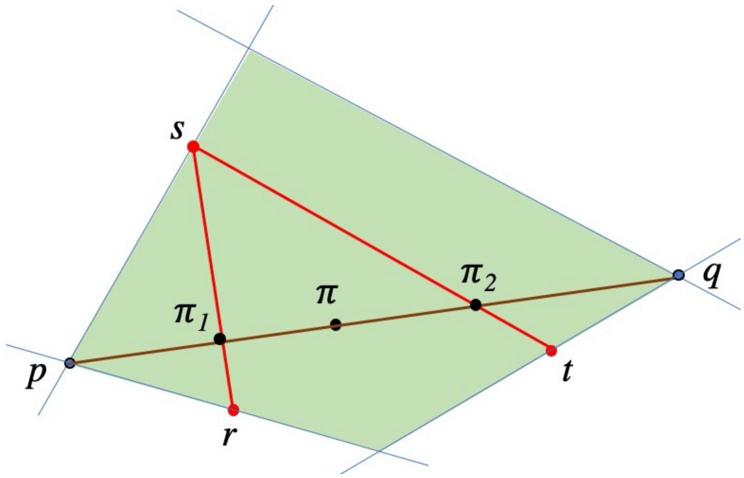

Figure 2: Illustration of decomposition of experiments.

can no longer prove that f dominates g. Our proof appeals to an indifference line that emanates from $(0,1)$ to define the upper right side of the LR box, with lower actions taken to the right of the line. However, then the low action a_1 is taken at the high signal s (and a_2 at r), in violation of signal-to-action monotonicity . Thus, this decision problem would not be permitted under such a restriction, and it would no longer be true that g is dominated by f. Our characterization thus differs from those of Athey and Levin (2018) and Kim (2021).

Our specification of the LR box and conclusion extends to the case when the action set is arbitrary. Suppose set A contains more than two actions. Consider two experiments (p,q) and (r,s), where (r,s) is inside the LR box of (p,q). Assume that the optimal actions are a and a' for experiment (r,s). By the earlier argument, (r,s) is dominated by (p,q) when we restrict the action set to be $\{a,a'\}$. Now enlarge the action set to A. This leaves payoff under experiment (r,s) unchanged but may only improve expected utility under (p,q), with a richer set of actions.

Our second key step shows that the conclusion also holds when the dominated experiment has more than two signals. For illustration of this step, consider a simple example where an experiment has three posteriors (r,s,t), all of them inside the LR box defined by (p,q). The trick is that both experiments can be regarded as composites of two binary experiments. Figure 2 shows the decomposition. Specifically, since the unconditional average of the three posteriors (r,s,t) should be the prior π, it must be that one posterior and the other two are separated by the $p-q$ segment. Suppose that posterior s is on the one side of the $p-q$ segment and r,t are on the other side. Let segments $s-r$ and $s-t$ intersect segment $p-q$ at π_1 and π_2. Experiment (r,s,t) with prior π is a composite of experiment (s,r) with prior π_1 and experiment (s,t) with prior π_2. That is, the expected utility of (r,s,t) is a weighted average of those of (s,r) and (s,t). The same holds of experiment (p,q) with prior π, which is a composite of experiments (p,q) with prior π_1 and (p,q)

with prior π_2. By the earlier binary-signal argument, (p,q) with prior π_1 dominates experiment (s,r), and (p,q) with prior π_2 dominates experiment (s,t). Thus, (p,q) dominates (r,s,t). When the dominated experiment has more than three signals, we can apply the same decomposition technique.

4 Discussion

Starting from a binary experiment f, Proposition 1 fully characterizes the experiments g which are less informative than f. We now discuss the relation of this result to the prior literature.

4.1 Garbling

Our notion of informativeness differs from Blackwell's, as we restrict attention to decision problems that satisfy IDO. As is well known, the binary experiment f dominates any experiment g where the support of posteriors under g is contained in the segment connecting the two posteriors under f.

Since there are fewer problems where we test whether f delivers higher expected utility than g, it is easier for any experiment g to be dominated by f. When f satisfies the MLR property, indeed we have found that f dominates any g with support in the LR box depicted in panel B of Figure 1. This result shows that our restriction to IDO problems considerably expands on the pairs of comparable experiments. This is not too surprising in light of Lehmann's results on the comparison of experiments in monotone decision problems.

On the other hand, when the binary f violates the MLR property, we find that our notion is no weaker than Blackwell's. Thus, the IDO class is sufficient for Blackwell's comparison of f to any g in this case. This fact may seem counterintuitive, since we are accustomed to combining IDO with the MLR assumption. But, as Figure 1 suggests, when f violates the MLR, there is an IDO problem where both posteriors under f sit on the indifference line, and perturbations of this problem allow to show that f cannot dominate experiments with support outside the interval between the two posteriors of f.[9]

4.2 Accuracy

Our restriction to IDO problems is in the spirit of Lehmann (1988), who considered the closely related class of problems introduced by Karlin and Rubin (1956). IDO preferences were introduced

[9]Actually, the expansion of the Blackwell interval to the LR box when f does satisfy the MLR is only possible because decision problems with indifference lines parallel to the Blackwell interval cannot satisfy IDO.

by Quah and Strulovici (2009), who also discuss the exact connection to Lehmann's work—see also Di Tillio et al. (2021).

Yet, the main characterization result by Lehmann restricts attention to the comparison of experiments f and g that both satisfy the MLR property. In part, Proposition 1 clarifies that if experiment f would violate the MLR, it could not dominate any g that satisfies the MLR. In part, we extend beyond Lehmann's characterization to show that when f satisfies the MLR, there also exist some dominated experiments g which violate the MLR.

When viewed in panel B of Figure 1, Lehmann's result only considers the binary experiments g with support lining up on rotations around π of the p, q interval, such that the line containing the support does not intersect the horizontal axis between 0 and 1. Rather than the full LR box, these rotations would thus depict a butterfly centered on prior π.

Lehmann's notion of accuracy is most easily expressed for continuously distributed signals.[10] Signals in both experiments are ordered (and real), with cumulatives $F(s|\theta)$ and $G(s'|\theta)$. Starting from arbitrary s and θ, it is of interest to match quantiles, i.e., to find s' that solves $G(s'|\theta) = F(s|\theta)$, i.e., $s' = G^{-1}(F(s|\theta)|\theta)$. Lehmann's accuracy notion requires that $G^{-1}(F(t|\theta)|\theta)$ is decreasing in θ for any t. Jewitt (2007) studies this notion in depth.

4.3 Monotone information order

Athey and Levin (2018) propose a generalization of Lehmann's approach to compare experiments. They relax the MLR assumption, and instead impose directly that experiments f and g should only be compared for the subset of decision problems where optimal decisions are monotone in signals. Towards this end, they combine various notions of signal ordering with appropriate notions of monotone decision problems. One example would be signals ranked by the MLR combined with IDO problems, but they point to other interesting possibilities. Their main theorem provides a characterization of their comparison in terms of a monotone information order which compares suitable averages of posterior beliefs.

More precisely, any such combination is defined by some set R of real functions of the state. The permitted decision problems are those where, for any actions $a' \succ a$, the payoff difference $u(a', \theta) - u(a, \theta)$ lies in R. An experiment f is R-ordered if its signals can be somehow ordered to satisfy single-crossing on R: for any $s' \succ s$ and for all $r \in R$, it holds that $E(r(\theta)|s') \geqslant 0 \Rightarrow E(r(\theta)|s) \geqslant 0$. The characterization theorem allows for comparison of two experiments which are both R-ordered.

While it is not clear whether such an R exists to define the set of IDO problems, perhaps we could work instead with the supermodular decision problems which are induced when R is the set

[10] An atomic signal can be represented by a continuous variable.

of increasing functions. Then the MLR implies the single-crossing condition. However, we do not see how the approach of Athey and Levin could be arranged to allow for the comparison of any binary f to any experiment g. The key advantage of our Proposition 1 is that it imposes no order on the two experiments.

4.4 Monotone quasi-garbling

Kim (2021) proposes to compare two experiments via the notion of monotone quasi-garbling (MQG). The signals of the weaker experiment must be ordered. Then the weaker experiment is obtained from a state-dependent garbling of the better experiment, where the garbling is state-monotone: For any fixed signal under the better experiment, the higher is the state, the first-order stochastically lower is the garbling's distribution (over signals in the weak experiment). This notion nicely generalizes both Blackwell's state-independent garbling, and Lehmann's deterministic, state-dependent map $G^{-1}(F(t|\theta)|\theta)$ between signals.

We may compare our results to those obtained through MQG in our setting with three states, and focusing on the case where f and g are binary experiments. Fix an arbitrary prior. The definition of MQG requires that the two possible posterior realizations of g are ordered, say $r \succ s$. Let S denote the Blackwell segment that connects the two posteriors p, q under f. It can then be shown that g is an MQG of f if and only if, the high posterior r is LR-dominated by some posterior in S and the low posterior s LR-dominates some posterior in S.

Our Proposition 1 then shows that if g is less informative than f then g can be obtained from f by MQG. In one case, f satisfies the MLR, and then $p \succ_{LR} r$ and $s \succ_{LR} q$. In another case, f violates the MLR, and then r, s are members of S.

However, the opposite implication fails. In case f satisfies the MLR, MQG allows for the possibility that $q \succ_{LR} r$ and/or $s \succ_{LR} p$, thereby placing support of g outside the LR-box. In case f violates the MLR, MQG permits the support of g to span a far larger set than Blackwell segment S, as the high posterior r can be anywhere LR-below S and the low posterior can be anywhere LR-above S.

Kim's main characterization result shows that if g is an MQG of f then f provides higher ex-ante utility than g in all decision problems which satisfy two conditions. For our binary experiments, the first condition is equivalent to IDO. The second condition restricts to decision problems which permit a monotone optimal map from signal to action under g. Our finding that the MQG notion allows f to dominate more experiments g follows from this restriction in the set of permissible decision problems.[11]

[11]Observe, however, that if we consider binary g with both possible orders of the signals, and require MQG of f for both orders, then we recover the set of g which are less informative than g according to Proposition 1.

Kim's approach allows for a non-transitive comparison of experiments. Take experiments f, g, h. Let D_g, D_h denote the decision problems with the required monotonicity property for g, h, respectively. Now, f is better than g if it delivers greater utility for all problems in D_g. Likewise, g is better than h if it delivers greater utility for all problems in D_h. But these two properties do not imply that f is better than h on domain D_h, unless $D_h \subseteq D_g$. The problem is real when considering MQG, for one can indeed have h as an MQG of g which is an MQG of f, and yet, h is not an MQG of f.[12]

A Proofs

A.1 Binary Signal

In this section we prove that Proposition 1 holds in the case where the weaker experiment g is binary. We proceed in steps.

Step 1. It suffices to consider binary decision problems where the lower action has $u(a_1, \theta) = 0$ for all $\theta \in \Theta$. In a monotone, non-trivial decision problem, then the higher action has $u(a_2, \theta_1) < 0 < u(a_2, \theta_3)$ while $u(a_2, \theta_2) \in \mathbb{R}$ is free.

Clearly, such decision problems must be considered. In the opposite direction, first when one action is weakly dominant, all experiments provide the same expected utility, and so they do not influence the comparison of experiments. Second, for any decision problem we can narrow down the action set to the one or two actions optimal under experiment g: if experiment f performs better in this restricted-action problem, it can only perform even better when the action set is again expanded. Third, we can normalize the utility of a_1 as claimed: starting from a more general utility function, for every $\theta \in \Theta$ and every $c \in A$, subtract $u(a_1, \theta)$ from $u(c, \theta)$. At any posterior belief r, the optimal action is the same. Ex ante, expected utility is reduced by $E[u(a_1, \theta)]$ under the prior, which is the same quantity in every experiment. The comparison of two experiments is therefore unaffected by this normalization.

Step 2. If experiment f dominates experiment g, there exists no decision problem where the posterior interval $p - q$ of f is strictly separated from a posterior point of g by the indifference line.

If such separation occurs, in this decision problem a constant action is optimal under f. Prior

[12] At prior $(.35, .3, .35)$, let the posteriors of f have support $p = (.3, .3, .4)$, $q = (.4, .3, .3)$ so $p \succ_{LR} q$. Let g have support $s = (.1, .3, .6)$, $r = (.6, .3, .1)$ and order $r \succ s$. Since $s \succ_{LR} p$ and $q \succ_{LR} r$, by our analysis, experiment g is an MQG of f. Concretely, one MQG is $\Pr(r|p, \theta_1) = 1 > \Pr(r|p, \theta_2) = 1/2 > \Pr(r|p, \theta_3) = 1/4$ and $\Pr(r|q, \theta_1) = 3/4 > \Pr(r|q, \theta_2) = 1/2 > \Pr(r|q, \theta_3) = 0$. Let h have support $(.2, .3, .5) \succ (.5, .3, .2)$. Since $(.5, .3, .2) \succ_{LR} r$ and $s \succ_{LR} (.2, .3, .5)$, h is an MQG of g. But $(.2, .3, .5) \succ_{LR} p$ (and $q \succ_{LR} (.5, .3, .2)$), so h is not an MQG of f. In this example, f is a Blackwell garbling of h which is a Blackwell garbling of g, but unlike Blackwell's property, MQG is robust to small perturbations in the supports of the three experiments.

π is an average of p, q, so experiment g must have posteriors strictly separated by the indifference line. Then the constant action is suboptimal under g, which therefore provides greater maximal utility than f.

Step 3. When experiment f does not satisfy the MLR property, experiment g is dominated precisely when it is a Blackwell garbling of f.

It follows from Blackwell's theorem that garbling suffices for dominance. Conversely, suppose it is not a garbling. As is well known (for binary experiments), for any full-support prior π some posterior of g is outside the posterior interval $p-q$ for f. Since p and q are not LR ordered, the line through p, q must intersect the base of the triangle. [Graphically, holding fixed p in panel B of Figure 1, q must be above the upward-sloping line through p or below the downward-sloping line through p; in either case, the line through p and q intersects the horizontal axis strictly between $(0,0)$ and $(1,0)$.] If some posterior r of experiment g is outside this line, it can be separated from the p, q line by a parallel line; by steps 1 and 2, experiment g would not be dominated. If posterior r lies on the line but outside the interval, there similarly exists a line through the base of the triangle that strictly separates r from interval $p-q$.

Step 4. For the remainder, fix prior π and suppose that $p \succ_{LR} q$. (By implication $p_1 < q_1$ and $p_3 > q_3$). If experiment g is dominated, it must have all its posteriors inside the LR box.

If experiment g has a posterior r strictly outside the LR box, then r can be separated from the LR box (and thus from p, q) by a line that intersects the base. Start by turning a diagonal defining the LR box until it strictly separates the LR box from r. Then make a slight parallel translation of this turned line, so it still separates, but intersects the base strictly between $(0,0)$ and $(1,0)$. By steps 1 and 2, experiment g is not dominated.

Step 5. It suffices to consider experiments g where both posteriors r, s are on the boundary of the LR box.

Due to the possibility of Blackwell garbling, it suffices to dominate the mentioned experiments. Other experiments in the LR box are garbled, and therefore cannot obtain higher maximal utility.

Step 6. Posterior pair (p, q) provides greater maximal utility than (r, s) if it further spreads the ex post utility from action a_2: Formally, let $u_2 = (u(a_2, \theta_1), u(a_2, \theta_2), u(a_2, \theta_3))$, and assume that $p \cdot u_2 \geqslant s \cdot u_2 \geqslant r \cdot u_2 \geqslant q \cdot u_2$ or that $p \cdot u_2 \geqslant r \cdot u_2 \geqslant s \cdot u_2 \geqslant q \cdot u_2$.

In both experiments, the average ex post utility from a_2 is $\pi \cdot u_2$, so the assumption provides a mean-preserving spread of ex post utility. The DM's posterior utility at arbitrary posterior belief t is the convex function $\max\{0, t \cdot u_2\}$. Hence, (p, q) is better.

Step 7. It suffices to consider decision problems where the iso-$t \cdot u_2$ line through $(0,0)$ intersects the LR box. That is, the slope satisfies $q_2/q_1 < (u(a_2, \theta_3) - u(a_2, \theta_1)) / (u(a_2, \theta_2) - u(a_2, \theta_3)) < p_2/p_1$.

First, the problems from step 1 with flatter upward sloping iso-utility lines are irrelevant, for then the indifference line cannot both cross the base (step 3) and separate r,s. Second, if the slopes are steeper, then moving parallel indifference lines along the base, p and q will be extremal points in the LR box, giving the situation from step 6. It suffices to consider upward-sloping indifference curves; the case with downward-sloping indifference curves is similar.

Step 8. It suffices to consider decision problems such that action a_2 is optimal at s (which is on the LR box boundary above the line from p to q). Otherwise, since r is below the line from p to q, we necessarily get the property $p \cdot u_2 \geqslant r \cdot u_2 \geqslant s \cdot u_2 \geqslant q \cdot u_2$ from step 6, and already know that experiment p,q is better. We conclude the proof by covering four possible positions of pair (r,s) on the boundary of the LR box.

Case 1. Posterior s is on the ray from $(0,0)$ through p, and r is on the ray from $(0,0)$ through q. This implies

$$\frac{p_2}{p_1} = \frac{s_2}{s_1} > \frac{r_2}{r_1} = \frac{q_2}{q_1}.$$

For the experiment r,s to be different from p,q, assume that $r_1, s_1 \in (p_1, q_1)$. Note that s sits above the ray from $(1,0)$ through p, that is, $p_3/p_2 > s_3/s_2$. Graphically, it is clear that the pair r,s sits outside the line connecting p with q. In experiment p,q let $\alpha \in (0,1)$ denote the probability of p. In experiment r,s, let $\beta \in (0,1)$ denote the probability of s.

The expected utility under experiment p,q is

$$\begin{aligned}
\alpha\left[p_1 u(a_2,\theta_1)+p_2 u(a_2,\theta_2)+p_3 u(a_2,\theta_3)\right] &= \alpha p_2\left[\frac{p_1}{p_2}u(a_2,\theta_1)+u(a_2,\theta_2)+\frac{p_3}{p_2}u(a_2,\theta_3)\right] \\
&> \alpha p_2\left[\frac{s_1}{s_2}u(a_2,\theta_1)+u(a_2,\theta_2)+\frac{s_3}{s_2}u(a_2,\theta_3)\right] \\
&= \frac{\alpha p_2}{\beta s_2}\beta\left[s_1 u(a_2,\theta_1)+s_2 u(a_2,\theta_2)+s_3 u(a_2,\theta_3)\right]
\end{aligned}$$

The inequality uses the comparisons of p and s mentioned before, and that $u(a_2,\theta_3) > 0$ and $\alpha > 0$. In the last line, the factor $(\alpha p_2)/(\beta s_2)$ multiplies the expected utility under experiment r,s. This expected utility must be positive if experiment r,s is a candidate to do better in this decision problem.

We prove that the factor is at least one, thus concluding that p,q yields a higher expected utility. The prior belief satisfies $\pi = \alpha p + (1-\alpha) q = \beta s + (1-\beta) r$, being the unique intersection of the p,q line with the r,s line. Thus, α, β solve

$$\begin{bmatrix} q_1 - p_1 & s_1 - r_1 \\ q_2 - p_2 & s_2 - r_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} q_1 - r_1 \\ q_2 - r_2 \end{bmatrix},$$

so that

$$\begin{aligned}\frac{\alpha}{\beta} &= \frac{(s_2-r_2)(q_1-r_1)+(r_1-s_1)(q_2-r_2)}{(p_2-q_2)(q_1-r_1)+(q_1-p_1)(q_2-r_2)} \\ &= \frac{s_2(q_1-r_1)-s_1(q_2-r_2)}{p_2(q_1-r_1)-p_1(q_2-r_2)} = \frac{s_2}{p_2}\end{aligned}$$

where the second equality uses $r_2(q_1-r_1) = r_1(q_2-r_2)$ and $q_2(q_1-r_1) = q_1(q_2-r_2)$, which hold because r lies on the ray from $(0,0)$ through q, i.e., $r_2/r_1 = q_2/q_1$, and the third equality follows because s lies on the ray from $(0,0)$ through p, i.e., $s_2/s_1 = p_2/p_1$.

Case 2. Suppose s remains on the ray from $(0,0)$ through p, but now let r sit on the ray from $(1,0)$ through p. It remains true that $p_1 < r_1 < q_1$. We rely on the same bound on utility as in case 1, but we need to extend the inequality derived from information. In the explicit solution for α/β, observe that both the numerator and the denominator are positive. Indeed, the denominator satisfies $\delta = (p_2-q_2)(q_1-r_1)+(q_1-p_1)(q_2-r_2) > 0$ iff

$$\frac{q_2-p_2}{q_1-p_1} < \frac{q_2-r_2}{q_1-r_1}.$$

This inequality holds: because r is located below the line from p to q, the line from r to q is steeper. Since α and β are positive, is follows that also the numerator satisfies $\nu = (s_2-r_2)(q_1-r_1)+(r_1-s_1)(q_2-r_2) > 0$. Suppose we add $\varepsilon = r_2(q_1-r_1)-r_1(q_2-r_2) = q_2(q_1-r_1)-q_1(q_2-r_2)$ to both denominator and numerator. Observe that the location of r above the ray from $(0,0)$ through q implies that $\varepsilon = r_2q_1-r_1q_2 > 0$. Then we see that $\alpha/\beta = \nu/\delta > (\nu+\varepsilon)/(\delta+\varepsilon)$ iff $\nu > \delta$. This is true, for

$$\nu-\delta = (s_2-p_2)(q_1-r_1)-(s_1-p_1)(q_2-r_2) > 0$$

iff

$$\frac{s_2-p_2}{s_1-p_1} > \frac{q_2-r_2}{q_1-r_1},$$

which is true: since r is located above the ray from $(0,0)$ to q, the line from r to q has lower slope than the line from p to s. Summing up, we have verified that

$$\frac{\alpha}{\beta} > \frac{s_2(q_1-r_1)-s_1(q_2-r_2)}{p_2(q_1-r_1)-p_1(q_2-r_2)} = \frac{s_2}{p_2}$$

where the last equation still follows from $s_2/s_1 = p_2/p_1$.

Case 3. Suppose s is now on the ray from $(1,0)$ through q, and go back to r located on the ray from $(0,0)$ through q. We can use the new property $s_2/s_1 < p_2/p_1$ to modify an equation from case 1 to the inequality

$$\frac{s_2(q_1-r_1)-s_1(q_2-r_2)}{p_2(q_1-r_1)-p_1(q_2-r_2)} > \frac{s_2}{p_2}.$$

This holds since $q_1 > r_1$ and $q_2 > r_2$ when r is located on the ray from $(0,0)$ through q. It might be noted that $p_2(q_1 - r_1) > p_1(q_2 - r_2)$ since the line from r to q has slope q_2/q_1.

Case 4. Finally, suppose s is on the ray from $(1,0)$ through q and r is on the ray from $(1,0)$ through p. We start by showing that there exists another binary experiment r', s' which does better than r, s, and which satisfies one of the above three cases. Consider $s'(a) = s + a(p-q)$ where $a \in [0, \hat{a}]$ as well as $r'(b) = r + b(q-p)$ where $b \in [0, \hat{b}]$. The critical numbers $\hat{a}$ and $\hat{b}$ are by definition those where the next boundary of the LR box is hit. Impose that the prior is unchanged for constant chance β, i.e., $\pi = \beta s'(a) + (1-\beta) r'(b)$ which is satisfied when $\beta a = (1-\beta) b$.

If $\beta\hat{a} < (1-\beta)\hat{b}$, we move point s to s' that is located on the ray from $(0,0)$ through p, while the corresponding $r'(\beta\hat{a}/(1-\beta))$ is interior to the LR box. By assumption, moving in direction $p-q$ implies moving to higher iso-utility lines. So, the expected utility from experiment $r'(\beta\hat{a}/(1-\beta))$ with s' is strictly higher than from experiment r, s. Finally, we can do even better by extending the line from s' through $r'(\beta\hat{a}/(1-\beta))$ until it hits the boundary of the LR box: let this intersection be our r'. We have improved r, s to an experiment r', s' where s' lives on the ray from $(0,0)$ through p, and hence we are in case 1 or case 2. But p, q is then better than r', s'.

If $\beta\hat{a} \geqslant (1-\beta)\hat{b}$, we likewise move until r' lives on the ray from $(0,0)$ through q. This will improve as above, and we end up in case 2 or case 3.

The proof of the claim is complete.

A.2 Three (and More) Actions and Signals

The main purpose of this section is to dispense with the restriction to two actions. We consider any IDO problem with a finite action set. As mentioned before, it is useful to decompose problems. This involves problems with fewer actions. Hence we start with the following result on the preservation of IDO.

Lemma 1. *If u is IDO on $A \times \{\theta_1, \theta_2, \theta_3\}$, and no action is dominated, then for any $B \subset A$, u is IDO on $B \times \{\theta_1, \theta_2, \theta_3\}$.*

Proof of Lemma 1. First, suppose some action $b \in B$ is optimal on some non-trivial interval $[b', b] \subset B$ at θ_1. If b fails interval-optimality (in B) at θ_2, IDO on A implies existence of $a \in \arg\max_{[b',b]} u(a, \theta_1)$ with $b' < a < b$ where $u(a, \theta_1) > u(b, \theta_1) \geqslant u(b', \theta_1)$. By definition, a is interval-optimal (in A) on $[b', a]$ at θ_1. By IDO on A, for all $\theta \in \Theta$, $u(a, \theta) > u(b', \theta)$, a contradiction to the assumption that no action is dominated. Second, suppose some $b \in B$ is optimal on $[b', b] \subset B$ at θ_2, but fails interval-optimality (in B) at θ_3. So there exists $b'' \in B$ with $b' \leqslant b'' < b$ where $u(b'', \theta_3) > u(b, \theta_3)$. Note that $b \in B$ is optimal on $[b'', b] \subset B$ at θ_2 but not at θ_3. IDO on A implies existence of $a \in \arg\max_{[b'',b]} u(a, \theta_2)$ with $b'' < a < b$ where $u(a, \theta_2) > u(b, \theta_2)$. By definition, a is interval-optimal (in A) on $[b'', a]$ at θ_2 and therefore also at θ_3. Since b, b'' are not dominated by a, we must have $u(b,\theta_1) > u(a,\theta_1)$ and $u(b'',\theta_1) > u(a,\theta_1)$. Then b cannot be optimal (in A) on $[a,b]$ at θ_1, implying the existence of $a' \in \arg\max_{[a,b]} u(a,\theta_1)$ with $a < a' < b$ where $u(a',\theta_1) > u(b,\theta_1)$. By construction, a' is optimal with $u(a',\theta_1) > u(a,\theta_1)$ on $[a,a']$ at θ_1 and hence also $u(a',\theta_2) > u(a,\theta_2)$. But this finally contradicts that $a \in \arg\max_{[b'',b]} u(a,\theta_2)$.

The remainder of this section proves Proposition 1 when experiment g has more than two signals. We thoroughly examine our main idea in the case with three signals. The final part of the section explains the extension to more signals.

Consider the new situation where, given some prior π, the experiment inside the LR box has its support on the three beliefs r,s,t, such that action a_1 is taken at t, action a_2 is taken at s, and action a_3 is taken at r. These actions are ordered $a_3 \succ a_2 \succ a_1$. Denote $u_r = r_1 u(a_3,\theta_1) + r_2 u(a_3,\theta_2) + r_3 u(a_3,\theta_3)$ and so on as the expected utilities from the optimal action at each posterior. The expected utility under this experiment is $V = \Pr(r)\,u_r + \Pr(s)\,u_s + \Pr(t)\,u_t$ and the prior satisfies $\pi = \Pr(r)\,r + \Pr(s)\,s + \Pr(t)\,t$.

With experiment (p,q), we take action 1 at q and action 3 at p for expected utility $U = \Pr(p)\,u_p + \Pr(q)\,u_q$, and $\pi = \Pr(p)\,p + \Pr(q)\,q$. By IDO, this must be optimal under f restricted to these three actions when $p \succ_{LR} r$ and $t \succ_{LR} q$.

Observe that at least two of the support points for r,s,t must be weakly on the same side of the p,q line. Suppose s sits alone on one side, strictly outside the line; this detail does not matter. As in Figure 2.

Lemma 2 (Decomposition)**.** *There exists $\lambda \in [0,1]$ such that $\pi^1 = [\Pr(r)\,r + \lambda \Pr(s)\,s]/[\Pr(r) + \lambda \Pr(s)]$ lies on the p,q line. Then also $\pi^2 = [\Pr(t)\,t + (1-\lambda)\Pr(s)\,s]\,/\,[\Pr(t) + (1-\lambda)\Pr(s)]$ lies on the p,q line.*

Proof of Lemma 2. With $\lambda = 0$, the point is on the r side of the line, with $\lambda = 1$ it will be on the s side because a convex combination with t is π on the line. For the second part, observe that $\pi = [\Pr(r) + \lambda \Pr(s)]\pi^1 + [\Pr(t) + (1-\lambda)\Pr(s)]\pi^2$, so π^2 must lie on the line that connects π and π^1.

Continuing the proof of Proposition 1, let $P^1 = \Pr(r) + \lambda \Pr(s) \in (0,1)$ and $P^2 = 1 - P^1$. Define $\mu^i \in (0,1)$ such that $\pi^i = \mu^i p + \left(1-\mu^i\right) q$. Note that $\Pr(p) = \mu^1 P^1 + \mu^2 P^2$ since $\pi = \Pr(p)\,p + \Pr(q)\,q = P^1\pi^1 + P^2\pi^2$. We are going to show that (i) $U = P^1U^1 + P^2U^2$ where U^i is the utility from the p,q experiment with prior π^i, (ii) $V = P^1V^1 + P^2V^2$ where V^1 is the utility from the r,s experiment with prior π^1 and V^2 is the utility from the s,t experiment with prior π^2, and (iii) $U^i \geqslant V^i$. Once we have (i)–(iii), $U \geqslant V$ follows, as desired to complete the proof of the proposition in the three-signal case.

Proof of (i): The p,q experiment with prior π^i has expected utility $U^i = \mu^i u_p + \left(1-\mu^i\right) u_q$. The result then follows from $\Pr(p) = \mu^1 P^1 + \mu^2 P^2$.

Proof of (ii): The r,s experiment at prior π^1 has expected utility $V^1 = [\Pr(r)\,u_r + \lambda \Pr(s)\,u_s]/P^1$. Likewise, $V^2 = [\Pr(t)\,u_t + (1-\lambda)\Pr(s)\,u_s]/P^2$. The result follows from the definition of V.

Proof of (iii): The binary r,s experiment with prior π^1 delivers utility V^1. By the established Proposition 1 for binary experiment g, the p,q experiment with prior π^1 provides utility no less than V^1, i.e., $U^1 \geqslant V^1$. Similarly, $U^2 \geqslant V^2$.

Finally, consider any finite number of actions. For any posterior r^i already in the p,q line, assign probability P^i to that, and compare to the p,q experiment with prior r^i. With that done, start from an arbitrary posterior outside the p,q line. Find an arbitrary posterior on the opposite side; draw the straight line through the p,q line and there is a binary experiment with prior at the intersection. Assign probability P^j as high as possible to this binary experiment in order to exhaust the probability of one of its support points. Then continue the process from the other end of the line, with its remaining chance. In a finite number of steps, this procedure decomposes the experiment inside the LR box into a distribution over conditional binary experiments. The p,q experiment can be similarly decomposed to have the same conditional priors, and points (i)–(iii) from before directly extend to prove the dominance of the p,q experiment. The proof of Proposition 1 is complete.

References

Agrawal, A., J. Gans, and A. Goldfarb (2018): *Prediction Machines*, Harvard Business Review Press.

Athey, S. and J. Levin (2018): "The value of information in monotone decision problems," *Research in Economics*, 72(1), 101–116.

Blackwell, D. (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 93–102.

——— (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 24(2), 265–272.

Di Tillio, A., M. Ottaviani, and P. N. Sørensen (2021): "Strategic Sample Selection," *Econometrica*, 89(2), 911–953.

Jewitt, I. (2007): "Information Order in Decision and Agency Problems," Oxford mimeo.

Karlin, S. and H. Rubin (1956): "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *Annals of Mathematical Statistics*, 27(2), 272–299.

Kim, Y. (2021): "Comparing Information in General Monotone Decision Problems," Duke University.

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

Milgrom, P. and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62(1), 157–180.

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

Quah, J. K.-H. and B. Strulovici (2009): "Comparative Statics, Informativeness, and the Interval Dominance Order," *Econometrica*, 77(6), 1949–1992.