

On the Monotonicity of Information Costs*

Xiaoyu Cheng[†]

Yonggyun Kim[‡]

August 11, 2025

Abstract

We study the monotonicity of information costs: more informative experiments must be more costly. As criteria for informativeness, we consider the standard information orders introduced by [Blackwell \(1951, 1953\)](#) and [Lehmann \(1988\)](#). We provide simple necessary and sufficient conditions for a cost function to be monotone with respect to each order, grounded in their garbling characterizations. Finally, we examine several well-known cost functions from the literature through the lens of these conditions.

JEL Classification: D81, D82, D83

Keywords: Statistical Experiments, Rational Inattention, Blackwell order, Lehmann order

*We thank Arjada Bardhi, Tommaso Denti, Teddy Kim, R. Vijay Krishna, Fei Li, Jeffrey Mensch, Luciano Pomatto, Todd Sarver, Bruno Strulovici, Curtis Taylor, João Thereze, Udayan Vaidya, Kun Zhang, Mu Zhang, and seminar participants at KAEA-VSS Micro Seminar, FSU, UNC, Duke, Mannheim, ITAM, and SEA 2024 for insightful comments.

[†]Department of Economics, Florida State University, Tallahassee, FL, USA. E-mail: xcheng@fsu.edu.

[‡]Department of Economics, Florida State University, Tallahassee, FL, USA. E-mail: ykim22@fsu.edu.

1 Introduction

Recent developments in economic theory have expanded the scope of decision-making by modeling information as a costly choice variable. This approach puts information on par with other key decision variables in economic models: just as consumers select consumption bundles and producers choose input combinations, agents actively decide which information to acquire, taking into account the cost of doing so. This framework underlies a growing body of work across economics, including macroeconomics (Sims, 2003), finance (Van Nieuwerburgh and Veldkamp, 2010), and games (Matějka and McKay, 2012; Yang, 2015; Ravid, 2020).¹

When modeling the cost of information, the most fundamental requirement is *monotonicity*, that is, more informative options should entail higher costs. Unlike consumption bundles or input combinations, for which the notion of “more” is straightforward, informativeness is more subtle and often defined indirectly. One natural approach is to assess informativeness by how much the information improves decision-making. Two classical information orders—introduced by Blackwell (1951, 1953) and Lehmann (1988)—formalize this idea and are widely recognized as standard criteria for comparing the informativeness of statistical experiments. Specifically, an experiment is more informative in the Blackwell sense if it yields higher expected payoffs in all decision problems, whereas it is more informative in the Lehmann sense if it does so across various classes of monotone decision problems—such as those with preferences satisfying the single-crossing property (Milgrom and Shannon, 1994) or the interval dominance order property (Quah and Strulovici, 2009).

This paper develops a unified framework for analyzing the monotonicity of information costs with respect to both the Blackwell and Lehmann orders. We derive necessary and sufficient conditions for a cost function to satisfy Blackwell and Lehmann monotonicity, respectively. These conditions take the form of simple first-order inequalities (in the space of experiments), which facilitate both the verification of monotonicity for a given cost function and the construction of new ones with desired properties. Our characterization of Blackwell monotonicity complements existing results and offers a new geometric perspective on the Blackwell order. For Lehmann monotonicity, we provide what is, to our knowledge, the first general characterization in the literature. This result fills an important gap in the literature, as previous analyses have largely focused on Blackwell monotonicity.

Our unified approach builds on the garbling representations of the two information orders. For the Blackwell order, it is well known that one experiment is more informative than another if the latter can be obtained by a garbling of the former, that is, by introducing additional noise. The simplest such garbling is the one that replaces one signal with another with some probability,

¹See Maćkowiak et al. (2023) for a recent and comprehensive survey of the literature.

while leaving other signals unchanged. Since this operation reduces informativeness, Blackwell monotonicity requires that the cost function decreases in the direction of such signal replacements. Taking the replacement probability to zero yields a natural first-order condition, which we call *decreasing in signal replacement*. Our characterization results (Theorems 1 and 3) show that, this local condition is sufficient for global Blackwell monotonicity, under additional natural conditions regulating the cost of equally informative experiments.²

The generality of the Blackwell order, while being its greatest strength, is also its primary limitation. By requiring an experiment to be superior to another across all decision problems, the criterion is often too demanding, rendering many pairs of experiments incomparable. This has led researchers to seek finer information orders by restricting the class of decision problems under consideration. A leading example of such a refinement is the Lehmann order, which focuses on the broad and economically significant class of monotone decision problems. This refinement makes it possible to rank experiments that are otherwise incomparable under the Blackwell order, providing a more discerning tool for this important set of economic applications.

Following the same logic as for the Blackwell order, our analysis of Lehmann monotonicity builds on its characterization in terms of garbling. This representation has been established recently by Kim (2023). It was shown that, for experiments satisfying the monotone likelihood ratio property (MLRP), one experiment is Lehmann more informative than another if the latter can be obtained from the former through a generalized form of garbling—one that introduces noise that is reversely monotone with respect to the underlying states. This generalized garbling gives rise to a family of *reverse signal replacement* operations, which replace a low (high) signal with a high (low) signal with some probability, but only in the states below (above) a certain threshold. Taking the replacement probability to zero again yields a first-order condition, which we call *decreasing in reverse signal replacement*. Our characterization results (Theorems 2 and 4) show that this condition plays the same central role for Lehmann monotonicity.

Our primary technical contribution is a method for constructing informativeness-reducing paths between two comparable experiments. The key to proving that our local monotonicity conditions imply global monotonicity lies in this construction, which allows us to decompose arbitrary information loss in the sense of the Blackwell and Lehmann orders into continuous paths of local signal replacements and local reverse signal replacements, respectively. The main challenge, particularly for the Lehmann order, is the requirement that all intermediate experiments along these paths must preserve the MLRP. This is a significant hurdle because the set of experiments satisfying the MLRP is known to be non-convex, making it difficult to apply standard analytical tools that often rely on

²Specifically, *permutation invariance* requires the costs remain unchanged when permuting the labels of the signals, while *splitting invariance* requires that splitting a signal into two copies does not change the cost. Both operations do not change the informativeness of an experiment. The results also require the cost function to be absolutely continuous, and in some cases, differentiable.

convexity.

Our path-construction method overcomes this challenge by ensuring the constructed paths lie entirely within the MLRP set, thus circumventing the non-convexity problem. We achieve this by leveraging novel geometric insights into the structure of the garbling representations for both the Blackwell and Lehmann orders. Specifically, our unified analysis draws on the zonotope representation for the Blackwell order (Bertschinger and Rauh, 2014) and the probability-probability (PP) plot representation for the Lehmann order (Jewitt, 2007).

To clarify these ideas, we present our analysis in two stages. We first examine the special case of binary-signal experiments (Section 2), where the simpler structure provides intuitions on how signal replacements serve as building blocks and how the MLRP can be preserved. We then extend this analysis to the more intricate case of arbitrary finite signal spaces (Section 3), demonstrating how the same core principles apply more generally.

Finally, to illustrate the usefulness of our characterizations, we apply them to several widely studied classes of information cost functions. For likelihood-separable and posterior-separable costs, where Blackwell monotonicity has been established (Caplin and Dean, 2015; Denti et al., 2022b), we derive conditions under which these costs also satisfy Lehmann monotonicity. This allows us to identify a broad subclass satisfying both notions of monotonicity, while also showing that Lehmann monotonicity does not hold in general. Fosgerau et al. (2020) introduce the class of Bregman information costs, which provide a rational inattention foundation for additive random utility models. These costs are not invariant to signal permutations and are thus known to violate Blackwell monotonicity globally. We show that they could also violate our local monotonicity conditions, and hence are neither Blackwell nor Lehmann monotone, even in a local sense.

Related Literature Our work contributes to the literature on information costs by unraveling the structure of **monotonicity**, a property considered to be a minimum requirement for any plausible information cost function.

This principle is typically formalized with respect to the information order by Blackwell (1951, 1953), and much of the prior literature on the topic has focused on ensuring costs are Blackwell monotone. A significant body of research, for instance, treats Blackwell monotonicity as a foundational axiom, combining it with other properties to characterize specific classes of cost functions (Mensch, 2018; Hébert and Woodford, 2021; Pomatto et al., 2023; Baker, 2023; Bloedel and Zhong, 2024). Another strand of the literature takes a decision-theoretic approach, aiming to derive a representation of utility and information costs from observable choice data. In identifying such costs, Blackwell monotonicity emerges as a necessary consequence of rational, cost-minimizing behavior as in, e.g., de Oliveira et al. (2017) and Chambers et al. (2020).³

³This revealed preference framework has been broadly applied to derive testable implications of stochastic choice

Unlike in these works, which use monotonicity as an ingredient to derive a full functional form, our analysis isolates the property itself. We ask what, on its own, Blackwell monotonicity requires of a cost function. The result is a necessary and sufficient condition in the form of a simple, local first-order inequality, which provides a more fundamental understanding of this core property.

The most closely related result to our characterization of Blackwell monotonicity is Claim 1 of [Ravid et al. \(2022\)](#), which likewise relies on directional derivatives of the cost function. A crucial distinction, however, arises from the domain of the cost function. Our framework defines cost directly on the space of experiments, yielding a prior-free characterization. In contrast, their analysis defines the cost function directly on the distribution of posterior means generated by an experiment. This approach is therefore inherently prior-dependent and implicitly treats any information not captured in the first moment of the posterior beliefs as costless. Due to this fundamental difference, the two characterizations are not directly translatable.

Subsequently, we extend our analysis to Lehmann monotonicity, a stronger notion of monotonicity that, to our knowledge, has not been systematically explored. The information order by [Lehmann \(1988\)](#) is the relevant criterion for the large and important class of monotone decision problems commonly found in economics ([Quah and Strulovici, 2009](#); [Chi, 2015](#)).⁴ By characterizing necessary and sufficient conditions for Lehmann monotonicity, we provide a method of examining whether an existing cost function satisfies Lehmann monotonicity or not. For instance, while the literature typically assumes that posterior- or likelihood-separable costs have “concave” or “sublinear” primitive functions to ensure Blackwell monotonicity ([Denti et al., 2022b](#)), these conditions alone do not guarantee Lehmann monotonicity. Our analysis fills this gap by providing a new set of conditions under which these costs also satisfy the more demanding Lehmann monotonicity (Proposition 2, 4).

2 Binary Experiments

We begin our analysis by focusing on the special case of *binary experiments*—those consisting of only two signals—to illustrate the key ideas of our characterization in a simpler setting. In this section, we provide necessary and sufficient conditions for information cost functions defined over binary experiments to satisfy Blackwell monotonicity and Lehmann monotonicity, respectively. Although this focus is restrictive, the results are not merely illustrative: they also offer practical

models and to characterize the behavioral foundations of various cost structures (e.g., [Caplin and Dean, 2015](#); [Denti, 2022](#); [Lipnowski and Ravid, 2023](#)).

⁴The Lehmann order is particularly relevant for applications with naturally ordered structures, such as in auctions ([Persico, 2000](#); [Bobkova, 2024](#)), finance ([Bond, 2023](#)), and screening ([Kim, 2023](#); [Asseeyer, 2025](#)). It also extends to decisions under ambiguity ([Li and Zhou, 2020](#)).

value in applications where the underlying signals or actions are binary, a common benchmark in many economic models.

2.1 Blackwell Monotonicity

2.1.1 Preliminaries

Let $\Omega = \{\omega_1, \dots, \omega_n\}$ be a finite set of states and $\mathcal{S} = \{s_L, s_H\}$ be a binary set of signals. A binary (statistical) experiment $f : \Omega \rightarrow \Delta(\mathcal{S})$ can be represented by an $n \times 2$ matrix:

$$f = \begin{bmatrix} 1 - f_1 & f_1 \\ \vdots & \vdots \\ 1 - f_n & f_n \end{bmatrix},$$

where $f_i = f(s_H|\omega_i)$ is the probability of observing the high signal (s_H) in state ω_i . For convenience, in this section only, we write $f = [f_1, \dots, f_n]^\top \in \mathcal{E}_2 \equiv [0, 1]^n$, and refer to the associated matrix as $[1 - f, f]$ where $\mathbf{1} = [1, \dots, 1]^\top$.

Blackwell Information Order An experiment f is said to be *Blackwell more informative* than another experiment g , denoted by $f \succeq_B g$, if and only if there exists a stochastic matrix M (i.e., $M_{ij} \geq 0$ and $\sum_j M_{ij} = 1$ for all i) such that $[1 - g, g] = [1 - f, f] M$. Such a matrix M is referred to as a *garbling matrix*. We say that f and g are *equally informative*, denoted $f \simeq_B g$, if both $f \succeq_B g$ and $g \succeq_B f$ hold. When both f and g are binary experiments, any garbling matrix M must be a 2×2 stochastic matrix. Let \mathcal{M}_2 denote the set of all such matrices.

Information Costs An *information cost function* defined over binary experiments is denoted by $C : \mathcal{E}_2 \rightarrow \mathbb{R}_+$. We assume C to be differentiable for discussions in the main text, but we will specify if this assumption is required in the statements of our theorems. Furthermore, say that $C : \mathcal{E}_2 \rightarrow \mathbb{R}_+$ is *absolutely continuous* if for all $f, g \in \mathcal{E}_2$ and $t \in [0, 1]$, the function $\varphi(t) = C(f + t(g - f))$ is absolutely continuous in t over $[0, 1]$.⁵ Or equivalently, the Fundamental Theorem of Calculus (FTC) holds, i.e., $\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t) dt$.

Blackwell Monotonicity Say that C is **Blackwell monotone** if for all $f, g \in \mathcal{E}_2$, $C(f) \geq C(g)$ whenever $f \succeq_B g$. For a given cost function C , let $S_C(f) = \{g \in \mathcal{E}_2 : C(f) \geq C(g)\}$ denote its sublevel set at f , and let $S_B(f) = \{g \in \mathcal{E}_2 : f \succeq_B g\}$ denote the set of experiments less

⁵There are multiple generalizations of absolute continuity from \mathbb{R} to \mathbb{R}^n emphasizing different aspects, see [Dymond et al. \(2017\)](#). We adopt the generalization which requires the restriction of C to any line segment is absolutely continuous. A sufficient condition for such absolute continuity is Lipschitz continuity.

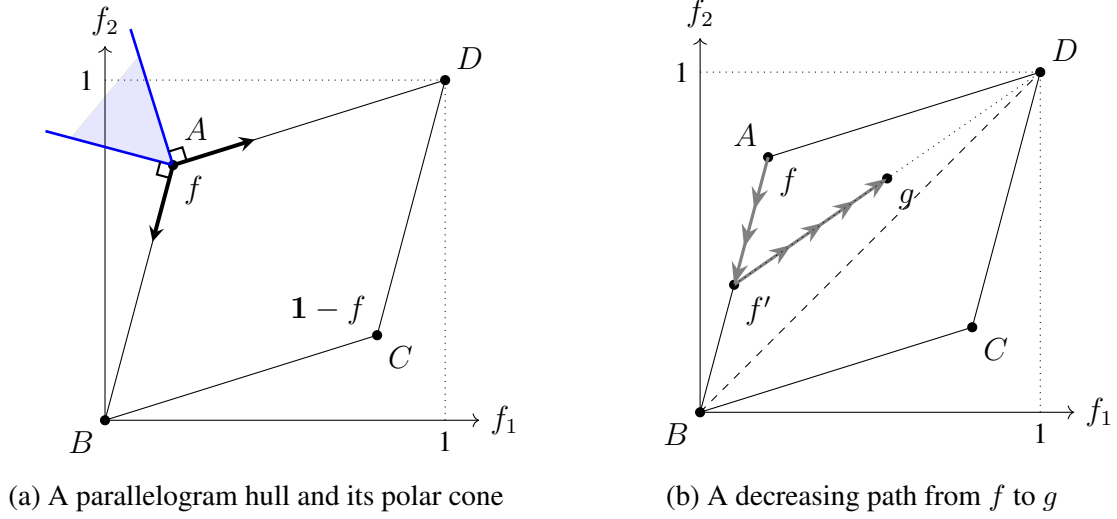


Figure 1: A Graphical Illustration with Binary States

informative than f in the Blackwell order. By definition, C is Blackwell monotone if and only if $S_C(f) \supseteq S_B(f)$ for all $f \in \mathcal{E}_2$.

2.1.2 Parallelogram Hull

To establish necessary and sufficient conditions for Blackwell monotonicity, we begin by characterizing the sublevel set of binary experiments, $S_B(f)$. For any $f, g \in \mathcal{E}_2$ with $f \succeq_B g$, there exists $M \in \mathcal{M}_2$ such that $[g, 1 - g] = [f, 1 - f]M$. Any stochastic matrix $M \in \mathcal{M}_2$ can be written, for some $(a, b) \in [0, 1]^2$, as

$$M = \begin{bmatrix} a & 1 - a \\ b & 1 - b \end{bmatrix},$$

which implies that $g = af + b(1 - f)$. This leads to the following characterization.

Lemma 1. *For any $f, g \in \mathcal{E}_2$, $f \succeq_B g$ if and only if g lies in the **parallelogram hull** of f and $1 - f$, defined as*

$$PARL(f, 1 - f) \equiv \{af + b(1 - f) : a, b \in [0, 1]\}.$$

*In other words, $S_B(f) = PARL(f, 1 - f)$.*⁶

When the state space is binary (i.e., $n = 2$), the parallelogram hull corresponds to the shaded region (ABCD) in Figure 1a. Specifically, for binary-state and binary-signal experiments, f is Blackwell more informative than g if and only if g lies within the parallelogram (ABCD).

⁶This geometric characterization of the Blackwell order coincides with the zonotope order in [Bertschinger and Rauh \(2014\)](#), which is further applied in [de Oliveira et al. \(2024\)](#)

2.1.3 The Characterization

Permutation Invariance Observe that $S_B(f) = S_B(\mathbf{1} - f) = \text{PARL}(f, \mathbf{1} - f)$, which implies that $f \succeq_B (\mathbf{1} - f) \succeq_B f$, or equivalently, $f \simeq_B (\mathbf{1} - f)$. Intuitively, $\mathbf{1} - f$ can be obtained by permuting the signals of f . This relabeling of signals should preserve the same information content. Therefore, any Blackwell monotone cost function C must satisfy $C(f) = C(\mathbf{1} - f)$. We refer to this property as *permutation invariance* and it serves as a necessary condition for Blackwell monotonicity.

Decreasing in Signal Replacement Consider the following two garbling matrices:

$$M_1 = \begin{bmatrix} 1 - \epsilon & \epsilon \\ 0 & 1 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 1 & 0 \\ \epsilon & 1 - \epsilon \end{bmatrix}.$$

The garbling induced by M_1 can be interpreted as replacing the low signal (s_L) with the high signal (s_H) with probability ϵ , while keeping s_H unchanged. Similarly, M_2 replaces s_H with s_L with probability ϵ , while keeping s_L unchanged. Applying these garbling to an experiment f should reduce its cost if C is Blackwell monotone:

$$C(f) \geq C(f + \epsilon(\mathbf{1} - f)), \quad C(f) \geq C(f - \epsilon f).$$

Notice these inequalities hold for all $\epsilon \in [0, 1]$. Then taking the limit as $\epsilon \rightarrow 0$, we obtain the following first-order conditions:

$$\langle \nabla C(f), \mathbf{1} - f \rangle = \sum_{i=1}^n \frac{\partial C}{\partial f_i} (1 - f_i) \leq 0, \tag{1}$$

$$\langle \nabla C(f), -f \rangle = \sum_{i=1}^n \frac{\partial C}{\partial f_i} (-f_i) \leq 0, \tag{2}$$

where ∇C denotes the gradient of C and $\langle \nabla C(f), h \rangle$ denotes the directional derivative of C at f in the direction of h . We say that the cost function C is *decreasing in signal replacement* if (1) and (2) hold for all $f \in \mathcal{E}_2$.⁷ Clearly, this property is also necessary for Blackwell monotonicity.

Geometrically, the vectors $-f$ and $\mathbf{1} - f$ corresponds to the two extreme directions of decreasing informativeness, \overrightarrow{AB} and \overrightarrow{AD} , respectively, in Figure 1a. Inequalities (1) and (2) together imply that $\nabla C(f)$ must lie in the polar cone of $\text{PARL}(f, \mathbf{1} - f)$, which is the blue shaded area in Figure 1a. In other words, Blackwell monotonicity imposes a constraint on the feasible directions of the gradient of C at each experiment.

⁷These derivatives are well-defined for all $f \in \mathcal{E}_2$ as $f + \epsilon(\mathbf{1} - f) \in \mathcal{E}_2$ and $f - \epsilon f \in \mathcal{E}_2$ for all $\epsilon \in [0, 1]$.

Our first main result shows that for binary experiments, these two properties are not only necessary but also sufficient for Blackwell monotonicity.

Theorem 1. *Suppose $C : \mathcal{E}_2 \rightarrow \mathbb{R}_+$ is absolutely continuous. Then, C is Blackwell monotone if and only if C is permutation invariant and decreasing in signal replacement.*

Proof Sketch for Sufficiency We use the binary-state case as in Figure 1 to illustrate the proof. Consider any experiment g lying inside the parallelogram $ABCD$, i.e., $f \succeq_B g$. If g is above the line BD as in Figure 1b, there exists a two-segment path from f to g , which moves only in the directions specified by (1) and (2): moving from f in the direction of $-f$ to reach f' and then moving from f' in the direction of $1 - f'$ to reach g . Thus, (1) and (2) imply that the directional derivatives are negative along this path, and applying the FTC leads to $C(g) \leq C(f)$.⁸

If g lies below the line BD , its permutation, $1 - g$, lies above the line BD and has the same cost as g , by permutation invariance. Then, the same argument applies to $1 - g$ implying $C(g) = C(1 - g) \leq C(f)$. \square

2.1.4 Further Characterizations with Binary States

When the state is binary, additional geometric insights can be drawn to provide further characterizations of Blackwell monotonicity. Given permutation invariance, it is without loss to focus on the following set of experiments:

$$\hat{\mathcal{E}}_2 \equiv \{(f_1, f_2) : 0 \leq f_1 \leq f_2 \leq 1\}.$$

For any $f, g \in \hat{\mathcal{E}}_2$, from the parallelogram in Figure 1b, notice that $f \succeq_B g$ if and only if the slope of AB for f is steeper than that for g , and the slope of AD for f is shallower than that for g . In other words, $f \succeq_B g$ if and only if

$$\frac{f_2}{f_1} \geq \frac{g_2}{g_1} \quad \text{and} \quad \frac{1 - f_1}{1 - f_2} \geq \frac{1 - g_1}{1 - g_2}. \quad (3)$$

Let $\alpha \equiv \frac{f_2}{f_1}$ and $\beta \equiv \frac{1 - f_1}{1 - f_2}$.⁹ The simple fact above yields the following characterization of Blackwell monotonicity which does not require any additional assumption on C .

Proposition 1. *$C : \hat{\mathcal{E}}_2 \rightarrow \mathbb{R}_+$ is Blackwell monotone if and only if $C\left(\frac{\beta - 1}{\alpha\beta - 1}, \frac{\alpha(\beta - 1)}{\alpha\beta - 1}\right)$ is increasing in α and β .*

⁸Since the proof concerns only the directional derivatives, differentiability of C is not required.

⁹Let $x/0 = +\infty$ for all $x > 0$ and $0/0 = 1$.

In other words, by a change of variables, Blackwell monotonicity in the binary-binary case is equivalent to the monotonicity of the function $\tilde{C}(\alpha, \beta) = C\left(\frac{\beta-1}{\alpha\beta-1}, \frac{\alpha(\beta-1)}{\alpha\beta-1}\right)$ in both variables. In Online Appendix OA.1, we provide examples of cost functions that can be verified easily using Proposition 1. It worth mentioning here that (3) also proves useful in establishing Lehmann monotonicity in the next section.

2.2 Lehmann Monotonicity

We now explore Lehmann monotonicity for binary experiments. [Lehmann \(1988\)](#) introduces an information order that refines the Blackwell order for experiments satisfying the monotone likelihood ratio property (MLRP), defined as:

$$f(s|\omega) \cdot f(s|\omega') \geq f(s|\omega') \cdot f(s'|\omega) \quad \text{for all } s' > s \text{ and } \omega' > \omega. \quad (4)$$

For binary experiments (with states and signals ordered in their index), the MLRP is equivalent to $f_1 \leq \dots \leq f_n$. Thus, we restrict our attention to the following set of experiments:

$$\mathcal{E}_2^{MLRP} \equiv \{f \in \mathcal{E}_2 : 0 \leq f_1 \leq \dots \leq f_n \leq 1\}.$$

Notice that while the MLRP is generally not preserved under convex combinations, \mathcal{E}_2^{MLRP} is indeed convex.

Lehmann Information Order Lehmann's order is originally defined on continuous signal spaces. This is without loss of generality, as he shows how to construct a continuous distribution from any distribution with discontinuous jumps. In our binary setup, for any $f \in \mathcal{E}_2$, we can associate a continuous experiment $\tilde{f} : \Omega \rightarrow \Delta([0, 2])$, where the probability distribution function given ω_i is $1 - f_i$ for $s \in [0, 1)$ and f_i for $s \in [1, 2]$. Formally, the cumulative distribution functions are given by:

$$\tilde{F}(x|\omega_i) = \begin{cases} (1 - f_i)x, & \text{if } 0 \leq x \leq 1, \\ (1 - f_i) + f_i(x - 1), & \text{if } 1 < x \leq 2. \end{cases} \quad (5)$$

For any $f, g \in \mathcal{E}_2^{MLRP}$, f is *Lehmann more informative* than g , denoted $f \succeq_L g$, if and only if for all $y \in [0, 2]$, $\tilde{F}^{-1}(\tilde{G}(y|\omega)|\omega)$ is increasing in ω . Under the MLRP, the Lehmann order is a strict refinement of the Blackwell order: $f \succeq_B g$ implies $f \succeq_L g$, but the reverse is not true.¹⁰

¹⁰For example, [Kim \(2023\)](#) (Appendix A.2) provides a pair of binary experiments that satisfy the MLRP and are comparable in the Lehmann order, but not in the Blackwell order. Moreover, it is also shown that when the MLRP is violated, $f \succeq_B g$ does not necessarily imply $f \succeq_L g$, i.e., the MLRP is a crucial assumption for this refinement to hold.

The following lemma provides a tractable characterization of the Lehmann order for binary experiments.

Lemma 2. *For any $f, g \in \mathcal{E}_2^{MLRP}$, $f \succeq_L g$ if and only if*

$$\frac{g_i}{f_i} \geq \frac{g_{i+1}}{f_{i+1}} \quad \text{and} \quad \frac{1 - g_{i+1}}{1 - f_{i+1}} \geq \frac{1 - g_i}{1 - f_i}, \quad (6)$$

for all $1 \leq i \leq n - 1$.

This condition is closely related to [Jewitt \(2007\)](#), which shows that for MLRP experiments, the Lehmann order coincides with the Blackwell order for all dichotomies. That is, when restricting the experiments to every pair of states, one is Blackwell more informative than the other. In particular, using the characterization of the Blackwell order for the binary-binary case in (3), we have $\frac{f'_i}{f_i} \geq \frac{g'_i}{g_i}$ and $\frac{1-f'_i}{1-f_i} \geq \frac{1-g'_i}{1-g_i}$ for all $i' > i$. This is equivalent to (6) for all $1 \leq i \leq n - 1$.

Lehmann Monotonicity Say that $C : \mathcal{E}_2^{MLRP} \rightarrow \mathbb{R}_+$ is **Lehmann monotone** if for all $f, g \in \mathcal{E}_2^{MLRP}$, $C(f) \geq C(g)$ whenever $f \succeq_L g$.

Since the Lehmann order allows more comparisons than the Blackwell order, Lehmann monotonicity requires stronger conditions.¹¹ Specifically, there are directions of perturbing an experiment that do not necessarily produce a Blackwell less informative experiment, yet they reduce informativeness in the Lehmann sense. The following lemma identifies a broader class of signal replacements that exhibit this behavior.

Lemma 3. *For any vector $h \in [0, 1]^n$ and for any $1 \leq l \leq n$, define*

$$(h)_{\leq l} \equiv [h_1, \dots, h_l, 0, \dots, 0]^\top \quad \text{and} \quad (h)_{\geq l} \equiv [0, \dots, 0, h_l, \dots, h_n]^\top.$$

For any $f \in \mathcal{E}_2^{MLRP}$ and $1 \leq l \leq n$, if $f_l < f_{l+1}$, there exists $\epsilon' \in (0, 1]$ such that $f \succeq_L f + \epsilon \cdot (\mathbf{1} - f)_{\leq l} \in \mathcal{E}_2^{MLRP}$ for all $\epsilon \in [0, \epsilon']$. Similarly, if $f_{l-1} < f_l$, there exists $\epsilon'' \in (0, 1]$ such that $f \succeq_L f + \epsilon \cdot (-f)_{\geq l} \in \mathcal{E}_2^{MLRP}$ for all $\epsilon \in [0, \epsilon'']$.¹²

Observe that $f + \epsilon \cdot (\mathbf{1} - f)_{\leq l}$ can be interpreted as replacing the low signal with the high signal with probability ϵ , but only in the lower states, specifically, those less than or equal to l . Likewise, $f + \epsilon \cdot (-f)_{\geq l}$ corresponds to replacing the high signal with the low signal with probability ϵ , but only in the higher states (those greater than or equal to l). These operations

¹¹However, permutation invariance—a key condition for Blackwell monotonicity—does not apply in the Lehmann setting, since the analysis with the Lehmann order is restricted to MLRP experiments where permuting signals can violate this property.

¹²We set $f_0 = 0$ and $f_{n+1} = 1$.

echo the monotone quasi-garbling characterization of the Lehmann order by [Kim \(2023\)](#): if f is Lehmann more informative than g , then g can be obtained by adding a *reversely monotone noise* to f , i.e., noise that is more likely to generate higher signals in lower states and lower signals in higher states.¹³

Taking the limit as $\epsilon \rightarrow 0$ yields the following first-order conditions that are necessary for Lehmann monotonicity by Lemma 3:

$$\langle \nabla C(f), (\mathbf{1} - f)_{\leq l} \rangle = \sum_{i=1}^l \frac{\partial C}{\partial f_i}(f) \cdot (1 - f_i) \leq 0, \quad \text{if } f_l < f_{l+1}; \quad (7)$$

$$\langle \nabla C(f), (-f)_{\geq l} \rangle = \sum_{i=l}^n \frac{\partial C}{\partial f_i}(f) \cdot (-f_i) \leq 0, \quad \text{if } f_{l-1} < f_l. \quad (8)$$

We say C is *decreasing in reverse signal replacement* if (7) and (8) hold for all $f \in \mathcal{E}_2^{MLRP}$ and $1 \leq l \leq n$. The following theorem shows that this condition also serves as a sufficient condition for Lehmann monotonicity over binary experiments.

Theorem 2. *Suppose $C : \mathcal{E}_2^{MLRP} \rightarrow \mathbb{R}_+$ is absolutely continuous. Then, C is Lehmann monotone if and only if C is decreasing in reverse signal replacement.*

Similar to the proof of Theorem 1, the sufficiency direction is proved by constructing a decreasing path from f to g , using only directions specified by (7) and (8). A key challenge absent in the Blackwell case is to ensure all experiments along the path remain within \mathcal{E}_2^{MLRP} , since (7) and (8) hold only for MLRP experiments. Our proof provides an explicit construction, where in each step, f_l is transformed to match g_l through the direction of either $(\mathbf{1} - f)_{\leq l}$ or $(-f)_{\geq l}$. Lemma 3 ensures all experiments along the path remain in \mathcal{E}_2^{MLRP} and are decreasing in Lehmann informativeness.

Observe that inequality (1) is equivalent to (7) with $l = n$, and (2) corresponds to (8) with $l = 1$. This implies that decreasing in reverse signal replacement implies decreasing in signal replacement. In other words, while Blackwell monotonicity is characterized by two inequalities, Lehmann monotonicity requires the satisfaction of $2n$ inequalities, making it a significantly stronger condition.

Remark 1. For $m = n = 2$, the Blackwell order and the Lehmann order are equivalent when restricted to \mathcal{E}_2^{MLRP} . This follows from the equivalence between the Lehmann order and the Blackwell dichotomies order, as established by [Jewitt \(2007\)](#). While Lehmann monotonicity in this case

¹³Unlike the state-independent noise in Blackwell's garbling, reversely monotone noise is state-dependent: the noise distribution for a lower state first-order stochastically dominates the distribution for a higher one. See Definition 3 of [Kim \(2023\)](#) for details.

involves 2×2 inequalities, two of these can be derived from the others:

$$\begin{aligned} \frac{1}{1-f_2} \cdot \sum_{l=1}^2 (1-f_l) \frac{\partial C}{\partial f_l} + \frac{1}{f_2} \cdot \sum_{l=1}^2 (-f_l) \frac{\partial C}{\partial f_l} \leq 0 & \implies (1-f_1) \frac{\partial C}{\partial f_1} \leq 0, \\ \frac{1}{1-f_1} \cdot \sum_{l=1}^2 (1-f_l) \frac{\partial C}{\partial f_l} + \frac{1}{f_1} \cdot \sum_{l=1}^2 (-f_l) \frac{\partial C}{\partial f_l} \leq 0 & \implies (-f_1) \frac{\partial C}{\partial f_1} \leq 0, \end{aligned}$$

given $f_1 < f_2$. Therefore, when $m = n = 2$, Lehmann monotonicity essentially requires the same two inequalities as Blackwell monotonicity, confirming their equivalence. This equivalence, however, does not extend beyond binary states.

3 Finite Experiments

In this section, we extend our characterization of Blackwell and Lehmann monotonicity to experiments with any finite number of signals. While the intuition for the necessary conditions carries over from the binary case, establishing their sufficiency is significantly more intricate. The core difficulty lies in constructing a continuous, cost-decreasing path from a more informative experiment to a less informative one. This challenge is particularly acute for Lehmann monotonicity, as the path must remain within the non-convex set of MLRP-satisfying experiments.¹⁴

3.1 Preliminaries

Let \mathcal{E}_m denote the set of experiments given a signal space $\mathcal{S} = \{s_1, \dots, s_m\}$. For any $f \in \mathcal{E}_m$, f can be represented by the $n \times m$ matrix where $f_i^j = f(s_j | \omega_i)$ is the probability of generating signal j in state i . Let $f^j = [f_1^j, \dots, f_n^j]^\top \in \mathbb{R}_+^n$ denote the j -th column vector of f . Using this notation, we can rewrite

$$f = [f^1, \dots, f^m] \in \mathbb{R}_+^{n \times m},$$

with the constraint that $\sum_{j=1}^m f^j = \mathbf{1}$. Let $\mathcal{E} = \bigcup_{2 \leq m < \infty} \mathcal{E}_m$ denote the set of all finite experiments. As \mathcal{E} is a disjoint union of the sets \mathcal{E}_m , we define a cost function $C : \mathcal{E} \rightarrow \mathbb{R}_+$ as a collection of functions $(C_m)_{m \geq 2}$ with $C_m : \mathcal{E}_m \rightarrow \mathbb{R}_+$ and $C(f) = C_m(f)$ for $f \in \mathcal{E}_m$. We say C is absolutely continuous (differentiable) if C_m is absolutely continuous (differentiable) for all $m \geq 2$.

¹⁴See Online Appendix [OA.2](#) for an example of violating convexity.

3.2 Blackwell Monotonicity

The Blackwell order on \mathcal{E} is defined as follows: given $f \in \mathcal{E}_m$ and $g \in \mathcal{E}_{m'}$, say that f is Blackwell more informative than g if and only if there exists an $m \times m'$ stochastic matrix M such that $g = fM$. In this section, we characterize necessary and sufficient conditions for a cost function $C : \mathcal{E} \rightarrow \mathbb{R}_+$ to be Blackwell monotone.

Permutation Invariance As in the binary case, relabeling signals does not change the informativeness of an experiment and the cost should remain the same. Hence, permutation invariance remains a necessary condition for Blackwell monotonicity. Formally, for an experiment $f \in \mathcal{E}_m$, a permutation of f can be represented by fP where P is an $m \times m$ permutation matrix—a stochastic matrix with exactly one non-zero entry in each row and each column.¹⁵ Permutation invariance requires that $C(f) = C(fP)$ for all $f \in \mathcal{E}_m$ and all permutation matrices P .

Split Invariance Consider a garbling operation in which, upon observing signal j , preserves the signal with probability $1 - \lambda$ and relabels it as a new, distinct signal with probability λ . This new experiment is Blackwell equivalent to the original one, since a garbling that merges the two signals together recovers the original experiment. Consequently, the cost should also remain the same. In matrix terms, this corresponds to splitting the column f^j into two columns, $(1 - \lambda)f^j$ and λf^j . Formally, a cost function $C : \mathcal{E} \rightarrow \mathbb{R}_+$ is *split invariant* if $C(f) = C(f')$ for any $f \in \mathcal{E}_m$ and $f' \in \mathcal{E}_{m+1}$ such that

$$f' = [f^1, \dots, (1 - \lambda)f^j, \lambda f^j, \dots, f^m] \quad (9)$$

for some $\lambda \in [0, 1]$ and $1 \leq j \leq m$. Note that when $\lambda = 0$ (or $\lambda = 1$), this operation simply adds a column 0 to f , effectively embedding f into \mathcal{E}_{m+1} . Thus, split invariance ensures that the cost remains unchanged when unobserved signals are added to an experiment.

Decreasing in Signal Replacement We now extend the “decreasing in signal replacement” property from the binary case to finite experiments. For $f \in \mathcal{E}_m$, define

$$f^{j \rightarrow k} \equiv \begin{bmatrix} 0 & \dots & \overbrace{-f^j}^{j\text{-th column}} & \dots & 0 & \dots & \overbrace{f^j}^{k\text{-th column}} & \dots & 0 \end{bmatrix}.$$

Observe that, for all $\epsilon \in [0, 1]$, $f + \epsilon f^{j \rightarrow k} \in \mathcal{E}_m$ is a garbling of f , in which signal j is replaced with signal k with probability ϵ . Since this operation makes the experiment less informative, any Blackwell monotone cost must decrease under it. By taking $\epsilon \rightarrow 0$, we obtain the first-order

¹⁵Observe that when P is a permutation matrix, its inverse, P^{-1} , is also a permutation matrix that restores the original experiment. Since both P and P^{-1} are stochastic matrices, we have $f \succeq_B fP \succeq_B fPP^{-1} = f$, which implies $f \simeq_B fP$.

condition:

$$\langle \nabla C_m(f), f^{j \rightarrow k} \rangle = \sum_{i=1}^n \left(-\frac{\partial C_m}{\partial f_i^j} + \frac{\partial C_m}{\partial f_i^k} \right) f_i^j \leq 0. \tag{10}$$

We say that $C : \mathcal{E} \rightarrow \mathbb{R}_+$ is *decreasing in signal replacement* if (10) holds for all $m \geq 2$ and all $1 \leq j \neq k \leq m$.

Our next theorem shows that these three conditions are necessary and sufficient for Blackwell monotonicity over finite experiments.

Theorem 3. *Suppose $C : \mathcal{E} \rightarrow \mathbb{R}_+$ is absolutely continuous and differentiable. Then, C is Blackwell monotone if and only if C is permutation invariant, split invariant, and decreasing in signal replacement.*

In the proof of Theorem 3, to construct a continuous path between any two comparable experiments $f \in \mathcal{E}_m$ and $g \in \mathcal{E}_{m'}$, we first embed both into $\mathcal{E}_{m+m'}$ and assign their signals to disjoint subsets. We then define a path by taking convex combinations of these embedded experiments within $\mathcal{E}_{m+m'}$. Along this path, the experiment evolves in a direction that can be expressed as a positive linear combination of signal replacements. The differentiability of the cost function implies that its directional derivative along this path is likewise a positive linear combination of its derivatives with respect to signal replacements, each of which is negative. Hence, the cost decreases along the path.

Notice that differentiability enables the construction of a path using positive linear combinations of signal replacements, rather than restricting to single-signal replacements as in the binary case. This relaxation is essential: as Proposition OA.1 in the Online Appendix shows, the argument used for binary experiments does not extend to more general cases. An alternative approach that does not rely on differentiability is discussed in the following remark.

Remark 2. The proof of Theorem 3 relies on the step of embedding the experiments into a larger space $\mathcal{E}_{m+m'}$ where the cost function must be well-defined. In some applications, however, it may be desirable to characterize Blackwell monotonicity for a cost function defined only over experiments with a fixed number of signals, say \mathcal{E}_m . In the Online Appendix OA.3, we show that when the cost function $C_m : \mathcal{E}_m \rightarrow \mathbb{R}_+$ is absolutely continuous and quasi-convex,¹⁷ C_m is Blackwell monotone if and only if it is permutation invariant and decreasing in signal replacement. To address the issue raised in Proposition OA.1, we show that every extreme point of the sublevel set of an experiment can be reached through a combination of single-signal replacements and permutations. This fact, combined with quasi-convexity, establishes the result.

¹⁶Since $f^1 + \dots + f^m = 1$, we can pick j^* and normalize $\partial C_m / \partial f_i^{j^*} = 0$. When $m = 2$ and $j^* = 1$, this definition corresponds to (1) and (2).

¹⁷ C_m is quasi-convex if for any $f, g \in \mathcal{E}_m$, $C(\lambda f + (1 - \lambda)g) \leq \max\{C(f), C(g)\}$ for all $\lambda \in [0, 1]$

3.3 Lehmann Monotonicity

We now turn to characterizing necessary and sufficient conditions for Lehmann monotonicity. Since the Lehmann order is defined over experiments that satisfy the MLRP, we focus on the following subset of \mathcal{E} :

$$\mathcal{E}^{MLRP} \equiv \{f \in \mathcal{E} : f \text{ satisfies the MLRP}\}.$$

Let \mathcal{E}_m^{MLRP} denote the subset of \mathcal{E}^{MLRP} of which elements have m signals.

The Lehmann order over \mathcal{E}^{MLRP} is defined analogously to the binary case. Given $f \in \mathcal{E}_m^{MLRP}$, we associate it with a continuous experiment $\tilde{f} : \Omega \rightarrow \Delta([0, m])$, as constructed in (5). Then, for any $f \in \mathcal{E}_m^{MLRP}$ and $g \in \mathcal{E}_{m'}^{MLRP}$, say that f is Lehmann more informative than g if and only if for all $y \in [0, m']$, $\tilde{F}^{-1}(\tilde{G}(y|\omega)|\omega)$ is increasing in ω .

Split Invariance Suppose that $f \in \mathcal{E}_m^{MLRP}$ and $f' \in \mathcal{E}_{m+1}$ are constructed as in (9) for some $\lambda \in [0, 1]$. Notice that f' also satisfies the MLRP, i.e., $f' \in \mathcal{E}_{m+1}^{MLRP}$. Since splitting an experiment does not change Blackwell informativeness, Lehmann informativeness also remains unchanged as the Lehmann order is weaker than the Blackwell order for MLRP experiments. Formally, it can be shown that for any $x \in [0, m]$ and $y \in [0, m+1]$, both $\tilde{F}^{-1}(\tilde{F}'(y|\omega)|\omega)$ and $\tilde{F}'^{-1}(\tilde{F}(x|\omega)|\omega)$ are constant across ω , thus, f and f' are equally informative in the Lehmann sense. Hence, split invariance serves as a necessary condition for Lehmann monotonicity.

Decreasing in Reverse Signal Replacement We extend “decreasing in reverse signal replacement” property from the binary case to finite experiments. Recall that $f_{\leq l}^j = [f_1^j, \dots, f_l^j, 0, \dots, 0]^\top$ and $f_{\geq l}^j = [0, \dots, 0, f_l^j, \dots, f_n^j]^\top$. For $f \in \mathcal{E}_m^{MLRP}$, define

$$\begin{aligned} f_{\leq l}^{j \rightarrow j+1} &\equiv [0, \dots, 0, 0, -f_{\leq l}^j, f_{\leq l}^j, 0, \dots, 0], \\ f_{\geq l}^{j \rightarrow j-1} &\equiv [0, \dots, 0, f_{\geq l}^j, -f_{\geq l}^j, 0, 0, \dots, 0]. \end{aligned}$$

Notice that, unlike in “decreasing in signal replacement” where signal j may be replaced by any signal k , here we only allow signal j to be replaced by its immediate neighbors, i.e., $j+1$ or $j-1$. This restriction is crucial for ensuring that the resulting experiment remains in \mathcal{E}_m^{MLRP} , as shown by the following lemma. More importantly, it shows that such perturbations reduce the Lehmann informativeness of the experiment.

Lemma 4. Suppose that $f \in \mathcal{E}_m^{MLRP}$. For any $1 \leq l \leq n$ and $1 \leq j \leq m-1$, if $f_l^j f_{l+1}^{j+1} > f_{l+1}^j f_l^{j+1}$,¹⁸ then there exists $\epsilon' \in (0, 1]$ such that $f \succeq_L f + \epsilon f_{\leq l}^{j \rightarrow j+1} \in \mathcal{E}_m^{MLRP}$ for all $\epsilon \in [0, \epsilon']$.

¹⁸Here, given j , we set $f_{n+1}^j = 0$ and $f_{n+1}^{j+1} = 1$.

Similarly, for any $1 \leq l \leq n$ and $2 \leq j \leq m$, if $f_{l-1}^{j-1} f_l^j > f_l^{j-1} f_{l-1}^j$,¹⁹ then there exists $\epsilon'' \in (0, 1]$ such that $f \succeq_L f + \epsilon \cdot f_{\geq l}^{j \rightarrow j-1} \in \mathcal{E}_m^{MLRP}$ for all $\epsilon \in [0, \epsilon'']$.

A cost function $C : \mathcal{E}^{MLRP} \rightarrow \mathbb{R}_+$ is said to be *decreasing in reverse signal replacement* if for any $m \geq 2$, experiment $f \in \mathcal{E}_m^{MLRP}$, and $1 \leq l \leq n$, the following conditions hold:

(a) if $1 \leq j \leq m-1$ and $f_i^j f_{i+1}^{j+1} > f_{i+1}^j f_i^{j+1}$,

$$\langle \nabla C_m(f), f_{\leq l}^{j \rightarrow j+1} \rangle = \sum_{i=1}^l \left(-\frac{\partial C_m}{\partial f_i^j} + \frac{\partial C_m}{\partial f_i^{j+1}} \right) f_i^j \leq 0, \quad (11)$$

(b) if $2 \leq j \leq m$ and $f_{i-1}^{j-1} f_i^j > f_i^{j-1} f_{i-1}^j$,

$$\langle \nabla C_m(f), f_{\geq l}^{j \rightarrow j-1} \rangle = \sum_{i=l}^n \left(-\frac{\partial C_m}{\partial f_i^j} + \frac{\partial C_m}{\partial f_i^{j-1}} \right) f_i^j \leq 0, \quad (12)$$

where C_m is the restriction of C to \mathcal{E}_m^{MLRP} .

The next theorem shows that this condition along with split invariance are necessary and sufficient conditions for Lehmann monotonicity.

Theorem 4. Suppose $C : \mathcal{E}^{MLRP} \rightarrow \mathbb{R}_+$ is absolutely continuous and differentiable. Then, C is Lehmann monotone if and only if C is split invariant and decreasing in reverse signal replacement.

While the core idea of the proof of Theorem 4 remains the construction of a continuous path of experiments, the additional requirement that every experiment along the path satisfies the MLRP introduces significant challenges—most notably because the MLRP is not preserved under convex combinations of experiments. As a result, the construction must be more intricate and draw on insights from multiple fronts. In the next section, we provide a detailed sketch of the proof, highlighting the key geometric ideas that underpin our approach.

3.3.1 Proof Sketch

Geometry of MLRP Experiments and the Lehmann Order We begin by revisiting the geometric characterization of MLRP experiments and the Lehmann order introduced by Jewitt (2007). He considers the probability-probability (PP) plots of experiments in two ordered states. Specifically, in our finite signal setup, we plot cumulative probabilities for adjacent states, e.g., the cumulative probabilities of signal j for an experiment f under states i and $i+1$, denoted by

¹⁹Similarly, we set $f_0^{j-1} = 1$ and $f_0^j = 0$.

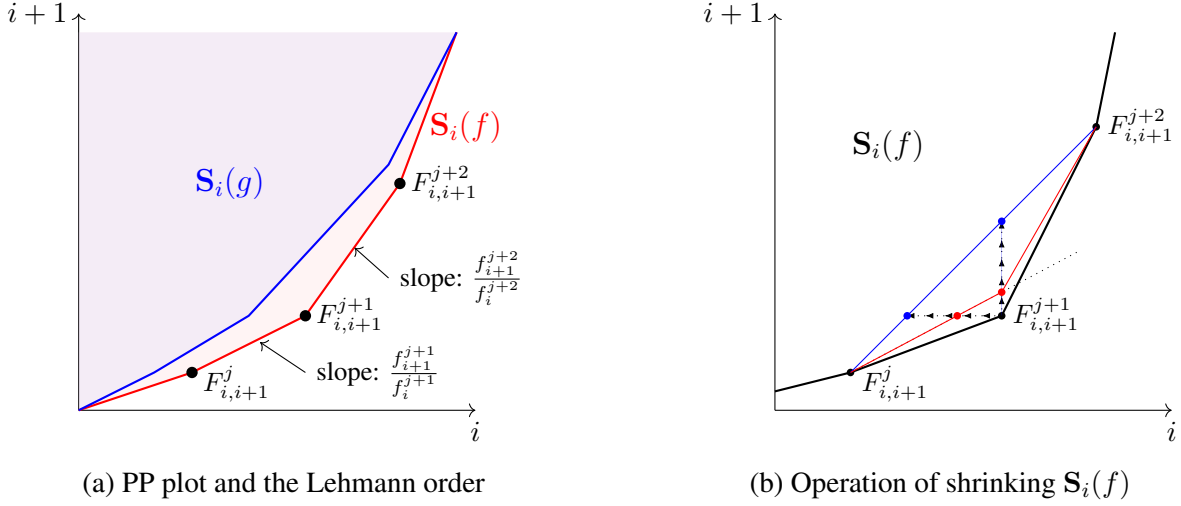


Figure 2: Geometry of MLRP experiments and the Lehmann order

$F_{i,i+1}^j \equiv (F_i^j, F_{i+1}^j)$. Observe that the slope of the line segment connecting $F_{i,i+1}^j$ and $F_{i,i+1}^{j+1}$ is f_{i+1}^{j+1}/f_i^{j+1} . Because of the MLRP, the slope is increasing, i.e., the PP plot is convex as illustrated in Figure 2a. Equivalently, the area in $[0, 1]^2$ above the PP plot, denoted by $S_i(f)$, is a convex set. The formal definition of $S_i(f)$ can be found in Appendix B.2.1.

Jewitt (2007) shows that $f \succeq_L g$ if and only if for each pair of states, the PP plot of f lies below that of g as illustrated in Figure 2a. This in turn becomes equivalent to $S_i(f) \supseteq S_i(g)$ for all $1 \leq i \leq n-1$, which we prove in Lemma B.1.

Path Construction For any $f, g \in \mathcal{E}^{MLRP}$ with $f \succeq_L g$, our goal is to construct a continuous path from f to g using the split and reverse signal replacement operations. Geometrically, since $S_i(f) \supseteq S_i(g)$, this amounts to continuously shrink $S_i(f)$ toward $S_i(g)$ for all $1 \leq i \leq n-1$. A key requirement is that the PP plots must remain convex throughout the transformation, thereby preserving the MLRP along the entire path.

We next illustrate that such a path can be constructed by iteratively applying a key operation that (i) shrinks $S_i(f)$, (ii) leaves $S_{i'}(f)$ unchanged for all $i' \neq i$, (iii) preserves the convexity of the PP plots throughout the transformation, and (iv) decreases the cost due to the decreasing in reverse signal replacement property.

The Key Operation Consider the PP plot of f in Figure 2b. The goal is to shrink $S_i(f)$ by removing the region below the blue segment connecting $F_{i,i+1}^j$ and $F_{i,i+1}^{j+2}$, while keeping $S_{i'}(f)$ unchanged for all $i' \neq i$. We show that this can be implemented through a combination of split and reverse signal replacement operations, preserving the MLRP throughout the transformation.

First, we insert a zero column between f^{j+1} and f^{j+2} . By split invariance, this does not affect

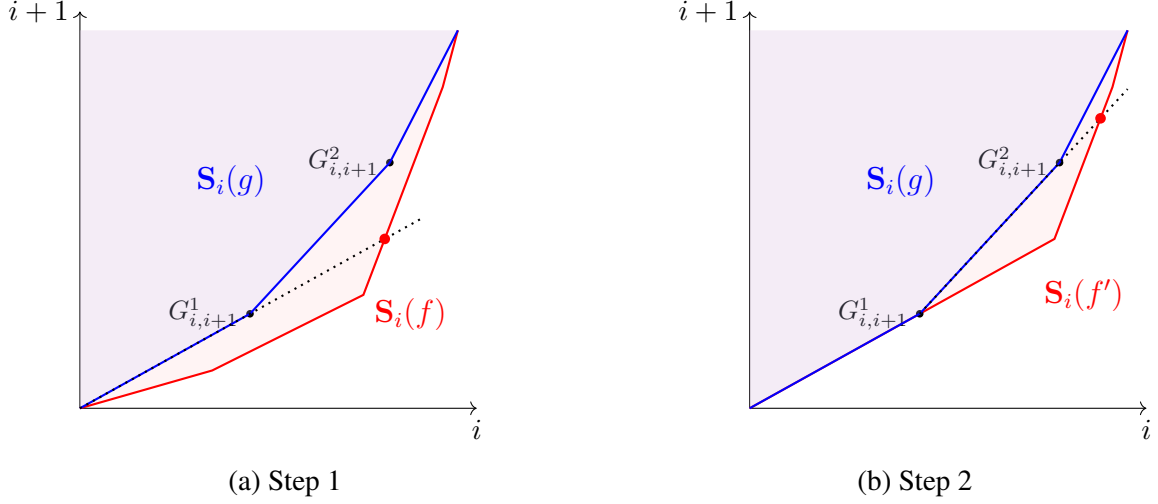


Figure 3: Shrinking Operations

the cost. Next, we repeatedly apply reverse signal replacements to transfer a portion of $f_{\leq i}^{j+1}$ and a portion of $f_{\geq i+1}^{j+2}$ into this new column. The transformation starts and ends as follows:

$$\begin{bmatrix} f_{\leq i}^{j+1} & \mathbf{0} & f_{\leq i}^{j+2} \\ f_{\geq i+1}^{j+1} & \mathbf{0} & f_{\geq i+1}^{j+2} \end{bmatrix} \rightarrow \dots \rightarrow \begin{bmatrix} (1-a)f_{\leq i}^{j+1} & af_{\leq i}^{j+1} & f_{\leq i}^{j+2} \\ f_{\geq i+1}^{j+1} & bf_{\geq i+1}^{j+2} & (1-b)f_{\geq i+1}^{j+2} \end{bmatrix} \equiv \tilde{f}.$$

At the endpoint \tilde{f} , notice that for all $i' \neq i$, this procedure merely splits a signal—either signal $j+1$ (if $i' < i$) or signal $j+2$ (if $i' > i$)—into two parts. Hence, $S_{i'}(\tilde{f})$ remains identical to $S_{i'}(f)$; geometrically, this corresponds to adding a dot along the same segment of the PP plot. For $S_i(\tilde{f})$, the first column represents a leftward movement of a dot from $F_{i,i+1}^j$, and the second corresponds to an upward movement of a dot from $F_{i,i+1}^{j+1}$, as indicated by the two blue dots in Figure 2b. The parameters a and b are chosen such that $S_i(\tilde{f})$ aligns precisely with the blue segment connecting $F_{i,i+1}^j$ and $F_{i,i+1}^{j+2}$.

Next, we detail how the incremental movements are constructed to ensure that the PP plot remains convex throughout the procedure. The red curve in Figure 2b illustrates the PP plot of an intermediate experiment during this process. The key is to maintain alignment of the first two red segments so that their common slope remains steeper than that of the third segment—thus preserving convexity. To achieve this, we construct the intermediate experiments by rotating the first two segments together around the point $F_{i,i+1}^j$.

Finally, we argue that every incremental movement is a positive linear combination of reverse signal replacements. This implies, by the decreasing-in-reverse-signal-replacement property and the differentiability of the cost function, that the overall cost decreases throughout the operation. The formal statement and proof for the key operation is given in Lemma B.2, where, importantly,

we further show that the key operation can be applied more generally to remove the region below the segment connecting any $F_{i,i+1}^j$ and $F_{i,i+1}^k$.

Applying the Key Operation Finally, we use Figure 3 to illustrate how the key operation can be applied to shrink $S_i(f)$ towards $S_i(g)$. Begin by extending the first segment of $S_i(g)$ until it intersects the boundary of $S_i(f)$. Then, split the corresponding signal of f so that the intersection becomes a new cumulative point as shown by the red dot in Figure 3a. This allows us to apply the key operation to remove the region below the extended segment. Once this step is complete, we proceed to the next segment of $S_i(g)$ and repeat the process as illustrated by Figure 3b. In this way, we iteratively shrink $S_i(f)$ toward $S_i(g)$. Throughout the procedure, $S_{i'}(f)$ remains unchanged for all $i' \neq i$, and the overall cost decreases. Repeating the process for each $1 \leq i \leq n-1$, we construct a path establishing that the cost of f is greater than an experiment that has the same PP plots as g . By splitting invariance, such an experiment shares the same cost as g , thus proving $C(f) \geq C(g)$. \square

4 Monotonicity Properties of Widely Used Costs

In this section, we apply our characterizations to several widely-used classes of information cost functions in the literature and analyze their monotonicity properties. While the Blackwell monotonicity of some of these classes is already known, our framework provides a unified approach for studying both Blackwell and Lehmann monotonicity. In particular, we demonstrate that Blackwell monotonicity of these costs does not guarantee Lehmann monotonicity, and we identify the additional conditions under which Lehmann monotonicity holds. Moreover, we show that some known costs may fail both Blackwell and Lehmann monotonicity, even in a local sense.

4.1 Likelihood Separable Costs

We define an information cost function $C : \mathcal{E} \rightarrow \mathbb{R}_+$ to be **likelihood separable**, a term introduced by Denti et al. (2022b), if there exists a differentiable function $\psi : [0, 1]^n \rightarrow \mathbb{R}_+$ such that for all m and $f \in \mathcal{E}_m$,

$$C(f) = \sum_{j=1}^m \psi(f^j) - \psi(\mathbf{1}).$$

Because likelihood separable costs are additively separable across signals, verifying the condition of decreasing in (reverse) signal replacement can be reduced to analyzing the behavior of the

function ψ . Specifically, it holds that

$$\langle \nabla C(f), f^{j \rightarrow k} \rangle = \sum_{i=1}^n \left(-\frac{\partial \psi(f^j)}{\partial f_i^j} + \frac{\partial \psi(f^k)}{\partial f_i^k} \right) f_i^j.$$

This observation leads to the following conditions for Blackwell and Lehmann monotonicity of likelihood separable costs.

Proposition 2. *The following statements are true:*

- (i) *A likelihood separable cost is Blackwell monotone if and only if ψ is sublinear.*²⁰
- (ii) *A likelihood separable cost with sublinear ψ is Lehmann monotone if, for all $l \in \{1, \dots, n\}$,*

$$\sum_{i=1}^l \left(-\frac{\partial \psi(h)}{\partial h_i} + \frac{\partial \psi(h')}{\partial h'_i} \right) h_i \leq 0, \text{ and } \sum_{i=l}^n \left(-\frac{\partial \psi(h')}{\partial h'_i} + \frac{\partial \psi(h)}{\partial h_i} \right) h'_i \leq 0,$$

for all $h, h' \in [0, 1]^n$ such that $h \leq_{MLRP} h'$, i.e., $h_i h'_{i'} \geq h_{i'} h'_i$ for all $i < i'$.

We note that results similar to (i) have been established in [Denti et al. \(2022b\)](#) and [Baker \(2023\)](#), but applying Theorem 3 yields a more direct proof. Specifically, to show that sublinear ψ implies decreasing in signal replacement,²¹ observe that

$$\begin{aligned} C(f + \epsilon f^{j \rightarrow k}) - C(f) &= \psi(f^k + \epsilon f^j) - \psi(f^j) + \psi((1 - \epsilon)f^j) - \psi(f^j) \\ &\leq \epsilon \psi(f^j) - \epsilon \psi(f^j) = 0, \end{aligned}$$

where the inequality follows from the sublinearity of ψ .

The conditions in (ii) are derived from our characterization of Lehmann monotonicity in Theorem 4. These conditions are sufficient as they imply decreasing in reverse signal replacement holds for MLRP experiments. However, they are not shown to be necessary with the only subtlety that, given a pair of $h \leq_{MLRP} h'$, it is not clear whether they always belong to an MLRP experiment. Nonetheless, to falsify the Lehmann monotonicity for a likelihood separable cost, it suffices to identify a single violation of these inequalities within some MLRP experiment. The proof of Proposition 3 provides such an example.

When seeking a Blackwell monotone cost function, Proposition 2 provides a convenient construction: any sublinear function ψ yields a likelihood separable cost that satisfies Blackwell monotonicity. This includes a broad class of functions, such as norms and seminorms. In contrast, the

²⁰That is, $\psi(\gamma \cdot h) = \gamma \cdot \psi(h)$ for all $h \in [0, 1]^n$ and $\gamma \geq 0$ such that $\gamma \cdot h \in [0, 1]^n$; and $\psi(h + l) \leq \psi(h) + \psi(l)$ for all $h, h' \in [0, 1]^n$ such that $h + h' \in [0, 1]^n$.

²¹Permutation invariance and split invariance follow directly from the definition and the sublinearity of ψ .

condition for Lehmann monotonicity is more restrictive. While not every sublinear ψ leads to a Lehmann monotone cost, we show below that a reasonably large and natural subclass of likelihood separable costs does satisfy Lehmann monotonicity.

Proposition 3. *The following statements are true:*

- (i) *There exist likelihood separable costs with sublinear ψ that are not Lehmann monotone.*
- (ii) *If $\psi(\cdot)$ is a weighted p -norm with $p > 1$, i.e., there exists $w_i > 0$ for all i such that*

$$\psi(h) = \left(\sum_{i=1}^n w_i h_i^p \right)^{1/p},$$

then the likelihood separable cost is Lehmann-monotone.

4.2 Posterior Separable Costs

When a full-support prior μ is given, information costs can be defined as functions of the random posteriors induced by an experiment.²² Specifically, for any $f \in \mathcal{E}$, let τ^j denote the total probability of observing signal s_j and let q^j denote the posterior distribution vector given signal s_j , i.e.,

$$\tau^j = \sum_{i'=1}^n \mu_{i'} f_{i'}^j, \quad \text{and} \quad q_i^j = \frac{\mu_i f_i^j}{\sum_{i'=1}^n \mu_{i'} f_{i'}^j}, \quad \text{for } i = 1, \dots, n.$$

Among such costs, a cost function is said to be **posterior separable** if there exists a function $H : [0, 1]^n \rightarrow \mathbb{R}$ representing a measure of uncertainty such that, for all m and $f \in \mathcal{E}_m$,

$$C_\mu(f) = H(\mu) - \sum_{j=1}^m \tau^j H(q^j).$$

It is well known that a posterior separable cost is Blackwell monotone if and only if H is concave (Caplin and Dean, 2015; Denti, 2022).

Notably, there is a dual relationship between posterior separable costs with concave H and likelihood separable costs with sublinear ψ .²³ As a result, not all such posterior separable costs

²²For extensive discussions on the relationship between experiment-based and posterior-based costs, see Gentzkow and Kamenica (2014), Mensch (2018), Morris and Strack (2019), Denti et al. (2022a) and Bloedel and Zhong (2024).

²³The general duality between experimental costs and posterior-based costs is established in Denti et al. (2022b): see Corollary 20. If we restrict attention to posterior separable or likelihood separable, they have a specific dual relationship (Proposition 37)

are Lehmann monotone, and the same sufficient condition for Lehmann monotonicity from Proposition 2 applies. When working with the posterior separable functional form, we can derive a further sufficient condition that is more conveniently expressed in terms of the derivatives of H with respect to the posteriors.

Proposition 4. *Suppose H is concave and differentiable. If, for all $l \in \{1, \dots, n\}$,*

$$\sum_{i=1}^l \left(\frac{\partial H(q)}{\partial q_i} - \frac{\partial H(q')}{\partial q'_i} \right) q_i \leq 0, \text{ and } \sum_{i=l}^n \left(\frac{\partial H(q')}{\partial q'_i} - \frac{\partial H(q)}{\partial q_i} \right) q'_i \leq 0,$$

for all $q \leq_{FOSD} q'$, i.e., $\sum_{i=1}^s q_i \geq \sum_{i=1}^s q'_i$ for all $s \in \{1, \dots, n\}$. Then the posterior separable cost is Lehmann monotone.

A well-known example of posterior separable costs is the **entropy cost** (Sims, 2003), defined by $H(q) = -\sum_{i=1}^n q_i \log q_i$. Its Blackwell monotonicity is well established, and we can also verify its Lehmann monotonicity using the sufficient condition in Proposition 4.

Proposition 5. *The entropy cost is Lehmann monotone.*

4.3 Bregman Information Costs

The entropy cost is well known to provide a rational inattention foundation for discrete choice models taking the multinomial logit form (Matějka and McKay, 2015). The previous section established its Lehmann monotonicity, thereby strengthening its theoretical appeal. As a generalization of this approach, Fosgerau et al. (2020) introduce the broader class of **Bregman information costs**, showing that they can support any additive random utility discrete choice model within a rational inattention framework. In this section, we examine the monotonicity properties of this generalized class of costs.

Given a prior μ , the Bregman information cost is defined with respect to a function $S : [0, 1]^m \rightarrow [0, 1]^m$ as follows:

$$C_\mu(f) = \sum_{j=1}^m \sum_{i=1}^n \mu_i f_i^j \log \left(\frac{\mu_i S_j(f_i^\top)}{S_j(\sum_{i'} \mu_{i'} f_{i'}^\top)} \right) - \sum_{i=1}^n \mu_i \log \mu_i,$$

where $f_i^\top \in [0, 1]^m$ denotes the i -th row of f (interpreted as a column vector), and $S_j(\cdot)$ is the j -th component of $S(\cdot)$. Notably, the entropy cost is a special case of the Bregman cost with $S(h) = h$ for all $h \in [0, 1]^m$.

Because S need not be symmetric, Bregman information costs are generally not permutation invariant and, hence, are not necessarily Blackwell monotone. For a common specification of S

corresponding to the nested logit model, we next demonstrate that the Bregman cost may violate decreasing in signal replacement, and thus fails to satisfy both Blackwell and Lehmann monotonicity, even in a *local* sense.

Let the signal space be partitioned into mutually exclusive nests, and let g_j denote the nest containing signal j . Consider the simplest nested logit structure, where the nesting parameter $\xi \in (0, 1]$ is constant across all nests. In this case, the Bregman information cost takes the form:

$$C_\mu(f) = \sum_{j=1}^m \sum_{i=1}^n \mu_i f_i^j \left(\xi \log \frac{\mu_i f_i^j}{\sum_{i'} \mu_{i'} f_{i'}^j} + (1 - \xi) \log \frac{\mu_i f_i^{g_j}}{\sum_{i'} \mu_{i'} f_{i'}^{g_j}} \right) - \sum_{i=1}^n \mu_i \log \mu_i,$$

where $f_i^{g_j} \equiv \sum_{j' \in g_j} f_i^{j'}$. The derivative of this cost with respect to f_i^j is given by

$$\frac{\partial C_\mu}{\partial f_i^j} = \mu_i (\xi \log q_i^j + (1 - \xi) \log q_i^{g_j}),$$

where $q_i^{g_j}$ denote the posterior given observing the nest g_j . Suppose for some experiment f , it holds that $q^j = q^k = q^{g_k} \neq q^{g_j}$. Then

$$\begin{aligned} \langle \nabla C_\mu(f), f^{j \rightarrow k} \rangle &= \sum_{i=1}^n \mu_i f_i^j (\xi \log q_i^k + (1 - \xi) \log q_i^{g_k} - \xi \log q_i^j - (1 - \xi) \log q_i^{g_j}) \\ &= (1 - \xi) \sum_{i=1}^n \mu_i f_i^j (\log q_i^j - \log q_i^{g_j}) > 0. \end{aligned}$$

Hence, the Bregman information cost violates decreasing in signal replacement. In other words, a signal replacement that reduces informativeness can lead to a higher information cost, showing that the Bregman cost could fail both Blackwell and Lehmann monotonicity under a local perturbation of the experiment.

4.4 State-wise Divergence Costs

Some recent papers in the literature such as [Pomatto et al. \(2023\)](#) and [Bordoli and Iijima \(2025\)](#) focus on information costs that are defined in terms of statistical divergences between probability distributions over signals across states. Specifically, let $D(f_i \| f_{i'})$ denote a statistical divergence between the probability distributions f_i and $f_{i'}$ over the signals. We say that an information cost function is a **state-wise divergence cost** if it is a monotone increasing function of $D(f_i \| f_{i'})$ for all $i, i' \in \{1, \dots, n\}$.

The Log-Likelihood Ratio (LLR) cost introduced in [Pomatto et al. \(2023\)](#) is a positive linear combination of Kullback-Leibler (KL) divergence between states. [Bordoli and Iijima \(2025\)](#)

consider both the Rényi divergence and the KL divergence across states and define costs as the maximum of the divergences across all pairs of states. Hence, both of these costs are state-wise divergence costs.

While our conditions may be used to verify Blackwell and Lehmann monotonicity of state-wise divergence costs, we note that their monotonicity can be established more directly using the properties of the divergence. A statistical divergence satisfies the *data-processing inequality* if the divergence between two distributions decreases under a garbling of the signals. This property is known to hold for many commonly used divergences, such as KL divergence, Rényi divergence, and more generally the f -divergence (Polyanskiy and Wu, 2022). As a result, whenever the data-processing inequality holds, the Blackwell monotonicity holds by definition and the Lehmann monotonicity can be established using the Blackwell on dichotomies characterization by Jewitt (2007).

5 Conclusion

As models of costly information acquisition become increasingly central to economic theory, so too does the need for well-founded cost functions. The principle of monotonicity—that more information should be more costly—is a bedrock assumption in virtually all such models. Yet, while this principle is a key ingredient in many characterizations of cost functions, its standalone implications, when isolated from other axioms, have been less explored. This gap in understanding is particularly acute for the Lehmann order; despite being the natural relaxation of the Blackwell criterion for the large class of monotone decision problems, the conditions required for Lehmann monotonicity have remained largely unexplored.

This paper has aimed to unpack this fundamental property. By isolating monotonicity and characterizing it through simple, local conditions, we provide a more transparent and workable foundation for the analysis of information costs. We hope that by clarifying what monotonicity requires on its own—and just as importantly, what it does not—our work can serve as a stepping stone. It offers a more flexible toolkit for researchers to build, verify, and apply cost functions in diverse economic environments, particularly for the large class of monotone problems where a full analysis was previously intractable. By strengthening the foundations, we hope to foster more robust and targeted applications of costly information acquisition in economics.

Appendix

A Proofs for Section 2

A.1 Absolute Continuity

All results in Section 2 do not need the cost function C to be differentiable. For any absolutely continuous cost function C , let $D^+C(f; h)$ denote its (one-sided) directional derivative at $f \in \mathcal{E}_2$ in the direction of $h \in \mathbb{R}^n$, if the following limit exists:

$$D^+C(f; h) \equiv \lim_{\epsilon \downarrow 0} \frac{C(f + \epsilon h) - C(f)}{\epsilon}.$$

In this case, we rewrite (1) and (2) as

$$D^+C(f; \mathbf{1} - f) \leq 0; \tag{A.1}$$

$$D^+C(f; -f) \leq 0. \tag{A.2}$$

Similarly, (7) and (8) can be rewritten as

$$D^+C(f; (\mathbf{1} - f)_{\leq l}) \leq 0, \quad \text{if } f_l < f_{l+1}; \tag{A.3}$$

$$D^+C(f; (-f)_{\geq l}) \leq 0, \quad \text{if } f_{l-1} < f_l. \tag{A.4}$$

Importantly, absolute continuity implies that the Fundamental Theorem of Calculus (FTC) holds, i.e., for $\varphi(t) = C(tg + (1 - t)f)$ for $t \in [0, 1]$, we have

$$C(g) - C(f) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t) dt = \int_0^1 D^+C(tg + (1 - t)f; g - f) dt. \tag{A.5}$$

A.2 Blackwell Monotonicity: Proof of Theorem 1

Lemma A.1. *For any $f, g \in \mathcal{E}_2$ such that $f \succeq_B g$, there exists $1 \geq a \geq b \geq 0$ such that either*

$$g = af + b(\mathbf{1} - f) \quad \text{or} \quad \mathbf{1} - g = af + b(\mathbf{1} - f). \tag{A.6}$$

Let g satisfy the first equation of (A.6) and $f' = \frac{a-b}{1-b}f$.²⁴ Then, for all $\lambda \in [0, 1]$,

$$f \succeq_B (1 - \lambda)f + \lambda f' \succeq_B f', \text{ and} \quad (\text{A.7})$$

$$f' \succeq_B (1 - \lambda)f' + \lambda g \succeq_B g. \quad (\text{A.8})$$

Proof of Lemma A.1. Recall that $f \succeq_B g$ implies that there exist $(a, b) \in [0, 1]^2$ such that $g = af + b(1 - f)$. If $a \geq b$, the first equation of (A.6) holds. If $a < b$, notice $1 - g = a'f + b'(1 - f)$ for $a' = 1 - a > 1 - b = b'$.

When $b = 1$, $a = 1$ and $g = f + (1 - f) = 1 = f'$. (A.8) holds. Notice that $(1 - \lambda)f + \lambda 1 = 1 \cdot f + \lambda(1 - f) \in \text{PARL}(f, 1 - f)$, thus, $f \succeq_B (1 - \lambda)f + \lambda 1$. Similarly, $(1 - \lambda)f + \lambda 1 \succeq_B 1$. (A.7) holds.

When $b < 1$, $\frac{a-b}{1-b} \in [0, 1]$ and $f \succeq_B f'$. For any $\lambda \in [0, 1]$, $f \succeq_B \lambda f + (1 - \lambda)f'$ follows from the convexity of $\text{PARL}(f, 1 - f)$. Next, notice that

$$f' = \frac{\frac{a-b}{1-b}}{1 - \lambda + \lambda \frac{a-b}{1-b}} ((1 - \lambda)f + \lambda f').$$

Since

$$\frac{\frac{a-b}{1-b}}{1 - \lambda + \lambda \frac{a-b}{1-b}} \in [0, 1],$$

we have $f' \in \text{PARL}(((1 - \lambda)f + \lambda f'), 1 - ((1 - \lambda)f + \lambda f'))$, and thus $(1 - \lambda)f + \lambda f' \succeq_B f'$.

From $g = af + b(1 - f)$,

$$g = \frac{a-b}{1-b}f + b \left(1 - \frac{a-b}{1-b}f \right) = f' + b(1 - f').$$

Thus $f' \succeq_B g$ and $g - f' = b(1 - f')$. Similarly, $f' \succeq_B (1 - \lambda)f' + \lambda g \succeq_B g$. \square

Proof of Theorem 1. Necessity is proved in the main text.

For sufficiency, take any $f \succeq_B g$. First, permute g if needed to have g satisfy the first equation of (A.6). Permutation invariance ensures the cost stays the same. Define $\varphi_1(\lambda) \equiv C((1 - \lambda)f + \lambda f')$ and $\varphi_2(\lambda) \equiv C((1 - \lambda)f' + \lambda g)$. Absolute continuity implies that φ_1 is differentiable almost everywhere, and satisfies

$$\varphi_1'(\lambda) = D^+C((1 - \lambda)f + \lambda f'; -f + f'),$$

²⁴When $b = 1$, define $f' = 1$.

when it is differentiable. On the other hand,

$$-f + f' = -\frac{\frac{1-a}{1-b}}{1-\lambda + \lambda \frac{a-b}{1-b}}((1-\lambda)f + \lambda f').$$

Therefore, $\varphi'_1(\lambda)$ has the same sign as $D^+C((1-\lambda)f + \lambda f'; -((1-\lambda)f + \lambda f'))$ and is negative by (A.2). The FTC implies

$$C(f') = \varphi_1(1) = \varphi_1(0) + \int_0^1 \varphi'_1(\lambda) d\lambda \leq \varphi_1(0) = C(f).$$

Similarly, observe that

$$\begin{aligned} \varphi'_2(\lambda) &= D^+C((1-\lambda)f' + \lambda g; -f' + g), \\ -f' + g &= b(1 - f') = \frac{b}{1-\lambda b} (1 - ((1-\lambda)f' + \lambda g)). \end{aligned}$$

Then, $\varphi'_2(\lambda)$ is non-positive since it has the same sign as $D^+C((1-\lambda)f' + \lambda g; 1 - ((1-\lambda)f' + \lambda g))$. By applying the FTC, $C(g) = \varphi_2(1) \leq \varphi_2(0) = C(f')$. Therefore, $C(g) \leq C(f)$. \square

A.3 Lehmann Monotonicity

A.3.1 Proof of Lemma 2

Proof of Lemma 2. Note that for $y \in [0, 1]$ and $1 \leq i \leq n$,

$$\tilde{F}^{-1}(\tilde{G}(y|\omega_i)|\omega_i) = \begin{cases} \frac{1-g_i}{1-f_i}y, & \text{if } y \leq \frac{1-f_i}{1-g_i}, \\ 1 + \frac{(1-g_i)y - (1-f_i)}{f_i}, & \text{if } y > \frac{1-f_i}{1-g_i}. \end{cases}$$

Additionally, if $y \in (1, 2]$,

$$\tilde{F}^{-1}(\tilde{G}(y|\omega_i)|\omega_i) = \begin{cases} \frac{1-g_i}{1-f_i} + \frac{g_i(y-1)}{1-f_i}, & \text{if } y \leq 2 - \frac{f_i}{g_i}, \\ 2 - \frac{g_i(2-y)}{f_i}, & \text{if } y > 2 - \frac{f_i}{g_i}. \end{cases}$$

In particular, when $f_i \leq g_i$, as illustrated in Figure 4a, $\tilde{F}^{-1}(\tilde{G}(\cdot|\omega_i)|\omega_i)$ is composed of three line segments connecting the vertices $(0, 0)$, $(1, \frac{1-g_i}{1-f_i})$, $(2 - \frac{f_i}{g_i}, 1)$ and $(2, 2)$. The curve passes through the lower-right unit square defined by coordinates $(1, 0)$ and $(2, 1)$.

When $f_i > g_i$, as illustrated in Figure 4b, $\tilde{F}^{-1}(\tilde{G}(\cdot|\omega_i)|\omega_i)$ is composed of three line segments connecting $(0, 0)$, $(\frac{1-f_i}{1-g_i}, 1)$, $(1, 2 - \frac{g_i}{f_i})$ and $(2, 2)$. The curve passes through the upper-left unit

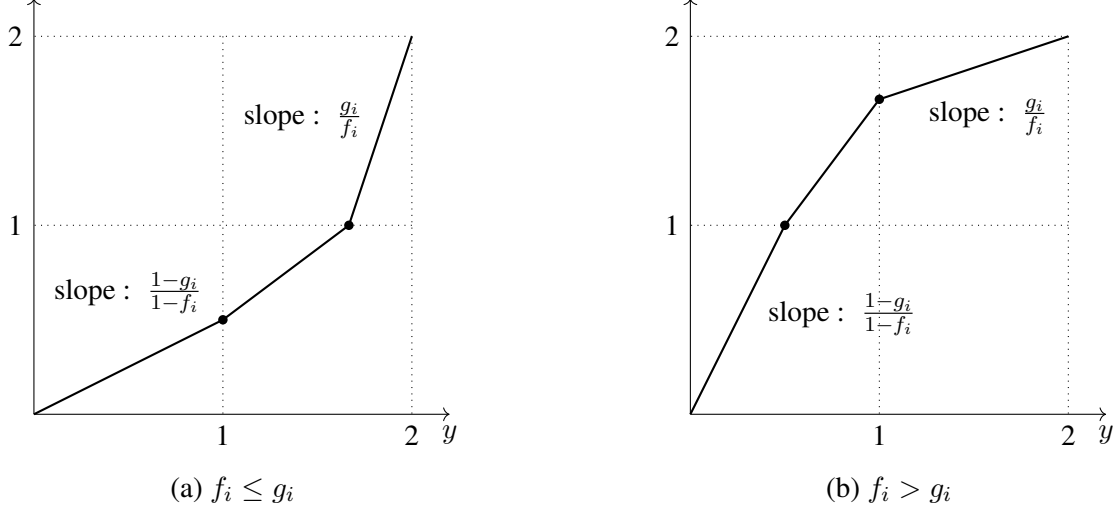


Figure 4: The graphs of $\tilde{F}^{-1}(\tilde{G}(y|\omega_i)|\omega_i)$

square defined by coordinates $(0, 1)$ and $(1, 2)$.

Notice that the slope of the first line segment is $\frac{1-g_i}{1-f_i}$ and that of the last line segment is $\frac{g_i}{f_i}$. The middle line segment is determined by the intersections of these lines and the lower-right or upper-left unit squares.

Using this property, we can see that $\tilde{F}^{-1}(\tilde{G}(y|\omega_i)|\omega_i) \leq \tilde{F}^{-1}(\tilde{G}(y|\omega_{i+1})|\omega_{i+1})$ for all $y \in [0, 2]$ if and only if the slopes of the first and last line segments under ω_{i+1} are steeper than those under ω_i , which is equivalent to condition (6). \square

A.3.2 Proof of Lemma 3

Proof of Lemma 3. Consider any $f \in \mathcal{E}_2^{MLRP}$ and $1 \leq l \leq n$ such that $f_l < f_{l+1}$. Define $\epsilon' \equiv \frac{f_{l+1}-f_l}{1-f_l} \in (0, 1]$, then we have $f_l + \epsilon' \cdot (1-f_l) = f_{l+1}$. For any $\epsilon \in (0, \epsilon']$, let g denote $f + \epsilon \cdot (1-f)_{\leq l}$. Then,

$$\frac{g_i}{f_i} = \begin{cases} 1 - \epsilon + \frac{\epsilon}{f_i}, & \text{if } i \leq l, \\ 1, & \text{if } i > l. \end{cases}$$

Since f_i is increasing in i , $\frac{g_i}{f_i}$ is decreasing in i . Also note that

$$\frac{1-g_i}{1-f_i} = \begin{cases} 1 - \epsilon, & \text{if } i \leq l, \\ 1, & \text{if } i > l. \end{cases}$$

This implies that $\frac{1-g_i}{1-f_i}$ is increasing in i , i.e., (6) holds for all $1 \leq i \leq n-1$. Therefore, $f \succeq_L g$ by Lemma 2.

Next, consider an experiment $f \in \mathcal{E}_2^{MLRP}$ and $1 \leq l \leq n$ such that $f_{l-1} < f_l$ and define $\epsilon'' = \frac{f_l - f_{l-1}}{f_l}$. Let g denote $f + \epsilon \cdot (-f)_{\geq l}$. Then,

$$\frac{g_i}{f_i} = \begin{cases} 1, & \text{if } i < l, \\ 1 - \epsilon, & \text{if } i \geq l, \end{cases}$$

thus, $\frac{g_i}{f_i}$ is decreasing in i . Also note that

$$\frac{1 - g_i}{1 - f_i} = \begin{cases} 1, & \text{if } i < l, \\ 1 + \frac{\epsilon}{1 - f_i}, & \text{if } i \geq l. \end{cases}$$

Since $1 - f_i$ is decreasing in i , thus, $\frac{1 - g_i}{1 - f_i}$ is increasing in i . Therefore, $f \succeq_L g$ by Lemma 2. \square

A.3.3 Proof of Theorem 2

Lemma A.2. Suppose that $f, g \in \mathcal{E}_2^{MLRP}$ and $f \succeq_L g$. Then, there exist $0 \leq k \leq n$, $1 \geq \epsilon_1 \geq \dots \geq \epsilon_k \geq 0$ and $1 \geq \epsilon_n \geq \dots \geq \epsilon_{k+1} \geq 0$ such that

$$[1 - g_i, g_i] = [1 - f_i, f_i] \begin{bmatrix} 1 - \epsilon_i & \epsilon_i \\ 0 & 1 \end{bmatrix} \quad (\text{A.9})$$

when $1 \leq i \leq k$, and

$$[1 - g_i, g_i] = [1 - f_i, f_i] \begin{bmatrix} 1 & 0 \\ \epsilon_i & 1 - \epsilon_i \end{bmatrix} \quad (\text{A.10})$$

when $k + 1 \leq i \leq n$.

Proof of Lemma A.2. Define $k \equiv \max(\{i : g_i \geq f_i\} \cup \{0\})$.

For $1 \leq i \leq k$, set $\epsilon_i = 1 - \frac{1 - g_i}{1 - f_i}$. Then, (A.9) holds by definition. Additionally, by $f \succeq_L g$ and Lemma 2, $\frac{1 - g_i}{1 - f_i}$ is increasing in i , thus, ϵ_i is decreasing in i . Last, by the definition of k , $\epsilon_k \geq 0$.

Next, for $k + 1 \leq i \leq n$, set $\epsilon_i = 1 - \frac{g_i}{f_i}$. Then, (A.10) holds by definition. Additionally, by $f \succeq_L g$ and Lemma 2, $\frac{g_i}{f_i}$ is decreasing in i , thus, ϵ_i is increasing in i . Last, by the definition of k , $\epsilon_{k+1} \geq 0$. \square

Proof of Theorem 2. Suppose that $f \succeq_L g$, $k \equiv \max(\{i : g_i \geq f_i\} \cup \{0\})$, and $(\epsilon_i)_{i=1}^n$ is defined

from Lemma A.2. Let $e_i = \frac{\epsilon_i - \epsilon_{i+1}}{1 - \epsilon_{i+1}}$ for all $1 \leq i \leq k-1$, $e_k = \epsilon_k$. Then, we have

$$\begin{bmatrix} 1 - \epsilon_i & \epsilon_i \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 - e_k & e_k \\ 0 & 1 \end{bmatrix} \cdots \begin{bmatrix} 1 - e_i & e_i \\ 0 & 1 \end{bmatrix}.$$

Likewise, let $e_{k+1} = \epsilon_{k+1}$ and $e_i = \frac{\epsilon_i - \epsilon_{i-1}}{1 - \epsilon_{i-1}}$ for all $k+1 \leq i \leq n$. Then, we also have

$$\begin{bmatrix} 1 & 0 \\ \epsilon_i & 1 - \epsilon_i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ e_{k+1} & 1 - e_{k+1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ e_{k+2} & 1 - e_{k+2} \end{bmatrix} \cdots \begin{bmatrix} 1 & 0 \\ e_i & 1 - e_i \end{bmatrix}.$$

Using these, define $h^k \equiv f + e_k \cdot (1 - f)_{\leq k}$, and for $1 \leq i \leq k-1$, define

$$h^i \equiv h^{i+1} + e_i \cdot (1 - h^{i+1})_{\leq i}$$

Then, from (A.3) and the FTC, we have

$$C(f) \geq C(h^k) \geq \cdots \geq C(h^1). \quad (\text{A.11})$$

Next, define $h^{k+1} \equiv h^1 + e_{k+1} \cdot (-h^1)_{\geq k+1}$, and for $k+2 \leq i \leq n$, define

$$h^i \equiv h^{i-1} + e_i \cdot (-h^{i-1})_{\geq i}.$$

Then, from (A.4) and the FTC, we have

$$C(h^1) \geq C(h^{k+1}) \geq \cdots \geq C(h^n).$$

Note that $h^n = g$. Then, from (A.11) and (A.3.3), we have $C(f) \geq C(g)$. \square

B Proofs for Section 3

B.1 Blackwell Monotonicity

B.1.1 Proof of Theorem 3

Proof of Theorem 3. Suppose that $f \in \mathcal{E}_m$, $g \in \mathcal{E}_{m'}$ and $f \succeq_B g$, i.e., there exists an $m \times m'$ stochastic matrix $g = fM$. Note that for any $1 \leq k \leq m'$,

$$g^k = \sum_{j=1}^m M_j^k f^j, \quad (\text{B.1})$$

and $\sum_{k=1}^m M_j^k = 1$. Define $\phi : [0, 1] \rightarrow \mathcal{E}_{m+m'}$ as follows:

$$\phi(t) \equiv \begin{bmatrix} (1-t)f & tg \end{bmatrix}.$$

By split invariance and permutation invariance, we have

$$C(f) = C(\phi(0)) \quad \text{and} \quad C(g) = C(\phi(1)).$$

For $t \in (0, 1)$, observe that from $g = fM$,

$$\begin{aligned} \phi'(t) = \begin{bmatrix} -f & g \end{bmatrix} &= \sum_{j=1}^m \sum_{k=1}^{m'} M_j^k \begin{bmatrix} \mathbf{0} & \dots & \overbrace{-f^j}^{j\text{-th column}} & \dots & \overbrace{f^j}^{m+k\text{-th column}} & \dots & \mathbf{0} \end{bmatrix} \\ &= \sum_{j=1}^m \sum_{k=1}^{m'} \frac{M_j^k}{(1-t)} \phi(t)^{j \rightarrow m+k}. \end{aligned}$$

Let $C_{m+m'}$ denote the restriction of C to $\mathcal{E}_{m+m'}$. Then, we have

$$\frac{d}{dt} C(\phi(t)) = \langle \nabla C_{m+m'}(\phi(t)), \phi'(t) \rangle = \sum_{j=1}^m \sum_{k=1}^{m'} \frac{M_j^k}{(1-t)} \langle \nabla C_{m+m'}(\phi(t)), \phi(t)^{j \rightarrow m+k} \rangle \leq 0,$$

where the inequality follows from $M_j^k \geq 0$ and $\langle \nabla C_{m+m'}(\phi(t)), \phi(t)^{j \rightarrow m+k} \rangle \leq 0$ holds by decreasing in signal replacement. Therefore, by FTC, we have

$$C(g) = C(\phi(1)) = C(\phi(0)) + \int_0^1 \frac{d}{dt} C(\phi(t)) dt \leq C(\phi(0)) = C(f).$$

□

B.2 Lehmann Monotonicity

B.2.1 Geometric Characterization of the Lehmann Order

Given an experiment $f \in \mathcal{E}_m^{MLRP}$, we begin by defining a continuous experiment $\tilde{f} : \Omega \rightarrow \Delta([0, m])$ associated with f as in the binary setup. For any $\omega_i \in \Omega$ and $x \in [0, m]$, the cumulative distribution of \tilde{f} is

$$\tilde{F}(x|\omega_i) = F_i^{\lfloor x \rfloor} + (x - \lfloor x \rfloor) \cdot f^{\lfloor x \rfloor + 1}.$$

Note that $\tilde{F}(0|\omega) = 0$ and $\tilde{F}(m|\omega) = 1$ for all $\omega \in \Omega$.

Given this associated continuous experiment, we can formally define $\mathbf{S}_i(f)$ for all $f \in \mathcal{E}^{MLRP}$:

$$\mathbf{S}_i(f) \equiv \left\{ (x, y) \in [0, 1]^2 : \tilde{F}^{-1}(y|\omega_{i+1}) \geq \tilde{F}^{-1}(x|\omega_i) \right\}. \quad (\text{B.2})$$

Lemma B.1 (Jewitt (2007)). Suppose that $f, g \in \mathcal{E}^{MLRP}$. $f \succeq_L g$ is equivalent to $\mathbf{S}_i(g) \subseteq \mathbf{S}_i(f)$ for all $1 \leq i \leq n-1$.

Proof of Lemma B.1. For any $x \in [0, 1]$, the y -coordinate of the lower boundary points for $\mathbf{S}_i(f)$ and $\mathbf{S}_i(g)$ are $\tilde{F}(\tilde{F}^{-1}(x|\omega_i)|\omega_{i+1})$ and $\tilde{G}(\tilde{G}^{-1}(x|\omega_i)|\omega_{i+1})$. Therefore, $\mathbf{S}_i(g) \subseteq \mathbf{S}_i(f)$ is equivalent to

$$\tilde{G}(\tilde{G}^{-1}(x|\omega_i)|\omega_{i+1}) \geq \tilde{F}(\tilde{F}^{-1}(x|\omega_i)|\omega_{i+1})$$

for all $x \in [0, 1]$. Substituting in $x = \tilde{G}(y|\omega_i)$ and applying $\tilde{F}^{-1}(\cdot|\omega_{i+1})$ to both sides yields

$$\tilde{F}^{-1}(\tilde{G}(y|\omega_i)|\omega_{i+1}) \geq \tilde{F}^{-1}(\tilde{G}(y|\omega_i)|\omega_i).$$

Therefore, $\mathbf{S}_i(g) \subseteq \mathbf{S}_i(f)$ for all $1 \leq i \leq n-1$ is equivalent to $\tilde{F}^{-1}(\tilde{G}(y|\omega_i)|\omega_i)$ is increasing in i , i.e., f is Lehmann more informative than g . \square

B.2.2 Proof of Lemma 4

Proof of Lemma 4. Suppose that $f \in \mathcal{E}_m^{MLRP}$ and $f_l^{j+1} f_{l+1}^{j+1} > f_{l+1}^j f_l^{j+1}$ for some l and j . Let $\epsilon' = \frac{f_l^j f_{l+1}^{j+1} - f_l^{j+1} f_{l+1}^j}{f_l^j (f_{l+1}^{j+1} + f_{l+1}^j)} \in (0, 1]$, where we $f_{n+1}^j = 0$ and $f_{n+1}^{j+1} = 1$. Then for all $\epsilon \in [0, \epsilon']$, $f' = f + \epsilon f_{\leq l}^{j \rightarrow j+1} \in \mathcal{E}_m^{MLRP}$. To see this, notice ϵ' is chosen such that the MLRP holds with equality for the i -th and $(i+1)$ -th row for $f + \epsilon' f_{\leq l}^{j \rightarrow j+1}$, and the MLRP for all other pairs of states and signals are preserved under this operation.

We next show $f \succeq_L f + \epsilon f_{\leq l}^{j \rightarrow j+1}$. Note that for any $i > l$, $f_i = f'_i$. Therefore, $\tilde{F}^{-1}(\tilde{F}'(y|\omega_i)|\omega_i) = y$ for all $i > l$ and $y \in [0, m]$. Next, consider the case with $i \leq l$.

1. If $y \leq j-1$ or $y \geq j+1$, $\tilde{F}'(y|\omega_i) = \tilde{F}(y|\omega_i)$, or equivalently, $\tilde{F}^{-1}(\tilde{F}'(y|\omega_i)|\omega_i) = y$, thus, $\tilde{F}^{-1}(\tilde{F}'(y|\omega_i)|\omega_i)$ is constant across i in this case.
2. If $j-1 < y \leq j$, $\tilde{F}'(y|\omega_i) = \tilde{F}(j-1|\omega_i) + (y-j+1)(1-\epsilon)f_i^j$, thus,

$$\tilde{F}^{-1}(\tilde{F}'(y|\omega_i)|\omega_i) = j-1 + (y-j+1)(1-\epsilon) = (j-1)\epsilon + y(1-\epsilon) < y.$$

Therefore, $\tilde{F}^{-1}(\tilde{F}'(y|\omega_i)|\omega_i)$ is increasing in i .

3. If $j < y < j+1$,

$$\tilde{F}'(y|\omega_i) = \tilde{F}(j-1|\omega_i) + (1-\epsilon)f_i^j + (y-j)(\epsilon f_i^j + f_i^{j+1}).$$

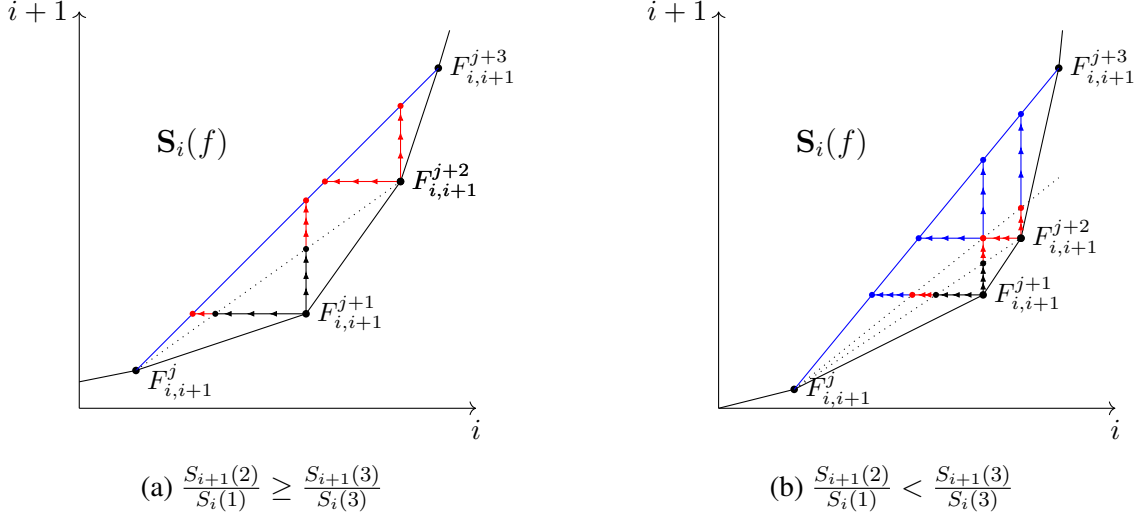


Figure 5: Operations of shrinking $S_i(f)$

Define $L_i = f_i^{j+1}/f_i^j$, then L_i is increasing in i .

$$\tilde{F}^{-1}(\tilde{F}'(y|\omega_i)|\omega_i) = \begin{cases} j - \epsilon + (y - j)(\epsilon + L_i), & \text{if } j < y \leq j + \frac{\epsilon}{\epsilon + L_i}, \\ y - (j + 1 - y)\frac{\epsilon}{L_i}, & \text{if } j + \frac{\epsilon}{\epsilon + L_i} < y < j + 1, \end{cases}$$

or equivalently, $\tilde{F}^{-1}(\tilde{F}'(y|\omega_i)|\omega_i) = \max\{j - \epsilon + (y - j)(\epsilon + L_i), y - (j + 1 - y)\frac{\epsilon}{L_i}\}$. Since both functions are increasing in L_i , $\tilde{F}^{-1}(\tilde{F}'(y|\omega_i)|\omega_i)$ is increasing for $i \leq l$. Additionally, for $j < y \leq j + \frac{\epsilon}{\epsilon + L_i}$, $y > j \geq j - \epsilon + (y - j)(\epsilon + L_i)$, and for $j + \frac{\epsilon}{\epsilon + L_i} < y < j + 1$, $y > y - (j + 1 - y)\frac{\epsilon}{L_i}$. Since $\tilde{F}^{-1}(\tilde{F}'(y|\omega_i)|\omega_i) = y$ for $i > l$, $\tilde{F}^{-1}(\tilde{F}'(y|\omega_i)|\omega_i) = y$ is increasing in i .

We can similarly prove the case with $f' = f + \epsilon f_{\geq l}^{j \rightarrow j-1}$ where $\epsilon'' = \frac{f_{l-1}^{j-1} f_l^j - f_l^{j-1} f_{l-1}^j}{f_l^j (f_{l-1}^{j-1} + f_{l-1}^j)} \in (0, 1]$. \square

B.2.3 Proof of Theorem 4

Lemma B.2. Suppose that $C : \mathcal{E}^{MLRP} \rightarrow \mathbb{R}_+$ is differentiable, split invariant, and decreasing in reverse signal replacement. For any $f \in \mathcal{E}_m^{MLRP}$, $1 \leq i \leq n$ and $1 \leq j+1 < k \leq m$, there exists $\tilde{f} \in \mathcal{E}^{MLRP}$ such that (i) $C(f) \geq C(\tilde{f})$; (ii) $S_{i'}(f) = S_{i'}(\tilde{f})$ for all $i' \neq i$; and (iii) $S_i(\tilde{f})$ is obtained from $S_i(f)$ by removing the region below the line segment connecting the points $F_{i,i+1}^j$ and $F_{i,i+1}^k$.

Proof of Lemma B.2. Define

$$S_i(p) \equiv \sum_{s=1}^p f_i^{j+s}, \quad S_{i+1}(p) \equiv \sum_{s=1}^p f_{i+1}^{j+s}.$$

First, consider the case where $k = j + 2$. Assume that $\frac{S_{i+1}(2)}{S_{i+1}(1)} > \frac{S_{i+1}(1)}{S_i(1)}$; otherwise, $F_{i,i+1}^j$, $F_{i,i+1}^{j+1}$ and $F_{i,i+1}^{j+2}$ already aligned, making this case trivial. Our goal is to construct a series of MLRP experiments, obtained through combinations of reverse signal replacements, that ultimately remove the region below the segment connecting $F_{i,i+1}^j$ and $F_{i,i+1}^{j+2}$, as shown by the black arrows in Figure 5.

Define

$$l_t \equiv (1 - t) \cdot \frac{S_{i+1}(1)}{S_i(1)} + t \cdot \frac{S_{i+1}(2)}{S_i(2)}.$$

Note that the slopes of $\overline{F_{i,i+1}^j F_{i,i+1}^{j+1}}$ and $\overline{F_{i,i+1}^j F_{i,i+1}^{j+2}}$ correspond to l_0 and l_1 , respectively.

Define $\phi : [0, 1] \rightarrow \mathcal{E}_{m+1}$ as follows:

$$\phi(t) \equiv \left[\dots, f^j, \begin{matrix} (1 - a_t)f_{\leq i}^{j+1} \\ + f_{\geq i+1}^{j+1} \end{matrix}, \begin{matrix} a_t f_{\leq i}^{j+1} \\ + b_t f_{\geq i+1}^{j+2} \end{matrix}, \begin{matrix} f_{\leq i}^{j+2} + \\ (1 - b_t) f_{\geq i+1}^{j+2} \end{matrix}, f^{j+3}, \dots \right]$$

where

$$a_t \equiv 1 - \frac{f_{i+1}^{j+1}/f_i^{j+1}}{l_t} \quad \text{and} \quad b_t \equiv \frac{l_t}{f_{i+1}^{j+2}/f_i^{j+1}} - \frac{f_{i+1}^{j+1}}{f_{i+1}^{j+2}}.$$

Let $\Phi(t)$ denote the cumulative distributions of $\phi(t)$: $\Phi(t)^l = \sum_{r=1}^l \phi(t)^r$. Note that $\Phi(t)^j = F^j$ and $\Phi(t)^{j+3} = F^{j+2}$.

The transformation $\phi(t)$ splits signal $j + 1$ in states $i' \leq i$, and splits signal $j + 2$ in states $i' > i$. Thus, for any $i' \neq i$, $\mathbf{S}_{i'}(\phi(t)) = \mathbf{S}_{i'}(f)$. It remains to analyze how $\mathbf{S}_i(\phi(t))$ changes with t .

Intuitively speaking, $\phi(t)$ represents a rotation of $\overline{F_{i,i+1}^j F_{i,i+1}^{j+1}}$ around $F_{i,i+1}^j$, with the slope being l_t . In the process, it splits the signal $j + 1$ into two parts. More formally, observe that $a_0 = b_0 = 0$, thus,

$$\phi(0) = [f^1, \dots, f^{j+1}, \mathbf{0}, f^{j+2}, \dots, f^m] \simeq_B f.$$

$\Phi(t)_{i,i+1}^{j+1}$ is the intersection of the line through $F_{i,i+1}^j$ with slope l_t and the horizontal line through $F_{i,i+1}^{j+1}$, while $\Phi(t)_{i,i+1}^{j+2}$ is its intersection with the vertical line through $F_{i,i+1}^{j+1}$. That is, the likelihood ratios for $\phi(t)_{i,i+1}^{j+1}$ and $\phi(t)_{i,i+1}^{j+2}$ are equal to l_t , while that for $\phi(t)_{i,i+1}^{j+3}$ is greater than or equal to l_t . Therefore, $\phi(t) \in \mathcal{E}_{m+1}^{MLRP}$ for all $t \in [0, 1]$. Last, when $t = 1$, l_t corresponds to the slope of $\overline{F_{i,i+1}^j F_{i,i+1}^{j+2}}$, i.e., $\mathbf{S}_i(\phi(1))$ is obtained from $\mathbf{S}_i(f)$ by removing the region below $\overline{F_{i,i+1}^j F_{i,i+1}^{j+2}}$.

We now show that $C(f) \geq C(\phi(1))$. Observe that

$$\begin{aligned} \dot{a}_t &= \frac{f_{i+1}^{j+1}/f_i^{j+1}}{l_t^2} \cdot \left(\frac{S_{i+1}(2)}{S_i(2)} - \frac{S_{i+1}(1)}{S_i(1)} \right) > 0, \\ \dot{b}_t &= \frac{f_i^{j+1}}{f_{i+1}^{j+2}} \cdot \left(\frac{S_{i+1}(2)}{S_i(2)} - \frac{S_{i+1}(1)}{S_i(1)} \right) > 0. \end{aligned}$$

Then, by decreasing in reverse signal replacement, we have

$$\begin{aligned} \frac{d}{dt}C_{m+1}(\phi(t)) &= \sum_{i'=1}^i \dot{a}_t \left(-\frac{\partial C_{m+1}}{\partial h_{i'}^{j+1}} + \frac{\partial C_{m+1}}{\partial h_{i'}^{j+2}} \right) f_{i'}^{j+1} + \sum_{i'=i+1}^n \dot{b}_t \left(-\frac{\partial C_{m+1}}{\partial h_{i'}^{j+3}} + \frac{\partial C_{m+1}}{\partial h_{i'}^{j+2}} \right) f_{i'}^{j+2} \\ &= \langle \nabla C_{m+1}(\phi(t)), \frac{\dot{a}_t}{1-a_t} \phi(t)_{\leq i}^{j+1 \rightarrow j+2} \rangle + \langle \nabla C_{m+1}(\phi(t)), \frac{\dot{b}_t}{1-b_t} \phi(t)_{\geq i+1}^{j+3 \rightarrow j+2} \rangle \leq 0. \end{aligned} \quad 25$$

Therefore,

$$C(\phi(1)) = C_{m+1}(\phi(0)) + \int_0^1 \frac{d}{dt}C_{m+1}(\phi(t))dt \leq C_{m+1}(\phi(0)) = C(f).$$

Next, we prove the case for $k = j + 3$. Using the argument from the $k = j + 2$ case, we can remove the region below $\overline{F_{i,i+1}^j F_{i,i+1}^{j+2}}$ and let $f' \in \mathcal{E}_{m+1}^{MLRP}$ denote the corresponding experiment. We consider two subcases: (i) $\frac{S_{i+1}(2)}{S_i(1)} \geq \frac{S_{i+1}(3)}{S_i(3)}$; (ii) $\frac{S_{i+1}(2)}{S_i(1)} < \frac{S_{i+1}(3)}{S_i(3)}$.

In the first case, define

$$\tilde{l}_t \equiv (1-t) \cdot \frac{S_{i+1}(2)}{S_i(2)} + t \cdot \frac{S_{i+1}(3)}{S_i(3)}.$$

Also define $\phi : [0, 1] \rightarrow \mathcal{E}_{m+2}^{MLRP}$:

$$\phi^k(t) \equiv \begin{cases} f^k, & \text{if } k \leq j, \\ (1-a_t^1)f_{\leq i}^{j+1} + f_{\geq i+1}^{j+1}, & \text{if } k = j+1, \\ a_t^1 f_{\leq i}^{j+1} + b_t^1 f_{\geq i+1}^{j+2}, & \text{if } k = j+2, \\ (1-a_t^2)f_{\leq i}^{j+2} + (1-b_t^1)f_{\geq i+1}^{j+2}, & \text{if } k = j+3, \\ a_t^2 f_{\leq i}^{j+2} + b_t^2 f_{\geq i+1}^{j+3}, & \text{if } k = j+4, \\ f_{\leq i}^{j+3} + (1-b_t^2)f_{\geq i+1}^{j+3}, & \text{if } k = j+5, \\ f^{k-2}, & \text{if } k = j+6, \end{cases} \quad (\text{B.3})$$

where

$$\begin{aligned} a_t^p &\equiv 1 - \frac{1}{f_i^{j+p}} \cdot \left(\frac{S_{i+1}(p)}{\tilde{l}_t} - S_i(p-1) \right) \quad \text{and} \\ b_t^p &\equiv \frac{1}{f_{i+1}^{j+p+1}} \cdot \left(\tilde{l}_t \cdot S_i(p) - S_{i+1}(p) \right), \end{aligned} \quad (\text{B.4})$$

²⁵Here, we use h to distinguish from the original experiment f .

for all $p \in \{1, 2\}$. This series of experiments is illustrated by the red arrows in Figure 5a. Again, $\phi(t)$ is constructed by rotating the black dashed line through $F_{i,i+1}^j$ with slope \tilde{l}_t . Thus, the MLRP holds.

Note that $\phi(0)$ is Blackwell equivalent to f' and $\phi(1)$ corresponds to the experiment after removing the region below $\overline{F_{i,i+1}^j F_{i,i+1}^{j+3}}$, i.e., $\phi(1) = \tilde{f}$. Using the MLRP of f and $\frac{S_{i+1}(2)}{S_i(1)} \geq \frac{S_{i+1}(3)}{S_i(3)}$, $1 \geq a_t^p \geq 0$ and $1 \geq b_t^p \geq 0$. Additionally, it can be shown that both \dot{a}_t and \dot{b}_t are nonnegative. Then, we apply the same argument as in $k = j + 2$ to establish that $C(f) \geq C(f') \geq C(\phi(1)) = C(\tilde{f})$.

Now we consider the case where $\frac{S_{i+1}(2)}{S_i(1)} < \frac{S_{i+1}(3)}{S_i(3)}$. The operation here takes two steps. We first keep rotating the black dashed line in Figure 5b until it reaches the point where the slope is equal to $\frac{S_{i+1}(2)}{S_i(1)}$, i.e., the intersection of the two red arrows. Formally, define

$$\tilde{l}_t \equiv (1 - t) \cdot \frac{S_{i+1}(2)}{S_i(2)} + t \cdot \frac{S_{i+1}(2)}{S_i(1)}$$

and $\phi : [0, 1] \rightarrow \mathcal{E}_{m+2}^{MLRP}$ using (B.3) and (B.4) with modified \tilde{l} . With the same argument as in the previous cases, it can be shown that $C(f) \geq C(\phi(1))$.

Next, we keep rotating the black dashed line as indicated by the blue arrows in Figure 5b. Formally, define

$$\hat{l}_t \equiv (1 - t) \cdot \frac{S_{i+1}(2)}{S_i(1)} + t \cdot \frac{S_{i+1}(3)}{S_i(3)}.$$

Then, consider a function $\hat{\phi} : [0, 1] \rightarrow \mathcal{E}_{m+2}^{MLRP}$ as follows:

$$\hat{\phi}^k(t) \equiv \begin{cases} f^k, & \text{if } k \leq j, \\ (1 - a_t^1)f_{\leq i}^{j+1} + f_{\geq i+1}^{j+1}, & \text{if } k = j + 1, \\ (a_t^1 - a_t^2)f_{\leq i}^{j+1} + f_{\geq i+1}^{j+2}, & \text{if } k = j + 2, \\ a_t^2 f_{\leq i}^{j+1} + b_t^2 f_{\geq i+1}^{j+3}, & \text{if } k = j + 3, \\ f_{\leq i}^{j+2} + (b_t^3 - b_t^2)f_{\geq i+1}^{j+3}, & \text{if } k = j + 4, \\ f_{\leq i}^{j+3} + (1 - b_t^3)f_{\geq i+1}^{j+3}, & \text{if } k = j + 5, \\ f^{k-2}, & \text{if } k = j + 6, \end{cases} \quad (\text{B.5})$$

where

$$\begin{aligned} a_t^p &\equiv 1 - \frac{S_{i+1}(p)/S_i(1)}{\hat{l}_t} \quad \text{and} \\ b_t^p &\equiv \frac{\hat{l}_t \cdot S_i(p-1) - S_{i+1}(2)}{f_{i+1}^{j+3}}. \end{aligned} \tag{B.6}$$

Since $\hat{\phi}(t)$ is constructed using the line through $F_{i,i+1}^j$ with slope \hat{l}_t , the MLRP holds. Additionally, $\hat{\phi}(0)$ is Blackwell equivalent to $\phi(1)$ and $\hat{\phi}(1)$ corresponds to \tilde{f} —the experiment such that $\mathbf{S}_i(\tilde{f})$ is obtained from $\mathbf{S}_i(f)$ by removing the region below the line segment connecting $\overline{F_{i,i+1}^j F_{i,i+1}^{j+3}}$.

Note that $1 > a_t^1 > a_t^2 \geq 0$, $1 > b_t^3 > b_t^2 \geq 0$ and $\dot{a}_t^p, \dot{b}_t^p > 0$. Again, by decreasing in reverse signal replacement, we have

$$\begin{aligned} \frac{d}{dt} C_{m+2}(\hat{\phi}(t)) &= \sum_{i'=1}^i \dot{a}_t^1 \left(-\frac{\partial C_{m+2}}{\partial h_{i'}^{j+1}} + \frac{\partial C_{m+2}}{\partial h_{i'}^{j+2}} \right) f_{i'}^{j+1} + \sum_{i'=1}^i \dot{a}_t^2 \left(-\frac{\partial C_{m+2}}{\partial h_{i'}^{j+2}} + \frac{\partial C_{m+2}}{\partial h_{i'}^{j+3}} \right) f_{i'}^{j+1} \\ &\quad + \sum_{i'=i+1}^n \dot{b}_t^2 \left(-\frac{\partial C_{m+2}}{\partial h_{i'}^{j+4}} + \frac{\partial C_{m+2}}{\partial h_{i'}^{j+3}} \right) f_{i'}^{j+3} + \sum_{i'=i+1}^n \dot{b}_t^3 \left(-\frac{\partial C_{m+2}}{\partial h_{i'}^{j+5}} + \frac{\partial C_{m+2}}{\partial h_{i'}^{j+4}} \right) f_{i'}^{j+3} \\ &= \langle \nabla C_{m+2}, \frac{\dot{a}_t^1}{1 - a_t^1} \hat{\phi}(t)_{\leq i}^{j+1 \rightarrow j+2} \rangle + \langle \nabla C_{m+2}, \frac{\dot{a}_t^2}{a_t^1 - a_t^2} \hat{\phi}(t)_{\leq i}^{j+2 \rightarrow j+3} \rangle \\ &\quad + \langle \nabla C_{m+2}, \frac{\dot{b}_t^2}{b_t^3 - b_t^2} \phi(t)_{\geq i+1}^{j+4 \rightarrow j+3} \rangle + \langle \nabla C_{m+2}, \frac{\dot{b}_t^3}{1 - b_t^3} \phi(t)_{\geq i+1}^{j+5 \rightarrow j+4} \rangle \leq 0. \end{aligned}$$

Using the same FTC argument, we have

$$C(\tilde{f}) = C(\hat{\phi}(1)) \leq C(\hat{\phi}(0)) = C(\phi(1)) \leq C(f') \leq C(f).$$

Finally, we argue that these arguments apply to all $k > j + 3$. As illustrated, the key is to construct a series of MLRP experiments by rotating the line through $F_{i,i+1}^j$ until it reaches a point like $F_{i,i+1}^{j+1}$ (as in case (i) above), or $(F_i^{j+1}, F_{i+1}^{j+2})$ (as in case (ii) above). Whenever such a point is encountered, we have shown ways to further split that point into two parts, thus allowing us to continue the rotation until it reaches the point $F_{i,i+1}^k$. At the end of this process, we obtain a new experiment $\tilde{f} \in \mathcal{E}^{MLRP}$ such that $\mathbf{S}_i(\tilde{f})$ is obtained from $\mathbf{S}_i(f)$ by removing the region below the line segment connecting $F_{i,i+1}^j$ and $F_{i,i+1}^k$. \square

Proof of Theorem 4. Consider two experiments $f, g \in \mathcal{E}^{MLRP}$ with $f \succeq_L g$, that is, $\mathbf{S}_i(g) \subseteq \mathbf{S}_i(f)$ for all $1 \leq i \leq n - 1$. Our goal is to show that $C(f) \geq C(g)$.

We begin by extending the line segment $\overline{G_{1,2}^0 G_{1,2}^1}$ in $S_1(g)$ until it intersects the boundary of $S_1(f)$. We can construct an experiment f' that is Blackwell equivalent to f and includes this intersection. Specifically, if the intersection lies on $\overline{F_{1,2}^{k-1} F_{1,2}^k}$, we can split the k -th signal of f so that the new k -th cumulative point $(F')_{1,2}^k$ matches the intersection. Then, by Lemma B.2, we can remove the region below $\overline{G_{1,2}^0 G_{1,2}^1}$ from $S_1(f)$, and reduce the cost at the same time.

This procedure can be applied iteratively along each segment $\overline{G_{1,2}^i G_{1,2}^{i+1}}$ for $1 \leq i \leq m-1$. Then, the resulting experiment, \hat{f} , has a strictly lower cost and satisfies $S_1(\hat{f}) = S_1(g)$. Repeating this for S_2, \dots, S_n , we can obtain an experiment \bar{f} that has a lower cost than f and has the same S_i as g for all $1 \leq i \leq n-1$. Then by Lemma B.1, we have $\bar{f} \simeq_L g$.

Finally, we show that \bar{f} and g can be made identical by splitting some of their signals. Suppose for some signal j and s such that $\bar{F}^j = G^s$ (which always exists as one start with $j = s = 0$). Suppose that for some i , $\bar{f}_i^{j+1} > g_i^{s+1}$ (if no such i exists, we move on to $j+1$ and $s+1$). Because $S_i(\bar{f}) = S_i(g)$, it must hold that $\frac{\bar{f}_{i+1}^{j+1}}{\bar{f}_i^{j+1}} = \frac{g_{i+1}^{s+1}}{g_i^{s+1}}$, which further implies $\bar{f}_{i+1}^{j+1} \geq g_i^{s+1}$, where equality holds only if both are zero. By applying the same reasoning inductively, we conclude the same holds for all i . Thus, a splitting of \bar{f}^{j+1} would yield a signal \hat{j} such that $\bar{F}^{\hat{j}} = G^{s+1}$. Repeating this argument inductively for the remaining signals establishes that \bar{f} and g can be made identical. Therefore, splitting invariance implies $C(g) = C(\bar{f}) \leq C(f)$. \square

C Proof for Section 4

C.1 Proof of Proposition 2

Proof of Proposition 2. Sufficiency of (i) is proved in the main text and necessity is proved in Online Appendix OA.4.1. For (ii), take any $f \in \mathcal{E}_m^{MLRP}$ and j , notice f^j and f^{j+1} satisfy the condition for h and h' , thus decreasing in reverse signal replacement holds. \square

C.2 Proof of Proposition 3

Proof of Proposition 3. (i) Consider the function $\psi(h) = \sqrt{h^\top A h}$. It is sublinear when A is symmetric and positive semi-definite. Consider the following A :

$$A = \begin{bmatrix} 10 & 10 & 10 \\ 10 & 20 & 10 \\ 10 & 10 & 20 \end{bmatrix}.$$

It's eigenvalues are approximately (37.3, 10, 2.7), thus is positive definite. Let

$$f = \begin{bmatrix} .78 & .1 & .1 & .02 \\ .2 & .3 & .4 & .1 \\ .05 & .1 & .3 & .55 \end{bmatrix} \in \mathcal{E}_4^{MLRP}.$$

Then we have

$$\langle \nabla C(f), f_{\leq 1}^{2 \rightarrow 3} \rangle = \left(-\frac{\partial \psi(f^2)}{\partial f_1^2} + \frac{\partial \psi(f^3)}{\partial f_1^3} \right) f_1^2 > 0.0008 > 0,$$

i.e., the likelihood separable cost function with this ψ does not satisfy decreasing in reverse signal replacement, and thus not Lehmann monotone.

(ii) Consider $\psi(h) = (\sum_{i=1}^n w_i h_i^p)^{1/p}$, notice that we have

$$\frac{\partial \psi(h)}{\partial h_i} = \frac{w_i h_i^{p-1}}{(\sum_{i=1}^n w_i h_i^p)^{(p-1)/p}}.$$

Then for any $h \leq_{MLRP} h'$ and for any $l \in \{1, \dots, n\}$, we have

$$\begin{aligned} \sum_{i=1}^l \left(\frac{\partial \psi(h')}{\partial h'_i} - \frac{\partial \psi(h)}{\partial h_i} \right) h_i &= \sum_{i=1}^l \left(\frac{w_i (h'_i)^{p-1}}{(\sum_{i=1}^n w_i (h'_i)^p)^{(p-1)/p}} - \frac{w_i h_i^{p-1}}{(\sum_{i=1}^n w_i h_i^p)^{(p-1)/p}} \right) h_i \\ &= \frac{\sum_{i=1}^l w_i (h'_i)^{p-1} h_i}{(\sum_{i=1}^n w_i (h'_i)^p)^{(p-1)/p}} - \frac{\sum_{i=1}^l w_i h_i^p}{(\sum_{i=1}^n w_i h_i^p)^{(p-1)/p}}. \end{aligned}$$

To show that this is negative, it is equivalent to showing that

$$\left(\sum_{i=1}^l w_i (h'_i)^{p-1} h_i \right)^p \left(\sum_{i=1}^n w_i h_i^p \right)^{p-1} \leq \left(\sum_{i=1}^l w_i h_i^p \right)^p \left(\sum_{i=1}^n w_i (h'_i)^p \right)^{p-1}.$$

Then by Hölder's inequality with weights (Theorem 4.7.2. in [Casella and Berger \(2002\)](#)), we have

$$\left(\sum_{i=1}^l w_i (h'_i)^{p-1} h_i \right)^p \leq \left(\sum_{i=1}^l w_i (h'_i)^p \right)^{p-1} \left(\sum_{i=1}^l w_i h_i^p \right).$$

Thus, it suffices to show that

$$\left(\sum_{i=1}^l w_i (h'_i)^p \right) \cdot \left(\sum_{i=1}^n w_i h_i^p \right) \leq \left(\sum_{i=1}^l w_i h_i^p \right) \cdot \left(\sum_{i=1}^n w_i (h'_i)^p \right).$$

By subtracting the common term $\left(\sum_{i=1}^l w_i(h'_i)^p\right) \left(\sum_{i=1}^l w_i(h_i)^p\right)$, we can rewrite the above inequality as

$$\left(\sum_{i=1}^l w_i(h'_i)^p\right) \left(\sum_{i=l+1}^n w_i h_i^p\right) \leq \left(\sum_{i=1}^l w_i h_i^p\right) \left(\sum_{i=l+1}^n w_i(h'_i)^p\right).$$

This inequality holds because $h \leq_{MLRP} h'$ implies that for any $k \leq l$ and $k' \geq l+1$, we have

$$h'_k h_{k'} \leq h_k h'_{k'}.$$

The other condition can be shown similarly. □

C.3 Proof of Proposition 4

Proof of Proposition 4. Fix any full-support prior μ and notice that if $f \in \mathcal{E}^{MLRP}$ then $q^j \leq_{FOSD} q^{j+1}$ for all j (Milgrom, 1981). Thus, it suffices to show that the condition in Proposition 4 implies decreasing in reverse signal replacement.

For a posterior separable cost C_μ given some H , we have

$$\frac{\partial C_\mu(f)}{\partial f_i^j} = -\mu_i H(q^j) - \tau^j \sum_{i'} \frac{\partial H(q^j)}{\partial q_{i'}^j} \cdot \frac{\partial q_{i'}^j}{\partial f_i^j}.$$

Then we can further derive:

$$\frac{\partial q_i^j}{\partial f_i^j} = \frac{\mu_i}{\tau^j} - \mu_i \frac{\mu_i f_i^j}{(\tau^j)^2}, \quad \frac{\partial q_i^j}{\partial f_s^j} = -\mu_s \frac{\mu_i f_i^j}{(\tau^j)^2},$$

and that

$$\begin{aligned} \tau^j f_i^j \sum_{i'} \frac{\partial H(q^j)}{\partial q_{i'}^j} \cdot \frac{\partial q_{i'}^j}{\partial f_i^j} &= \tau^j f_i^j \left(\frac{\partial H(q^j)}{\partial q_i^j} \left(\frac{\mu_i}{\tau^j} - \mu_i \frac{\mu_i f_i^j}{(\tau^j)^2} \right) + \sum_{i' \neq i} \frac{\partial H(q^j)}{\partial q_{i'}^j} \left(-\mu_i \frac{\mu_{i'} f_{i'}^j}{(\tau^j)^2} \right) \right) \\ &= \mu_i f_i^j \frac{\partial H(q^j)}{\partial q_i^j} - \mu_i f_i^j \sum_{i'} \frac{\partial H(q^j)}{\partial q_{i'}^j} q_{i'}^j. \end{aligned}$$

Thus, for any j and l , substituting in the above expressions yields

$$\begin{aligned} \sum_{i=1}^l \left(\frac{\partial C_\mu(f)}{\partial f_i^{j+1}} - \frac{\partial C_\mu(f)}{\partial f_i^j} \right) f_i^j &= \sum_{i=1}^l \left(H(q^j) - H(q^{j+1}) - \sum_{i'} \frac{\partial H(q^{j+1})}{\partial q_{i'}^{j+1}} (q_{i'}^j - q_{i'}^{j+1}) \right) \mu_i f_i^j \\ &\quad + \tau^j \sum_{i=1}^l \left(\frac{\partial H(q^j)}{\partial q_i^j} - \frac{\partial H(q^{j+1})}{\partial q_i^{j+1}} \right) q_i^j. \end{aligned}$$

Note that the first term is negative from concavity of H and the second term is negative from the condition in the proposition. Thus, (11) holds for any j and l . A similar argument shows (12) also holds for all j and l . \square

C.4 Proof of Proposition 5

Proof of Proposition 5. For the entropy cost, we have

$$H(q^j) = - \sum_{i=1}^n q_i^j \log q_i^j, \text{ and } \frac{\partial H(q^j)}{\partial q_i^j} = -\log q_i^j - 1.$$

Then for any l and $q^j \leq_{FOSD} q^{j+1}$, we have

$$\begin{aligned} \sum_{i=1}^l q_i^j \left(\frac{\partial H(q^j)}{\partial q_i^j} - \frac{\partial H(q^{j+1})}{\partial q_i^{j+1}} \right) &= \sum_{i=1}^l q_i^j (\log q_i^{j+1} - \log q_i^j) \\ &= \sum_{i=1}^l q_i^j (\log q_i^{j+1} - \log q_i^j) + \left(\sum_{i=l+1}^n q_i^j \right) \left(\log \sum_{i=l+1}^n q_i^{j+1} - \log \sum_{i=l+1}^n q_i^j \right) \\ &\quad + \left(\sum_{i=l+1}^n q_i^j \right) \left(\log \sum_{i=l+1}^n q_i^j - \log \sum_{i=l+1}^n q_i^{j+1} \right). \end{aligned}$$

Observe that the sum of the first and second term is negative because

$$\sum_{i=1}^l q_i^j \log \frac{q_i^{j+1}}{q_i^j} + \left(\sum_{i=l+1}^n q_i^j \right) \log \frac{\sum_{i=l+1}^n q_i^{j+1}}{\sum_{i=l+1}^n q_i^j} \leq \log \sum_{i=1}^n q_i^{j+1} = 0$$

by $\sum_{i=1}^n q_i^j = \sum_{i=1}^n q_i^{j+1} = 1$ and the concavity of log function. The third term is also negative from $\sum_{i=l+1}^n q_i^{j+1} \geq \sum_{i=l+1}^n q_i^j$ as $q^j \leq_{FOSD} q^{j+1}$. The other condition can be shown similarly. \square

References

- Asseyer, Andreas (2025) “Information orders in screening problems,” *Journal of Economic Theory*, 225, 105993.
- Baker, Christian (2023) “Economies and diseconomies of scale in the cost of information,” *Working Paper*.

- Bertschinger, Nils and Johannes Rauh (2014) “The Blackwell relation defines no lattice,” in *2014 IEEE International Symposium on Information Theory*, 2479–2483, [10.1109/ISIT.2014.6875280](#).
- Blackwell, David (1951) “Comparison of experiments,” in *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, 2, 93–103, University of California Press.
- (1953) “Equivalent comparisons of experiments,” *The annals of mathematical statistics*, 265–272.
- Bloedel, Alexander W and Weijie Zhong (2024) “The cost of optimally acquired information,” *Working Paper*.
- Bobkova, Nina (2024) “Information choice in auctions,” *American Economic Review*, 114 (7), 1883–1915.
- Bond, Philip (2023) “Ordering information content using the quantile function.”
- Bordoli, Davide and Ryota Iijima (2025) “Convex Cost of Information via Statistical Divergence,” *Working Paper*.
- Cao, Lei, Darian McLaren, and Sarah Plosker (2022) “Centrosymmetric stochastic matrices,” *Linear and Multilinear Algebra*, 70 (3), 449–464, [10.1080/03081087.2020.1733461](#).
- Caplin, Andrew and Mark Dean (2015) “Revealed preference, rational inattention, and costly information acquisition,” *American Economic Review*, 105 (7), 2183–2203.
- Casella, George and Roger L. Berger (2002) *Statistical Inference*, Pacific Grove, CA: Duxbury Thomson Learning, 2nd edition.
- Chambers, Christopher P, Ce Liu, and John Rehbeck (2020) “Costly information acquisition,” *Journal of Economic Theory*, 186, 104979.
- Chi, Chang Koo (2015) “The Value of Information and Dispersion.”
- Denti, Tommaso (2022) “Posterior separable cost of information,” *American Economic Review*, 112 (10), 3215–3259.
- Denti, Tommaso, Massimo Marinacci, and Aldo Rustichini (2022a) “Experimental cost of information,” *American Economic Review*, 112 (9), 3106–3123, [10.1257/aer.20210879](#).
- (2022b) “The experimental order on random posteriors.”

- Dymond, Michael, Beata Randrianantoanina, and Huaqiang Xu (2017) “On Interval Based Generalizations of Absolute Continuity for Functions on \mathbb{R}^n ,” *Real Analysis Exchange*, 42 (1), 49 – 78.
- Fosgerau, Mogens, Emerson Melo, Andre De Palma, and Matthew Shum (2020) “Discrete choice and rational inattention: A general equivalence result,” *International economic review*, 61 (4), 1569–1589.
- Gentzkow, Matthew and Emir Kamenica (2014) “Costly persuasion,” *American Economic Review: Papers & Proceedings*, 104 (5), 457–462.
- Hébert, Benjamin and Michael Woodford (2021) “Neighborhood-Based Information Costs,” *American Economic Review*, 111 (10), 3225–55, [10.1257/aer.20200154](https://doi.org/10.1257/aer.20200154).
- Jewitt, Ian (2007) “Information order in decision and agency problems,” *Nuffield College*.
- Kim, Yonggyun (2023) “Comparing Information in General Monotone Decision Problems,” *Journal of Economic Theory*, 211, 105679, [10.1016/j.jet.2023.105679](https://doi.org/10.1016/j.jet.2023.105679).
- Lehmann, EL (1988) “Comparing Location Experiments,” *The Annals of Statistics*, 521–533.
- Li, Jian and Junjie Zhou (2020) “Information order in monotone decision problems under uncertainty,” *Journal of Economic Theory*, 187, 105012.
- Lipnowski, Elliot and Doron Ravid (2023) “Predicting choice from information costs,” *arXiv preprint arXiv:2205.10434*.
- Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt (2023) “Rational inattention: A review,” *Journal of Economic Literature*, 61 (1), 226–273.
- Matějka, Filip and Alisdair McKay (2012) “Simple market equilibria with rationally inattentive consumers,” *American Economic Review*, 102 (3), 24–29.
- (2015) “Rational inattention to discrete choices: A new foundation for the multinomial logit model,” *American Economic Review*, 105 (1), 272–298.
- Mensch, Jeffrey (2018) “Cardinal representations of information,” *Available at SSRN 3148954*.
- Milgrom, Paul R. (1981) “Good News and Bad News: Representation Theorems and Applications,” *The Bell Journal of Economics*, 12 (2), 380–391, <http://www.jstor.org/stable/3003562>.

- Milgrom, Paul and Chris Shannon (1994) “Monotone Comparative Statics,” *Econometrica*, 62 (1), 157–180, <http://www.jstor.org/stable/2951479>.
- Morris, Stephen and Philipp Strack (2019) “The wald problem and the relation of sequential sampling and ex-ante information costs,” *Available at SSRN 2991567*.
- de Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek (2017) “Rationally inattentive preferences and hidden information costs,” *Theoretical Economics*, 12 (2), 621–654.
- de Oliveira, Henrique, Yuhta Ishii, and Xiao Lin (2024) “Robust Aggregation of Correlated Information.”
- Persico, Nicola (2000) “Information acquisition in auctions,” *Econometrica*, 68 (1), 135–148.
- Polyanskiy, Yury and Yihong Wu (2022) “Information theory: From coding to learning.”
- Pomatto, Luciano, Philipp Strack, and Omer Tamuz (2023) “The Cost of Information: The Case of Constant Marginal Costs,” *American Economic Review*, 113 (5), 1360–1393, [10.1257/aer.20190185](https://doi.org/10.1257/aer.20190185).
- Quah, John K-H and Bruno Strulovici (2009) “Comparative statics, informativeness, and the interval dominance order,” *Econometrica*, 77 (6), 1949–1992.
- Ravid, Doron (2020) “Ultimatum Bargaining with Rational Inattention,” *American Economic Review*, 110 (9), 2948–63, [10.1257/aer.20170620](https://doi.org/10.1257/aer.20170620).
- Ravid, Doron, Anne-Katrin Roesler, and Balázs Szentes (2022) “Learning before trading: on the inefficiency of ignoring free information,” *Journal of Political Economy*, 130 (2), 346–387.
- Sims, Christopher A (2003) “Implications of rational inattention,” *Journal of monetary Economics*, 50 (3), 665–690.
- Van Nieuwerburgh, Stijn and Laura Veldkamp (2010) “Information acquisition and underdiversification,” *The Review of Economic Studies*, 77 (2), 779–805.
- Yang, Ming (2015) “Coordination with flexible information acquisition,” *Journal of Economic Theory*, 158, 721–738.

Online Appendix for “On the Monotonicity of Information Costs”

Xiaoyu Cheng Yonggyun Kim

OA.1 Examples for the Binary-Binary Case

In this section, we provide examples of information costs in the binary-binary case. We begin by examining examples using Proposition 1.

Example OA.1. Consider two information cost functions defined over $\hat{\mathcal{E}}_2$:

$$C_1(f_1, f_2) \equiv \left(\frac{f_2}{f_1} - 1 \right)^2 \left(1 - \frac{1 - f_2}{1 - f_1} \right), \quad C_2(f_1, f_2) \equiv \frac{f_2(1 - f_2)}{f_1(1 - f_1)} - 1.$$

By using the definition of \tilde{C} , we have

$$\tilde{C}_1(\alpha, \beta) \equiv (\alpha - 1)^2 \left(1 - \frac{1}{\beta} \right), \quad \tilde{C}_2(\alpha, \beta) \equiv \frac{\alpha}{\beta} - 1.$$

Then, from $\alpha, \beta \geq 1$, \tilde{C}_1 is increasing in both α and β , whereas \tilde{C}_2 is not increasing in β . Therefore, it follows that C_1 is Blackwell monotone, but C_2 is not.

Next, imagine that in Figure 1a, we draw a curve passing through the point A to illustrate a potential isocost curve, indicating the same information cost of a smooth cost function. When C is differentiable at f , (1) and (2) can be rewritten as follows:

$$\underbrace{\frac{f_2}{f_1}}_{\text{the slope of } \overline{AB}} \geq \underbrace{-\frac{\partial C / \partial f_1}{\partial C / \partial f_2}}_{\text{the slope of the isocost curve}} \geq \underbrace{\frac{1 - f_2}{1 - f_1}}_{\text{the slope of } \overline{AD}}.^{26} \quad (\text{OA.1.1})$$

The slope of the isocost curve can be considered as the *marginal rate of information transformation (MRIT)*. Thus, this inequality says that the MRIT of a Blackwell monotone cost function should fall between the two likelihood ratios provided by the experiment.

We first provide examples of verifying Blackwell monotonicity using the marginal rate of information transformation described in (OA.1.1).

²⁶With some algebra, we can show that $f_2 \geq f_1$ and (1) and (2) imply $\frac{\partial C}{\partial f_2} \geq 0 \geq \frac{\partial C}{\partial f_1}$.

Example OA.2. Consider two information cost functions defined over $\hat{\mathcal{E}}_2$:

$$C_3(f_1, f_2) \equiv (f_2 - f_1)^2, \quad C_4(f_1, f_2) \equiv f_2 - 2f_1.$$

Notice that

$$MRIT_3 \equiv -\frac{\partial C_3 / \partial f_1}{\partial C_3 / \partial f_2} = 1, \quad MRIT_4 \equiv -\frac{\partial C_4 / \partial f_1}{\partial C_4 / \partial f_2} = 2.$$

Then, for C_3 , (OA.1.1) holds all $(f_1, f_2) \in \hat{\mathcal{E}}_2$, but not so for C_4 , e.g., when $f_1 = .5$ and $f_2 = .6$. Therefore, we can conclude that C_3 is Blackwell monotone, but C_4 is not.

OA.2 Non-Convexity of the MLRP set

In this section, we show that the set of MLRP experiments, \mathcal{E}^{MLRP} , is not convex by providing a pair of experiments such that both experiments satisfy the MLRP, while a convex combination of them does not.

Let $n = 3$ and define $f, g \in \mathcal{E}_3^{MLRP}$:

$$f = \begin{bmatrix} 0.04 & 0.36 & 0.60 \\ 0.02 & 0.18 & 0.80 \\ 0.02 & 0.18 & 0.80 \end{bmatrix} \quad \text{and} \quad g = \begin{bmatrix} 0.60 & 0.04 & 0.36 \\ 0.40 & 0.06 & 0.54 \\ 0.40 & 0.06 & 0.54 \end{bmatrix}.$$

Observe that f and g satisfy the MLRP.

Let h be the average of f and g :

$$h = \frac{1}{2}f + \frac{1}{2}g = \begin{bmatrix} 0.32 & 0.20 & 0.48 \\ 0.21 & 0.12 & 0.67 \\ 0.21 & 0.12 & 0.67 \end{bmatrix}$$

The combined experiment h fails to satisfy the MLRP. To see this, consider the likelihood ratio for ω_2 vs. ω_1 .

$$\frac{.21}{.32} > \frac{.12}{.2} < \frac{.67}{.48}.$$

Therefore, the set of MLRP experiments is not convex.

OA.3 Blackwell Monotonicity with Limited Signals

OA.3.1 Blackwell Monotonicity under Quasiconvexity

In this section, we explore Blackwell monotonicity in the case where experiments are limited to a maximum number of signals. Specifically, we consider a cost function $C : \mathcal{E}_m \rightarrow \mathbb{R}_+$ for a given m , where experiments with fewer than m signals are embedded in \mathcal{E}_m by adding zero columns.

Under this restriction, split invariance is no longer relevant, as splitting an experiment would place it outside of the domain. However, permutation invariance and decreasing in signal replacement remain necessary conditions for Blackwell monotonicity.

The key step in establishing sufficiency for Blackwell monotonicity in Theorem 1 and 3 is to construct a decreasing path connecting any $f \succeq_B g$. For \mathcal{E}_2 , we were able to construct such a path within \mathcal{E}_2 itself, whereas for \mathcal{E} , we relied on split invariance and introduced additional signals to construct the path.

However, when the number of signals is limited to some $m > 2$, such a path within the space \mathcal{E}_m does not always exist, as shown by the following proposition.

Proposition OA.1. *Suppose that $n = m = 3$ and let*

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \succeq_B g = \begin{bmatrix} 4/5 & 1/5 & 0 \\ 0 & 4/5 & 1/5 \\ 1/5 & 0 & 4/5 \end{bmatrix} \in \mathcal{E}_3.$$

If $f \in \mathcal{E}_3$ is Blackwell more informative than g , then f is a permutation of I_3 or g .

Proposition OA.1 suggests that there is no continuous path in \mathcal{E}_3 connecting I_3 and g along which Blackwell informativeness decreases. Because if such a path existed, there would have to be an experiment, other than permutations of I_3 or g , that is more informative than g but less informative than I_3 , which is impossible according to the proposition.

We overcome the issue by imposing quasiconvexity on the cost function. Let $C \in \mathcal{C}_m$ be defined as *quasiconvex* if for any $f, g \in \mathcal{E}_m$ and $\lambda \in [0, 1]$,

$$C(\lambda f + (1 - \lambda)g) \leq \max\{C(f), C(g)\}.$$

In other words, a mixture of two experiments cannot be more costly than both of them.²⁷

²⁷To make sense of this property, notice a mixture of two experiments, $\lambda f + (1 - \lambda)g$, can be replicated by running experiment f with probability λ , and experiment g with probability $1 - \lambda$, then reporting the realized signal without indicating which experiment was conducted. Thus, if the cost of $\lambda f + (1 - \lambda)g$ is higher than $\max\{C(f), C(g)\}$, one could make an arbitrage from this replication.

To see how quasiconvexity is able to address the difficulty raised in Proposition OA.1, observe that $g = \frac{4}{5}I_3 + \frac{1}{5}I_3P$ for some permutation matrix P . When C is quasiconvex, $C(g) \leq \max\{C(I_3), C(I_3P)\} = C(I_3)$. Our next result shows that, under quasiconvexity, permutation invariance and decreasing in signal invariance serve as necessary and sufficient conditions for Blackwell monotonicity over \mathcal{E}_m .

Theorem OA.1. *Suppose $C \in \mathcal{C}_m$ is absolutely continuous and quasiconvex. Then, C is Blackwell monotone if and only if C is permutation invariant and decreasing in signal replacement.*

Proof of Theorem OA.1 Necessity is established in the main text. For sufficiency, we begin by considering the set of all $m \times m$ stochastic matrix, denoted by \mathcal{M}_m . Then, for any $f, g \in \mathcal{E}_m$, $f \succeq_B g$ if and only if there exists $M \in \mathcal{M}_m$ such that $g = fM$.

Notice \mathcal{M}_m is a convex subset of $\mathbb{R}_+^{m \times m}$ and its extreme points are given by the matrices with exactly one non-zero entry in each row (see e.g., Cao et al. (2022)). Let $\mathbf{ext}(\mathcal{M}_m) = \{E_1, \dots, E_m\}$ denote the set of all extreme points of \mathcal{M}_m . Then, since any $M \in \mathcal{M}_m$ can be written as a convex combination of elements in $\mathbf{ext}(\mathcal{M}_m)$, the quasiconvexity of C implies that

$$C(g) \leq \max\{C(fE) : E \in \mathbf{ext}(\mathcal{M}_m)\},$$

Therefore, it suffices to show that $C(f) \geq C(fE)$ for all $E \in \mathbf{ext}(\mathcal{M}_m)$.

For any $k \leq m$, let $\mathbf{ext}_k(\mathcal{M}_m)$ denote extreme-point matrices with rank k . Note that $\mathbf{ext}_m(\mathcal{M}_m)$ corresponds to the set of all permutation matrices. Thus, permutation invariance implies that $C(f) = C(fE)$ for all $E \in \mathbf{ext}_m(\mathcal{M}_m)$. The following lemma shows that decreasing in signal replacement implies that for any $1 \leq k \leq m$ and $E \in \mathbf{ext}_{k-1}(\mathcal{M}_m)$, there exists $E' \in \mathbf{ext}_k(\mathcal{M}_m)$ such that fE' is at least as costly as fE . By repeatedly applying this, we can see that for any $E \in \mathbf{ext}(\mathcal{M}_m)$, there exists $\tilde{E} \in \mathbf{ext}_m(\mathcal{M}_m)$ such that $C(fE) \leq C(f\tilde{E}) = C(f)$, which completes the proof.

Lemma OA.3.1. *Suppose $C \in \mathcal{C}_m$ is absolutely continuous and decreasing in signal replacement. Then for any $1 \leq k \leq m$ and $E \in \mathbf{ext}_{k-1}(\mathcal{M}_m)$, there exists $E' \in \mathbf{ext}_k(\mathcal{M}_m)$ such that for all $\lambda \in [0, 1]$,*

$$fE' \succeq_B (1 - \lambda)fE' + \lambda fE \succeq_B fE. \quad (\text{OA.3.1})$$

And it further implies $C(fE') \geq C(fE)$.

Proof of Lemma OA.3.1. Since E is not a full rank matrix, there exists a column e^i such that at least two entries are equal to 1. Let $e_j^i = e_{j'}^i = 1$. Additionally, there are $n - k + 1$ columns such that all the entries are equal to zero. Let one of such columns be $e^{i'}$. Let E' be a matrix such

that $e_{j'}^{i'} = 1$, $e_{j'}^i = 0$ and all other entries are same as E . Note that E' has exactly $n - k$ empty columns, i.e., $E' \in \mathbf{ext}_k(\mathcal{M}_m)$.

Let B denote a matrix an $m \times m$ matrix such that $b_{i'}^{i'} = -1$, $b_{i'}^i = 1$, and all other entries are equal to zero. Observe that when $l \neq i, i'$, l -th column of $fB_{i'}^{i'}$ is equal to $\mathbf{0}$. Additionally, i -th column of $fB_{i'}^{i'}$ is $-f^i$ and i' -th column of $fB_{i'}^{i'}$ is f^i . Thus, $fB_{i'}^{i'}$ is equal to $f^{i \rightarrow i'}$.

Note that when I_m is the identity matrix of size m , $I_m + \lambda B$ is a stochastic matrix for any $\lambda \in [0, 1]$. Observe that $B^2 = -B$ and $(I_m + \lambda B) \cdot (I_m + B) = I_m + B$. Additionally, $E'(I_m + B) = E$ and $E'(I_m + \lambda B) = (1 - \lambda)E' + \lambda E$. Therefore, we have

$$\begin{aligned} (1 - \lambda)fE' + \lambda fE &= fE'(I_m + \lambda B), \\ fE &= fE'(I_m + B) = fE'(I_m + \lambda B) \cdot (I_m + B). \end{aligned}$$

Since $I_m + \lambda B$ and $I_m + B$ are stochastic matrices, (OA.3.1) holds.

Next, from $fB = f^{i' \rightarrow i}$ and decreasing in signal replacement, we have that for all $\lambda \in [0, 1]$,

$$\begin{aligned} &D^+(C((1 - \lambda)fE' + \lambda fE), fE - ((1 - \lambda)fE' + \lambda fE)) \\ &= D^+(C((1 - \lambda)fE' + \lambda fE), ((1 - \lambda)fE' + \lambda fE)B) \leq 0. \end{aligned} \quad (\text{OA.3.2})$$

Finally, we show for such E and E' ,

$$C(fE') \geq C(fE).$$

For $\lambda \in [0, 1]$, define the function $\varphi(\lambda) = C((1 - \lambda)fE' + \lambda fE)$. By absolute continuity, φ is differentiable almost everywhere on $[0, 1]$ and satisfy

$$\varphi'(\lambda) = D^+C((1 - \lambda)fE' + \lambda fE; fE - fE').$$

Then, the FTC implies that

$$\begin{aligned} C(fE) - C(fE') &= \varphi(1) - \varphi(0) = \int_0^1 \varphi'(\lambda) d\lambda = \int_0^1 D^+((1 - \lambda)fE' + \lambda fE; fE - fE') d\lambda \\ &= \int_0^1 \frac{1}{1 - \lambda} D^+(C((1 - \lambda)fE' + \lambda fE), fE - ((1 - \lambda)fE' + \lambda fE)) d\lambda \leq 0, \end{aligned}$$

where the second last equality uses positive homogeneity of $D^+C(f; \cdot)$ and the last inequality follows from that (OA.3.2) holds for all $\lambda \in [0, 1]$. \square

OA.3.2 Remarks

Remark 1. Theorem OA.1 characterizes necessary and sufficient conditions for Blackwell monotonicity under the presence of quasiconvexity. This raises the question of whether quasiconvexity is necessary for Blackwell monotonicity. The following example illustrates a cost function over binary experiments that is Blackwell monotone but not quasiconvex.

Example OA.3. Suppose $n = m = 2$. Denote any experiment $f \in \mathcal{E}_2$ by $f = [f_1, f_2]^\top$. As before, we restrict attention to the set $\hat{\mathcal{E}}_2 = \{(f_1, f_2) : 0 \leq f_1 \leq f_2 \leq 1\}$. Consider $C : \hat{\mathcal{E}}_2 \rightarrow \mathbb{R}_+$ defined by

$$C(f) = \min \left\{ \frac{f_2}{f_1}, \frac{1-f_1}{1-f_2} \right\}.$$

By using (OA.1.1), we can easily see that $f \succeq_B g$ implies $C(f) \geq C(g)$, i.e., C is Blackwell monotone.

Consider $f = [0, 1/2]^\top$ and $g = [1/2, 1]^\top$ with costs $C(f) = C(g) = 2$. For the one-half mixture of them, given by $h = [1/4, 3/4]^\top$, the cost is $C(h) = 3 > C(f) = C(g)$. Hence, this cost function is not quasiconvex.

Remark 2. Quasiconvexity is not needed in establishing Blackwell monotonicity over binary experiments. However, when quasiconvexity is imposed in this case, it is almost sufficient for Blackwell monotonicity.

Recall that any binary experiment can be represented by $f = [f_1, \dots, f_n]^\top \in [0, 1]^n$, and $\mathbf{0}$ and $\mathbf{1}$ are completely uninformative experiments. Let C be *non-null* if for any $f \in [0, 1]^n$, $C(f) \geq C(\mathbf{1}) = C(\mathbf{0})$.

Proposition OA.2. *If $C \in \mathcal{E}_2$ is quasiconvex, permutation invariant, and non-null, then C is Blackwell monotone.*

Remark 3. We provide a weaker version of quasiconvexity, which can also serve as a necessary condition for Blackwell monotonicity.

Definition OA.1. $C \in \mathcal{C}_m$ is **garbling-quasiconvex** if for all $f \in \mathcal{E}$, any finite collection of its garblings, $\{g_1, \dots, g_n\}$, and $\lambda_0, \dots, \lambda_n \in [0, 1]$ with $\sum_{i=0}^n \lambda_i = 1$,

$$C \left(\lambda_0 f + \sum_{i=1}^n \lambda_i g_i \right) \leq \max \{C(f), C(g_1), \dots, C(g_n)\}.$$

Proposition OA.3. *Suppose $C \in \mathcal{C}_m$ is absolutely continuous. Then, C is Blackwell monotone if and only if C is permutation invariant, garbling-quasiconvex and decreasing in signal replacement.*

The necessity of garbling-quasiconvexity follows from the fact that $f \succeq_B \lambda_0 f + \sum_i \lambda_i g_i$ for all such configurations. For sufficiency, the proof proceeds almost identically to that of Theorem OA.1, with additional steps required to show that garbling-quasiconvexity, together with continuity, is sufficient to establish the final step.

OA.3.3 Proofs

OA.3.3.1 Proof of Proposition OA.2

Proof of Proposition OA.2. By Lemma 1, $f \succeq_B g$ if and only if $g = af + b(1 - f)$ for $(a, b) \in [0, 1]^2$. If $a \geq b$, $g = (1 - a) \cdot \mathbf{0} + (a - b) \cdot f + b \cdot \mathbf{1}$; and if $a < b$, $g = (1 - b) \cdot \mathbf{0} + (b - a) \cdot (1 - f) + a \cdot \mathbf{1}$. From quasiconvexity and non-nullness, we have $C(f) \geq C(g)$ or $C(1 - f) \geq C(g)$. Then, by permutation invariance, $C(f) = C(1 - f)$, thus, $C(f) \geq C(g)$. \square

OA.3.3.2 Proof of Theorem OA.3

Proof of Theorem OA.3. The necessity is already addressed in the main text.

For sufficiency, take any $f \succeq_B g$. By the same argument as in the proof of Theorem OA.1, all extreme points of $S_B(f)$ are in $S_C(f)$. By convexity of $S_B(f)$, g can be written as a convex combination of these extreme points, denoted by $g = \sum_{i=1}^n \lambda_i g_i$. Moreover, for all $\epsilon > 0$, $g_\epsilon \equiv \epsilon f + (1 - \epsilon)g \in S_B(f)$. By garbling-quasiconvexity,

$$C(g_\epsilon) \leq \max\{C(f), C(g_1), \dots, C(g_n)\} \leq C(f),$$

for all $\epsilon > 0$, where the last inequality follows because g_i 's are extreme points of $S_B(f)$. Taking the limit as $\epsilon \rightarrow 0$, by continuity, we have $C(f) \geq C(g)$, and thus C is Blackwell monotone. \square

OA.4 Omitted Proofs

OA.4.1 Proof of Proposition 2 (i)

Proof of Necessity for Proposition 2 (i). Suppose C is likelihood separable. First, given any $\hat{f} \in [0, 1]^n$. For any $k \in \mathbb{N}$, consider the following experiments,

$$f = \begin{bmatrix} \hat{f} & 0 & \dots & 0 & 1 - \hat{f} \end{bmatrix} \in \mathcal{E}_{k+1},$$

and

$$g = \begin{bmatrix} \frac{1}{k}\hat{f} & \cdots & \frac{1}{k}\hat{f} & 1 - \hat{f} \end{bmatrix} \in \mathcal{E}_{k+1}.$$

Observe that

$$f \begin{bmatrix} 1/k & \cdots & 1/k & 0 \\ \vdots & \ddots & \vdots & 0 \\ 1/k & \cdots & 1/k & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} = g \quad \text{and} \quad g \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} = f,$$

that is, $f \succeq_B g \succeq_B f$. Thus, Blackwell monotonicity implies that $C(f) = C(g)$. Then it holds that $\psi\left(\frac{1}{k}\hat{f}\right) = \frac{1}{k}\psi(\hat{f})$.

Next, for any $\ell \in \mathbb{N}$ such that $\ell\hat{f} \in [0, 1]^n$. Consider the following experiments,

$$f = \begin{bmatrix} \ell\hat{f} & \mathbf{1} - \ell\hat{f} \end{bmatrix} \in \mathcal{E}_2,$$

and

$$g = \begin{bmatrix} \hat{f} & \cdots & \hat{f} & \mathbf{1} - \ell\hat{f} \end{bmatrix} \in \mathcal{E}_{\ell+1}.$$

By the same argument, Blackwell monotonicity implies that $C(f) = C(g)$, thus, $\psi(\ell\hat{f}) = \ell\psi(\hat{f})$. Together it implies that, for all $\hat{f} \in [0, 1]^n$, for all $z \in \mathbb{Q}$ such that $z\hat{f} \in [0, 1]^n$, $\psi(z\hat{f}) = z\psi(\hat{f})$. By the density of \mathbb{Q} in \mathbb{R} and the continuity of $\psi(\cdot)$, we have positive homogeneity of ψ over $[0, 1]^n$.

Next, we show subadditivity, i.e., for any $\hat{f}, \hat{g} \in [0, 1]^n$ such that $\hat{f} + \hat{g} \in [0, 1]^n$, then

$$\psi(\hat{f} + \hat{g}) \leq \psi(\hat{f}) + \psi(\hat{g}).$$

Consider the following experiments,

$$f = \begin{bmatrix} \hat{f} & \hat{g} & 1 - \hat{f} - \hat{g} \end{bmatrix} \in \mathcal{E}_3,$$

and

$$g = \begin{bmatrix} \hat{f} + \hat{g} & 1 - \hat{f} - \hat{g} \end{bmatrix} \in \mathcal{E}_2.$$

As g is obtained by merging the first two signals in f , we have $f \succeq_B g$. Thus, Blackwell monotonicity implies that $C(f) \geq C(g)$, and thus sublinearity of ψ holds. \square