Blackwell-Monotone Information Costs*

Xiaoyu Cheng[†] Yonggyun Kim[‡]

July 10, 2024

Abstract

A Blackwell-monotone information cost function assigns higher costs to Blackwell more informative experiments. This paper provides simple necessary and sufficient conditions for a cost function to be Blackwell monotone over finite experiments. The key condition involves a system of linear differential inequalities. By using this characterization, we show that when a cost function is additively separable, it is Blackwell monotone if and only if it is the sum of sublinear functions. This identifies a wide range of practical information cost functions. Finally, we apply our results to bargaining and persuasion problems with costly information, broadening and strengthening earlier findings.

JEL Classification: C78, D81, D82, D83

Keywords: Statistical Experiments, Rational Inattention, Additively Separable Costs, Bargaining, Costly Persuasion

*We thank Teddy Kim, R.Vijay Krishna, Fei Li, Luciano Pomatto, Todd Sarver, Bruno Strulovici, Curtis Taylor, João Thereze, Udayan Vaidya, Kun Zhang, and seminar participants at KAEA-VSS Micro Seminar, FSU, UNC and Duke 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

A central agenda in contemporary economic theory is the integration of costly information across various fields and applications. This raises the question of which information cost functions *should* or *could* be used. One principle that is widely accepted as the minimum requirement for sensible information costs is *Blackwell monotonicity*: a statistical experiment is more costly than another if it is more informative according to the classical information order by Blackwell (1951, 1953). Despite its importance, the conditions for information costs to satisfy Blackwell monotonicity remain underexplored.

In this paper, we characterize simple necessary and sufficient conditions for a cost function to be Blackwell monotone over finite experiments.[1] This provides a tractable method to verify Blackwell monotonicity when an arbitrary information cost function is considered. Using this characterization, we provide a functional representation of all information cost functions that satisfy: (i) Blackwell monotonicity; (ii) being the sum of functions of probabilities regarding each signal; and (iii) assigning zero costs to uninformative experiments.

We first provide a rationale for imposing Blackwell monotonicity on information costs using an example. Consider consumers who seek to acquire information about their health status (e.g., regarding COVID-19). The health status can either be positive $(+)$ or negative $(-)$. There are two types of tests available in a competitive market, denoted as A and B, with their testing probabilities as follows:

Test A =

		signal n	signal p
state	$-$	80%	20%
state	$+$	20%	80%

,

Test B =

		signal n	signal p
state	$-$	60%	40%
state	$+$	15%	85%

.

This means that test A (B) reports a negative result ('n') with 80% (60%) when the true state is negative, and reports a positive result ('p') with a probability of 80% (85%) when the true state is positive.

Suppose that the prices of tests A and B are \$10 and \$12, respectively. Given these prices and the provided probabilities, a producer can make an arbitrage by replicating test B using test A, which is less costly. After performing test A, the producer can implement

[1] A finite experiment stands for a map from finite states to probability distributions over finite signals.

the following process: (i) if the result is 'p', report 'p'; (ii) if the result is 'n', flip a coin twice, report 'p' only if both flips are heads; otherwise report 'n.' This procedure produces exactly the same probabilities as test B. Notice that the replication procedure generating the arbitrage is a *garbling* of test A, and this procedure is feasible if and only if test A is more informative than test B (Blackwell, 1951, 1953). Hence, Blackwell monotonicity ensures that there is no such 'information arbitrage' via the garbling of a more informative experiment.

The above garbling process also motivates the key condition in our characterization of Blackwell monotonicity. Specifically, we focus on the type of garbling that replaces a signal with another signal for some probability ϵ, while keeping the generation of other signals unchanged. As this worsens Blackwell informativeness, any Blackwell-monotone cost function should be decreasing in the direction corresponding to this change. By sending ϵ to zero, we obtain a first-order condition that is necessary for Blackwell monotonicity: the directional derivative of the cost function in such garbling directions should be non-positive provided it exists.

Our main results establish the sufficiency (as well as the necessity) of this first-order condition for Blackwell monotonicity under absolute continuity and permutation invariance (relabeling signals results in the same cost). Theorem 1 shows that the first-order condition is necessary and sufficient for Blackwell monotonicity of information cost functions defined over experiments with two signals. Theorem 2 extends this result to more than two signals when the cost function is quasiconvex, meaning that any mixture of two experiments costs less than at least one of the individual experiments.

As a crucial step in establishing sufficiency, we provide a novel geometric characterization of the set of experiments that are less Blackwell informative than a given experiment (Lemma 1 and 2). In particular, these lemmas imply that the garbling directions specified in the first-order condition are exactly the extreme directions of worsening informativeness. In other words, any direction of worsening informativeness can be represented by a positive linear combination of these extreme directions. As a result, when the cost function is quasiconvex (so its sublevel sets are convex), if it decreases in all extreme directions, then it will decrease in any direction of worsening informativeness, thereby ensuring Blackwell monotonicity.

Next, we apply our results to the case where costs are additively separable across signals, referred to as likelihood separable (LS). This means that costs can be represented as a sum of component functions of likelihoods for each signal. Using our main results, we show that any LS cost is Blackwell monotone if and only if the component function is sublinear (Theorem 3). This result characterizes a class of *grounded sublinear likelihood separable* (GSLS) information costs—introduced by Denti et al. (2022b)—as the class of information cost functions that satisfy Blackwell monotonicity, likelihood separability, and groundedness (ensuring zero cost for any uninformative experiments). This class of information cost functions encompasses well-known cost functions such as entropy costs (Sims, 2003; Matějka and McKay, 2015) and log-likelihood ratio (LLR) costs (Pomatto et al., 2023), as well as novel ones such as the norm costs and absolute-linear costs (Section 4.2). Furthermore, we show that an information cost function is GSLS if and only if it has a posterior separable representation under any full support prior (Proposition 4).

Equipped with these new tools, we revisit two economic applications involving costly information. In these applications, assumptions other than Blackwell monotonicity are often imposed on the information cost. Our characterizations allow us to study these problems in a more general framework without the necessity of these additional assumptions and thus help to strengthen and broaden the existing insights.

We first consider the ultimatum bargaining model of Chatterjee et al. (2024) where the buyer can acquire costly information about the unknown value of the seller's object before accepting an offer. To model costly information, the authors exogenously restrict the buyer's feasible set of experiments and define an information cost function over this restricted set. Using our characterization, we are able to extend their cost function to a Blackwell-monotone cost function over all experiments. This allows us to examine their results in a more general setting where the buyer's ability to acquire information is unrestricted. We show that while the exogenous restriction is crucial for the main result, its main insight remains true in the general setting when considering a different Blackwell-monotone cost function (Proposition 6, 7).

As another application, we consider the costly persuasion problem proposed by Gentzkow and Kamenica (2014). They extend their celebrated Bayesian persuasion model (Kamenica and Gentzkow, 2011) by assuming that it is costly for the sender to commit to a persuasion policy, which is in the form of statistical experiments. To apply the concavification technique, they focus on cases where the information cost function is uniformly posterior separable, and the literature follows this tradition. We propose another way of solving the costly persuasion problem (without using concavification) that can be applied to any Blackwell-monotone information costs. As an illustration, we provide a solution for the classic prosecutor-judge example with a specific non-posterior separable cost.

1.1 Related Literature

The classical information order by Blackwell (1951, 1953) has recently regained prominence in light of the rapid expansion of the information design and costly information acquisition literatures.[2] Therefore, it has become essential to understand how Blackwell's information order can be integrated into the cost of information. When defining information costs, there are mainly two approaches: posterior-based costs (defined over distributions of posteriors); and experiment-based costs (defined over statistical experiments). We refer readers to the introduction of Denti et al. (2022a) for a formal discussion.

Ever since Sims (2003) introduced entropy cost to economics, posterior-based information costs have been widely applied. Blackwell monotonicity of such costs (with a concave measure of uncertainty, e.g., entropy) is implied by one of Blackwell's sufficient conditions related to the convex order.[3] These costs have been applied to diverse problems including bargaining (Ravid, 2020) and dynamic information acquisition (Zhong, 2022), among many others. However, some recent papers find that the results derived under posterior-based costs can change qualitatively when experiment-based costs are employed instead (Denti et al., 2022a; Ramos-Mercado, 2023). Furthermore, as Mensch (2018) and Denti et al. (2022a) point out, because experiment-based costs are prior independent, they may be more compatible with applications where priors are endogenously determined in equilibrium.

All these observations highlight the importance of determining which experiment-based costs ought to be utilized, with Blackwell monotonicity as a consideration. As a pioneering work in this direction, Pomatto et al. (2023) show that an information cost takes the form of log-likelihood ratio (LLR) costs if and only if (along with technical assumptions) it satisfies: (i) Blackwell monotonicity; (ii) the dilution axiom—the cost of generating an experiment with a half probability is half of generating it with probability one; and (iii) the independence axiom—the cost of generating independent experiments equals the sum of their individual costs.

We contribute to the literature on experiment-based costs by (i) deriving the necessary and sufficient condition for Blackwell monotonicity alone; and (ii) characterizing the class of GSLS costs that include a wide range of information costs including LLR costs. Ad-

[2] For example, Mu et al. (2021) study Blackwell dominance in large samples, and Khan et al. (2024) investigate Blackwell's theorem with infinite states.

[3] See, for example, the discussion of Assumption 1 in Gentzkow and Kamenica (2014) and Lemma 6 of Denti et al. (2022b).

ditionally, we show that GSLS costs satisfy the dilution axiom but not the independence axiom. We also present an example where the independence axiom might not be applicable (Section 4.3).

Finally, there is a strand of literature in decision theory that focuses on the rationalization of revealed choice data with information costs. There, Blackwell monotonicity often appears as a central property in the preference representations. See, for example, Caplin and Dean (2015); de Oliveira et al. (2017); Chambers et al. (2020); Bloedel and Zhong (2021); Caplin et al. (2022); and Denti (2022). Our paper differs methodologically from these as our primitives are experiments instead of choice data.

1.2 Outline

Section 2 introduces the formal framework. Section 3 provides the main characterization results of Blackwell-monotone information costs. Section 4 applies the characterization to study the class of likelihood separable costs. Section 5 presents two applications. Section 6 provides additional discussion about quasiconvexity and Section 7 concludes. All omitted proofs can be found in the Appendix and Online Appendix.

2 Preliminaries

Let $\Omega = \{\omega_1, \cdots, \omega_n\}$ be a finite set of states. Fix a finite set of signals $\mathcal{S} = \{s_1, \cdots, s_m\}$, a *statistical experiment* $f : \Omega \to \Delta(\mathcal{S})$ is represented by the $n \times m$ matrix

$$f = \begin{bmatrix} f_1^1 & \cdots & f_1^m \\ \vdots & \ddots & \vdots \\ f_n^1 & \cdots & f_n^m \end{bmatrix},$$

where $f_i^j = f(s_j|\omega_i)$ is the probability of generating signal s_j in state ω_i. Let $f^j = [f_1^j, \cdots, f_n^j]^\intercal \in \mathbb{R}_+^n$ denote the j-th column vector of f for $j = 1, \cdots m$. Using this notation, we can rewrite

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

Notice that $\sum_{j=1}^m f^j = \mathbf{1}$ where $\mathbf{1} = [1, \cdots, 1]^\intercal$. Let $\mathcal{E}_m$ denote the set of all experiments generating at most m signals. Notice it is without loss to use the same signal set $\mathcal{S}$ for all such experiments and thus we can embed $\mathcal{E}_m$ into $\mathbb{R}^{n\times m}$ equipped with the Euclidean topology. Let $\mathcal{E} = \bigcup_{m<\infty} \mathcal{E}_m$ denote the *class* of all (finite) experiments.

For any $f, g \in \mathcal{E}_m$ and $\lambda \in [0,1]$, we define the *(state-wise) mixture* of experiments by the convex combination of their matrix representations, i.e., $\lambda f + (1-\lambda)g \in \mathcal{E}_m$. Namely, in each state ω, the distribution of signals under $\lambda f + (1-\lambda)g$ is the λ-mixture of those under f and g.

Blackwell Informativeness An experiment f is *Blackwell more informative* than another experiment g, denoted by $f \succeq_B g$, if there exists a *stochastic matrix* M (i.e., $M_{ij} \geq 0$ and $\sum_j M_{ij} = 1$ for all i) such that $g = fM$. This matrix M is also called a *garbling matrix*. f and g are said to be *equally informative* and denoted by $f \simeq_B g$ if both $f \succeq_B g$ and $g \succeq_B f$ hold.

When both f and g are in $\mathcal{E}_m$, any potential garbling matrix is an $m \times m$ square stochastic matrix. Let $\mathcal{M}_m$ denote the set of all such stochastic matrices. 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 $\mathbf{ext}(\mathcal{M}_m) = \{E_1, \cdots, E_{m^m}\}$ denote the set of all extreme points of $\mathcal{M}_m$. For any $k \leq m$, let $\mathbf{ext}_k(\mathcal{M}_m)$ denote those extreme-point matrices with rank k.

A *permutation matrix* P is a stochastic matrix with exactly one non-zero entry in each row and each column. Let $\mathcal{P}_m$ be the set of all $m \times m$ permutation matrices and observe that $\mathbf{ext}_m(\mathcal{M}_m) = \mathcal{P}_m$. Since when $P \in \mathcal{P}$, P^{-1} is also a permutation matrix, we have $f \succeq_B fP \succeq_B fPP^{-1} = f$, namely $f \simeq_B fP$. Intuitively, permuting an experiment involves merely relabeling signals, so it should remain equally informative after permutation.

Information Costs We define an *information cost function* as $C : \mathcal{E}_m \to \mathbb{R}_+$, i.e., defined over the set of experiments with a fixed number of possible signals. Let $\mathcal{C}_m$ denote the set of all such functions. Under this formulation, each $C \in \mathcal{C}_m$ is a mapping on Euclidean space which facilitates analysis. All applicable results can be carried over to information cost functions defined over $\mathcal{E}$ by considering their restriction to $\mathcal{E}_m$ for all m.

For each $C \in \mathcal{C}_m$, let $D^+C(f;h)$ denote its (one-sided) directional derivative at $f \in \mathcal{E}_m$ in the direction of $h \in \mathbb{R}^{n\times m}$, if the following limit exists:

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

When C is differentiable at f, let $\nabla C(f) \in \mathbb{R}^{n\times m}$ denote its gradient and we have

$$D^+C(f;h) = \langle \nabla C(f), h\rangle.$$

In addition, it is convenient to define $\nabla^j C(f) \in \mathbb{R}^n$ as the j-th column vector of $\nabla C(f)$, i.e., $\nabla^j C(f) = [\partial C(f)/\partial f_1^j, \cdots, \partial C(f)/\partial f_n^j]^\intercal$. Thus, we can similarly write

$$\nabla C(f) = [\nabla^1 C(f), \cdots, \nabla^m C(f)].$$

Because for each $f \in \mathcal{E}_m$, $\sum_{j=1}^m f^j = \mathbf{1}$, it is without loss to let $\nabla^m C(f) = 0$ if needed.

Functional Assumptions The weakest possible continuity assumption required for our results is absolute continuity. Say that $C \in \mathcal{C}_m$ is *absolutely continuous* if for all $f, g \in \mathcal{E}_m$ and $t \in [0,1]$, the function $\varphi(t) = C(f + t(g-f))$ is absolutely continuous in t over $[0,1]$.[4] Equivalently, it says that $\varphi(\cdot)$ is differentiable almost everywhere and the Fundamental Theorem of Calculus (FTC) holds, i.e.,

$$C(g) - C(f) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt = \int_0^1 D^+C(tg + (1-t)f; g-f)dt. \quad (1)$$

Notice that a sufficient condition for absolute continuity is Lipschitz continuity over $\mathcal{E}_m$.

Finally, say that $C \in \mathcal{C}_m$ is *permutation invariant* if for any $f \in \mathcal{E}_m$, $C(f) = C(fP)$ for all $P \in \mathcal{P}_m$. This property is naturally required under the Blackwell information order, as $f \simeq_B fP$.

3 Blackwell-Monotone Information Costs

In this section, we provide our main characterization results of information cost functions that are consistent with the Blackwell information order.

Definition 1. *An information cost function $C \in \mathcal{C}_m$ is **Blackwell monotone** if for all $f, g \in \mathcal{E}_m$, $C(f) \geq C(g)$ whenever $f \succeq_B g$.*

For any $C \in \mathcal{C}_m$, let $S_C(f) = \{g \in \mathcal{E}_m : C(f) \geq C(g)\}$ denote its sublevel set at $f \in \mathcal{E}_m$. In addition, let $S_B(f) = \{g \in \mathcal{E}_m : f \succeq_B g\}$ denote the sublevel set under the

[4] There are multiple generalizations of absolute continuity from $\mathbb{R}$ to $\mathbb{R}^n$ emphasizing different aspects. See Dymond et al. (2017) for a reference. We adopt the generalization which requires the restriction of C to any line segment is absolutely continuous, corresponding to their definition of 0-absolute continuity.

Blackwell information order. By definition, Blackwell monotonicity is equivalent to

$$S_C(f) \supseteq S_B(f) \text{ for all } f \in \mathcal{E}_m.$$[5]

Our characterization of Blackwell monotonicity relies on a novel geometric characterization of the set $S_B(f)$ and then links it to the set $S_C(f)$. To better illustrate the key idea, we start by characterizing Blackwell monotonicity over $\mathcal{E}_2$, the *binary experiments*.

3.1 Blackwell Monotonicity over Binary Experiments

For any $f \in \mathcal{E}_2$, since $f^1 + f^2 = \mathbf{1}$, f is uniquely identified by the vector f^1. For simplicity, we use the column vector to denote a binary experiment as $f = [f_1, \cdots, f_n]^\intercal \in [0,1]^n$. Similarly, for any $C \in \mathcal{C}_2$, we let $C : [0,1]^n \to \mathbb{R}_+$. Our next result shows that, under the aforementioned functional assumptions, Blackwell monotonicity can be characterized by a pair of linear differential inequalities.

Theorem 1. *Suppose $C \in \mathcal{C}_2$ is absolutely continuous and permutation invariant. C is Blackwell monotone if and only if for all $f \in \mathcal{E}_2$,*

$$D^+C(f; -f) \leq 0 \text{ and } D^+C(f; \mathbf{1} - f) \leq 0, \text{whenever exists.} \tag{2}$$

When C is differentiable at f, (2) simplifies to

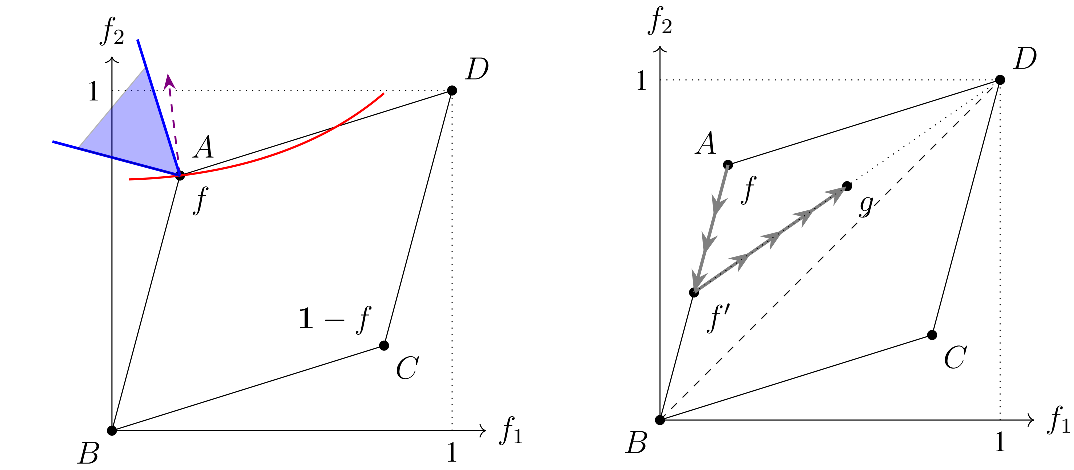
$$\langle \nabla C(f), f \rangle \geq 0 \geq \langle \nabla C(f), \mathbf{1} - f \rangle. \tag{3}$$

In the following, we present our geometric characterization of the set $S_B(f)$ for binary experiments and use it to illustrate the proof sketch of Theorem 1. After the sketch, we show that the geometric characterization also leads to a further characterization of Blackwell monotonicity in the case of binary states which does not require any continuity assumption.

3.1.1 Proof Sketch

Parallelogram Hull For any $f, g \in \mathcal{E}_2$ with $f \succeq_B g$, there exists $M \in \mathcal{M}_2$ such that $[g, \mathbf{1} - g] = [f, \mathbf{1} - f]M$. For a stochastic matrix $M \in \mathcal{M}_2$, it can be written as, for some

[5]From this, we can easily see that any monotone transformation of a Blackwell-monotone information cost is also Blackwell monotone.



(a) A parallelogram hull and its polar cone

(b) A decreasing path from f to g

Figure 1: A Graphical Illustration with Binary States

$(a, b) \in [0, 1]^2$, that

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

which implies

$$g = af + b(\mathbf{1} - f).$$

This key observation leads to the following geometric characterization of the set $S_B(f)$ for binary experiments.

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

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

In other words, $S_B(f) = PARL(f, \mathbf{1} - f)$.

A parallelogram hull in the case of binary states, i.e., $n = 2$, is depicted by the parallelogram (ABCD) in Figure 1a. We next show this geometric characterization implies the necessity of (2).

Necessity The parallelogram hull highlights two extreme directions of decreasing informativeness: $-f$ and $\mathbf{1} - f$. In Figure 1a, they are shown by $\overrightarrow{AB}$ and $\overrightarrow{AD}$, respectively.

For any $\epsilon > 0$, notice that moving from f in the direction of $-f$ results in the following experiment:

$$[f - \epsilon f, \mathbf{1} - f + \epsilon f].$$

From the perspective of garbling, this experiment is derived from f by applying the following garbling: whenever s_1 is generated, there is a probability ϵ of generating s_2 instead, while the generation of s_2 is unchanged. Conversely, $\mathbf{1} - f$ represents the opposite type of garbling: s_2 is occasionally replaced by s_1.

If C is Blackwell monotone, then $C(f - \epsilon f) \leq C(f)$ for all $\epsilon \geq 0$. As a result, when $D^+C(f; -f)$ exists, we must have

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

By the same argument, we also have $D^+C(f; \mathbf{1} - f) \leq 0$.

Remark 1. When C is differentiable at f, $D^+C(f; h)$ is linear in h and equals $\langle \nabla C(f), h \rangle$. In this case, notice (3) is equivalent to

$$\langle \nabla C(f), g - f \rangle \leq 0, \quad \forall g \in \text{PARL}(f, \mathbf{1} - f) = S_B(f).$$

Geometrically, this says that $\nabla C(f)$ lies in the *polar cone* of $S_B(f)$ at f, depicted in Figure 1a by the blue cone. In other words, when C is differentiable at f, Blackwell monotonicity imposes a constraint on the feasible directions of its gradient at f.

For a graphical demonstration, 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. As the gradient of such a function is tangent to its isocost curve, the gradient of this cost function (the purple arrow) lies outside the polar cone of $S_B(f)$ and thus violates Blackwell monotonicity. This is confirmed by noticing that the cost increases in the direction of $\overrightarrow{AD}$ near A.

Sufficiency Because (2) is only a local property, sufficiency requires additional regularity conditions on the cost function. Permutation invariance is necessarily needed and absolute continuity ensures the Fundamental Theorem of Calculus (FTC) applies. Consider any experiment g lying inside the parallelogram $ABCD$, i.e., $f \succeq_B g$. If g is above the line BD, as illustrated in Figure 1b, we can find a two-segment path from f to g, which moves only in the extreme directions required by (2): moving from f in the direction of $-f$ to reach f' and then moving from f' in the direction of $\mathbf{1} - f'$ to reach g. Thus, applying (2) implies the directional derivatives are non-positive along this path, and applying (1) leads to $C(g) \leq C(f)$.

If g lies below the line BD, its permutation, namely gP, lies above the line BD and has the same cost as g, following from permutation invariance. Then, the same argument applies to gP implying $C(g) = C(gP) \leq C(f)$. Lemma A.1 formally shows this argument.

3.1.2 A Further Characterization with Binary States

Consider the binary-binary case ($n = m = 2$) and restrict attention to the following set of experiments:[6]

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

For any $f, g \in \hat{\mathcal{E}}_2$, from the parallelogram in Figure 1b, we have $f \succeq_B g$ 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

$$\alpha \equiv \frac{f_2}{f_1} \geq \frac{g_2}{g_1} \equiv \alpha' \quad \text{and} \quad \beta \equiv \frac{1 - f_1}{1 - f_2} \geq \frac{1 - g_1}{1 - g_2} \equiv \beta'.^{7} \tag{4}$$

Note that α is the likelihood ratio for generating signal s_1 and $1/\beta$ is the likelihood ratio for signal s_2. Thus, (4) implies that if both α and β increase, Blackwell informativeness increases. Also note that α and β can take any value in $[1, +\infty]$ and

$$f_1 = \frac{\beta - 1}{\alpha\beta - 1} \quad \text{and} \quad f_2 = \frac{\alpha(\beta - 1)}{\alpha\beta - 1}.^{8}$$

Define $\tilde{C} : [1, \infty]^2 \to \mathbb{R}_+$ as follows:

$$\tilde{C}(\alpha, \beta) \equiv C\left(\tfrac{\beta-1}{\alpha\beta-1}, \tfrac{\alpha(\beta-1)}{\alpha\beta-1}\right). \tag{5}$$

Thus, we obtain the following characterization of Blackwell monotonicity which does not require any continuity assumption.

[6] By applying permutation invariance, the cost for the other half-piece will be properly defined.

[7] Let $x/0 = +\infty$ for all $x > 0$ and $0/0 = 1$.

[8] If $\alpha = +\infty$, then $f_1 = 0$ and $f_2 = \frac{\beta-1}{\beta}$. If $\alpha = \beta = 1$, let $f_1 = f_2 = 0$.

Proposition 1. *For any $C : \hat{\mathcal{E}}_2 \to \mathbb{R}_+$, C is Blackwell monotone if and only if $\tilde{C}$ as defined in (5) is increasing in α and β.*

Proposition 1 provides a simple way to check Blackwell monotonicity for binary-binary experiments. We illustrate this with two examples.

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

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

By using (5), we have

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

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.

The likelihood ratios (α and $1/\beta$) can also be used to provide an interpretation of Blackwell monotonicity in this case. Notice when C is differentiable and $\frac{\partial C}{\partial f_2} \neq 0$, (3) can be rewritten as

$$\alpha = \underbrace{\frac{f_2}{f_1}}_{\text{the slope of } \overline{AB}} \geq \underbrace{-\frac{\partial C/\partial f_1}{\partial C/\partial f_2}}_{\substack{\text{the slope of} \\ \text{the isocost curve}}} \geq \underbrace{\frac{1-f_2}{1-f_1}}_{\text{the slope of } \overline{AD}} = \frac{1}{\beta}.^{9} \tag{6}$$

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.

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

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

Notice that

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

[9] With some algebra, we can show that $f_2 \geq f_1$ and (3) imply $\frac{\partial C}{\partial f_2} \geq 0 \geq \frac{\partial C}{\partial f_1}$.

Then, for C_3, (6) 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.

3.2 Blackwell Monotonicity over Finite Experiments

In this section, we characterize Blackwell monotonicity for $\mathcal{E}_m$ with arbitrary m. For necessity, much of the intuition from the binary case carries over, but the geometric characterization of the set $S_B(f)$ becomes more intricate.

Lemma 2. *For any $f, g \in \mathcal{E}_m$, $f \succeq_B g$ if and only if*

$$g - f \in \left\{ \sum_{j=1}^{m} \lambda_j h_j : \lambda_j \in [0,1], h_j \in co(\{f^j(k) : k \neq j\}), \forall j \right\}, \tag{7}$$

where $co(\cdot)$ denote the convex hull and $f^j(k) \in \mathbb{R}^{n\times m}$ is the matrix with f^j in the k-th column, $-f^j$ in the j-th column and zeros elsewhere, i.e.,

$$f^j(k) \equiv \begin{bmatrix} 0 & \cdots & \underbrace{-f^j}_{j\text{-th column}} & \cdots & 0 & \cdots & \underbrace{f^j}_{k\text{-th column}} & \cdots & 0 \end{bmatrix}.$$

In other words,

$$S_B(f) = \left\{ f + \sum_{j=1}^{m} \lambda_j h_j : \lambda_j \in [0,1], h_j \in co(\{f^j(k) : k \neq j\}), \forall j \right\}.$$

Lemma 2 implies that, for each $f \succeq_B g$, the direction $g - f$ is a positive linear combination of $f^j(k)$'s, an analogue to the implication of parallelogram hull in the binary case. In other words, these $f^j(k)$'s identify the extreme directions of decreasing informativeness. Observe that for any $\epsilon \in [0,1]$, $f + \epsilon f^j(k)$ belongs to $\mathcal{E}_m$ and is obtained by applying the type of garbling that, with a probability ϵ, s_j is replaced by s_k, i.e., merging the signal s_j into s_k. Consequently, Blackwell monotonicity requires the cost function to be decreasing along these extreme directions, implying similar necessary conditions.

When establishing sufficiency for binary experiments, a key step is to construct a decreasing path connecting any $f \succeq_B g$. With more than two signals, however, such a path within the space $\mathcal{E}_m$ does not always exist, as shown by the following proposition.

Proposition 2. *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 2 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 next show this issue can be addressed 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.[10]

To see how quasiconvexity is able to address the difficulty raised in Proposition 2, observe that $g = \frac{4}{5} I_3 + \frac{1}{5} I_3 P$ for some permutation matrix P. When C is quasiconvex, $C(g) \leq \max\{C(I_3), C(I_3 P)\} = C(I_3)$. Our next result shows that, with quasiconvexity, the same type of necessary and sufficient condition as in Theorem 1 can be established for all experiments.

Theorem 2. *Suppose $C \in \mathcal{C}_m$ is absolutely continuous, permutation invariant, and quasiconvex. C is Blackwell monotone if and only if for all $f \in \mathcal{E}_m$,*

$$D^+ C(f; f^j(k)) \leq 0, \quad \forall j \neq k, \text{ whenever exists.} \tag{8}$$

When C is differentiable at f, (8) simplifies to

$$\langle \nabla^k C(f) - \nabla^j C(f), f^j \rangle = \sum_{i=1}^{n} \left(\frac{\partial C}{\partial f_i^k} - \frac{\partial C}{\partial f_i^j} \right) f_i^j \leq 0, \quad \forall j \neq k. \tag{9}$$

[10] By using a replication argument, we can also provide a no-arbitrage justification for imposing quasiconvexity on information cost functions. A mixture of two experiments, $\lambda f + (1 - \lambda) g$, can be replicated by running experiment f with a probability λ, and experiment g with a 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 can always make an arbitrage from this replication.

First of all, notice that when $m = 2$, the condition reduces to those in Theorem 1 by letting $f^1 = f$, $f^2 = \mathbf{1} - f$ and $C(f) \equiv C(f, \mathbf{1} - f)$.

In the proof of Theorem 2, the key step is to show that any extreme point of $S_B(f)$ is either a permutation of f or can be reached from f by a sequence of segments in the directions of $f^j(k)$'s required by (8) (Lemma A.4). For an intuition, these directions of merging signals are precisely those from an extreme point $E \in \mathbf{ext}_k(\mathcal{M}_m)$ to another extreme point with a lower rank in $\mathbf{ext}_{k-1}(\mathcal{M}_m)$. Once this is established, similar to the proof of sufficiency in Theorem 1, applying FTC along these segments implies that all extreme points of $S_B(f)$ are in $S_C(f)$. Then, quasiconvexity ensures that the entire set $S_B(f)$ is in $S_C(f)$, and thus C is Blackwell monotone.

Remark 2. Given Theorem 2, checking Blackwell monotonicity over non-binary experiments requires one more step: verifying quasiconvexity of C. It is worth noting that when C is twice differentiable, quasiconvexity can be checked by verifying the determinants of every order of its bordered Hessian matrices are non-positive, similar to checking convexity. See Arrow and Enthoven (1961) and also Proposition 3.4.4 in Osborne (2016) for references.

Remark 3. Quasiconvexity is not necessary for Blackwell monotonicity (see an example in Section 6.1). In Section 6.3, we identify a weaker (though less standard) quasiconvexity condition that, along with the first-order condition, serves as necessary and sufficient condition for Blackwell monotonicity. Nevertheless, we present the main theorem with the standard quasiconvexity for two reasons: (i) it is easier to verify, especially given Remark 2; (ii) it is a standard and convenient condition in cost minimization problems, ensuring that local optimality implies global optimality.

4 Likelihood Separable Costs

The results in Section 3 provide valuable and tractable means to verify whether an arbitrary information cost function adheres to Blackwell monotonicity, essentially by checking the differential inequalities. This becomes even simpler when the cost function is additively separable across signals, as the differential inequalities specify directions where only two signals are changing. In this section, we characterize Blackwell monotonicity of such cost functions.

4.1 Characterization

We begin by formally defining likelihood separable costs.

Definition 2. *An information cost function $C : \mathcal{E} \to \mathbb{R}_+$ is* ***likelihood separable*** *if there exist a constant a and an absolutely continuous function $\psi : [0,1]^n \to \mathbb{R}$ with $\psi(\mathbf{0}) = 0$ such that, for all m and $f \in \mathcal{E}_m$,*

$$C(f) = a + \sum_{j=1}^{m} \psi(f^j). \tag{10}$$

Let $\mathcal{C}^{LS}$ denote the class of likelihood separable cost functions.

Notice that the definition of likelihood separability already incorporates permutation invariance (same component function ψ for all signals) and $\psi(\mathbf{0}) = 0$, which is naturally required by Blackwell monotonicity.[11] These facilitate defining the likelihood separable costs over all experiments in $\mathcal{E}$ as its restriction to $\mathcal{E}_m$ is obvious. In this case, Blackwell monotonicity over $\mathcal{E}$ is equivalent to Blackwell monotonicity over every $\mathcal{E}_m$.

The following theorem characterizes Blackwell monotonicity for the class of likelihood separable cost functions.

Theorem 3. *For any $C \in \mathcal{C}^{LS}$, C is Blackwell monotone if and only if ψ is sublinear, i.e.,*

(i) positively homogeneous: $\psi(\gamma \cdot h) = \gamma \cdot \psi(h)$ for all $h \in [0,1]^n$ and $\gamma > 0$ with $\gamma \cdot h \in [0,1]^n$; and

(ii) subadditive: $\psi(h + l) \leq \psi(h) + \psi(l)$ for all $h, l \in [0,1]^n$ with $h + l \in [0,1]^n$.

We next present the proof of sublinearity implying Blackwell monotonicity, which highlights the application of sufficient conditions in Theorem 2. The proof of necessity is in Appendix B.[12]

[11] For $\psi(\mathbf{0}) = 0$, adding columns of zeros does not change the informativeness of an experiment and thus should not change its cost.

[12] Another way to prove the sufficiency direction is to apply Lemma 6 of Denti et al. (2022b). Our proof is different in that we utilize the garbling characterization and the first-order approach in Theorem 2, whereas they use the convex order approach. Furthermore, the necessity direction cannot be implied by their results.

Proof of Sufficiency in Theorem 3. If ψ is sublinear, then for any $f, g \in \mathcal{E}_m$ and $\lambda \in (0,1)$,

$$
\begin{aligned}
C(\lambda f + (1-\lambda)g) &= a + \sum_{j=1}^{m} \psi(\lambda f^j + (1-\lambda)g^j) \\
&\leq a + \sum_{j=1}^{m} (\lambda \cdot \psi(f^j) + (1-\lambda) \cdot \psi(g^j)) = \lambda C(f) + (1-\lambda)C(g).
\end{aligned}
$$

Therefore, C is convex, and thus quasiconvex. Then, it remains to verify (8). For any $f \in \mathcal{E}_m$, any $j \neq k$, and $\epsilon > 0$, we have

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

where the first equality follows from $C \in \mathcal{C}^{LS}$ and the inequality follows from sublinearity of ψ. As this holds for all $\epsilon > 0$, it follws that $D^+C(f; f^j(k)) \leq 0$ whenever exists, thus establishing Blackwell monotonicity by Theorem 2. □

Likelihood separability plays a key role in the proof, as it implies the difference $C(f + \epsilon f^j(k)) - C(f)$ depends only on the j-th and k-th column vectors in both experiments. In other words, a crucial property of likelihood separable costs is that, when holding the generation of other signals fixed, the change in costs by any operation involving only two signals is independent of the other signals. While this property might seem strong, it is actually shared by many well-known information costs, as we show in the next section.

Another desirable property for an information cost function is to have zero cost for any uninformative experiments. If it is the case, we say that the information cost is *grounded*. Note that any uninformative experiment takes the form of $f_\circ \equiv [\lambda_1 \mathbf{1}, \cdots, \lambda_m \mathbf{1}]$ with $\sum_{j=1}^m \lambda_j = 1$. When C satisfies (10) with some sublinear function ψ, we have

$$
C(f_\circ) = a + \sum_{j=1}^{m} \psi(\lambda_j \mathbf{1}) = a + \sum_{j=1}^{m} \lambda_j \cdot \psi(\mathbf{1}) = a + \psi(\mathbf{1}).
$$

Therefore, groundedness is equivalent to $a = -\psi(\mathbf{1})$.

Based on this observation and Theorem 3, we can conclude that a class of information cost functions introduced by Denti et al. (2022b) is exactly the class of costs satisfying likelihood separability, Blackwell monotonicity, and groundedness.

Definition 3. *(Denti et al. (2022b)) For an information cost function $C : \mathcal{E} \to \mathbb{R}_+$, if it can be represented as*

$$C(f) = \sum_{j=1}^{m} \psi(f^j) - \psi(\mathbf{1}), \tag{11}$$

for some absolutely continuous and sublinear function $\psi : [0,1]^n \to \mathbb{R}_+$,[13] *we call that C is a grounded sublinear likelihood separable (GSLS) information cost function. Let $\mathcal{C}^{GSLS}$ denote the class of GSLS cost functions.*[14]

Corollary 1. *Let $\mathcal{C}^{BM}$ denote the class of Blackwell-monotone cost functions and $\mathcal{C}^{G}$ denote the class of grounded cost functions. Then, the following holds:*

$$\mathcal{C}^{GSLS} = \mathcal{C}^{LS} \cap \mathcal{C}^{BM} \cap \mathcal{C}^{G}.$$

4.2 Subclasses of GSLS Costs

When seeking a Blackwell-monotone cost function, the results of the previous subsection enable one to construct a GSLS cost function by simply finding a sublinear function ψ, which includes all norms and seminorms. Here, we present notable subclasses of GSLS costs.

Norm Costs Norms are natural choices of sublinear functions. For any norm $\|\cdot\|$ on $\mathbb{R}^n$, the following cost function is Blackwell monotone and grounded:

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

Among norm costs, the supnorm can be used to construct probably the simplest example of a Blackwell-monotone cost function:

$$C(f) = \sum_{j=1}^{m} \max_i f_i^j - 1,$$

[13] $\psi(\mathbf{0}) = 0$ follows from the sublinearity.

[14] Denti et al. (2022b) incorporated groundedness and sublinearity in their definition of likelihood separability (Definition 36). Therefore, our GSLS cost functions correspond to their definition of LS cost functions.

For binary experiments represented by a single vector $f \in \mathbb{R}^n_+$, where $f^1 = \mathbf{1} - f$ and $f^2 = f$, the supnorm cost function is simply given by

$$C(f) = \max_i f_i - \min_i f_i.$$

Absolute-Linear Costs The absolute value of a linear function is a seminorm, and thus sublinear. Therefore, given any $a \in \mathbb{R}^n$, the following cost function is Blackwell monotone and grounded:

$$C(f) = \sum_{j=1}^{m} |\langle a, f^j \rangle| - |\langle a, \mathbf{1} \rangle| = \sum_{j=1}^{m} \left| \sum_{i=1}^{n} a_i f_i^j \right| - \left| \sum_{i=1}^{n} a_i \right|.$$

Notice that when $a \in \mathbb{R}^n$ is arbitrary, this cost function can potentially be a constant over all experiments. For example, when every entry of vector a is nonnegative, we have

$$C(f) = \sum_{j=1}^{m} |\langle a, f^j \rangle| - |\langle a, \mathbf{1} \rangle| = \sum_{j=1}^{m} \langle a, f^j \rangle - \langle a, \mathbf{1} \rangle = 0,$$

for all $f \in \mathcal{E}$.

To avoid this issue, say that a Blackwell-monotone cost function is *strictly grounded* if it assigns zero costs only to uninformative experiments. Our next result identifies a necessary and sufficient condition for strict groundedness of absolute-linear costs.

Proposition 3. *An absolute-linear cost function $C(f) = \sum_{j=1}^{m} |\langle a, f^j \rangle| - |\langle a, \mathbf{1} \rangle|$ is strictly grounded if and only if $\sum_{i=1}^{n} a_i = 0$ and $a_i \neq 0$ for all i.*

Absolute-linear costs are useful in applications as they provide a tractable and general class of Blackwell-monotone costs to start with. Especially, in the case of binary states and signals, as an experiment can be represented by $f = [f_1, f_2]^\intercal$, any strictly grounded absolute-linear cost function is given by

$$C(f) = \lambda |f_2 - f_1|, \text{ for some } \lambda > 0.$$

Furthermore, any monotone (not necessarily linear) transformation of this cost function is still Blackwell monotone, providing some freedom in choosing functional forms in applications. For example, the following quadratic cost function proves useful in our later applications:

$$C(f) = \lambda(f_2 - f_1)^2, \text{ for some } \lambda > 0.$$

Linear ϕ-divergence Costs For any i, i', let $\phi_{ii'} : [0, \infty] \to \mathbb{R} \cup \{+\infty\}$ be a convex function with $\phi_{ii'}(1) = 0$ and $\beta_{ii'} \geq 0$. The linear ϕ-divergence cost function is defined as

$$C(f) = \sum_{j=1}^{m} \sum_{i,i'} \beta_{ii'} f_{i'}^j \phi_{ii'} \left(\frac{f_i^j}{f_{i'}^j} \right) = \sum_{i,i'} \beta_{ii'} \sum_{j=1}^{m} f_{i'}^j \phi_{ii'} \left(\frac{f_i^j}{f_{i'}^j} \right), \tag{12}$$

where $\sum_{j=1}^{m} f_{i'}^j \phi_{ii'} \left(\frac{f_i^j}{f_{i'}^j} \right)$ is the ϕ-divergence (with generator $\phi_{ii'}$) between the probability distributions over signals in state ω_i and $\omega_{i'}$. The LLR cost axiomatized in Pomatto et al. (2023) is a special case with $\phi_{ii'}(x) = x \log x$. Blackwell monotonicity of linear ϕ-divergence costs is already well known, following from the *data-processing inequality* for ϕ-divergence (see Theorem 7.4 in Polyanskiy and Wu (2022)).

Another way to verify Blackwell monotonicity of linear ϕ-divergence costs is to show that they are GSLS costs. Define $\psi(f) \equiv \sum_{i,i'} \beta_{ii'} f_{i'}^j \phi_{ii'} \left(\frac{f_i^j}{f_{i'}^j} \right)$. Then, the sublinearity of ψ follows from Jensen's inequality given the convexity of $\phi_{ii'}$ (See Theorem 2.7.1 in Cover and Thomas (2006)). Additionally, we have $\psi(\mathbf{1}) = 0$ from $\phi_{ii'}(1) = 0$ for all i, i'. Therefore, the first expression of (12) shows that C is a GSLS cost.

Entropy Costs and Posterior-Separable Costs As already mentioned, another popular strand of defining costs of information is based on distribution over posteriors. This type of information cost is prior-dependent and thus cannot be directly defined as a function over only experiments. Nonetheless, once a full-support prior is fixed, it would induce a cost function over experiments.

Among such cost functions, the most popular cost function is the entropy cost (Sims, 2003; Matějka and McKay, 2015). Given a full support prior belief $\mu \in \Delta(\Omega)$, define $\tau_\mu : [0, 1]^n \to [0, 1]$ and $q_\mu : [0, 1]^n \to \Delta(\Omega)$ as follows:

$$\tau_\mu(h) \equiv \sum_{i=1}^{n} \mu_i \cdot h_i \qquad \text{and} \qquad q_\mu(h) \equiv \left(\frac{\mu_1 h_1}{\tau_\mu(h)}, \cdots, \frac{\mu_n h_n}{\tau_\mu(h)} \right)^\intercal.$$

According to these definitions, for an experiment $f : \Omega \to \Delta(\mathcal{S})$, $q_\mu(f^j)$ is the posterior belief upon receiving s_j and $\tau_\mu(f^j)$ is the probability of receiving s_j. The entropy cost function is defined as follows:

$$C_\mu(f) = H(\mu) - \sum_{j=1}^{m} \tau_\mu(f^j) \cdot H(q_\mu(f^j)), \tag{13}$$

where $H(\nu) \equiv -\sum_{i=1}^{n} \nu_i \log(\nu_i)$. By defining $\psi(h) \equiv -\tau_\mu(h) \cdot H(q_\mu(h))$, we can see that the entropy cost is likelihood separable. In fact, this type of cost with a concave H is known as the posterior-separable cost (Caplin and Dean, 2015; Denti, 2022).

Definition 4. *A cost function $C \in \mathcal{C}$ has a* ***posterior separable (PS) representation*** *at a full support prior belief $\mu \in \Delta(\Omega)$ if there exists a concave and absolutely continuous function $H : \Delta(\Omega) \to \mathbb{R}$ such that it satisfies (13) for any m and $f \in \mathcal{E}_m$. Let $\mathcal{C}_\mu^{PS}$ denote the class of cost functions that have PS representations at μ.*

As in the case of the entropy cost, any cost that has PS representation is likelihood separable. Additionally, by using the concavity of H, ψ can be shown to be sublinear and $\psi(\mathbf{1}) = -H(\mu)$, thus, any such cost is a GSLS cost. Moreover, we can show that the converse is also true: any GSLS cost has a PS representation.[15]

Proposition 4. *For any full support prior $\mu \in \Delta(\Omega)$, $\mathcal{C}^{GSLS} = \mathcal{C}_\mu^{PS}$.*

As the literature on information costs has been primarily focused on posterior separable costs, it is worth highlighting Blackwell-monotone costs that do not have posterior separable representations. Since posterior separability is a cardinal property, while Blackwell monotonicity is an ordinal property, any non-linear monotone transformation of posterior separable costs would preserve Blackwell monotonicity but would no longer maintain posterior separability.

This raises another question of whether there exists a Blackwell-monotone cost where any monotone transformation of it does not allow posterior separable representations.

Proposition 5. *Let $C_1 : \hat{\mathcal{E}}_2 \to \mathbb{R}_+$ be the cost function defined in Example 1. Then, C_1 cannot be represented as a monotone transformation of any PS cost.*

Recall that C_1 is shown to be Blackwell monotone in Example 1. However, according to the above proposition, it is not a PS cost, even in an ordinal sense. Therefore, the class of all PS costs (or equivalently, GSLS costs) and their monotone transformations remains a proper subset of Blackwell-monotone costs.

[15] A similar result is also shown by Proposition 37 of Denti et al. (2022b).

4.3 Discussion of the Axioms of Pomatto et al. (2023)

Given that the LLR cost characterized by Pomatto et al. (2023) is a special case of GSLS costs, we discuss how GSLS costs relate to the axioms presented in Pomatto et al. (2023). Their first axiom states that experiments that are Blackwell equivalent should have the same cost—this is implied by Blackwell monotonicity. Their fourth axiom pertains to the uniform continuity of cost functions. Next, we discuss their dilution and independence axioms.

The Dilution Axiom This axiom implies that diluting an experiment proportionally scales its cost by the dilution probability. Given an experiment $f \equiv [f^1, \cdots, f^m]$, a diluted experiment $\alpha \cdot f$ is an experiment that runs the experiment f with probability α, and sends an uninformative signal with probability $1-\alpha$:

$$\alpha \cdot f \equiv \left[\alpha f^1, \cdots, \alpha f^m, (1-\alpha)\mathbf{1}\right].$$

When C is a GSLS cost function with a component function ψ, from the sublinearity of ψ, we have

$$\begin{aligned} C(\alpha \cdot f) &= \sum_{j=1}^{m} \psi(\alpha f^j) \ + \psi((1-\alpha)\mathbf{1}) - \psi(\mathbf{1}) \\ &= \alpha \sum_{j=1}^{m} \psi(f^j) + (1-\alpha)\psi(\mathbf{1}) - \psi(\mathbf{1}) \ = \alpha \left(\sum_{j=1}^{m} \psi(f^j) - \psi(\mathbf{1}) \right) = \alpha C(f). \end{aligned}$$

Therefore, any GSLS cost function satisfies the dilution axiom.

The Independence Axiom This axiom states that the cost of performing two independent experiments is the sum of their costs, which captures the idea of constant marginal cost of each experiment. Most of the cost functions in Section 4.2, except for the LLR cost function, however, do not satisfy the independence axiom. For example, consider a supnorm cost defined in (4.2) and the following experiments:

$$f = \begin{array}{c|cc} & n & p \\ \hline - & 3/4 & 1/4 \\ + & 1/2 & 1/2 \end{array}, \qquad f \otimes f = \begin{array}{c|cccc} & nn & np & pn & pp \\ \hline - & 9/16 & 3/16 & 3/16 & 1/16 \\ + & 1/4 & 1/4 & 1/4 & 1/4 \end{array}.$$

Note that the cost of f is $1/4$, while the cost of $f \otimes f$ is $5/16$. Thus, the cost of $f \otimes f$ is not twice the cost of f, violating the independence axiom.

Another case that violates the independence axiom is when the completely informative experiment—where for each signal corresponds to exactly one possible state—has a finite (but nonzero) cost. Let f be such an experiment. Note that $f \otimes f$ is equally informative to f since both are completely informative. Thus, Blackwell monotonicity implies that they should have the same cost, however, the independence axiom implies that $f \otimes f$ should cost twice as much as f.

For this reason, in the setup of Pomatto et al. (2023), a completely informative experiment is considered infinitely costly. Therefore, such experiment can never be chosen as an optimal experiment under the independence axiom. Hence, if one considers constructing a model with a finite cost on completely informative experiment (as in the following application section), the independence axiom needs to be dropped.

5 Applications

In this section, we study two applications with costly information. Importantly, we highlight how our characterization of Blackwell monotonicity and likelihood separable costs can provide a general framework and tractable tools to analyze these problems.

5.1 Bargaining with Information Acquisition

Chatterjee et al. (2024) study an ultimatum bargaining model where the buyer can acquire costly information about the unknown value of an object before accepting the seller's offer.[16] To model costly information, they exogenously restrict the buyer's feasible set of experiments and define an information cost function over the restricted set. Using our characterization, we are able to extend their cost functions to Blackwell-monotone cost functions over all experiments, allowing us to examine their results in a more general setting. In this section, we show that while the exogenous restriction is crucial for their main result (Theorem 1), the same conclusion can still be obtained in the general setting when using a different Blackwell-monotone information cost function.

[16] Their model differs from Ravid (2020) in that the buyer observes the seller's offer and acquires information about his valuation, whereas in Ravid (2020), the buyer chooses an attention strategy, which is a map from the valuation and the offer to the acceptance probability.

Formally, a seller possesses an object (which holds zero value to herself) that has two possible values (for the buyer), denoted as $v \in V = \{H, L\}$ with $H > L > 0$, according to a common prior $\pi = Pr(H) \in (0, 1)$. The seller observes the value of the object and offers a price $p \in \mathbb{R}$ to the buyer. The buyer observes only the price p and can acquire information about v with experiments. Under Blackwell monotonicity, it is without loss of generality to restrict attention to binary experiments.[17] Let $f = [f_L, f_H]^\intercal \in \mathcal{E}_2$ where f_L and f_H denote the probability of generating signal h in states L and H, respectively. Let the information cost be denoted by $C_\lambda(f)$, parametrized by $\lambda > 0$.

Chatterjee et al. (2024) exogenously restrict the buyer's feasible set of experiments to *H*-focused experiments, i.e., $f_L = 0$.[18] With this restriction, they could define information cost functions simply as an increasing function of f_H. One example of their cost function is given by

$$C_\lambda((f_H, 0)) = \frac{\lambda}{2} f_H^2.$$

Notice this cost function can be extended to over all experiments by the quadratic cost function, which is Blackwell monotone:

$$C_\lambda(f) = \frac{\lambda}{2}(f_H - f_L)^2.$$

This enables us to examine their results without restricting the buyer's ability to acquire information.

Let $\sigma : V \to \Delta(\mathbb{R}_+)$ denote the seller's strategy. After the buyer observes the seller's offer p, the buyer forms a belief $\mu \in [0, 1]$ about the value of the object, chooses an experiment $f \in \mathcal{E}_2$ and takes an action contingent on the signal. The buyer's optimal strategy f^* given (p, μ) can be solved by the following program:

$$\max_{[f_L, f_H]^\intercal \in \mathcal{E}_2} \mu f_H(H - p) + (1 - \mu) f_L(L - p) - C_\lambda(f), \tag{14}$$

under which the buyer accepts the offer if receives signal h and rejects otherwise. A strategy profile $(\sigma^*, f^*(p, \mu))$ and a belief system $(\mu_p)_{p \in \mathbb{R}_+}$ constitute a Perfect Bayesian equilib-

[17] Suppose that the buyer utilizes an experiment with more than two signals. Then, consider a garbling of the signals, where signals inducing the buyer to accept the offer are assigned to h, and signals inducing the buyer to decline the offer are assigned to l. After applying this garbling, the expected material payoff remains the same, but it is less costly since it is less Blackwell informative.

[18] They also separately consider cases with *L*-focused experiments with $f_H = 1$, and a mix of both that does not span the space of all experiments.

rium if:[19]

1. $f^*(p,\mu_p)$ solves (14) given all (p,μ_p) ; σ^* is optimal given f^*.

2. μ_p is obtained via Bayes' rule on path.

Theorem 1 in Chatterjee et al. (2024) claims that when only H-focused experiments are available to the buyer, as information becomes arbitrarily cheap, i.e., $\lambda \to 0$, all equilibria are pooling equilibrium under which both types of seller offer the same price close to L and the buyer accepts the offer without acquiring any information. Once the buyer's feasible experiment is not restricted, however, this result no longer holds under (at least) the quadratic cost function as other equilibria would emerge:

Proposition 6. *When $C_\lambda(f) = \frac{\lambda}{2}(f_H - f_L)^2$, for all $\lambda < \pi(1-\pi)H$, there always exists non-pooling equilibria under which the buyer acquires information.*

The main intuition of Proposition 6 is similar to Proposition 1 in Chatterjee et al. (2024) where the buyer is restricted to acquire information using L-focused experiments. When $\lambda < \pi(1-\pi)H$, we show that there always exists a semi-separating equilibrium where the buyer acquires information using an experiment with $1 = f_H^* > f_L^* > 0$. This confirms that their intuition holds even when the restriction of L-focused experiments is lifted. More importantly, this also implies that their Theorem 1 hinges crucially on the exogenous restriction of using only H-focused experiments.

Despite this fact, our next proposition reestablishes their Theorem 1 without any restriction on experiments by considering cost functions that are not covered in their model, the absolute-linear costs.

Proposition 7. *When $C(f) = \lambda|f_H - f_L|$, for any $\lambda > 0$, there exists $\epsilon > 0$ such that every equilibrium is a pooling equilibrium where*

1. *$\sigma(L) = \sigma(H) = \delta_{p^*}$ with $p^* \in [L, L+\epsilon)$.*

2. *On the equilibrium path, $[f_L^*, f_H^*] = [1,1]$, i.e., the buyer acquires no information and buys at price p^* with certainty.*

Moreover, $\epsilon \to 0$ as $\lambda \to 0$, and thus eventually, the buyer extracts the full surplus.

[19] Same as in their model, we focus on Pareto-undominated equilibria.

For an intuition of Proposition 7, under absolute-linear costs, we show that the buyer's optimal information acquisition is either no information or full information, a special feature driven by linearity (Lemma OA.3.3). This fact removes the possibility of semi-separating equilibrium as in Proposition 6 and thus ensures all equilibria are pooling.

5.2 Costly Persuasion

Consider the classical prosecutor-judge example with costly information provision studied in Gentzkow and Kamenica (2014). The judge (Receiver) chooses between two actions: either **a**quits or **c**onvicts. There are two states of the world: the defendant is innocent ($\omega = i$) or guilty ($\omega = g$). The payoff of the prosecutor (Sender) is state-independent with $u_S(c) = 1$ and $u_S(a) = 0$, whereas the judge's payoff is to match the state and the action: $u_R(a,i) = u_R(c,g) = 1$ and $u_R(c,i) = u_R(a,g) = 0$.

The prosecutor commits to a persuasion policy at some Blackwell-monotone information cost C. Since the judge's action is binary, by using the same argument in footnote 17, it is without loss to consider binary experiments $(f_1, f_2) \in \mathcal{E}_2$ where $f_2 = \Pr(c|g)$ and $f_1 = \Pr(c|i)$. When the prior belief is $\mu = \Pr(g)$, the prosecutor's problem is

$$\max_{[f_1,f_2]^\intercal \in \mathcal{E}_2} \mu \cdot f_2 + (1-\mu) \cdot f_1 - C(f_1, f_2)$$

subject to the posterior belief upon receiving c is greater than or equal to $1/2$:

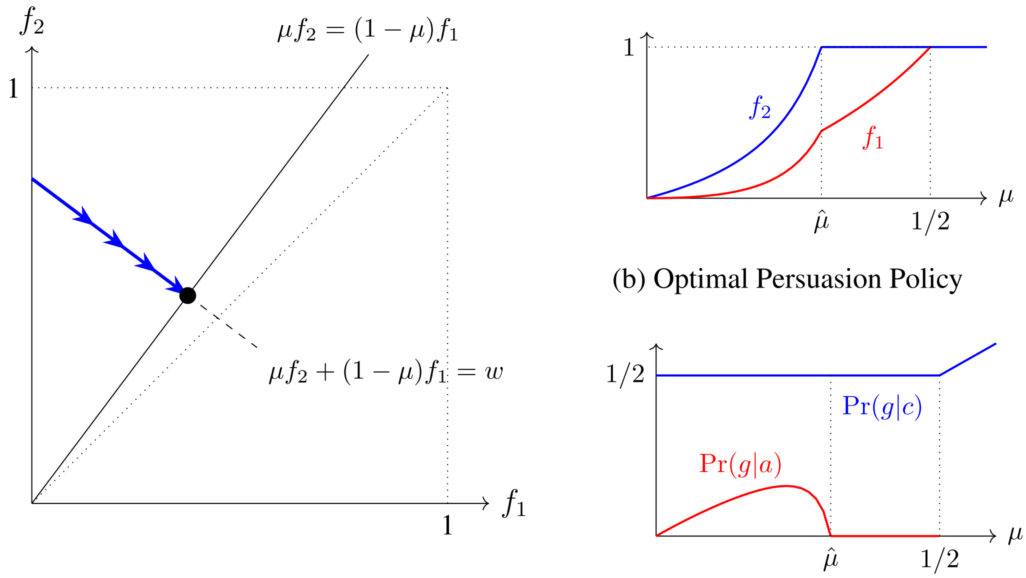
$$\frac{\mu f_2}{\mu f_2 + (1-\mu) f_1} \geq \frac{1}{2} \quad \Leftrightarrow \quad \mu f_2 \geq (1-\mu) f_1. \tag{15}$$

When $\mu \geq 1/2$, setting $f_1 = f_2 = 1$ yields the highest material payoff and the least information cost, and satisfies (15), i.e., always sending the signal c is optimal.

Now assume that $\mu < 1/2$. As an intermediate step to solve the problem, we consider an auxiliary cost minimization problem:

$$\min_{[f_1,f_2]^\intercal \in \mathcal{E}_2} C(f) \qquad \text{s.t.} \quad \mu \cdot f_2 + (1-\mu) \cdot f_1 = w \quad \text{and} \quad (15). \tag{16}$$

In other words, the auxiliary problem is to solve for the least costly information needed for the prosecutor to achieve a given material payoff level w. Note that from (15), the range of w is $[0, 2\mu]$.



(a) Cost Minimization

(b) Optimal Persuasion Policy

(c) Posteriors

Figure 2: Costly Persuasion with $C(f_1, f_2) = (f_2 - f_1)^2$

Lemma 3. *When C is Blackwell monotone and $1 > 2\mu \geq w \geq 0$, (15) binds for the solution of (16).*

This lemma is illustrated in Figure 2a. From the results of Section 3.1.2, as f_1 increases, Blackwell informativeness decreases along the line of $\mu f_2 + (1-\mu) f_1 = w$. Therefore, to minimize the Blackwell-monotone cost, (15) needs to bind:

$$f_1 = \frac{w}{2(1-\mu)} \quad \text{and} \quad f_2 = \frac{w}{2\mu}. \tag{17}$$

Intuitively, when (15) is non-binding, it implies that the posterior belief after receiving c is greater than $1/2$. Thus, the prosecutor can save on information cost by making the experiment *less* persuasive, while still ensuring the judge convicts.

Next, by plugging in (17), the prosecutor's problem becomes

$$\max_{0 \leq w \leq 2\mu} \; w - C\left(\frac{w}{2(1-\mu)}, \frac{w}{2\mu}\right). \tag{18}$$

Therefore, given the cost function, the prosecutor's problem becomes a one-dimensional maximization problem. As an example, the following proposition characterizes the optimal persuasion policy under the quadratic cost function.

Proposition 8. *Suppose $C(f_1, f_2) = (f_2 - f_1)^2$. The prosecutor's optimal persuasion policy is given by: for some $0 < \hat{\mu} < 1/2$,*

$$f_1 = \begin{cases} 1, & \text{if } \mu \geq 1/2, \\ \frac{\mu}{1-\mu}, & \text{if } \mu \in (\hat{\mu}, 1/2), \\ \frac{\mu^2(1-\mu)}{(1-2\mu)^2} & \text{if } \mu \leq \hat{\mu}, \end{cases} \quad \text{and} \quad f_2 = \begin{cases} 1, & \text{if } \mu \geq 1/2, \\ 1, & \text{if } \mu \in (\hat{\mu}, 1/2), \\ \frac{\mu(1-\mu)^2}{(1-2\mu)^2} & \text{if } \mu \leq \hat{\mu}. \end{cases} \tag{19}$$

This result is illustrated in Figure 2b and 2c. When $\mu \geq \hat{\mu}$, the optimal persuasion policy is the same as the one without the cost: the posterior belief is either $1/2$ or 0. In this case, the prosecutor always convicts guilty defendants and, with some positive probability, convicts innocent defendants. When $\mu < \hat{\mu}$, this policy is no longer optimal as it becomes too expensive. Instead, the prosecutor sacrifices the probability of convicting the guilty defendant to lower the costs. Observe that the posterior belief upon receiving a depends on μ, while the posterior belief upon receiving c is constant $(1/2)$. This result differs qualitatively from the one with uniformly posterior separable costs, where the posterior beliefs are independent of the prior belief whenever the information is provided.

6 Additional Results regarding Quasiconvexity

6.1 Non-necessity of Quasiconvexity

Theorem 2 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 3. Suppose $n = m = 2$. Denote any experiment $f \in \mathcal{E}_2$ by $f = [f_1, f_2]^\intercal$. 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 $C : \hat{\mathcal{E}}_2 \to \mathbb{R}_+$ defined by

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

By using (6), we can easily see that $f \succeq_B g$ implies $C(f) \geq C(g)$, i.e., C is Blackwell monotone.

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

6.2 Binary Experiments with Quasiconvexity

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 $f = [f_1, \cdots, f_n]^\intercal \in [0,1]^n$, 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 9. *If $C \in \mathcal{E}_2$ is quasiconvex, permutation invariant, and non-null, then C is Blackwell monotone.*

6.3 A Weaker Quasiconvexity Condition

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

Definition 5. *$C \in \mathcal{C}_m$ is **garbling-quasiconvex** if for all $f \in \mathcal{E}$, any finite collection of its garblings, $\{g_1, \cdots, g_n\}$, and $\lambda_0, \cdots, \lambda_n \in [0,1]$ with $\sum_{i=0}^n \lambda_i = 1$,*

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

Theorem 4. *Suppose $C \in \mathcal{C}_m$ is absolutely continuous and permutation invariant. C is Blackwell monotone if and only if (i) C is garbling-quasiconvex and (ii) for all $f \in \mathcal{E}_m$,*

$$D^+C(f; f^j(k)) \leq 0, \quad \forall j \neq k, \textit{ whenever exists}.$$

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 is almost the identical to the proof of Theorem 2, with additional steps needed to show that garbling-quasiconvexity (with continuity) is also sufficient for establishing the final step of the proof.

7 Concluding Remarks

Information is costly and *more* information should cost more. Building upon this premise, this paper characterizes necessary and sufficient conditions for information cost functions to be monotone when informativeness is compared using Blackwell's information order. This characterization allows us to study the implications of Blackwell monotonicity in various economic applications. For some applications exhibiting monotonicity between signals and actions, another well-known information order proposed by Lehmann (1988), which refines the Blackwell order, becomes more relevant. We believe the methodology developed in this paper can be extended to characterize Lehmann-monotone costs. We leave this for future research.

References

Arrow, Kenneth J. and Alain C. Enthoven (1961) "Quasi-Concave Programming," *Econometrica*, 29 (4), 779–800, http://www.jstor.org/stable/1911819.

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 (2021) "The cost of optimally acquired information."

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.

Caplin, Andrew, Mark Dean, and John Leahy (2022) "Rationally inattentive behavior: Characterizing and generalizing Shannon entropy," *Journal of Political Economy*, 130 (6), 1676–1715.

Chambers, Christopher P, Ce Liu, and John Rehbeck (2020) "Costly information acquisition," *Journal of Economic Theory*, 186, 104979.

Chatterjee, Kalyan, Miaomiao Dong, and Tetsuya Hoshino (2024) "Bargaining and Information Acquisition," mimeo, ITAM.

Cover, Thomas M. and Joy A. Thomas (2006) *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*, USA: Wiley-Interscience.

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.

Gentzkow, Matthew and Emir Kamenica (2014) "Costly persuasion," *American Economic Review*, 104 (5), 457–462.

Kamenica, Emir and Matthew Gentzkow (2011) "Bayesian persuasion," *American Economic Review*, 101 (6), 2590–2615.

Khan, M Ali, Haomiao Yu, and Zhixiang Zhang (2024) "On comparisons of information structures with infinite states," *Journal of Economic Theory*, 105841.

Lehmann, EL (1988) "Comparing Location Experiments," *The Annals of Statistics*, 521–533.

Matějka, Filip and Alisdair McKay (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*.

Mu, Xiaosheng, Luciano Pomatto, Philipp Strack, and Omer Tamuz (2021) "From Blackwell Dominance in Large Samples to Rényi Divergences and Back Again," *Econometrica*, 89 (1), 475–506, https://doi.org/10.3982/ECTA17548.

de Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek (2017) "Rationally inattentive preferences and hidden information costs," *Theoretical Economics*, 12 (2), 621–654.

Osborne, Martin J (2016) "Mathematical methods for economic theory," *University of Toronto*.

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.

Ramos-Mercado, Jorge D (2023) "Optimal Sequential Experimentation."

Ravid, Doron (2020) "Ultimatum Bargaining with Rational Inattention," *American Economic Review*, 110 (9), 2948–63, 10.1257/aer.20170620.

Sajbura, Katarzyna (2005) "Level sets of continuous functions increasing with respect to each variable," *Discussiones Mathematicae, Differential Inclusions, Control and Optimization*, 25 (1), 19–26, `http://eudml.org/doc/271441`.

Sims, Christopher A (2003) "Implications of rational inattention," *Journal of monetary Economics*, 50 (3), 665–690.

Zhong, Weijie (2022) "Optimal dynamic information acquisition," *Econometrica*, 90 (4), 1537–1582.

A Proofs for Section 3

A.1 Proof of Theorem 1

A.1.1 A Lemma

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

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

Without loss, let g satisfy the first equation of (A.1) and $f' = \frac{a-b}{1-b} f$.[20] *Then, for all $\lambda \in [0, 1]$, the followings hold:*

$$f \succeq_B (1 - \lambda)f + \lambda f' \succeq_B f', \text{ and} \tag{A.2}$$

$$f' \succeq_B (1 - \lambda)f' + \lambda g \succeq_B g. \tag{A.3}$$

Moreover, $f' - f$ points in the direction of $-f$, and $g - f'$ points in the direction of $\mathbf{1} - f'$.

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

When $b = 1$, a is also equal to 1 and $g = f + (\mathbf{1} - f) = \mathbf{1} = f'$. Then, (A.3) trivially holds. Notice that $(1 - \lambda)f + \lambda\mathbf{1} = 1 \cdot f + \lambda(\mathbf{1} - f) \in \text{PARL}(f, \mathbf{1} - f)$, thus, $f \succeq_B (1 - \lambda)f + \lambda\mathbf{1}$. Next, we have $\mathbf{1} = 1 \cdot ((1 - \lambda)f + \lambda\mathbf{1}) + 1 \cdot \{\mathbf{1} - ((1 - \lambda)f + \lambda\mathbf{1})\}$, which implies $(1 - \lambda)f + \lambda\mathbf{1} \succeq_B \mathbf{1}$. Therefore, (A.2) holds.

When $b < 1$, we have $\frac{a-b}{1-b} \in [0, 1]$ and $f \succeq_B f'$. For any $\lambda \in [0, 1]$, $f \succeq_B \lambda f + (1 - \lambda)f'$ simply follows from convexity of $\text{PARL}(f, \mathbf{1} - f)$. On the other hand, 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 $f' \in \text{PARL}(((1 - \lambda)f + \lambda f'), \mathbf{1} - ((1 - \lambda)f + \lambda f'))$, and thus $(1 - \lambda)f + \lambda f' \succeq_B f'$.

[20] When $b = 1$, define $f' = \mathbf{1}$.

From $g = af + b(\mathbf{1} - f)$, we have

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

thus $f' \succeq_B g$ and $g - f' = b(\mathbf{1} - f')$. By a similar argument, $f' \succeq_B (1-\lambda)f' + \lambda g \succeq_B g$. The last statement also follows from the above argument. $\square$

A.1.2 Proof

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

For sufficiency, take any $f \succeq_B g$. First, permutate and relabel g if needed to have g satisfy the first equation of (A.1). Permutation invariance ensures that 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, when differentiable,

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

On the other hand, observe that

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

Therefore, $\varphi_1'(\lambda)$ has the same sign as $D^+C((1-\lambda)f + \lambda f'; -((1-\lambda)f + \lambda f'))$ and it is non-positive from (2). Then, the FTC implies That

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

Similarly, observe that

$$\varphi_2'(\lambda) = D^+C((1-\lambda)f' + \lambda g; -f' + g),$$
$$-f' + g = b(\mathbf{1} - f') = \frac{b}{1-\lambda b}\left(\mathbf{1} - ((1-\lambda)f' + \lambda g)\right).$$

Then, $\varphi_2'(\lambda)$ is non-positive since it has the same sign with $D^+C((1-\lambda)f' + \lambda g; \mathbf{1} - ((1-\lambda)f' + \lambda g))$. By applying the FTC, we also have $C(g) = \varphi_2(1) \leq \varphi_2(0) = C(f')$. Therefore, we have $C(f) \leq C(g)$, thus C is Blackwell monotone. $\square$

A.2 Proof of Theorem 2

A.2.1 Proof of Lemma 2

Proof of Lemma 2. If $f \succeq_B g$, then there exists a stochastic matrix $M \in \mathcal{M}_m$ such that $g = fM$. Then, we have

$$\begin{aligned}
g - f &= f(M - I) \\
&= \begin{bmatrix} f^1 & \cdots & f^m \end{bmatrix} \begin{bmatrix} m_1^1 - 1 & \cdots & m_1^m \\ \vdots & \ddots & \vdots \\ m_m^1 & \cdots & m_m^m - 1 \end{bmatrix} \\
&= \begin{bmatrix} f^1 & \cdots & f^m \end{bmatrix} \begin{bmatrix} -\sum_{k=2}^m m_1^k & \cdots & m_1^m \\ \vdots & \ddots & \vdots \\ m_m^1 & \cdots & -\sum_{k=1}^{m-1} m_m^k \end{bmatrix} \\
&= \sum_{j=1}^m \sum_{k \neq j} m_j^k f^j(k)
\end{aligned}$$

with $m_j^k \geq 0$ and $\sum_{k=1}^m m_j^k = 1$ for all j. It thus can be further written as

$$g - f = \sum_{j=1}^m \left(\sum_{k \neq j} m_j^k \right) \left(\sum_{k \neq j} \frac{m_j^k}{\sum_{k \neq j} m_j^k} f^j(k) \right).$$

Notice that $\sum_{k \neq j} m_j^k \in [0, 1]$ and the term in the second parentheses is a convex combination of $f^j(k)$ for $k \neq j$. Thus, we have

$$g - f = \sum_{j=1}^m \lambda_j h_j,$$

for $\lambda_j \in [0, 1]$ and $h_j \in \text{co}(\{f^j(k) : k \neq j\})$.

Conversely, it suffices to reverse the above steps to construct a stochastic matrix M such that $g = fM$. In particular, we can take $m_j^k = \lambda_j \cdot \mu_j^k$ where $\sum_{k \neq j} \mu_j^k f_k^j = h_j$ and let $m_j^j = 1 - \sum_{k \neq j} m_j^k$. Then, such M is a stochastic matrix and $g = fM$. $\square$

A.2.2 Useful Lemmas

Lemma A.2. *If $C \in \mathcal{C}_m$ is Blackwell monotone, then for all $f \in \mathcal{E}_m$ and all $j \neq k$,*

$$D^+C(f; f^j(k)) \leq 0, \text{ whenever exists.}$$

Proof of Lemma A.2. For any $f \in \mathcal{E}_m$ and any $j \neq k$, let

$$f_\lambda \equiv f + \lambda f^j(k), \quad \lambda \in [0, 1].$$

Since $f_\lambda - f$ satisfies (7) for all λ, Blackwell monotonicity implies

$$C(f) \geq C(f_\lambda), \quad \forall \lambda \in [0, 1].$$

If $D^+C(f; f^j(k))$ exists, then

$$D^+C(f; f^j(k)) = \lim_{\lambda \downarrow 0} \frac{C(f_\lambda) - C(f)}{\lambda} \leq 0.$$

□

Lemma A.3. *Let B_{jk} be an $m \times m$ matrix such that $b_{jj} = -1$, $b_{jk} = 1$, and all other entries are equal to zero. Then for any $f \in \mathcal{E}_m$, $fB_{jk} = f^j(k)$.*

Proof of Lemma A.3. When $i \neq j, k$, i-th column of fB_{jk} is equal to $\mathbf{0}$. Additionally, j-th column of fB_{jk} is $-f^j$ and k-th column of fB_{jk} is f^j. Thus, fB_{jk} is equal to $f^j(k)$. □

Lemma A.4. *Suppose $C \in \mathcal{C}_m$ is absolutely continuous and satisfies (8) in Theorem 2. Then for any $1 \leq k \leq m$ and $E \in \textbf{ext}_{k-1}(\mathcal{M}_m)$, there exists $E' \in \textbf{ext}_k(\mathcal{M}_m)$ such that for all $\lambda \in [0, 1]$,*

$$fE' \succeq_B (1 - \lambda) fE' + \lambda fE \succeq_B fE. \tag{A.4}$$

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

Proof of Lemma A.4. For any $1 \leq k \leq m$ and $E \in \textbf{ext}_{k-1}(\mathcal{M}_m)$, we show that there exists $E' \in \textbf{ext}_k(\mathcal{M}_m)$ such that (A.4) holds.

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 $e_{ji} = e_{j'i} = 1$. 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., $E' \in \textbf{ext}_k(\mathcal{M}_m)$.

Let B denote $B_{i'i}$ as defined in Lemma A.3. 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, (A.4) holds.

Recall by Lemma A.3,

$$
fB = f^{i'}(i).
$$

(8) then implies that for all $\lambda \in [0,1]$,

$$
\begin{aligned}
&D^+(C((1-\lambda)fE' + \lambda fE), fE - ((1-\lambda)fE' + \lambda fE))\\
=&D^+(C((1-\lambda)fE' + \lambda fE), ((1-\lambda)fE' + \lambda fE)B) \ \ \leq 0.
\end{aligned}
\tag{A.5}
$$

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 (A.5) holds for all $\lambda \in [0,1]$. $\square$

A.2.3 Proof

Proof of Theorem 2. Necessity follows directly from Lemma A.2.

For sufficiency, permutation invariance and Lemma A.4 together imply that $C(f) \geq C(fE)$ for all $E \in \textbf{ext}(\mathcal{M}_m)$. Take any $f, g \in \mathcal{E}_m$ with $f \succeq_B g$ so that $g = fM$. By quasiconvexity of C, we have

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

thus, C is Blackwell monotone. $\square$

B Proofs for Section 4

B.1 Proof of Necessity in Theorem 3

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

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

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 additive separability implies

$$\psi\left(\frac{1}{k}\hat{f}\right) = \frac{1}{k}\psi(\hat{f}).$$

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 $C(f) = C(g)$, and 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 $z\hat{f} \in [0,1]^n$,

$$\psi(\hat{f}) = z\psi(\hat{f}).$$

By density of $\mathbb{Q}$ in $\mathbb{R}$ and 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. $\square$

B.2 Proof of Proposition 3

Proof of Proposition 3. Let $\underline{\mathcal{E}}_m \subset \mathcal{E}_m$ denote the set of uninformative experiments in $\mathcal{E}_m$. Notice that $f \in \underline{\mathcal{E}}_m$ if and only if

$$f^j \in \{\lambda \mathbf{1} : \lambda \in [0,1]\}, \quad \forall j.$$

For sufficiency, given $\sum_{i=1}^n a_i = 0$ and $a_i \neq 0$ for all i, then for any $\hat{f} \in [0,1]^n$,

$$\langle a, \hat{f} \rangle = 0 \Leftrightarrow \hat{f} \in \{\lambda \mathbf{1} : \lambda \in [0,1]\}.$$

This implies that $C(f)$ is strictly grounded.

For necessity, $C(f) = 0$ for all $f \in \underline{\mathcal{E}}_m$ implies that $\sum_{i=1}^n a_i = 0$. Next, towards a contradiction, suppose that $\sum_{i=1}^n a_i = 0$ yet $a_1 = 0$. Then the following experiment,

$$f = \begin{bmatrix} \lambda & 1-\lambda \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix},$$

is not in $\underline{\mathcal{E}}_2$ for $\lambda \neq 0$ and $C(f) = 0$. Thus, a contradiction. □

B.3 Proof of Proposition 4

Proof of Proposition 4. **[$\mathcal{C}^{GSLS} \supseteq \mathcal{C}_\mu^{PS}$]** Suppose that given a prior $\mu \in \Delta(\Omega)$, $C \in \mathcal{C}_\mu^{PS}$ with H_μ. Define $\psi : [0,1]^n \to \mathbb{R}$ as follows: for any $h \in [0,1]^n$,

$$\psi(h) \equiv -\tau_\mu(h) \cdot H_\mu\left(q_\mu(h)\right).$$

Observe that $\tau_\mu(\mathbf{1}) = 1$, $q_\mu(h) = \mu$ and $\psi(\mathbf{1}) = -H_\mu(\mu)$. Therefore, (13) can be rewritten as a form of (11).

It remains to show that ψ is sublinear. For positive homogeneity, from $\tau(\gamma \cdot h) = \gamma \cdot \tau(h)$, we have $\psi(\gamma \cdot h) = \gamma \cdot \psi(h)$ whenever $\gamma \cdot h \in [0,1]^n$. Additionally, $\tau(h+l) = \tau(h) + \tau(l)$ and the concavity of H_μ imply that

$$\begin{aligned}\psi(h) + \psi(l) = &- \tau(h) \cdot H_\mu\left[\left(\frac{\mu_i f_i^j}{\tau(h)}\right)_i\right] - \tau(l) \cdot H_\mu\left[\left(\frac{\mu_i f_i^k}{\tau(l)}\right)_i\right] \\ \geq &- (\tau(h) + \tau(l)) \cdot H_\mu\left[\left(\frac{\mu_i f_i^j + \mu_i l_i}{\tau(h+l)}\right)_i\right] = \psi(h+l),\end{aligned}$$

thus, ψ is subadditive. Therefore, C is a GSLS cost.

[$\mathcal{C}^{GSLS} \subseteq \mathcal{C}_\mu^{PS}$] Suppose that $C \in \mathcal{C}^{GSLS}$ with ψ. Fix a full support prior $\mu \in \Delta(\Omega)$. Let

$p = (p_1, \cdots, p_n)^\intercal \in \Delta(\Omega)$. Observe that $q_\mu(h) = p$ if and only if

$$\frac{h}{\tau_\mu(h)} = \left(\frac{p_1}{\mu_1}, \cdots, \frac{p_n}{\mu_n}\right)^\intercal.$$

Now consider a small enough $\epsilon > 0$ such that

$$\left(\frac{p_1}{\mu_1}\epsilon, \cdots, \frac{p_n}{\mu_n}\epsilon\right)^\intercal \in [0,1]^n.$$

Then, define

$$H_\mu(p) \equiv -\frac{\psi\left(\left(\frac{p_1}{\mu_1}\epsilon, \cdots, \frac{p_n}{\mu_n}\epsilon\right)^\intercal\right)}{\epsilon}.$$

This function is well defined due to the sublinearity of ψ. Additionally, the concavity of H_μ can be obtained by the sublinearity of ψ: for any $p, p' \in \Delta(\Omega)$,

$$\begin{aligned}
H_\mu(\lambda p + (1-\lambda)p') &= -\frac{\psi\left(\left(\frac{\lambda p_1 + (1-\lambda)p'_1}{\mu_1}\epsilon, \cdots, \frac{\lambda p_n + (1-\lambda)p'_n}{\mu_n}\epsilon\right)^\intercal\right)}{\epsilon} \\
&\geq -\frac{\psi\left(\left(\frac{\lambda p_1}{\mu_1}\epsilon, \cdots, \frac{\lambda p_n}{\mu_n}\epsilon\right)^\intercal\right)}{\epsilon} - \frac{\psi\left(\left(\frac{(1-\lambda)p'_1}{\mu_1}\epsilon, \cdots, \frac{(1-\lambda)p'_n}{\mu_n}\epsilon\right)^\intercal\right)}{\epsilon} \\
&= \lambda H_\mu(p) + (1-\lambda)H_\mu(p').
\end{aligned}$$

Also observe that $\psi(h) = -\tau_\mu(h) \cdot H_\mu(q_\mu(h))$ for any $h \in [0,1]^n$. Moreover, $H_\mu(\mu) = -\psi(\mathbf{1})$. Therefore, (11) can be rewritten as a form of (13), i.e., C can be represented as a PS cost at μ. □

C Proofs for Section 6

C.1 Proof of Proposition 9

Proof of Proposition 9. By Lemma 1, $f \succeq_B g$ if and only if $g = af + b(\mathbf{1} - f)$ 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 $a < b$, $g = (1-b)\cdot\mathbf{0} + (b-a)\cdot(\mathbf{1}-f) + a\cdot\mathbf{1}$. From quasiconvexity and non-nullness, we have $C(f) \geq C(g)$ or $C(\mathbf{1}-f) \geq C(g)$. Then, by permutation invariance, $C(f) = C(\mathbf{1}-f)$, thus, $C(f) \geq C(g)$. □

C.2 Proof of Theorem 4

Proof of Theorem 4. 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 2, 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,

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

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

Online Appendix for "Blackwell-Monotone Information Costs"

Xiaoyu Cheng and Yonggyun Kim

OA.1 Strict Blackwell Monotonicity

In this section, we provide a sufficient condition for strict Blackwell monotonicity in the case of binary experiments.

For any $f, g \in \mathcal{E}_m$, if $f \succeq_B g$ but $g \not\succeq_B f$, then f is *strictly more informative* than g, denoted by $f \succ_B g$.

Definition OA.1.1. *A Blackwell-monotone cost function $C \in \mathcal{C}_m$ is* ***strictly Blackwell monotone*** *if for all $f, g \in \mathcal{E}_m$, $C(f) > C(g)$ whenever $f \succ_B g$.*

Theorem OA.1.1. *Suppose $C \in \mathcal{C}_2$ is absolutely continuous, permutation invariant, and Blackwell monotone. C is strictly Blackwell monotone if the inequalities in (2) hold strictly whenever $f \notin \{\lambda \mathbf{1} : \lambda \in [0,1]\}$.*

Proof of Theorem OA.1.1. First, the following lemma provides a characterization of when $f \succ_B g$.

Lemma OA.1.1. *For $f \neq g$, if $f \notin \{\lambda \mathbf{1} : \lambda \in [0,1]\}$, then $f \simeq_B g$ if and only if $g = \mathbf{1} - f$, otherwise $f \succeq_B g$ if and only if $f \simeq_B g$. In other words, $f \succ_B g$ if and only if $f \notin \{\lambda \mathbf{1} : \lambda \in [0,1]\}$ and $g \neq \mathbf{1} - f$.*

Proof of Lemma OA.1.1. Recall that $f \succeq_B g$ if and only if there exists $(a,b) \in [0,1]^2$ such that $g = af + b(\mathbf{1} - f)$. Thus, $f \simeq_B g$ if and only if there exists $(a,b) \in [0,1]^2$ and $(a',b') \in [0,1]^2$ such that

$$g = af + b(\mathbf{1} - f) \quad \text{and} \quad f = a'g + b'(\mathbf{1} - g).$$

Plugging the first equation into the second, we have

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

This equation holds only when either one of the following holds:

(i) $a = 1$ and $b = 0$, i.e., $g = f$; or

(ii) $a = 0$ and $b = 1$, i.e., $g = \mathbf{1} - f$; or

(iii) $f \in \{\lambda\mathbf{1} : \lambda \in [0, 1]\}$.

Notice in the third case, $f \succeq_B g$ if and only if $g \in \{\lambda\mathbf{1} : \lambda \in [0, 1]\}$. Consequently, it implies that $f \succeq_B g$ if and only if $f \simeq_B g$. □

Consider any $f \succ_B g$, Lemma OA.1.1 implies that $f \notin \{\lambda\mathbf{1} : \lambda \in [0, 1]\}$ and $g \neq \mathbf{1} - f$. Then, after a permutation if needed, there exists a path from f to g as proved in Lemma A.1 and that every experiment along this path is not in $\{\lambda\mathbf{1} : \lambda \in [0, 1]\}$. Since the inequalities in (2) are strict, FTC implies along this path implies $C(f) > C(g)$. □

OA.2 Sufficiency via Higher Dimensions

In Theorem 2, sufficiency for Blackwell monotonicity is established by imposing quasiconvexity on the cost function. In this section, we present an alternative approach to establishing the sufficiency of the first-order condition by embedding the space $\mathcal{E}_m$ into a higher-dimensional space.

Recall Proposition 2 shows an example in $\mathcal{E}_3$ where $f = I_3 \succeq_B g$ but there does not exist a continuous path in $\mathcal{E}_3$ connecting f and g along which Blackwell informativeness decreases. Nevertheless, when both f and g are considered as experiments in $\mathcal{E}_6$, a decreasing path can actually be found. To see this, first embed f and g into $\mathcal{E}_6$ by adding three columns of zeros. Then consider the following experiment

$$\overline{f} = \begin{bmatrix} 4/5 & 0 & 0 & 1/5 & 0 & 0 \\ 0 & 4/5 & 0 & 0 & 1/5 & 0 \\ 0 & 0 & 4/5 & 0 & 0 & 1/5 \end{bmatrix}.$$

It is not hard to see that $f \simeq_B \overline{f}$ by finding stochastic matrices connecting them. Thus, $\overline{f} \succeq_B g$. Then for any $\lambda \in [0, 1]$, let

$$f_\lambda = (1-\lambda)\overline{f} + \lambda g = \begin{bmatrix} 4/5 & \lambda/5 & 0 & (1-\lambda)/5 & 0 & 0 \\ 0 & 4/5 & \lambda/5 & 0 & (1-\lambda)/5 & 0 \\ \lambda/5 & 0 & 4/5 & 0 & 0 & (1-\lambda)/5 \end{bmatrix}.$$

It can also be shown that $\overline{f} \succeq_B f_\lambda \succeq_B g$. Thus, a decreasing path connecting $\overline{f}$ and g now can be found in $\mathcal{E}_6$.

The previous observation actually holds for all $f \succeq_B g$ in $\mathcal{E}_m$. That is, there always exists a decreasing path connecting f and g in the space $\mathcal{E}_{2m}$ (Lemma OA.2.1). Thus, we can establish sufficiency by connecting every pair of experiments in $\mathcal{E}_m$ by a higher-dimensional path. However, such a decreasing path does not necessarily move in the extreme direction $f^j(k)$. Therefore, we need to rely on the linearity of directional derivatives, which requires a stronger differentiability assumption on C:

Theorem OA.2.1. *Suppose $C \in \mathcal{C}_{2m}$ is continuously differentiable and permutation invariant. If (9) holds for all $f \in \mathcal{E}_{2m}$, then C is Blackwell monotone over $\mathcal{E}_m$.*

OA.2.1 Proof of Theorem OA.2.1

Proof of Theorem OA.2.1. For any $f, g \in \mathcal{E}_m$ and $\lambda \in [0,1]$, define experiment

$$\lambda f \oplus (1-\lambda) g \in \mathcal{E}_{2m}$$

to be that with probability λ, it generates signals in $\{s_1, \cdots, s_m\}$ according to f and with probability $1-\lambda$, it generates signals in $\{s_{m+1}, \cdots, s_{2m}\}$ according to g. Notice we can write such an experiment as

$$\begin{bmatrix} \lambda f & (1-\lambda) g \end{bmatrix} \in \mathcal{E}_{2m}.$$

We next present a lemma showing that for every $f, g \in \mathcal{E}_m$, if $f \succeq_B g$, then there always exists a decreasing path from f to g in $\mathcal{E}_{2m}$.

Lemma OA.2.1. *For any $f, g \in \mathcal{E}_m$, if $f \succeq_B g$, then for all $\lambda \in [0,1]$,*

$$f \succeq_B \lambda f \oplus (1-\lambda) g \succeq_B g.$$

Proof of Lemma OA.2.1. We first show the lemma holds when $f = g$. Notice that for all $\lambda \in [0,1]$,

$$\begin{bmatrix} f & 0 \end{bmatrix} \begin{bmatrix} \lambda I & (1-\lambda) I \\ 0 & I \end{bmatrix} = \begin{bmatrix} \lambda f & (1-\lambda) f \end{bmatrix},$$

and

$$\begin{bmatrix} \lambda f & (1-\lambda) f \end{bmatrix} \begin{bmatrix} I & 0 \\ I & 0 \end{bmatrix} = \begin{bmatrix} f & 0 \end{bmatrix}.$$

Next, consider any $f \succeq_B g$. Let $g = fM$ for some stochastic matrix $M \in \mathcal{M}_m$. Consider the following stochasitc matrix in $\mathcal{M}_{2m}$:

$$\begin{bmatrix} \lambda I & 0 \\ 0 & M \end{bmatrix}.$$

Then we have

$$\begin{bmatrix} \lambda f & (1-\lambda)f \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix} = \begin{bmatrix} \lambda f & (1-\lambda)g \end{bmatrix}.$$

Thus, we have establishes that $f \succeq_B \lambda f \oplus (1-\lambda)g$ for all $\lambda \in [0,1]$.

Consider another stochastic matrix in $\mathcal{M}_{2m}$:

$$\begin{bmatrix} M & 0 \\ 0 & I \end{bmatrix}.$$

Then for all $\lambda \in [0,1]$, notice that

$$\begin{bmatrix} \lambda f & (1-\lambda)g \end{bmatrix} \begin{bmatrix} M & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} \lambda g & (1-\lambda)g, \end{bmatrix}$$

where the last matrix is Blackwell equivalent to g. Thus, we have also established that $\lambda f \oplus (1-\lambda)g \succeq_B g$ for all $\lambda \in [0,1]$. $\square$

Suppose $C \in \mathcal{C}_{2m}$ is continuously differentiable and satisfies (i) and (ii) in Theorem 2 over $\mathcal{E}_{2m}$. By differentaibility of C, we have for any $f \in \mathcal{E}_{2m}$,

$$D^+C(f; f^j(k)) = \langle \nabla C(f), f^j(k) \rangle \leq 0, \forall j \neq k.$$

Then by Lemma A.2, for any $g \in \mathcal{E}_{2m}$ with $f \succeq_B g$, we have

$$\langle \nabla C(f), g - f \rangle \leq 0,$$

where the inequality follows as it is a positive linear combination of $\langle \nabla C(f), f^j(k) \rangle$.

Finally, for any $f, g \in \mathcal{E}_m$ with $f \succeq_B g$, applying the FTC along the decreasing path from f to g in $\mathcal{E}_{2m}$ identified in Lemma OA.2.1 implies $C(f) \geq C(g)$, and thus establishes the conclusion. $\square$

OA.3 Omitted Proofs

OA.3.1 Proof of Proposition 2

Proof of Proposition 2. Suppose that $f \succeq_B g$, i.e., there exists a 3×3 stochastic matrix $B = (b_i^j)$ such that $fB = g$.

Observe that at least one of f_1^1, f_1^2 and f_1^3 is positive—if not, every entry of the first row of fB is equal to zero. Without loss of generality, let f_1^1 be positive (we can obtain it by permuting f). Note that $f_1^1 b_1^3 + f_1^2 b_2^3 + f_1^3 b_3^3 = 0$. Since every entry of f and B are nonnegative, $b_1^3 = 0$.

Next, observe that $4/5 = f_2^1 b_1^3 + f_2^2 b_2^3 + f_2^3 b_3^3$. From $b_1^3 = 0$, at least one of f_2^2 and f_2^3 is positive. Without loss, let f_2^2 be positive. Then, from $g_2^1 = 0$, we have $b_2^1 = 0$. Then, it gives us $f_2^1 b_1^1 + f_2^3 b_3^1 = 0$. We consider two cases: $b_1^1 = 0$ or $f_2^1 = 0$.

1. $b_1^1 = 0$**:** From $b_1^1 = b_1^3 = 0$, we have $b_1^2 = 1$. From $g_3^2 = 0$ and $b_1^2 > 0$, we have $f_3^1 = 0$.

Additionally, we have $f_1^3 b_3^1 = 4/5$, $f_2^3 b_3^1 = 0$ and $f_3^3 b_3^1 = 1/5$. Therefore, $b_3^1, f_1^3, f_3^3 \neq 0$ and $f_2^3 = 0$. From $g_1^3 = 0$ and $f_1^3 \neq 0$, we have $b_3^3 = 0$. Likewise, from $g_3^2 = 0$ and $f_3^3 > 0$, $b_3^2 = 0$. Then, it gives us $b_3^1 = 1$.

From $b_1^1 = 0, b_2^1 = 0$ and $b_3^1 = 1$, we have $f_1^3 = 4/5$ and $f_3^3 = 1/5$. From $f_3^1 = 0$, $f_3^2 = 4/5$. From $g_3^2 = 0$ and $f_3^2 > 0$, we have $b_2^2 = 0$. It gives us $b_2^3 = 1$. Therefore, B is a permutation matrix and f is a permutation of g.

2. $b_1^1 > 0$ **and** $f_2^1 = 0$**:** Observe that $f_2^3 b_3^1 = 0$, $f_2^2 b_2^2 + f_2^3 b_3^2 = 4/5$ and $f_2^2 b_2^3 + f_2^3 b_3^3 = 1/5$. We consider two subcases: $f_2^3 = 0$ or $b_3^1 = 0$.

(a) $f_2^3 = 0$**:** From $f_2^1 = f_2^3 = 0$, we have $f_2^2 = 1$. Additionally, we have $b_2^2 = 4/5$ and $b_2^3 = 1/5$. From $0 = g_1^3 = f_1^2 b_2^3 + f_1^3 b_3^3$ and $0 = g_3^2 = f_3^1 b_1^2 + f_3^2 b_2^2 + f_3^3 b_3^2$, we have $f_1^2 = f_3^2 = 0$. Observe that $0 = g_1^3 = f_1^3 b_3^3$ and $4/5 = g_3^3 = f_3^3 b_3^3$. Then, we have $b_3^3 > 0$ and $f_1^3 = 0$. From $f_1^2 = f_1^3 = 0$, we have $f_1^1 = 1$. This also gives $b_1^1 = 4/5$ and $b_1^2 = 1/5$. Again from $0 = g_3^2$ and $b_1^2 = 1/5$, we have $f_3^1 = 0$. Therefore, from $f_3^1 = f_3^2 = 0$, we have $f_3^3 = 1$, i.e., f is I_3.

(b) $b_3^1 = 0$**:** From $b_2^1 = b_3^1 = 0$, we have $f_1^1 \cdot b_1^1 = 4/5$ and $f_3^1 \cdot b_1^1 = 1/5$. Therefore, $b_1^1, f_1^1, f_3^1 > 0$. Next, $0 = g_3^2 = f_3^1 b_1^2 + f_3^2 b_2^2 + f_3^3 b_3^2$ gives $b_1^2 = 0$. From $b_1^2 = b_1^3 = 0$, we have $b_1^1 = 1$.

Suppose that both b_2^2 and b_3^2 are positive. Then, from $0 = g_3^2 = f_3^2 b_2^2 + f_3^3 b_3^2$, we have $f_3^2 = f_3^3 = 0$. It contradicts $4/5 = g_3^3 = f_3^1 b_1^3 + f_3^2 b_2^3 + f_3^3 b_3^3$ since $b_1^3 = f_3^2 = f_3^3 = 0$. Therefore, at least one of b_2^2 and b_3^2 is equal to zero. Likewise, if both b_2^3 and b_3^3 are positive, we have $f_1^2 = f_1^3 = 0$ from $g_1^3 = 0$, but it contradicts $g_1^2 = 1/5 > 0$. Thus, at least one of b_2^3 and b_3^3 is equal to zero. Also, note that B needs to be a full rank matrix (as g has a full rank). To have that, there are two possibilities: (i) $b_2^2 = b_3^3 = 1$ and $b_2^3 = b_3^2 = 0$; or (ii) $b_2^3 = b_3^2 = 1$ and $b_2^2 = b_3^3 = 0$. Then, B is either I_3 or a permutation of I_3. Therefore, f is g or a permutation of g. □

OA.3.2 Proof of Proposition 5

OA.3.2.1 Overview

In this section, we provide the high-level overview of the proof of Proposition 5. Our goal is to show that the following Blackwell-monotone cost function (defined in Example 1) is not a monotone transformation of any GSLS (and thus PS) cost function:

$$C(f_1, f_2) = \left(\frac{f_2}{f_1} - 1\right)^2 \left(1 - \frac{1 - f_2}{1 - f_1}\right).$$

Recall for any convex function $\phi : [0, \infty] \to \mathbb{R} \cup \{+\infty\}$ with $\phi(1) = 0$, the ϕ-divergence between the distributions in state ω_1 and ω_2 is given by

$$D_\phi(f_2 \| f_1) = f_1 \phi\left(\frac{f_2}{f_1}\right) + (1 - f_1)\phi\left(\frac{1 - f_2}{1 - f_1}\right).$$

Denti et al. (2022b) establish the equivalence between ϕ-divergence and GSLS costs (Proposition 39).

In the next subsection, we provide a necessary condition for being a monotone transformation of ϕ-divergence. In Section OA.3.2.3, we show that C_1 does not satisfy this necessary condition. Therefore, C_1 is not a monotone transformation of any GSLS cost function.

OA.3.2.2 Necessary Condition for ordinal ϕ-divergence

Let $\Omega = \{\omega_1, \omega_2\}$ and we consider only experiments with $f_2 > f_1 > 0$. To facilitate the proof, we introduce a change of variable: Let $\hat{\alpha} \equiv f_2/f_1 - 1 \in (0, \infty)$ and $\hat{\beta} \equiv$

$1-(1-f_2)/(1-f_1) \in (0,1)$ and define

$$\hat{C}(\hat{\alpha},\hat{\beta}) = C\left(\frac{\hat{\beta}}{\hat{\alpha}+\hat{\beta}}, \frac{(1+\hat{\alpha})\hat{\beta}}{\hat{\alpha}+\hat{\beta}}\right).$$

By algebra, it can be shown that if C is differentiable and Blackwell monotone, then

$$\begin{aligned}\frac{\partial \hat{C}(\hat{\alpha},\hat{\beta})}{\partial \hat{\alpha}} &= \frac{1}{\hat{\alpha}+\hat{\beta}}\left[\frac{f_2-f_1}{1-f_1}\frac{\partial C}{\partial f_2} - f_1\frac{\partial C}{\partial f_2} - f_2\frac{\partial C}{\partial f_2}\right] \\ &= \frac{1}{\hat{\alpha}+\hat{\beta}}\left[\frac{f_1 f_2 - f_1}{1-f_1}\frac{\partial C}{\partial f_2} - f_1\frac{\partial C}{\partial f_1}\right] \geq 0,\end{aligned}$$

where the last inequality follows by noticing that $\frac{\partial C}{\partial f_2} \geq 0$ when $f_2 > f_1$ and plugging in $(1-f_1)\frac{\partial C}{\partial f_1} + (1-f_2)\frac{\partial C}{\partial f_2} \leq 0$. Similarly, we can show that

$$\frac{\partial \hat{C}(\hat{\alpha},\hat{\beta})}{\partial \hat{\beta}} \geq 0.$$

Let C be strictly Blackwell monotone, i.e., the inequalities hold strictly.

Next, we apply the same change of variables to $D_\phi(f_2 \| f_1)$:

$$D_\phi(f_2\|f_1) = \frac{\hat{\beta}}{\hat{\alpha}+\hat{\beta}}\phi(\hat{\alpha}+1) + \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}}\phi(1-\hat{\beta}),$$

for some convex function $\phi : [0,\infty] \to \mathbb{R}\cup\{+\infty\}$ with $\phi(1)=0$. By Proposition 7.2 in Polyanskiy and Wu (2022), any such ϕ is equivalent, in terms of ϕ-divergence, to another ϕ with $\phi(\cdot) \geq 0$. Thus, it is without loss of generality to assume $\phi(\cdot) \geq 0$. We further define

$$\begin{aligned}\phi_+(\hat{\alpha}) &\equiv \phi(\hat{\alpha}+1) = \phi\left(\frac{f_2}{f_1}\right), \\ \phi_-(\hat{\beta}) &\equiv \phi(1-\hat{\beta}) = \phi\left(\frac{1-f_2}{1-f_1}\right).\end{aligned}$$

Notice that $\phi_+(0) = 0$, $\phi_+(\cdot)$ is convex and strictly increasing if and only if $\phi^+(\cdot) > 0$ (That is, ϕ^+ can be a constant only in the interval $[0,\alpha_0]$ for some $\alpha_0 \in [0,\infty]$). Similarly, $\phi_-(0)=0$, $\phi_-(\cdot)$ is convex and strictly increasing if and only if $\phi_-(\cdot) > 0$. Moveover, by convexity, both ϕ_+ and ϕ_- are continuous in the interior of their domains. Thus, we define and rewrite

$$D_\phi(\hat{\alpha},\hat{\beta}) \equiv D_\phi(f_2\|f_1) = \frac{\hat{\beta}}{\hat{\alpha}+\hat{\beta}}\phi_+(\hat{\alpha}) + \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}}\phi_-(\hat{\beta}).$$

For each $c \in \mathbb{R}_+$, define the (interior) level set of $\hat{C}$ by

$$\mathcal{L}_{\hat{C}}(c) \equiv \left\{(\alpha,\beta) \in (0,\infty)\times(0,1) : \hat{C}(\hat{\alpha},\hat{\beta}) = c\right\},$$

and let $\mathcal{L}_{\hat{C}}$ denote the collection of all its level sets.[21] Similarly, let $\mathcal{L}_\phi$ denote the collection of all level sets of $D_\phi(\hat{\alpha},\hat{\beta})$.

Lemma OA.3.1. *Suppose $\hat{C}(\hat{\alpha},\hat{\beta})$ is strictly increasing in $\hat{\alpha}$ and $\hat{\beta}$. If it can be represented by a monotone transformation of $D_\phi(\hat{\alpha},\hat{\beta})$, then $\mathcal{L}_{\hat{C}} = \mathcal{L}_\phi$.*

Proof of Lemma OA.3.1. If $\hat{C}(\hat{\alpha},\hat{\beta})$ is a monotone transformation of $D_\phi(\hat{\alpha},\hat{\beta})$, then for all $(\hat{\alpha},\hat{\beta})$ and $(\hat{\alpha}',\hat{\beta}')$, we have

$$\hat{C}(\hat{\alpha},\hat{\beta}) > \hat{C}(\hat{\alpha}',\hat{\beta}') \Rightarrow D_\phi(\hat{\alpha},\hat{\beta}) > D_\phi(\hat{\alpha}',\hat{\beta}'), \tag{OA.3.1}$$

and

$$D_\phi(\hat{\alpha},\hat{\beta}) = D_\phi(\hat{\alpha}',\hat{\beta}') \Rightarrow \hat{C}(\hat{\alpha},\hat{\beta}) = \hat{C}(\hat{\alpha}',\hat{\beta}'). \tag{OA.3.2}$$

When $\hat{C}(\hat{\alpha},\hat{\beta})$ is strictly increasing in $\hat{\alpha}$ and $\hat{\beta}$, (OA.3.1) implies $D_\phi(\hat{\alpha},\hat{\beta})$ is also strictly increasing in $\hat{\alpha}$ and $\hat{\beta}$. (OA.3.2) implies that, for each $c \in \mathbb{R}_+$, if $\mathcal{L}_{\hat{C}}(c)$ is non-empty, then there exists $c' \in \mathbb{R}_+$ such that $\mathcal{L}_\phi(c') \subseteq \mathcal{L}_{\hat{C}}(c)$. Then it remains to show that, for all such c and c', it is the case that $\mathcal{L}_{\hat{C}}(c) \subseteq \mathcal{L}_\phi(c')$. , i.e., $\mathcal{L}_{\hat{C}}(c) = \mathcal{L}_\phi(c')$.

Towards a contradiction, suppose there exists $(\hat{\alpha},\hat{\beta})$ and $(\hat{\alpha}',\hat{\beta}')$ such that $\hat{C}(\hat{\alpha},\hat{\beta}) = \hat{C}(\hat{\alpha}',\hat{\beta}')$ but $D_\phi(\hat{\alpha},\hat{\beta}) > D_\phi(\hat{\alpha}',\hat{\beta}')$. By strict monotonicity of $\hat{C}$, it cannot be the case that either $(\hat{\alpha},\hat{\beta}) \leq (\hat{\alpha}',\hat{\beta}')$ or $(\hat{\alpha},\hat{\beta}) \geq (\hat{\alpha}',\hat{\beta}')$. Then without loss of generality, let $\hat{\alpha} < \hat{\alpha}'$ and $\hat{\beta} > \hat{\beta}'$. By continuity and strict monotonicity of D_ϕ, there always exists $(\hat{\alpha}'',\beta'')$ such that $\hat{\alpha}'' > \hat{\alpha}'$, $\hat{\beta}'' > \hat{\beta}'$, and $D_\phi(\hat{\alpha}'',\hat{\beta}'') = D_\phi(\hat{\alpha},\hat{\beta})$. Then by (OA.3.2), we have $\hat{C}(\hat{\alpha}'',\hat{\beta}'') = \hat{C}(\hat{\alpha},\hat{\beta})$. However, by strict monotonicity of $\hat{C}$, we must have $\hat{C}(\hat{\alpha}'',\hat{\beta}'') > \hat{C}(\hat{\alpha}',\hat{\beta}')$, a contradiction. □

For any continuous function $f : \mathbb{R}^2 \to \mathbb{R}$ that is strictly increasing in both variables, Corollary 1 in Sajbura (2005) says that all its non-empty level sets are "arcs" with an empty interior. This further implies that, if f is differentiable at point $(\hat{\alpha},\hat{\beta})$, then its gradient

[21] Since we are deriving a necessary condition here, it is valid to consider only interior points in the domain.

$\nabla f(\hat{\alpha}, \hat{\beta})$ is perpendicular to the tangent hyperplane of the level set $\mathcal{L}_f(f(\alpha, \beta))$ at point $(\hat{\alpha}, \hat{\beta})$. Notice this applies to both $\hat{C}$ and the D_ϕ that represents it as $\hat{C}$ is differentiable and D_ϕ is differentiable almost everywhere. Therefore, their gradients must be in the same direction almost everywhere in the interior of their domains. From this, we can derive our necessary conditions:

Lemma OA.3.2. *Suppose $\hat{C}(\hat{\alpha}, \hat{\beta})$ is strictly increasing in $\hat{\alpha}$ and $\hat{\beta}$. If it can be represented by a monotone transformation of $D_\phi(\hat{\alpha}, \hat{\beta})$, then let $K(\hat{\alpha}, \hat{\beta}) \equiv \frac{\hat{\alpha}\partial\hat{C}/\partial\hat{\alpha}}{\hat{\beta}\partial\hat{C}/\partial\hat{\beta}}$, we have for almost every $(\hat{\alpha}, \hat{\beta})$,*

$$\frac{\phi_+(\hat{\alpha}) - \phi_-(\hat{\beta})}{\hat{\alpha} + \hat{\beta}} = \frac{\phi'_+(\hat{\alpha}) + K(\hat{\alpha}, \hat{\beta})(-\phi'_-(\beta))}{1 + K(\hat{\alpha}, \hat{\beta})}. \tag{OA.3.3}$$

Proof of Lemma OA.3.2. Suppose $\hat{C}(\hat{\alpha}, \hat{\beta})$ can be represented by a monotone transformation of D_ϕ. Then notice that

$$\frac{\partial D_\phi(\alpha, \hat{\beta})}{\partial \hat{\alpha}} = \frac{\hat{\beta}}{(\hat{\alpha} + \hat{\beta})^2} \cdot \left[\phi_-(\hat{\beta}) - (\phi_+(\hat{\alpha}) - \phi'_+(\hat{\alpha})(\hat{\alpha} + \hat{\beta}))\right] > 0,$$
$$\frac{\partial D_\phi(\hat{\alpha}, \hat{\beta})}{\partial \hat{\beta}} = \frac{\hat{\alpha}}{(\hat{\alpha} + \hat{\beta})^2} \cdot \left[\phi_+(\hat{\alpha}) - (\phi_-(\hat{\beta}) - \phi'_-(\hat{\beta})(\hat{\alpha} + \hat{\beta}))\right] > 0.$$

Then (OA.3.3) can be derived from

$$\frac{\partial \hat{C}/\partial \hat{\alpha}}{\partial \hat{C}/\partial \hat{\beta}} = \frac{\hat{\beta}\left[\phi_-(\hat{\beta}) - (\phi_+(\hat{\alpha}) - \phi'_+(\hat{\alpha})(\hat{\alpha} + \hat{\beta}))\right]}{\hat{\alpha}\left[\phi_+(\hat{\alpha}) - (\phi_-(\hat{\beta}) - \phi'_-(\hat{\beta})(\hat{\alpha} + \hat{\beta}))\right]}.$$

□

OA.3.2.3 Proof

Proof of Proposition 5. Observe that C_1 in Example 1 can be rewritten as $\hat{C}(\hat{\alpha}, \hat{\beta}) = \hat{\alpha}^2\hat{\beta}$. It is strictly increasing in $\hat{\alpha}$ and $\hat{\beta}$ and is differentiable. Notice that for this cost function,

$$K(\hat{\alpha}, \hat{\beta}) = \frac{\hat{\alpha}\partial\hat{C}/\partial\hat{\alpha}}{\hat{\beta}\partial\hat{C}/\partial\hat{\beta}} = \frac{\hat{\alpha} \cdot 2\hat{\alpha}\hat{\beta}}{\hat{\beta} \cdot \hat{\alpha}^2} = 2.$$

Then plug $K(\hat{\alpha}, \hat{\beta}) = 2$ into the differential equation (OA.3.3), we can solve that

$$\phi_+(\hat{\alpha}) = \phi_-(\hat{\beta}) - (\hat{\alpha} + \hat{\beta})\phi'_-(\hat{\beta}) + \kappa_1(\hat{\alpha} + \hat{\beta})^3.$$

for some constant κ_1. For this $\phi_+(\hat{\alpha})$ to be part of a convex function ϕ, it cannot depend on $\hat{\beta}$, i.e., $d\phi_+(\hat{\alpha})/d\hat{\beta} = 0$. Thus, we can further derive the following differential equation,

$$(\hat{\alpha} + \hat{\beta})(3(\hat{\alpha} + \hat{\beta})\kappa_1 - \phi''_-(\hat{\beta})) = 0.$$

Solving this differential equation gives

$$\phi_-(\hat{\beta}) = \frac{1}{2}\hat{\beta}^2(3\hat{\alpha} + \hat{\beta})\kappa_1 + \hat{\beta}\kappa_2 + \kappa_3.$$

for some κ_2 and κ_3. Notice that for $\phi_-(\hat{\beta})$ to not depend on $\hat{\alpha}$ and $\phi_-(0) = 0$, one needs to have $\kappa_1 = \kappa_3 = 0$, i.e., $\phi_-(\hat{\beta}) = \hat{\beta}\kappa_2$. Then by plugging $\phi_-(\hat{\beta})$ back to $\phi_+(\hat{\alpha})$, we have

$$\phi_+(\hat{\alpha}) = -\kappa_2 \cdot \hat{\alpha}.$$

Notice that there is no κ_2 such that both ϕ_+ and ϕ_- are increasing, thus there is no ϕ-divergence cost function that can represent this cost function. □

OA.3.3 Bargaining with Information Acquisition

OA.3.3.1 Lemma

Before proving Proposition 6 and 7, we first show the following lemma which characterizes the buyer's optimal information acquisition strategy (we break indifference towards accepting the offer).

Lemma OA.3.3. *For all $p \in [L, H]$ and $\mu \in (0, 1)$,*

(i) If $C_\lambda(f) = \lambda(f_H - f_L)^2/2$, then the buyer's optimal strategy satisfies either $f^*_H = 1$ or $f^*_L = 0$, or both.

(ii) If $C_\lambda(f) = \lambda|f_H - f_L|$, then the buyer's optimal strategy is either full information acquisition, i.e., $[f^*_L, f^*_H] = [0, 1]$ or no information acquisition, i.e., $[f^*_L, f^*_H] = [0, 0]$ or $[1, 1]$.

Notice that when $p \notin [L, H]$ or $\mu = 0$ or 1, the buyer's optimal strategy is always no information acquisition.

Proof of Lemma OA.3.3. Recall the buyer's problem is to solve, for all p and μ,

$$\max_{[f_L, f_H]^\intercal \in \mathcal{E}_2} \mu f_H(H-p) + (1-\mu) f_L(L-p) - C_\lambda(f),$$

under which the buyer accepts the offer after observing the signal h and rejects otherwise.

First, observe that for all $\mu \in (0,1)$, if $p = L$, then the buyer's optimal strategy is no information acquisition and always accepts, i.e., $f_L^* = f_H^* = 1$; If $p = H$, then the buyer's optimal strategy is no information acquisition and always rejects, i.e., $f_L^* = f_H^* = 0$.

Next, we consider the case where $p \in (L, H)$. Consider the auxiliary cost minimization problem that solves the minimum cost needed to achieve a given material payoff level w:

$$\min_{[f_L, f_H]^\intercal \in \mathcal{E}_2} C_\lambda(f) \quad \text{s.t.} \quad \mu f_H(H-p) + (1-\mu) f_L(L-p) = w.$$

The feasible levels of w is given by $[\underline{w}, \overline{w}]$ where $\underline{w} = \max\{0, \mu(H-p) + (1-\mu)(L-p)\}$ and $\overline{w} = \mu(H-p)$. That is, $\underline{w}$ is the optimal payoff level under no information, and $\overline{w}$ is the optimal payoff level under full information.

Given $w \in [\underline{w}, \overline{w}]$, the set of experiments that achieve w is given by the line segment

$$f_H = -\frac{(1-\mu)(L-p)}{\mu(H-p)} f_L + \frac{w}{\mu(H-p)}.$$

Because $p \in (L, H)$ and $\mu \in (0,1)$, the slope of this segment is positive and the intercept $\frac{w}{\mu(H-p)}$ is non-negative.

Notice that for both cost functions, their isocost curves are the same and in the form of $f_H - f_L = c$. Therefore, the buyer's optimal experiment must be on the boundary with either $f_H^* = 1$ or $f_L^* = 0$. Specifically,

$$[f_L^*, f_H^*] = \begin{cases} [\frac{\mu(H-p)-w}{(1-\mu)(p-L)}, 1] & \text{if } \mu(H-p) + (1-\mu)(L-p) \geq 0, \\ [0, \frac{w}{\mu(H-p)}] & \text{if } \mu(H-p) + (1-\mu)(L-p) < 0. \end{cases} \tag{OA.3.4}$$

This proves (i).

For (ii), let $C_\lambda(f) = \lambda |f_H - f_L|$. By (OA.3.4), it is without loss to restrict attention to experiments with $f_H \geq f_L$ and thus the cost function can be simplified to $C_\lambda(f) = \lambda(f_H - f_L)$.

Let $[f_L^*(w), f_H^*(w)]^\intercal$ denote the optimal experiment that achieves payoff w solved from

(OA.3.4). The buyer's problem is then to solve

$$\max_{w\in[\underline{w},\overline{w}]} w - \lambda(f_H^*(w) - f_L^*(w)).$$

Notice that, both $f_L^*(w)$ and $f_H^*(w)$ are linear in w. We conclude that the optimal w is either $\underline{w}$, under which $[f_L^*, f_H^*] = [1, 1]$ or $[0, 0]$; or $\overline{w}$, under which $[f_L^*, f_H^*] = [0, 1]$. $\square$

OA.3.3.2 Proof of Proposition 6

Proof of Proposition 6. Let

$$C_\lambda(f) = \frac{\lambda}{2}(f_H - f_L)^2.$$

Then notice that the buyer's optimal information acquisition strategy solves

$$\max_{w\in[\underline{w},\overline{w}]} w - \frac{\lambda}{2}(f_H^*(w) - f_L^*(w))^2.$$

Suppose $\mu(H-p) + (1-\mu)(L-p) \geq 0$ and substituting (OA.3.4), we have

$$\max_{w\in[\underline{w},\overline{w}]} w - \frac{\lambda}{2}\left(1 - \frac{\mu(H-p) - w}{(1-\mu)(p-L)}\right)^2.$$

From this we can solve that the optimal w^* is given by

$$w^* = \begin{cases} \overline{w} & \text{if } (1-\mu)(p-L) \geq \lambda, \\ \underline{w} + \frac{(1-\mu)^2(L-p)^2}{\lambda} & \text{if } (1-\mu)(p-L) < \lambda. \end{cases}$$

Notice in the second case, the buyer acquires information in the optimal strategy. Specifically, the optimal experiment is given by $f_H^* = 1$ and

$$f_L^* = 1 - \frac{(1-\mu)(p-L)}{\lambda}.$$

The low-type seller is indifferent between offering p and L if and only if

$$p\left(1 - \frac{(1-\mu)(p-L)}{\lambda}\right) = L,$$

which is equivalent to

$$p(1-\mu) = \lambda.$$

In summary, we claim that there is a non-pooling equilibrium where the buyer acquires information when there exists p and μ such that

$$p(1-\mu)=\lambda, \mu H \geq p, \text{ and } \mu > \pi.$$

In this case, let $\sigma(H)=\delta_p$, $\sigma(L)(p)=\frac{\pi(1-\mu)}{\mu(1-\pi)}$ and $\sigma(L)(L)=1-\sigma(L)(p)$. These are well-defined as $\mu > \pi$. Then on the equilibrium path, if the buyer observes price p, the buyer's belief is exactly μ, and the other two conditions imply that

$$\mu(H-p) \geq p(1-\mu) = \lambda \geq (1-\mu)(p-L).$$

The previous discussions thus imply that in this case, the buyer's optimal information acquisition exactly makes the low-type seller indifferent between offering p and L.

Finally, we show that it is always possible to find p and μ such that

$$p(1-\mu)=\lambda, \mu H \geq p, \text{ and } \mu > \pi,$$

when $\lambda < \pi(1-\pi)H$. Letting $p=\lambda/(1-\mu)$, notice the second condition implies

$$\lambda \leq \mu(1-\mu)H.$$

As $\lambda < \pi(1-\pi)H$, by continuity, one can always find $\mu > \pi$ such that the above conditions hold. □

OA.3.3.3 Proof of Proposition 7

Proof of Proposition 7. By Lemma OA.3.3, it is without loss to focus on experiments with $f_H \geq f_L$. In this case, we have

$$C_\lambda(f)=\lambda(f_H-f_L).$$

We first show that such pooling equilibria are possible. Suppose both types of sellers offer price $p^* \geq L$. Then by Lemma OA.3.3, the buyer's optimal strategy is $[f_L^*, f_H^*]=[1,1]$ if and only if

$$\pi(H-p^*)+(1-\pi)(L-p^*) \geq 0,$$

and

$$\pi(H - p^*) + (1-\pi)(L - p^*) \geq \pi(H - p^*) - \lambda,$$

where the first condition follows from (OA.3.4) and the second follows from achieving $\underline{w} = \pi(H - p^*) + (1-\pi)(L - p^*) \geq 0$ is more optimal than achieving $\overline{w}$. From these two conditions, one can derive that $p^* \leq L + \epsilon$ where

$$\epsilon = \min\left\{\pi(H - L), \frac{\lambda}{1-\pi}\right\}. \tag{OA.3.5}$$

This can be supported as a PBE by letting the buyer's off-path belief satisfy $\mu_p = 0$ for all $p \neq p^*$ and thus the buyer accepts the offer only when $p \leq L$. This gives the seller a worse payoff than offering p^*.

Next, we argue that there cannot be any separating equilibria. Suppose there is a separating equilibrium where the two types of sellers offer different prices $H \geq p_H > p_L \geq L$. In this case, the buyer will not acquire any information and always accepts the offers. Then the low-type seller can profitably deviate by offering p_H instead of p_L, a contradiction.

Finally, we argue that there cannot be any equilibria where any type of seller mixes between two different prices. If the buyer always accepts both offers, then the seller will not be indifferent between these two prices, a contradiction. Thus, the only possibility is that the buyer does not accept with probability 1 under one of the offers, i.e., the buyer acquires information. By Lemma OA.3.3, the buyer must acquire full information and accept only the high-type seller's offer. As a result, neither type of seller would be indifferent between the two offers they possibly randomize, a contradiction.

For the last statement of the proposition, notice (OA.3.5) implies that $\epsilon \to 0$ as $\lambda \to 0$. □

OA.3.4 Costly Persuasion

OA.3.4.1 Proof of Lemma 3

Proof of Lemma 3. Consider α and β defined in Section 3.1.2 as functions of f_1 along the line $\mu f_2 + (1-\mu) f_1 = w$:

$$\alpha(f_1) \equiv \frac{f_2}{f_1} = \frac{w - (1-\mu)f_1}{\mu \cdot f_1}, \quad \beta(f_1) \equiv \frac{1-f_1}{1-f_2} = \frac{\mu(1-f_1)}{\mu - w + (1-\mu)f_1}.$$

By taking derivatives and using $w \leq 2\mu < 1$, we have

$$\alpha'(f_1) = -\frac{w}{\mu f_1^2} < 0, \quad \beta'(f_1) = -\frac{(1-w)\mu}{(\mu - w + (1-\mu)f_1)^2} < 0.$$

Then, from (4), Blackwell informativeness decreases along the line as f_1 increases. Then, from Blackwell monotonicity, C is minimized when f_1 is maximized. To achieve this maximization, (15) needs to bind. □

OA.3.4.2 Proof of Proposition 8

Proof of Proposition 8. For $\mu \geq 1/2$, we show that $f_1 = f_2 = 1$ is optimal in the main text.

Now assume that $\mu < 1/2$. By plugging the cost function in, (18) is equivalent to

$$\max_{0 \leq w \leq 2\mu} w - \frac{w^2}{4\mu \cdot h(\mu)} \qquad \text{where} \quad h(\mu) \equiv \frac{\mu(1-\mu)^2}{(1-2\mu)^2}. \tag{OA.3.6}$$

Observe that for all $0 < \mu < 1/2$

$$h'(\mu) = \frac{2\mu + (1-2\mu)(1+\mu^2)}{(1-2\mu)^3} > 0.$$

Additionally, $h(0) = 0$ and $\lim_{\mu \to 1/2} h(\mu) = \infty$. Therefore, there exists $\hat{\mu}$ such that $h(\hat{\mu}) = 1$. Then, the solution of the minimization problem (OA.3.6) subject to $0 \leq w \leq 2\mu$ is

$$w^* = \begin{cases} 2\mu, & \text{if } \mu \in (\hat{\mu}, 1/2), \\ 2\mu \cdot h(\mu), & \text{if } \mu \leq \hat{\mu}. \end{cases}$$

By plugging this into (17), we have (19). □