Blackwell-Monotone Information Costs*

Xiaoyu Cheng† Yonggyun Kim‡

April 23, 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 Blackwell monotonicity over finite experiments. The key condition is a system of linear differential inequalities that are convenient to check given an arbitrary cost function. When the cost function is additively separable across signals, our characterization implies that Blackwell monotonicity is equivalent to sublinearity. This identifies a wide range of practical information cost functions. Finally, we apply our results to bargaining and persuasion problems with costly information.

JEL Classification: C78, D81, D82, D83

Keywords: Blackwell Information Order, Costly Information, Additively Separable Cost, Bargaining, Costly Persuasion

---

*We thank Teddy Kim, R.Vijay Krishna, Kun Zhang, and seminar participants at FSU and KAEA-VSS Micro Seminar 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. This raises the question of which information cost function *should* or *could* be used. One common principle, shared by almost all recent developments of information cost functions,[1] 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). While Blackwell monotonicity is recognized as a desirable feature and has been extensively studied, it is usually analyzed in combination with other properties. Therefore, the implications and requirements of *solely* imposing Blackwell monotonicity on information costs remain underexplored.

In this paper, we characterize simple necessary and sufficient conditions of Blackwell monotonicity for information cost functions defined over finite experiments.[2] This characterization provides a tractable method for verifying Blackwell monotonicity given an arbitrary information cost function. Additionally, it enables us to construct a novel and broad class of Blackwell-monotone information cost functions that are additively separable across signals. Notably, this class encompasses well-known existing information costs such as all cost functions mentioned in footnote 1.

Our characterization is motivated by the garbling interpretation of the Blackwell order. Blackwell (1951, 1953) shows that for any Bayesian decision problem, the expected payoff under a statistical experiment f is no less than that under another experiment g if and only if g can be generated by adding some noise to f. That is, g is a garbling of f. If this condition holds, f is called to be Blackwell more informative than g.

Consider the Euclidean subspace of statistical experiments with n states and m signals, and a Blackwell-monotone information cost function C defined over this space. An experiment f in this space can be represented as an $n \times m$ matrix, where each entry f_{ij} denotes the probability of generating signal s_j when the state is ω_i. Now, imagine the following garbling process: whenever signal s_j is generated, s_j is replaced by another signal s_k with probability ϵ, while the generation of other signals remains unchanged. Since this process worsens Blackwell informativeness, any Blackwell-monotone cost function should result in a lower cost after its application. By sending ϵ to zero, we can derive a local necessary

[1] These include entropy costs (Sims, 2003; Matějka and McKay, 2015), posterior separable costs (Caplin et al., 2022; Denti, 2022), and log-likelihood ratio (LLR) costs (Pomatto et al., 2023), among many others.

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

condition for Blackwell monotonicity when the following limit exists:

$$D^+C(f; f^j(k)) \equiv \lim_{\epsilon \downarrow 0} \frac{C(f + \epsilon \cdot f^j(k)) - C(f)}{\epsilon} \leq 0, \quad \forall j \neq k, \tag{$*$}$$

where f^j is the j-th column vector of f and $f^j(k) = [\mathbf{0}, \cdots, -f^j, \cdots, f^j, \cdots, \mathbf{0}]$ denotes an $n \times m$ matrix where j-th column is $-f^j$, k-th column is f^j, and all other entries are zero. Notice $f^j(k)$ represents the direction of the aforementioned garbling. When C is differentiable at f, ($*$) reduces to simple first-order inequalities:

$$\sum_{i=1}^{n} \left(\frac{\partial C}{\partial f_{ik}} - \frac{\partial C}{\partial f_{ij}} \right) f_{ij} \leq 0, \quad \forall j \neq k.$$

Our main results establish that this local condition is not only necessary but also sufficient for Blackwell monotonicity when combined with additional technical conditions. When the signal is binary, under Lipschitz continuity, we show that an information cost function is Blackwell monotone if and only if it satisfies ($*$) and permutation invariance, i.e., relabeling signals results in the same cost (Theorem 1). When there are more than two signals, we additionally impose quasiconvexity and show that the local condition and permutation invariance serve as necessary and sufficient conditions for Blackwell monotonicity (Theorem 2).[3]

The local condition ($*$) becomes especially useful when the cost function is additively separable across signals, i.e., when there exists $\psi : [0,1]^n \to \mathbb{R}_+$ such that $C(f) = \sum_{j=1}^{m} \psi(f^j)$. Using our characterization result, we show that an additively separable cost function is Blackwell monotone if and only if its component function ψ is sublinear, i.e., positively homogeneous and subadditive (Theorem 3). This result enables the construction of Blackwell-monotone information cost functions by simply choosing a sublinear function ψ. As a demonstration, we identify novel Blackwell-monotone information cost functions with simple functional forms, the norm costs and absolute-linear costs.

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 the existing insights.

[3]Quasiconvexity is imposed to address a technical issue that arises when extending to the case of more than two signals (see Proposition 2 for details).

We first consider Chatterjee et al. (2024)'s ultimatum bargaining model where the buyer can acquire costly information about the unknown value of the seller's object before accepting an offer. To model costly information, they 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 their main result, its main insight remains true in the general setting when considering a different Blackwell-monotone information cost function (Proposition 4, 5).

As another application, we consider the costly persuasion problem proposed by Gentzkow and Kamenica (2014). They extended 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 posterior separable, and the literature follows this tradition. We propose another way of solving the costly persuasion problem (without using concavification) that can be applicable to any Blackwell-monotone information costs. By using this, we provide a solution for the standard 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 information design and costly information acquisition literature.[4] Therefore, it has brought more attention 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 to the introduction of Denti et al. (2022) for a formal discussion.

Ever since Sims (2003) introduced the entropy cost to the economic literature, posterior-based information costs have gained widespread recognition. Blackwell monotonicity of such costs (with a concave measure of uncertainty, e.g., entropy) is implied by one of Black-

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

well's sufficient condition.[5] 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 found that their results derived from posterior-based costs can qualitatively change when experiment-based costs are employed instead (Denti et al., 2022; Ramos-Mercado, 2023). Furthermore, as Mensch (2018) and Denti et al. (2022) point out, experiment-based costs may be more compatible with applications where priors are endogenously determined.

These observations highlight the importance of determining which experiment-based costs ought to be utilized, with Blackwell monotonicity in consideration. As a pioneering work in this direction, Pomatto et al. (2023) characterize the log-likelihood ratio (LLR) cost with Blackwell monotonicity and several other axioms. We contribute to the literature on experiment-based costs by deriving necessary and sufficient conditions for Blackwell monotonicity alone.[6]

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); Caplin et al. (2022); and Denti (2022). Our paper differs methodologically from these papers 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 additively separable costs. Section 5 presents two applications. Section 6 provides additional discussions and Section 7 concludes. All omitted proofs can be found in the Appendix.

[5] See, for example, the discussion of Assumption 1 in Gentzkow and Kamenica (2014).

[6] Additionally, the LLR cost is not defined over experiments that fully reveal some state as it would incur an infinite cost. Our characterization does not have this limitation, and it turns out that such boundary experiments play important roles in our applications.

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

$$f = \begin{bmatrix} f_{11} & \cdots & f_{1m} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nm} \end{bmatrix},$$

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

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

and sometimes write f_{ij} as f_i^j. Notice that $\sum_{j=1}^m f^j = \mathbf{1}$ where $\mathbf{1} = [1, \cdots, 1]^\intercal$. Let $\mathcal{E}_m \subset \mathbb{R}^{n\times m}$ denote the space of all experiments with m possible signals, equipped with the Euclidean topology. It is worth noting that, for all $m' \leq m$, $\mathcal{E}_{m'}$ can be embedded into $\mathcal{E}_m$. Let $\mathcal{E} = \bigcup_{m<\infty} \mathcal{E}_m$ denote the *class* of all (finite) experiments.[7]

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*.

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 $\textbf{ext}(\mathcal{M}_m) = \{E_1, \cdots, E_{m^m}\}$ denote the set of all extreme points of $\mathcal{M}_m$. For any $k \leq m$, let $\textbf{ext}_k(\mathcal{M}_m)$ 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 $\textbf{ext}_m(\mathcal{M}_m) = \mathcal{P}_m$. Since when $P \in \mathcal{P}$, P^{-1} is also a permutation matrix, we have $f \succeq_B fP \succeq_B fPP^{-1} = f$. Therefore, f and fP are equivalent in Blackwell informativeness, denoted by $f \simeq_B fP$. Intuitively, permuting an experiment involves merely relabeling signals, so it should remain equally informative after permutation.

[7] The terminology of class follows from Pomatto et al. (2023).

Information Costs An *information cost function* is given by $C : \mathcal{E}_m \to \mathbb{R}_+$, i.e., defined over the space of experiments with m signals. Let $\mathcal{C}_m$ denote the set of all such functions. Under this formalization, each $C \in \mathcal{C}_m$ is a mapping over an 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 every $\mathcal{E}_m$.

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 exists. 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 of generality to let $\nabla^m C(f) = 0$ if needed.

Our results require a few standard functional assumptions: $C \in \mathcal{C}_m$ is *Lipschitz* if there exists a constant $L > 0$ such that for all $f, g \in \mathcal{E}_m$, $|C(f) - C(g)| \le L\|f - g\|$.[8] It is *quasiconvex* if for any $f, g \in \mathcal{E}_m$ and $\lambda \in [0, 1]$, $C(\lambda f + (1 - \lambda)g) \le \max\{C(f), C(g)\}$.

Finally, say that it is *permutation invariant* if for any $f \in \mathcal{E}_m$, $C(f) = C(fP)$ for all $P \in \mathcal{P}_m$.

3 Blackwell-Monotone Information Costs

In this section, we provide our main characterizations 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 $f, g \in \mathcal{E}_m$, $C(f) \ge C(g)$ whenever $f \succeq_B g$.*

[8] Local Lipschitz continuity is sufficient for all our results. Since $\mathcal{E}_m$ is compact, these two notions are equivalent for C over $\mathcal{E}_m$. However, this distinction becomes important when considering information costs defined over only the interior of $\mathcal{E}_m$, i.e., experiments where $f_{ij} > 0$ for all i, j, as in Pomatto et al. (2023). All our results still hold in this case under local Lipschitz continuity, thus including their LLR cost.

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

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

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

3.1 Blackwell Monotonicity over Binary Experiments

For any $f \in \mathcal{E}_2$, since $f^1 + f^2 = \mathbf{1}$, it suffices to represent f by the vector f^1. For simplicity, 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 can also let $C : [0,1]^n \to \mathbb{R}_+$.

Theorem 1. *Suppose $C \in \mathcal{C}_2$ is Lipschitz. C is Blackwell monotone if and only if it is (i) permutation invariant; and (ii) for all $f \in \mathcal{E}_2$,*

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

When C is differentiable at f, (1) simplifies to

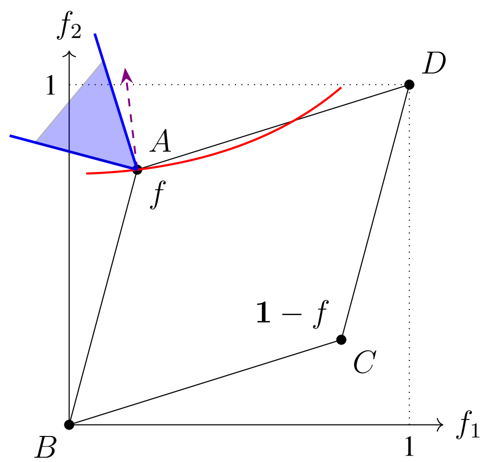
$$\langle \nabla C(f), f \rangle \geq 0 \geq \langle \nabla C(f), \mathbf{1} - f \rangle. \tag{2}$$

In the following, we provide a 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 leads to a further characterization of Blackwell monotonicity in the case of binary states that do not require any functional assumption.

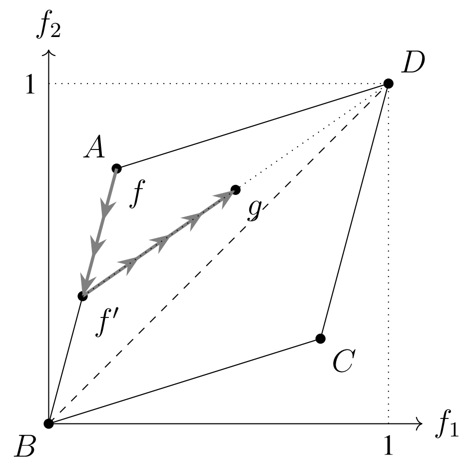
3.1.1 Proof Sketch

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 $M \in \mathcal{M}_2$, there exists $[a, b]^\intercal \in [0,1]^2$ such that

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



(a) The parallelogram hull and its polar cone at f



(b) A path from f to g

Figure 1: A Graphical Illustration with Binary States

which implies

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

In other words, $f \succeq_B g$ if and only if g is in the *parallelogram hull* of the vectors, f and $\mathbf{1} - f$, defined by

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

This parallelogram hull is precisely the sublevel set $S_B(f)$ when considered as a subset in $\mathbb{R}^n$. In the case of binary states, i.e., $n = 2$, it is depicted in Figure 1a by the parallelogram (ABCD).

The parallelogram highlights two extreme directions of decreasing informativeness: $-f$ and $\mathbf{1} - f$, depicted by $\overrightarrow{AB}$ and $\overrightarrow{AD}$, respectively. For any $\epsilon > 0$, notice moving from f in the direction of $-f$ resulting 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, with probability ϵ, generate s_2 instead, while s_2 continues to generate itself. The other direction, $\mathbf{1} - f$, captures the opposite type of garbling: s_2 is sometimes replaced by s_1.

Necessity Permutation invariance follows straightforwardly from Blackwell monotonicity. If $D^+C(f;-f)$ exists, then

$$D^+C(f;-f) = \lim_{\epsilon \downarrow 0} \frac{C(f-\epsilon f) - C(f)}{\epsilon} \leq 0, \tag{3}$$

where the inequality follows from $C(f-\epsilon f) \leq C(f)$ for all $\epsilon > 0$ under Blackwell monotonicity, as illustrated above. Similarly, $D^+C(f;\mathbf{1}-f) \leq 0$ also holds. $\square$

Remark 1. When C is differentiable at f, $D^+C(f;h)$ is linear in h and given by $\langle \nabla C(f), h\rangle$. In this case, notice (2) 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 can be translated into a local constraint on the feasible directions of its gradient.

For a graphical illustration, 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 (1) is a local property, sufficiency requires additional regularity conditions on the cost function. Lipschitz continuity ensures that when restricting the cost function to any line segment, it is differentiable almost everywhere and the Fundamental Theorem of Calculus (FTC) applies. Consider any $f \succeq_B g$, i.e., g lies inside the parallelogram $ABCD$. 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 (1): moving f in the direction of $-f$ to reach f', and then moving f' in the direction of $\mathbf{1}-f'$ to reach g. Then, as derivatives of the function when restricting to line segments coincide with the corresponding directional derivatives, (1) implies it is non-positive almost everywhere along this path. Therefore, applying FTC implies $C(f) \geq C(g)$. If g lies below the line BD, the permutation of g, namely gP, lies above the line BD and has the same cost as g, follows from permutation invariance. Then, the same argument applies to gP implying

$C(f) \geq C(gP) = C(g)$. Lemma A.1 in the appendix formally shows this argument.

3.1.2 A Further Characterization with Binary States

In the binary-binary case ($n = m = 2$), restrict attention to the following set of experiments (the same cost function can be applied to the other experiments by permutation invariance):

$$\hat{\mathcal{E}}_2 \equiv \{(f_1, f_2) \ : 0 \leq f_1 \leq f_2 \leq 1\}. \tag{4}$$

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

$$\alpha \equiv \frac{f_2}{f_1} \geq \frac{g_2}{g_1} \equiv \alpha' \quad \text{and} \quad \beta \equiv \frac{1-f_1}{1-f_2} \geq \frac{1-g_1}{1-g_2} \equiv \beta'.\text{[9]} \tag{5}$$

Note that α is the likelihood ratio for generating the signal s_1 and $1/\beta$ is the likelihood ratio for signal s_2. Thus, (5) 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[10]

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

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

$$\tilde{C}(\alpha, \beta) \equiv C\left(\tfrac{\beta-1}{\alpha\beta-1}, \tfrac{\alpha(\beta-1)}{\alpha\beta-1}\right). \tag{6}$$

Thus, we have the following proposition.

Proposition 1. *When $n = m = 2$, for any $C : \hat{\mathcal{E}}_2 \to \mathbb{R}_+$, C is Blackwell monotone if and only if $\tilde{C}$ as defined in (6) is increasing in α and β.*

These likelihood ratios can also be used to provide a different interpretation of Black-

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

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

well monotonicity. Notice when C is differentiable and $\frac{\partial C}{\partial f_2} \neq 0$, (2) can be rewritten as

$$\alpha = \underbrace{\frac{f_2}{f_1}}_{\text{the slope of } \overline{AB}} \geq \underbrace{-\frac{\partial C/\partial f_1}{\partial C/\partial f_2}}_{\substack{\text{the slope of} \\ \text{the isocost curve}}} \geq \underbrace{\frac{1-f_2}{1-f_1}}_{\text{the slope of } \overline{AD}} = \frac{1}{\beta}.\text{[11]} \tag{7}$$

The slope of the isocost curve can be considered as the *marginal rate of information transformation (MRIT)*. Thus, inequality (7) provides another interpretation for Blackwell monotonicity: the MRIT should fall between the two likelihood ratios provided by the experiment.

3.2 Blackwell Monotonicity over Finite Experiments

In this section, we characterize Blackwell monotonicity over $\mathcal{E}_m$ with an arbitrary number m of signals. For necessity, much of the intuition from binary experiments carries over to the general case. Permutation invariance is clearly required for Blackwell monotonicity. An analogous condition to (1) can be derived again by considering the extreme directions of decreasing informativeness in the space $\mathcal{E}_m$ (Lemma A.2). Specifically, such extreme directions are characterized by the following vectors:

Given $f \in \mathcal{E}_m$, for $j \neq k$, define $f^j(k) \in \mathbb{R}^{n\times m}$ as the matrix with f^j in the k-th column, $-f^j$ in the j-th column and zeros elsewhere, i.e.,

$$f^j(k) \equiv \begin{bmatrix} 0 & \cdots & -f^j & \cdots & 0 & \cdots & f^j & \cdots & 0 \end{bmatrix}.$$

Observe that for any $\epsilon \in [0,1]$, $f + \epsilon f^j(k) \in \mathcal{E}_m$ and is obtained again by the garbling where with a probability ϵ, s_j is replaced by s_k, i.e., merging the signal s_j into s_k. Thus, Blackwell monotonicity would require the cost function to be decreasing along these extreme directions.

When establishing sufficiency over binary experiments, the key step was to construct a path along which Blackwell informativeness decreases. 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.

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

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 path in $\mathcal{E}_3$ connecting I_3 and g along which Blackwell informativeness decreases. Because had there exist such a path, there must exist some experiment other than permutations of I_3 or g that is more informative than g but less informative than I_3, which contradicts the proposition.

In this paper, we identify two approaches to overcome this issue and establish sufficiency. One approach is to embed $\mathcal{E}_m$ into a higher-dimensional space, $\mathcal{E}_{2m}$, where a decreasing path connecting experiments in $\mathcal{E}_m$ can be shown to always exist. In other words, we can use the necessary condition in $\mathcal{E}_{2m}$ to characterize sufficiency in $\mathcal{E}_m$. This characterization, however, might be difficult to apply when it is not obvious how to extend a given cost function defined on $\mathcal{E}_m$ to $\mathcal{E}_{2m}$. For this reason, we delay the presentation of this idea to Section 6.3.

The more convenient and direct approach, still within the original space $\mathcal{E}_m$, is to impose quasiconvexity, a standard condition in cost minimization problems ensuring the local minimum is also global.[12] To see how quasiconvexity is able to address the problem raised in Proposition 2, observe that there is a permutation matrix P such that

$$g = \frac{4}{5} I_3 + \frac{1}{5} I_3 P.$$

As permutation invariance implies $C(I_3) = C(I_3 P)$, we have $C(g) \leq C(I_3)$ when C is quasiconvex. 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 Lipschitz and quasiconvex. C is Blackwell monotone if and only if it is (i) permutation invariant, and (ii) for all $f \in \mathcal{E}_m$,*

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

[12] However, quasiconvexity is strictly non-necessary for Blackwell monotonicity. See Section 6.1.

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

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

First of all, notice that when $m = 2$, condition (ii) 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 (ii) (Lemma A.4). Intuitively, these directions of merging signals are precisely those directions from an extreme point $E \in \mathbf{ext}_k(\mathcal{M}_m)$ to some extreme point in $E \in \mathbf{ext}_{k-1}(\mathcal{M}_m)$. Once this is established, similar to the proof of sufficiency in Theorem 1, FTC along these segments along with (i) and (ii) together imply 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. From Theorem 2, checking Blackwell monotonicity over non-binary experiments requires one more step. Namely, one needs to first verify whether C is quasiconvex. 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, a similar procedure as checking convexity. See Arrow and Enthoven (1961) and also Proposition 3.4.4 in Osborne (2016) for a reference.

4 Additively Separable Costs

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 local condition. 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 for such additively separable information costs and identify a wide range of such costs including both new and well-known existing forms.

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

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

Notice the definition of additive separability has already built-in permutation invariance (same ψ for all signals) and $\psi(\mathbf{0}) = 0$, naturally required by Blackwell monotonicity.[13] As a result, an additively separable cost function can be easily defined over all experiments $\mathcal{E}$ as its restriction to $\mathcal{E}_m$ is obvious. In this case, Blackwell monotonicity over $\mathcal{E}$ is equivalent to Blackwell monotonicity over all $\mathcal{E}_m$.

For additively separable costs, Blackwell monotonicity can be characterized by a simple condition on the component function ψ.

Theorem 3. *Suppose $C : \mathcal{E} \to \mathbb{R}_+$ is additively separable. C is Blackwell monotone if and only if ψ is sublinear, i.e.,*

(i) ψ is positive homogeneous, i.e., $\psi(\gamma \cdot h) = \gamma \cdot \psi(h)$ for all $h \in [0,1]^n$ and $\gamma > 0$ with $\gamma \cdot h \in [0,1]^n$;

(ii) ψ is subadditive, i.e., $\psi(h+l) \leq \psi(h) + \psi(l)$ 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.

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

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

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

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

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

where the first equality follows from additive separability of C 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. $\square$

Additive 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 additively separable costs is that, when holding the other signals fixed, the decrease in information cost by garbling between two signals is independent of all 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.

4.1 Subclasses of Additively Separable Costs

Theorem 3 enables the construction of Blackwell-monotone information costs by simply finding a sublinear component function ψ. Sublinear functions include all norms, seminorms, and linear functions. Additionally, any pointwise maximum of sublinear functions is also sublinear. As an illustration, we highlight the following subclasses of such 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:

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

where subtracting $\|\mathbf{1}\|$ is to ensure that the cost is *grounded*, i.e., the cost of an uninformative experiment is zero. Among norm costs, the supnorm can be used to construct probably the simplest example of a Blackwell-monotone cost function:

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

For binary experiments represented by a single vector $f \in \mathbb{R}^n_+$, 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:

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

Notice that when $a \in \mathbb{R}^n$ is arbitrary, this cost function can potentially be a constant over all experiments. For example, when $a \in \mathbb{R}^n_+$, one has $C(f) = \sum_{j=1}^{m} |\langle a, f^j \rangle| = \sum_{j=1}^{m} \langle a, f^j \rangle = \langle a, \mathbf{1} \rangle$.

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 necessary and sufficient conditions to ensure strict groundedness of absolute-linear costs.

Proposition 3. *An absolute-linear cost function* $C(f) = \sum_{j=1}^{m} |\langle a, f^j \rangle|$ *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 prove instrumental 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 freedom in choosing functional forms in applications. For example, the following quadratic cost function will be used in our applications:

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

Linear ϕ-divergence Costs 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

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

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 such costs is well-known following from the *data-processing inequality* for ϕ-divergence (see Theorem 7.4 in Polyanskiy and Wu (2022)). Here, as we show it can be rewritten as an additively separable cost, Blackwell monotonicity can also be established by sublinearity of $\sum_{i,i'} \beta_{ii'} f_{i'}^{j} \phi_{ii'} \left(\frac{f_i^j}{f_{i'}^j} \right)$, which follows from Jensen's inequality given convexity of $\phi_{ii'}$ (See Theorem 2.7.1 in Cover and Thomas (2006)).

Entroy 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 the prior is fixed, it would induce a cost function over experiments.

Among such cost functions, the entropy cost studied in Sims (2003) and Matějka and McKay (2015) can be shown to be additively separable: Let $\mu \in \Delta(\Omega)$ be a given prior,

$$C_\mu(f) = \lambda \cdot \sum_{j=1}^{m} \sum_{i=1}^{n} \mu_i f_i^j \cdot \log \left(\frac{\mu_i f_i^j}{\tau(f^j)} \right) - \lambda \left(\sum_{i=1}^{n} \mu_i \log \mu_i \right).$$

where $\tau(f^j)$ is the probability of receiving signal j, i.e., $\tau(f^j) \equiv \sum_{i=1}^{n} \mu_i \cdot f_i^j$. Notice for each j, the term $\sum_{i=1}^{n} \mu_i f_i^j \cdot \log \left(\frac{\mu_i f_i^j}{\tau(f^j)} \right)$ depends only on the column f^j, thus the entropy cost is additively separable.

Moreover, a generalization of the entropy cost, the posterior-separable cost, is defined by letting H (the measure of uncertainty) be any concave and upper semi-continuous function (see, e.g., Caplin and Dean (2015) and Denti (2022)). Then notice that

$$C_\mu(f) = H(\mu) - \mathbb{E}[H(\mu(\cdot|s_j))] = H(\mu) + \sum_{j=1}^{m} \psi(f^j)$$

where $\psi : [0,1]^n \to \mathbb{R}$ is defined by

$$\psi(f^j) \equiv -\tau(f^j) \cdot H \left[\left(\frac{\mu_i f_i^j}{\tau(f^j)} \right)_i \right]. \tag{10}$$

Thus, posterior-separable costs are also additively separable.

To verify whether $C_\mu(f)$ is Blackwell-monotone, it is sufficient to check the sublinearity of ψ. For positive homogeneity, from $\tau(\gamma \cdot f^j) = \gamma \cdot \tau(f^j)$, we have $\psi(\gamma \cdot f^j) = \gamma \cdot \psi(f^j)$ whenever $\gamma \cdot f^j \in [0,1]^n$. Additionally, $\tau(f^j + f^k) = \tau(f^j) + \tau(f^k)$ and the concavity of H imply that

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

thus, ψ is subadditive.

To summarize, we identify the class of additively separable costs that encompasses a wide range of well-known information costs. Theorem 3 provides a simple characterization of Blackwell monotonicity for such costs. This not only provides a simple method to construct Blackwell-monotone costs, but also, as highlighted in linear ϕ-divergence costs and posterior-separable costs, provides an alternative approach to verify Blackwell monotonicity of costs in this form.

5 Applications

In this section, we study two applications with costly information. Importantly, we highlight how our characterization of Blackwell monotonicity and additively 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.[14] 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 function to a Blackwell-monotone cost

[14] 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.

function 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, their Theorem 1, the same conclusion can still be shown in the general setting under a different Blackwell-monotone information cost function.

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.[15] 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., $f_L = 0$.[16] 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

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

Based on discussions in the previous section, this cost function can be extended to a Blackwell-monotone cost function over all experiments as follows:

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

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^*

[15] 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.

[16] 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.

given (p, μ) can be solved by the following program:

$$\max_{[f_L, f_H]^\intercal \in \mathcal{E}_2} \mu f_H(H-p) + (1-\mu) f_L(L-p) - C_\lambda(f), \tag{11}$$

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 equilibrium if:[17]

1. $f^*(p, \mu_p)$ solves (11) 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 4. *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 4 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 5. *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 $p^* \in [L, L+\epsilon)$.

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

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.*

For an intuition of Proposition 5, 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 C.1). This fact removes the possibility of semi-separating equilibrium as in Proposition 4 and thus ensures all equilibria are pooling.

5.2 Costly Persuasion

Consider the standard prosecutor-judge example with costly information provision 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 15, 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

$$\max_{[f_1,f_2]^\intercal \in \mathcal{E}_2} \mu \cdot f_2 + (1-\mu)\cdot f_1 - C(f_1, f_2) \tag{12}$$

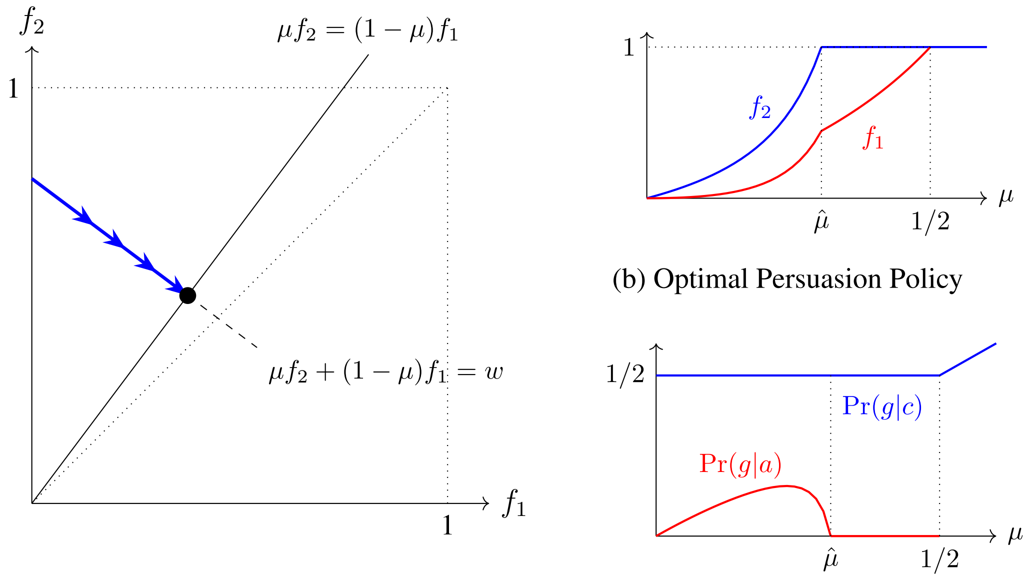
subject to the posterior belief upon receiving c is greater than or equal to $1/2$:

$$\frac{\mu f_2}{\mu f_2 + (1-\mu) f_1} \geq \frac{1}{2} \quad \Leftrightarrow \quad \mu f_2 \geq (1-\mu) f_1. \tag{13}$$

When $\mu \geq 1/2$, setting $f_1 = f_2 = 1$ yields the highest material payoff and the least information cost, and satisfies (13), 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:

$$\min_{[f_1,f_2]^\intercal \in \mathcal{E}_2} C(f) \qquad \text{s.t.} \quad \mu \cdot f_2 + (1-\mu)\cdot f_1 = w \;\; \text{and} \;\; (13). \tag{14}$$



(a) Cost Minimization

(b) Optimal Persuasion Policy

(c) Posteriors

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

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 (13), the range of w is $[0, 2\mu]$.

Lemma 1. *When C is Blackwell monotone and $1 > 2\mu \geq w \geq 0$,* (13) *binds for the solution of* (14).

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, (13) needs to bind:

$$f_1 = \frac{w}{2(1-\mu)} \quad \text{and} \quad f_2 = \frac{w}{2\mu}. \tag{15}$$

Intuitively, when (13) 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 (15) in, the prosecutor's problem becomes

$$\max_{0 \leq w \leq 2\mu} \; w - C\left(\frac{w}{2(1-\mu)}, \frac{w}{2\mu}\right). \tag{16}$$

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

Proposition 6. *Suppose that $C(f_1, f_2) = (f_2 - f_1)^2$. There exists $0 < \hat{\mu} < 1/2$ such that the prosecutor's optimal persuasion policy is as follows:*

$$f_1 = \begin{cases} 1, & \textit{if } \mu \geq 1/2, \\ \frac{\mu}{1-\mu}, & \textit{if } \mu \in (\hat{\mu}, 1/2), \\ \frac{\mu^2(1-\mu)}{(1-2\mu)^2} & \textit{if } \mu \leq \hat{\mu}, \end{cases} \quad \textit{and} \quad f_2 = \begin{cases} 1, & \textit{if } \mu \geq 1/2, \\ 1, & \textit{if } \mu \in (\hat{\mu}, 1/2), \\ \frac{\mu(1-\mu)^2}{(1-2\mu)^2} & \textit{if } \mu \leq \hat{\mu}. \end{cases} \tag{17}$$

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 posterior separable costs, where the posterior beliefs are independent of prior belief whenever the information is provided.

6 Additional Results and Discussions

6.1 Non-necessity of Quasiconvexity

As already mentioned, quasiconvexity is stirctly non-necessary. The following example illustrates a cost function over binary experiments that is Blackwell monotone but not quasiconvex.

Example 1. 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 (7), 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 average convex combination 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 was not needed in establishing Blackwell monotonicity over binary experiments $\mathcal{E}_2$. However, when impose quasiconvexity in this case, we can establish a sufficient condition for Blackwell monotonicity without imposing Lipschitz continuity and the local condition in Theorem 1:

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. Say that C is *non-null* if for any $f \in [0,1]^n$, $C(f) \geq C(\mathbf{1}) = C(\mathbf{0})$.

Proposition 7. *If $C \in \mathcal{E}_2$ is quasiconvex, permutation invariant, and non-null, then C is Blackwell monotone.*

Proof of Proposition 7. As an intermediate result of Theorem 1, we have $S_B(f) = \text{PARL}(f, \mathbf{1} - f)$, that is, $f \succeq_B g$ implies that $g = af + b(\mathbf{1} - f)$ for some $(a,b) \in [0,1]^2$. Then, 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, $C(f) \geq C(g)$. $\square$

6.3 Sufficiency via Higher Dimensions

Proposition 2 demonstrates an example in $\mathcal{E}_3$ where $f = I_3 \succeq_B g$ but there does not exist a 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 be found. To see this, we 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 D.1). However, such a decreasing path does not necessarily move in the extreme directions, $f^j(k)$'s. Therefore, to establish a similar argument as in Theorem 1 in this case, we need to rely on the linearity of directional derivaitives, and thus requires a stronger functional assumption on C:

Theorem 4. *Suppose $C \in \mathcal{C}_{2m}$ is continuously differentiable. Conditions (i) and (ii) in Theorem 2 over $\mathcal{E}_{2m}$ are sufficient for C to be Blackwell monotone over $\mathcal{E}_m$.*

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 also extended to characterize Lehmann-monotone costs. However, 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.

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 Hoshino Tetsuya (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 (2022) "Experimental cost of information," *American Economic Review*, 112 (9), 3106–3123, 10.1257/aer.20210879.

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.

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 $f' = \frac{a-b}{1-b} f$.[18] *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, $f' - f$ is in the direction of $-f$, and $g - f'$ is in the direction of $1 - f'$.

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

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

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

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 $(1-\lambda)f+\lambda f' \succeq_B f'$.

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. □

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 $\phi_1(\lambda) \equiv C((1-\lambda)f + \lambda f')$ and $\phi_2(\lambda) \equiv C((1-\lambda)f' + \lambda g)$. They are both Lipschitz continuous function in λ as Lipschitz continuity is preserved under composition. Therefore, ϕ_1 is differentiable almost everywhere and satisfies, when differentiable,

$$\phi_1'(\lambda) = D^+C((1-\lambda)f + \lambda f'; -f + f').$$

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, $\phi_1'(\lambda)$ has the same sign as $D^+C((1-\lambda)f + \lambda f'; -((1-\lambda)f + \lambda f'))$ and it is non-positive from (1). Lipschitz continuity in λ enables FTC, which gives:

$$C(f') = \phi_1(1) = \phi_1(0) + \int_0^1 \phi_1'(\lambda)d\lambda \leq \phi_1(0) = C(f). \tag{A.4}$$

Similarly, observe that

$$\phi_2'(\lambda) = D^+C((1-\lambda)f' + \lambda g; -f' + g),$$
$$-f' + g = b(\mathbf{1} - f') = \frac{b}{1-\lambda b}\left(\mathbf{1} - ((1-\lambda)f' + \lambda g)\right).$$

Then, $\phi_2'(\lambda)$ 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 $C(g) = \phi_2(1) \leq \phi_2(0) = C(f')$. Therefore, we have $C(f) \leq C(g)$, thus C is Blackwell monotone. $\square$

A.2 Proof of Proposition 2

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

Observe that at least one of $f_{11},\ f_{12}$ and f_{13} is positive—if not, every entry of the first row of fB is equal to zero. Without loss of generality, let f_{11} be positive (we can obtain it by permuting f). Note that $f_{11}b_{13} + f_{12}b_{23} + f_{13}b_{33} = 0$. Since every entry of f and B are nonnegative, $b_{13} = 0$.

Next, observe that $4/5 = f_{21}b_{13} + f_{22}b_{23} + f_{23}b_{33}$. From $b_{13} = 0$, at least one of f_{22} and f_{23} is positive. Without loss, let f_{22} be positive. Then, from $g_{21} = 0$, we have $b_{21} = 0$. Then, it gives us $f_{21}b_{11} + f_{23}b_{31} = 0$. We consider two cases: $b_{11} = 0$ or $f_{21} = 0$.

1. $b_{11} = 0$**:** From $b_{11} = b_{13} = 0$, we have $b_{12} = 1$. From $g_{32} = 0$ and $b_{12} > 0$, we have $f_{31} = 0$.

Additionally, we have $f_{13}b_{31} = 4/5$, $f_{23}b_{31} = 0$ and $f_{33}b_{31} = 1/5$. Therefore, $b_{31}, f_{13}, f_{33} \neq 0$ and $f_{23} = 0$. From $g_{13} = 0$ and $f_{13} \neq 0$, we have $b_{33} = 0$. Likewise, from $g_{32} = 0$ and $f_{33} > 0$, $b_{32} = 0$. Then, it gives us $b_{31} = 1$.

From $b_{11} = 0, b_{21} = 0$ and $b_{31} = 1$, we have $f_{13} = 4/5$ and $f_{33} = 1/5$. From $f_{31} = 0$, $f_{32} = 4/5$. From $g_{32} = 0$ and $f_{32} > 0$, we have $b_{22} = 0$. It gives us $b_{23} = 1$. Therefore, B is a permutation matrix and f is a permutation of g.

2. $b_{11} > 0$ **and** $f_{21} = 0$**:** Observe that $f_{23}b_{31} = 0$, $f_{22}b_{22} + f_{23}b_{32} = 4/5$ and $f_{22}b_{23} + f_{23}b_{33} = 1/5$. We consider two subcases: $f_{23} = 0$ or $b_{31} = 0$.

(a) $f_{23} = 0$**:** From $f_{21} = f_{23} = 0$, we have $f_{22} = 1$. Additionally, we have $b_{22} = 4/5$ and $b_{23} = 1/5$. From $0 = g_{13} = f_{12}b_{23} + f_{13}b_{33}$ and $0 = g_{32} = f_{31}b_{12} + f_{32}b_{22} + f_{33}b_{32}$, we have $f_{12} = f_{32} = 0$. Observe that $0 = g_{13} = f_{13}b_{33}$ and $4/5 = g_{33} = f_{33}b_{33}$. Then, we have $b_{33} > 0$ and $f_{13} = 0$. From $f_{12} = f_{13} = 0$, we have $f_{11} = 1$. This also gives $b_{11} = 4/5$ and $b_{12} = 1/5$. Again from $0 = g_{32}$ and $b_{12} = 1/5$, we have $f_{31} = 0$. Therefore, from $f_{31} = f_{32} = 0$, we have $f_{33} = 1$, i.e., f is I_3.

(b) $b_{31} = 0$**:** From $b_{21} = b_{31} = 0$, we have $f_{11} \cdot b_{11} = 4/5$ and $f_{31} \cdot b_{11} = 1/5$. Therefore, $b_{11}, f_{11}, f_{31} > 0$. Next, $0 = g_{32} = f_{31}b_{12} + f_{32}b_{22} + f_{33}b_{32}$ gives $b_{12} = 0$. From $b_{12} = b_{13} = 0$, we have $b_{11} = 1$.

Suppose that both b_{22} and b_{32} are positive. Then, from $0 = g_{32} = f_{32}b_{22} + f_{33}b_{32}$, we have $f_{32} = f_{33} = 0$. It contradicts $4/5 = g_{33} = f_{31}b_{13} + f_{32}b_{23} + f_{33}b_{33}$ since $b_{13} = f_{32} = f_{33} = 0$. Therefore, at least one of b_{22} and b_{32} is equal to zero. Likewise, if both b_{23} and b_{33} are positive, we have $f_{12} = f_{13} = 0$ from $g_{13} = 0$, but it contradicts $g_{12} = 1/5 > 0$. Thus, at least one of b_{23} and b_{33} 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) $b_{22} = b_{33} = 1$ and $b_{23} = b_{32} = 0$; or (ii) $b_{23} = b_{32} = 1$ and $b_{22} = b_{33} = 0$. Then, B is either I_3 or a permutation of I_3. Therefore, f is g or a permutation of g. □

A.3 Proof of Theorem 2

A.3.1 Useful Lemmas

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

$$g - f \in \left\{ \sum_{j=1}^{m} \lambda_j h_j : \lambda_j \in [0,1], h_j \in co(\{f^j(k) : k \neq j\}), \forall j \right\}, \tag{A.5}$$

where $co(\cdot)$ denote the convex hull.

Moreover, 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 the first part, for any $f, g \in \mathcal{E}_m$. If $f \succeq_B g$, then there exists a stochastic matrix $M \in \mathcal{M}_m$ such that $g = fM$. Then we have

$$\begin{aligned}
g - f &= f(M - I) \\
&= \begin{bmatrix} f^1 & \cdots & f^m \end{bmatrix} \begin{bmatrix} m_{11} - 1 & \cdots & m_{1m} \\ \vdots & \ddots & \vdots \\ m_{m1} & \cdots & m_{mm} - 1 \end{bmatrix} \\
&= \begin{bmatrix} f^1 & \cdots & f^m \end{bmatrix} \begin{bmatrix} -\sum_{k=2}^m m_{1k} & \cdots & m_{1m} \\ \vdots & \ddots & \vdots \\ m_{m1} & \cdots & -\sum_{k=1}^{m-1} m_{mk} \end{bmatrix} \\
&= \sum_{j=1}^m \sum_{k \neq j} m_{jk} f^j(k)
\end{aligned}$$

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

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

Notice that $\sum_{k \neq j} m_{jk} \in [0, 1]$ 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 $m_{jk} = \lambda_j \cdot \mu_{jk}$ where $\sum_{k \neq j} \mu_{jk} f_k^j = h_j$ and let $m_{jj} = 1 - \sum_{k \neq j} m_{jk}$. Then, such M is a stochastic matrix and $g = fM$.

For the second part, fix any $f \in \mathcal{E}_m$ and any $j \neq k$, let

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

Since $f_\lambda - f$ satisfies (A.5) 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

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

$\square$

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 $f \in \mathcal{E}_m$,*

$$fB_{jk} = f^j(k).$$

The proof of Lemma A.3 is by algebra and thus omitted.

Lemma A.4. *Suppose $C \in \mathcal{C}_m$ is Lipschitz and satisfies (ii) 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 $\lambda \in [0, 1]$,*

$$fE' \succeq_B (1-\lambda)fE' + \lambda fE \succeq_B fE. \tag{A.6}$$

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.6) 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.6) holds.

Recall by Lemma A.3,

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

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

$$\begin{aligned}
&D^+(C((1-\lambda)fE' + \lambda fE), fE - ((1-\lambda)fE' + \lambda fE)) \\
=&D^+(C((1-\lambda)fE' + \lambda fE), ((1-\lambda)fE' + \lambda fE)B) \;\; \leq 0.
\end{aligned} \tag{A.7}$$

Finally, we show for such E and E',

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

For $\lambda \in [0,1]$, define the function $\phi(\lambda) = C((1-\lambda)fE' + \lambda fE)$. Notice that $\phi(0) = C(fE')$ and $\phi(1) = C(fE)$. ϕ is Lipschitz continuous over $[0,1]$ as a composition of Lipschitz continuous functions. Thus ϕ is differentiable almost everywhere on $[0,1]$ and satisfy

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

Moreover, Lipschitz continuity implies absolute continuity and thus the Fundamental Theorem of Calculus (FTC) holds for ϕ on $[0,1]$, which further implies that

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

where the second last equality uses positive homogeneity of $D^+C(f;\cdot)$ and the last inequality follows from that (A.7) holds for all $\lambda \in [0,1]$. $\square$

A.3.2 Proof

Proof of Theorem 2. Necessity follows directly from the definition and Lemma A.2.

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

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

thus, C is Blackwell monotone. $\square$

B Proofs for Section 4

B.1 Proof of Necessity in Theorem 3

Proof. Suppose C is additively 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

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

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,

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

and

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

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. □

C Proofs for Section 5

C.1 Bargaining with Information Acquisition

Before proving Proposition 4 and 5, we first show the following lemma which characterizes the buyer's optimal information acquisition strategy (we break indifference towards accepting the offer).

Lemma C.1. *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 C.1. Recall the buyer's problem is to solve, for all p and μ,

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

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,

$$[f_L^*, f_H^*] = \begin{cases} [\frac{\mu(H-p)-w}{(1-\mu)(p-L)}, 1] & \text{if } \mu(H-p) + (1-\mu)(L-p) \geq 0, \\ [0, \frac{w}{\mu(H-p)}] & \text{if } \mu(H-p) + (1-\mu)(L-p) < 0. \end{cases} \tag{C.2}$$

This proves (i).

For (ii), let $C_\lambda(f) = \lambda |f_H - f_L|$. By (C.2), 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 (C.2). 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 $[f_L^*, f_H^*] = [0, 1]$. □

Proof of Proposition 4. 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 (C.2), 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. □

Proof of Proposition 5. By Lemma C.1, 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 C.1, 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 (C.2) 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

$$\epsilon = \min\left\{\pi(H-L), \frac{\lambda}{1-\pi}\right\}. \tag{C.3}$$

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 C.1, 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 (C.3) implies that $\epsilon \to 0$ as $\lambda \to 0$. □

C.2 Costly Persuasion

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

$$\alpha(f_1) \equiv \frac{f_2}{f_1} = \frac{w-(1-\mu)f_1}{\mu \cdot f_1}, \quad \beta(f_1) \equiv \frac{1-f_1}{1-f_2} = \frac{\mu(1-f_1)}{\mu - w + (1-\mu)f_1}. \tag{C.4}$$

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

$$\alpha'(f_1) = -\frac{w}{\mu f_1^2} < 0, \quad \beta'(f_1) = -\frac{(1-w)\mu}{(\mu - w + (1-\mu)f_1)^2} < 0. \tag{C.5}$$

Then, from (5), 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, (13) needs to bind. □

Proof of Proposition 6. 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, (16) is equivalent to

$$\max_{0 \leq w \leq 2\mu} w - \frac{w^2}{4\mu \cdot h(\mu)} \qquad \text{where} \quad h(\mu) \equiv \frac{\mu(1-\mu)^2}{(1-2\mu)^2}. \tag{C.6}$$

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

$$h'(\mu) = \frac{2\mu + (1-2\mu)(1+\mu^2)}{(1-2\mu)^3} > 0. \tag{C.7}$$

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 (C.6) subject to $0 \le w \le 2\mu$ is

$$w^* = \begin{cases} 2\mu, & \text{if } \mu \in (\hat{\mu}, 1/2), \\ 2\mu \cdot h(\mu), & \text{if } \mu \le \hat{\mu}. \end{cases} \tag{C.8}$$

By plugging this into (15), we have (17). $\square$

D Proofs for Section 6

D.1 Proof of Theorem 4

Proof of Theorem 4. 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 D.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 D.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 $\lambda \in [0,1]$. □

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

$$\langle \nabla C(f), g - f \rangle \leq 0, \tag{D.1}$$

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 FTC along the decreasing path from f to g in $\mathcal{E}_{2m}$ identified in Lemma D.1 implies $C(f) \geq C(g)$, and thus establishes the conclusion. □