

Meaningful Information*

Luca Braghieri

September 2026

Abstract

I study how to measure and rank information transmission when agents have misspecified priors, misinterpret signals, or update beliefs in a non-Bayesian way. I axiomatize and characterize measures of *meaningful* information transmission that account for both the objective features of a signal structure and agents' (mis)interpretations of it. Meaningful information measures coincide with canonical ones under correct Bayesian beliefs, but otherwise extend them in a disciplined way grounded in the same underlying decision geometry. I then connect this class of measures to the robust order over misinterpreted experiments from [Morris \(1991\)](#) and derive a new equivalent portfolio representation of that order. Finally, in an empirical application to political correctness on college campuses that leverages the experiment in [Braghieri \(2024\)](#), I show that canonical and meaningful information measures and orders can differ sharply.

*Braghieri: Bocconi University and NHH, CEPR, CESifo, and IGIER. Email: luca.braghieri@unibocconi.it. I thank Martino Banchio, Francesco Bilotta, Alex Frankel, Massimo Marinacci, Stephen Morris, Marco Ottaviani, Collin Raymond, Joel Sobel, Chris Turansick, and Mark Whitmeyer for helpful discussions. I thank Shuige Liu and Alessandro Pinzi for careful proofreading and Francesco Doga for excellent research assistance.

Interpreting information is often difficult. Doctors misread diagnostic tests (Chan, Gentzkow and Yu, 2022), investors react to irrelevant news (Rashes, 2001; Balashov and Nikiforov, 2019), experts misinterpret AI-generated decision aids (Agarwal et al., 2025; Dreyfuss and Hoong, 2025), and people frequently deviate from Bayesian updating (Benjamin, 2019). These examples raise a natural question: how should we measure and rank information transmission when agents misinterpret signals or update beliefs in a non-Bayesian way? Canonical measures of information transmission and orders over information structures offer little guidance, because they do not take into account the ways in which agents interpret the signals they receive.

To make the discussion concrete, consider the problem of political correctness on college campuses analyzed experimentally in Braghieri (2024). Students may be reluctant to express politically sensitive views in public when those views could damage their social image. As a result, public statements need not faithfully reflect privately held opinions: social-image concerns can distort what people say and thereby make public speech less informative about their underlying political views. At the same time, interpreting public speech is itself difficult. A listener who observes a politically sensitive statement must reason not only about the speaker’s underlying views, but also about the incentives that shaped what the speaker chose to say. If listeners misunderstand those incentives, for instance by taking public statements at face value, the beliefs they form may differ substantially from the inferences warranted by the actual relationship between political views and public speech. Thus, how much information public discourse transmits depends not only on what speakers say, but also on how listeners (mis)interpret what they hear.

In this paper, I study how to measure and rank information transmission when agents have a misspecified prior, misinterpret signals, or update their beliefs in a non-Bayesian way. I introduce and axiomatize measures of *meaningful information transmission* that account for both the objective features of the signals agents observe and the way they interpret them. These measures coincide with canonical ones under correct Bayesian beliefs, but otherwise extend them in a disciplined way grounded in the same underlying decision geometry. I then relate these measures to the robust order over misinterpreted experiments developed by Morris (1991) and show that this order admits a new equivalent characterization based on the same geometry that underpins the cardinal measures. Finally, I apply the framework to the experiment on political correctness in Braghieri (2024) and show that canonical and meaningful information can diverge sharply.

To formalize the idea of meaningful information transmission, I begin by revisiting canonical ex-ante measures of information. Following the terminology of [Frankel and Kamenica \(2019\)](#), who axiomatize canonical ex-post measures, I refer to an ex-ante measure of information as valid if it coincides with the ex-ante instrumental value of information in some decision problem. Valid ex-ante measures are characterized by a simple set of axioms and admit a posterior-separable representation related to results in the rational inattention literature.

The goal of the first part of the paper is to extend that logic to measures of information transmission that account both for the objective features of the signal structure and for the way agents interpret signals. Extending the canonical framework to this setting requires three ingredients: a notion of validity for environments with both objective uncertainty and subjective interpretations, a corresponding set of axioms, and a representation theorem tying the two together.

For canonical ex-ante measures of information, the notion of validity is built around the value of information: the expected change in an agent’s welfare that results from making decisions conditional on the realization of a signal rather than on the prior alone. In environments with both objective uncertainty and subjective interpretations, the value of information can be defined in an analogous way, but with a different underlying notion of welfare. In particular, I assume that when objective and subjective interpretations coexist, agents choose actions based on their subjective beliefs, but that the welfare consequences of those actions are determined by the objective probabilities.¹ I refer to the value of information built around this notion of welfare as the *behavioral value of information*; furthermore, I refer to a measure of meaningful information as *behaviorally valid* if it corresponds to the behavioral value of information in some decision problem.

[Theorem 2](#) establishes that a measure of meaningful information transmission is behaviorally valid if and only if it satisfies five axioms and admits a transparent representation. The first three axioms extend the ones that underpin canonical ex-ante measures of information to environments with both objective uncertainty and subjective interpretations. The remaining two axioms are simple and intuitive. *Invariance to null updating* requires that meaningful information measures be zero if the agent does not update her beliefs at all in response to a signal. *Dominance at correct posteriors* requires that meaningful information measures be maximized when the agent holds

¹As noted by [Brunnermeier, Simsek and Xiong \(2014\)](#), this objective-outcome welfare criterion is a “standard approach” in models with distorted beliefs, though it is of course not the only possible benchmark. See [Section 1.2](#) for further discussion.

correct posterior beliefs. For an agent with a correct prior, imposing this axiom is tantamount to requiring that any misinterpretations or non-Bayesian updating reduce the amount of meaningful information transmitted to the agent.

The axioms above yield a transparent representation that clarifies where meaningful and canonical measures diverge. The representation is additively separable into three components. The first corresponds to a canonical ex-ante measure of information. The second, which enters the expression positively, reflects the potential benefit that the agent could in principle reap if sampling from the signal structure corrected her prior misspecification. The third captures the expected loss from incorrect posteriors, which might result from holding an incorrect prior, from having misspecified beliefs about the signal structure, or from non-Bayesian updating.

The representation yields two stark qualitative implications. First, meaningful information can be negative: if a signal induces sufficiently distorted posterior beliefs, the agent may be worse off than if she had ignored the signal altogether and continued to act on her prior. Second, measures of meaningful information need not be monotone in the Blackwell order. When an agent systematically misreads signals, making the objective signal structure more informative can push her beliefs further in the wrong direction and reduce the amount of meaningful information transmitted.

The next result in the paper, [Theorem 3](#), formalizes the relationship between measures of meaningful information and canonical measures. In particular, following in spirit [Frankel and Kamenica \(2019\)](#), I introduce a notion of joint validity that requires that a measure of meaningful information and a canonical measure arise from the same underlying decision problem. Under joint validity, the theorem shows that, if a canonical measure I and a meaningful measure M are jointly valid, then M is obtained from I by a simple, disciplined adjustment based on the underlying geometry of I : add a term that rewards correcting prior misspecifications and subtract a term that penalizes posterior misspecification. Moreover, this adjustment characterizes M : among all candidates, M is the unique jointly valid extension of I to subjective interpretations that satisfies the paper’s axioms. Thus, measures of meaningful information are a structural extension of canonical measures that leverage the same underlying decision geometry.

The results in [Theorem 2](#) and [Theorem 3](#) are, by construction, decision-problem dependent: different decision problems generate different cardinal measures of information and may therefore rank experiments differently. This is a well-understood feature of the canonical framework. For

canonical ex-ante measures of information, Blackwell’s order identifies the comparisons between signal structures that hold across all decision problems. The related robust-comparison problem for misinterpreted experiments was studied by [Morris \(1991\)](#).

[Theorem 4](#) integrates Morris’s robust comparison of interpreted experiments with the framework developed in this paper. First, it shows that Morris’s order is exactly the unanimity order generated by the class of measures of meaningful information axiomatized in [Theorem 2](#). Second, it provides a new equivalent portfolio representation of that order. Each elementary asset in the portfolio corresponds to avoiding a pairwise mistake: taking an action that is optimal at one belief when another belief is the relevant distribution of states. Whenever one interpreted experiment robustly dominates another, the difference between them can be represented by non-negative exposures to these assets. The economic interpretation is that a dominating experiment can be represented as avoiding a non-negative portfolio of pairwise mistakes relative to the dominated experiment. The differential exposures to pairwise mistakes depend on the interpreted experiments themselves; particular measures of meaningful information simply determine how those exposures are priced.

The picture that emerges from the theoretical analysis is summarized in [Figure 1](#). When beliefs are correctly specified and updating is Bayesian, there is an axiomatically-defined class of canonical measures of information, each corresponding to the value of information in some decision problem; agreement across all such measures gives rise to Blackwell’s order. When beliefs may be misspecified or updating non-Bayesian, there is an analogous axiomatically-defined class of measures of *meaningful* information, each corresponding to the *behavioral* value of information in some decision problem; agreement across all such measures gives rise to the *meaningful-information* order. The two rows of the figure therefore provide complementary ways of evaluating information: cardinal measures quantify how much information an experiment transmits in the welfare units of a particular decision problem, while robust orders identify comparisons between experiments that hold independently of the decision problem.

In the empirical part of the paper, I apply the theoretical framework to the experiment on political correctness in [Braghieri \(2024\)](#). The experiment documents two distinct distortions in political communication on college campuses. The first is that social-image concerns affect what students are willing to say in public, driving a wedge between their privately held opinions and their

public statements. This distortion changes the objective mapping from underlying political type (Democrat vs. Republican) to political speech, making public speech less informative than private speech in the Blackwell order. The second is that other students do not always correctly account for this distortion when interpreting what their peers say in public. Specifically, many students interpret public statements at face value, unless the role of social image in shaping political speech is made salient to them. If only the first distortion were present (i.e., if students correctly understood how political types map into both private and public speech) the empirical information ranking in the experiment would be unambiguous: private speech would be Blackwell more informative than public speech, which in turn would be Blackwell more informative than silence.

Once receivers' actual interpretations are taken into account, however, this ranking no longer holds for most receivers. Applying the meaningful-information order, which incorporates both the actual mapping from political types into speech and receivers' interpretations of that speech, the full private–public–silence ordering remains valid across all measures of meaningful information for only 5% of receivers. Moreover, 19% exhibit at least one robust reversal of the objective Blackwell ranking: for these receivers, a Blackwell inferior communication regime transmits *more* meaningful information than a Blackwell superior one under every measure of meaningful information. For the remaining 76%, the meaningful-information order is incomplete, so the ranking genuinely depends on the chosen cardinal measure or, equivalently, on the downstream decision problem. To complete these comparisons, I consider the meaningful-information extensions of two familiar benchmarks, mutual information and quadratic scoring. Under the Shannon geometry that underlies mutual information as well as its meaningful information extension, 37% of receivers satisfy the full private–public–silence ranking; under the quadratic geometry, 36% do so.

The cardinal measures of meaningful information also make it possible to decompose the loss in information transmission between two margins: the objective loss in informativeness caused by moving from private to public speech and the additional loss caused by receivers' misinterpretation of public speech. Under both geometries, misinterpretations account for the lion's share of meaningful information loss: of the average shortfall between correctly interpreted private speech and actually interpreted public speech, 39% is attributable to the change in what speakers say and 61% to how receivers interpret what they hear. Overall, the application shows that interpretation can fundamentally alter objective informativeness rankings, distinguishes conclusions that are robust

across measures of meaningful information from those that depend on the particular measure selected, and uses cardinal measures to quantify how much information is lost because social-image concerns alter the mapping from political views into public speech and how much additional information is lost because receivers misinterpret that public speech.

This paper makes a fourfold contribution to the literature. First, it develops decision-theoretic foundations for, and axiomatizes, cardinal measures of information transmission that account for both the objective features of a signal structure and the agent’s interpretation of it. These interpretations may involve prior misspecification, misspecified beliefs about how signals are generated, or non-Bayesian updating. Second, it shows that canonical measures of information and measures of meaningful information share the same underlying convex geometry, and that each canonical measure admits a unique extension to meaningful information that satisfies the paper’s axioms. Third, it shows that the robust order in [Morris \(1991\)](#) coincides with the unanimity order generated by the class of measures of meaningful information axiomatized in the paper, and derives a new equivalent characterization of that order in terms of the same convex geometry that underpins the cardinal measures. Fourth, it shows how the cardinal and robust branches of the framework can be fruitfully brought to empirical data.

The paper thus sits at the intersection of three strands of the literature. The first strand examines the costs ([Sims, 2003](#); [Caplin and Dean, 2013](#); [Matějka and McKay, 2015](#); [Denti, 2022](#); [Zhong, 2019](#), among others) and benefits ([Blackwell, 1951, 1953](#); [Lehmann, 1988](#); [Athey and Levin, 2018](#); [Frankel and Kamenica, 2019](#), among others) of information acquisition. On the benefits side, the closest paper to the cardinal-measure results is [Frankel and Kamenica \(2019\)](#), who axiomatize ex-post measures of information transmission. This paper shares their goal of providing decision-theoretic foundations for cardinal measures of information, but it allows agents to hold misspecified priors, misperceive the signal structure, and update beliefs in a non-Bayesian way. The closest antecedent to the paper’s robust-comparison result is [Morris \(1991\)](#), who characterizes when one misinterpreted experiment is uniformly preferred to another across decision problems.²

A second strand examines learning and decision making under misspecified models, where agents may hold incorrect prior beliefs and incorrect beliefs about how signals are generated

²[Frick, Iijima and Ishii \(2024\)](#) build on [Morris \(1991\)](#) in their static analysis and derive simple sufficient conditions for the robust comparisons of interpreted experiments. Their main focus, however, is to develop a dynamic ranking.

(Esponda and Pouzo, 2016; Bohren and Hauser, 2021; Heidhues, Kőszegi and Strack, 2021; Fudenberg, Lanzani and Strack, 2021; Frick, Iijima and Ishii, 2023; Gossner and Steiner, 2018).³ Much of this literature emphasizes long-run learning, studies equilibrium behavior, or explores applications such as overconfidence and discrimination. My focus is different: I study how to measure and rank the amount of information transmitted when a signal is interpreted through a potentially misspecified prior or likelihood model. This focus is also motivated by a growing empirical literature documenting that people frequently deploy incorrect mental models of the data-generating process, leading them to draw systematically distorted inferences; for example, correlation neglect (Enke and Zimmermann, 2019), insufficient skepticism under disclosure (Jin, Luca and Martin, 2021), and misinterpretation of diagnostic information (Chan, Gentzkow and Yu, 2022), among others.

A third strand studies systematic departures from Bayesian updating. On the theory side, classic models formalize particular updating biases; for instance, representativeness and conservatism (Barberis, Shleifer and Vishny, 1998), and confirmation bias (Rabin and Schrag, 1999). More recent perspectives develop and compare alternatives to Bayesian updating (Ortoleva, 2024). On the empirical side, a long line of work documents robust deviations from Bayesian inference in laboratory and field settings (e.g., Grether, 1980).⁴ I contribute to this literature by providing a unified axiomatic framework that accommodates both misspecification and non-Bayesian updating rules, and by showing how the resulting measures can be implemented to quantify meaningful information transmission in empirical applications.

The remainder of the paper is organized as follows. Section 1 develops the theory of cardinal measures of information: it revisits canonical ex-ante measures of information (Section 1.1), characterizes measures of meaningful information (Section 1.2), and relates the two classes of measures (Section 1.3). Section 2 then turns to decision-problem-independent comparisons and, thus, to a Blackwell-style order for interpreted experiments. Section 3 presents the empirical application to the political correctness experiment in Braghieri (2024). Section 4 concludes. All proofs are relegated to the appendix.

³See Bohren and Hauser (2025) for a recent review.

⁴See Benjamin (2019) for a review.

1. Cardinal Measures of Information

Before turning to meaningful information, which allows the agent to have a misspecified prior, to misunderstand signals, and to deviate from Bayesian updating, it is helpful to review the canonical setting.

1.1 Canonical Measures of Information Transmission

Information can be evaluated from two complementary perspectives. An *ex-ante* perspective asks how much information a signal structure conveys in expectation, before any signal is realized. An *ex-post* perspective asks how informative a particular signal realization is, once it has been observed. The ex-post characterization was developed by Frankel and Kamenica (2019). The corresponding ex-ante characterization has, to the best of my knowledge, not been stated explicitly, although it is closely related to results in the rational-inattention literature (e.g., Theorem 10 in Zhong 2019). This subsection makes the ex-ante characterization explicit, links it to the ex-post measures from Frankel and Kamenica (2019), and provides a foothold to develop measures of meaningful information transmission.

Setup. Let $\Omega = \{\omega_1, \dots, \omega_n\}$ denote a finite set of mutually exclusive and exhaustive states of the world. The state ω is drawn according to a prior distribution $p \in \Delta(\Omega)$. The agent does not observe ω directly but instead receives a signal generated by a *signal structure*. A signal structure consists of a set of possible signal realizations S and a likelihood function $\sigma : \Omega \rightarrow \Delta(S)$, where $\sigma(s | \omega)$ denotes the probability of signal s conditional on state ω and $\Pr \in \Delta(S)$ denotes the unconditional distribution of signals. For each $s \in S$, the pair (p, σ) induces a posterior $p(\cdot | s) \in \Delta(\Omega)$. Finally, I denote the space of all finite signal structures by Σ , identifying a signal structure with its likelihood function σ .

Definition 1. A (primitive) *ex-ante measure of information* is a functional $I : \Delta(\Omega) \times \Sigma \rightarrow \mathbb{R}$ that assigns to each prior p and signal structure σ a real number summarizing the degree of informativeness of the signal structure.

Two approaches can be used to motivate what properties a measure of information should satisfy: an instrumental approach and an axiomatic approach.

Instrumental Approach. The instrumental approach starts from the premise that information is valuable because it improves decision-making. Hence, a measure of information should correspond to the *instrumental value of information* in some decision problem. Following Frankel and Kamenica (2019), I call such measures *valid*. Formally, I is valid if there exists a decision problem in which $I(p, \sigma)$ equals the value of information produced by the signal structure σ for an agent with prior p . I denote a decision problem by a tuple (A, u) consisting of an action space A that is potentially uncountably infinite and a utility function $u : A \times \Omega \rightarrow \mathbb{R} \cup \{-\infty\}$.⁵ I assume that the agent’s decision problem is well-defined, by requiring $\arg \max_{a \in A} \mathbb{E}_{\omega \sim p}[u(a, \omega)]$ to be non-empty for all beliefs $p \in \Delta(\Omega)$ and for the maximum to take a finite value. Furthermore, I adopt the convention $-\infty \times 0 = 0$, so that an action that yields $-\infty$ in some state is not costly if the probability of that state occurring is zero. Lastly, I treat a fixed tie-breaking rule $q \mapsto a_q$ from the optimal-action correspondence as part of the decision problem. Henceforth, whenever I refer to a decision problem, I implicitly include such a tie-breaking rule.⁶

Axiomatic Approach. The axiomatic approach starts from an intuitive set of properties that any reasonable measure of information should satisfy. Three such properties are: positivity, posterior dependence, and path independence. *Positivity* captures the intuition that a signal cannot contain a negative amount of information. *Posterior dependence* refers to the idea that informativeness depends on the signal structure only through the distribution of posteriors it induces. *Path independence*, a term I borrow from Zhong (2019), requires that informativeness satisfy a chain-rule decomposition across stages: the informativeness of a compound signal equals the informativeness of the first stage plus the expected informativeness of the continuation experiment.

The main result of this section shows that the instrumental and axiomatic approaches coincide. While I state the result formally in the present setting, the representation is closely related to the characterization of posterior-separable, path-independent information measures in the rational inattention literature. The theorem’s role here is to recast that logic in an ex-ante framework and to show that it coincides with the validity-based approach adopted in this paper.

To state the result formally, it is first necessary to introduce notation for a compound signal

⁵The reason for assuming that u takes value in $\mathbb{R} \cup \{-\infty\}$ is that some common measures of uncertainty, such as negative Shannon entropy, require this convention.

⁶See Footnote 9 for the explanation behind this convention.

structure. Consider a first signal structure σ_0 . Suppose that, after observing a realization $s_0 \in S_0$, the agent draws a second signal from a structure $\sigma_1^{s_0}$ with signal space $S_1^{s_0}$ that depends on the first realization. The *compound signal structure* obtained by first sampling from σ_0 and then from $\sigma_1^{s_0}$ is denoted $\sigma_1 \circ \sigma_0$. Its support is $\{(s_0, s_1) \mid s_0 \in S_0, s_1 \in S_1^{s_0}\}$, and its likelihood function is $\sigma((s_0, s_1) \mid \omega) = \sigma_1^{s_0}(s_1 \mid \omega, s_0) \sigma_0(s_0 \mid \omega)$.

Below, I state the axioms formally.

Axiom 1 (Positivity). $I(p, \sigma) \geq 0$ for all $p \in \Delta(\Omega)$, $\sigma \in \Sigma$.

Axiom 2 (Posterior dependence). I depends on (p, σ) only through the distribution of posteriors $\tau_{p, \sigma} \in \Delta(\Delta(\Omega))$ induced by (p, σ) , where

$$\tau_{p, \sigma}(B) = \Pr(\{s \in S \mid p(\cdot \mid s) \in B\})$$

for all Borel sets $B \subseteq \Delta(\Omega)$.

Axiom 3 (Path independence). For any compound signal structure that first reveals s_0 and then s_1 conditional on s_0 ,

$$I(p, \sigma_1 \circ \sigma_0) = I(p, \sigma_0) + \mathbb{E}_{s_0 \sim \Pr_0} [I(p(\cdot \mid s_0), \sigma_1^{s_0})]$$

where $\mathbb{E}_{s_0 \sim \Pr_0}$ is the expectation with respect to the unconditional distribution of signals induced by (p, σ_0) .

Characterization of Ex-ante Measures of Information. The following theorem characterizes canonical ex-ante measures of information in terms of validity, the axioms introduced above, and a posterior-separable representation. Its main role is to provide the canonical benchmark for the analysis that follows.

Theorem 1 (Characterization of ex-ante measures of information). *Let $I : \Delta(\Omega) \times \Sigma \rightarrow \mathbb{R}$ be a (primitive) ex-ante measure of information. The following are equivalent:*

(i) (Validity) *There exists a decision problem (A, u) such that, for all (p, σ) ,*

$$I(p, \sigma) = \mathbb{E}_{s \sim \Pr} \left[\max_{a \in A} \mathbb{E}_{\omega \sim p(\cdot \mid s)} [u(a, \omega)] \right] - \max_{a \in A} \mathbb{E}_{\omega \sim p} [u(a, \omega)].$$

(ii) (Axioms) I satisfies Axioms 1, 2, and 3.

(iii) (Representation) There exists a convex function $G : \Delta(\Omega) \rightarrow \mathbb{R}$ such that, for all (p, σ) ,

$$I(p, \sigma) = \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p).$$

Moreover, if G and $\tilde{G}$ both represent I , then there exist $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^\Omega$ such that

$$\tilde{G}(p) = G(p) + \alpha + \beta \cdot p \quad \text{for every } p \in \Delta(\Omega).$$

The proof of [Theorem 1](#) can be found in [Appendix A.1](#). Here I simply provide some intuition. The convex function $G(p)$, henceforth referred to as the *convex potential*, can be interpreted as the value function in a decision problem for an agent with beliefs p . When useful, I write I^G for the canonical ex-ante measure generated by the potential G . The quantity $I^G(p, \sigma)$ is the expected improvement in the value generated by the signal structure σ . Because G is the upper envelope of linear functions, each representing the expected value of a specific action, G is convex. Convexity ensures that information is never harmful: the expected value from drawing signals and updating beliefs weakly exceeds the value of staying uninformed.

The convex potential G also implicitly encodes a penalty for acting on incorrect beliefs. This penalty is central to both canonical measures of information and measures of meaningful information. To build intuition for why, consider first the two-state case illustrated in [Figure 2](#). With two states, a belief can be represented by a single probability, so the value function G is a convex curve. Consider an action a_p that is optimal at belief p . Its expected payoff is linear in beliefs and is represented in the figure by the dashed line that touches G at p . The fact that the line touches G at p reflects the optimality of a_p at that belief. Indeed, in the two-state case, the values of this line at the endpoints 0 and 1 are precisely the state-contingent payoffs $u(a_p, \omega_2)$ and $u(a_p, \omega_1)$, respectively.

Now suppose that the relevant distribution of states is instead q , but the decision-maker acts on the basis of p and therefore chooses a_p . Her expected payoff is given by the height of the dashed line at q . Had she acted on the correct belief q , she would instead obtain $G(q)$. The vertical gap between these two objects, labeled $d^{(G,g)}(q, p)$ in [Figure 2](#), is therefore the welfare loss from taking

an action that is optimal at p when q is the relevant distribution of states.

The same construction extends directly to arbitrary finite state spaces. Let a_p denote the action selected at belief p , and let $g(p) = u(a_p, \cdot)$ denote its vector of state-contingent payoffs. Because a_p is optimal at p , this payoff vector defines a supporting hyperplane to G at p ; equivalently, $g(p) \in \partial G(p)$, where $\partial G(p)$ denotes the subdifferential of G at p .⁷ If the relevant distribution of states is q , the expected payoff from taking a_p is $G(p) + g(p) \cdot (q - p)$. In contrast, acting optimally at q yields $G(q)$. Hence the welfare loss from acting on p when q is the relevant distribution of states is

$$d^{(G,g)}(q, p) = G(q) - G(p) - g(p) \cdot (q - p).$$

This quantity is the *Bregman divergence* associated with the convex potential G and the selected subgradient $g(p)$, and it is always non-negative. Geometrically, $d^{(G,g)}(q, p)$ is the gap between $G(q)$ and the supporting hyperplane to G at p , evaluated at q . When G is differentiable relative to the simplex at p , the supporting hyperplane, and hence $d^{(G,g)}(q, p)$, is uniquely determined. Thus, when G is differentiable relative to the simplex at p , the subgradient selector g can be omitted.⁸ Conversely, when G is not differentiable at p , the divergence depends on the selected subgradient $g(p)$ or, equivalently, on the tie-breaking rule selecting among actions that are optimal at p .⁹ Thus, the convex potential G , together with the subgradient selector g , determines the welfare cost of acting on an incorrect belief.

The Bregman penalty for incorrect beliefs connects to canonical ex-ante measures of information via [Frankel and Kamenica \(2019\)](#)'s ex-post measures of information. Specifically, let us think of the “correct belief” after signal realization s as being $p(\cdot | s)$. Under this interpretation, acting on the prior p rather than on the correct belief $p(\cdot | s)$ generates the ex-post loss

$$d^{(G,g)}(p(\cdot | s), p) = G(p(\cdot | s)) - G(p) - g(p) \cdot (p(\cdot | s) - p),$$

which is precisely the ex-post measure of information characterized by [Frankel and Kamenica](#)

⁷The subdifferential of G at p , denoted $\partial G(p)$, is the set of subgradients of G at p . A subgradient is a vector $g(p)$ such that $G(q) \geq G(p) + g(p) \cdot (q - p)$ for every $q \in \Delta(\Omega)$. Geometrically, each subgradient defines a supporting hyperplane to G : the affine function $q \mapsto G(p) + g(p) \cdot (q - p)$ lies weakly below G everywhere and coincides with G at p . In the decision-theoretic interpretation, this supporting hyperplane gives the expected payoff, as beliefs vary, from taking an action that is optimal at p .

⁸Since G is convex, non-differentiability occurs only on a Lebesgue-null subset of beliefs relative to the simplex.

⁹ I treat the tie-breaking rule as part of the decision problem so that the decision problem pins down both G and g .

(2019).¹⁰ Taking expectations of the ex-post measure above with respect to the unconditional distribution of signals yields a canonical ex-ante measure of information:

$$I^G(p, \sigma) = \mathbb{E}_{s \sim \text{Pr}}[d^{(G,g)}(p(\cdot | s), p)] = \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p),$$

where the second equality follows from Bayes plausibility, $\mathbb{E}_{s \sim \text{Pr}}[p(\cdot | s)] = p$.¹¹

The discussion above yields two takeaways. First, the convex potential G does more than simply help represent the canonical value of information: together with the selected supporting hyperplane, it also encodes the decision-relevant penalty from acting on incorrect beliefs. Second, canonical measures of information can be understood as the expected reduction in that penalty when the agent moves from the prior to the correct posterior. Importantly, this canonical ex-ante measure depends only on G , even when the underlying ex-post penalty depends on the selector g .

This perspective provides the natural benchmark for measures of meaningful information. Once agents are allowed to misinterpret signals or update beliefs in a non-Bayesian way, the relevant question is no longer how much the signal structure could reduce the penalty for incorrect beliefs under correct interpretation, but rather how much of that reduction is actually realized under the beliefs the agent actually forms. As I show in the next section, measures of meaningful information leverage the geometry of G to quantify the penalty from acting on incorrect beliefs arising from misspecification and non-Bayesian updating.

Two examples of canonical ex-ante measures of information might help solidify intuition. The first measure is mutual information. Let G be the negative of Shannon’s entropy function, $G(p) = \sum_{\omega} p(\omega) \log p(\omega)$. The associated Bregman divergence is the Kullback-Leibler divergence $d^{(G,g)}(p, q) = \text{KL}(p \| q)$, and the measure of information transmission is mutual information $I^{\text{Sh}}(p, \sigma) = \mathbb{E}_{s \sim \text{Pr}}[\text{KL}(p(\cdot | s) \| p)]$ (Kullback and Leibler, 1951).¹² The second measure relates to mean-squared forecast error (Brier, 1950). Let $G(p) = \|p\|^2$, where $\|\cdot\|$ denotes the Euclidean norm on the simplex. Then $d^{(G,g)}(x, y) = \|x - y\|^2$, so $I^Q(p, \sigma) = \mathbb{E}_{s \sim \text{Pr}}[\|p(\cdot | s) - p\|^2]$ is the

¹⁰Frankel and Kamenica (2019) use a concave uncertainty index H and its associated Bregman divergence based on supergradients; setting $G = -H$ flips the sign but yields the same object.

¹¹Note that, although the ex-post Bregman penalty may depend on the selected subgradient, its expectation under correct Bayesian updating does not: the selector-dependent linear term vanishes by Bayes plausibility.

¹²For the meaningful-information extension of mutual information, I restrict attention to the natural support-compatible domain on which all relevant KL divergences are finite. Otherwise, the meaningful information extension of mutual information might be undefined.

expected reduction in mean-squared forecast error.

1.2 Measures of Meaningful Information Transmission

Canonical measures of information transmission quantify the expected improvement in an agent's decisions when the agent has an accurate understanding of the information environment and updates beliefs according to Bayes' rule. In many real-world settings, however, agents might have a misspecified prior, might misinterpret signals, or might deviate from Bayesian updating.

This section develops measures of information transmission that preserve the decision-theoretic foundation of the canonical approach while allowing for misspecification and non-Bayesian belief updating. I refer to these new measures as measures of *meaningful* information transmission in order to highlight how the preconditions for meaning to be conveyed are encoded in the signal structure, but the actual transmission of meaning depends on the agent interpreting (decoding) that structure correctly.¹³

The political-correctness application helps illustrate the core idea. Suppose students' public statements are shaped by social-image concerns, so that what they say in public is less closely related to their privately held political views than what they say in private. A listener may nevertheless misperceive this mapping; for instance, by taking public statements at face value or, conversely, by overcorrecting for the effect of social-image concerns. By misunderstanding how underlying political views map into public speech, the listener is misspecified about the signal structure she faces. It is therefore unclear how much information she actually extracts from what others say. On the one hand, public statements may objectively contain valuable information about speakers' political views; on the other, listeners who misinterpret those statements form incorrect beliefs and may be misled by observing them. To capture the amount of information listeners actually extract from political speech, one needs to take into account both the features of the signal structure that map latent views to political speech and the agents' interpretations of that speech.

In order to allow agents to misunderstand the information environment, I introduce a distinction between *objective uncertainty*, which refers to the true data-generating process, and *subjective interpretations*, which refer to the agent's subjective beliefs about that process.

¹³This terminology is meant to evoke a theory of communication akin to the one in [Lewis \(1969\)](#). See also [Skyrms \(2010\)](#).

Objective Uncertainty and Subjective Interpretations. I adopt the notation from the previous subsection to describe objective uncertainty. Let $\Omega = \{\omega_1, \dots, \omega_n\}$ be a finite set of states of the world and let $p \in \Delta(\Omega)$ be the *objective* distribution of states. Let $\sigma \in \Sigma$ be a finite *objective* signal structure with finite signal space S . For each $s \in S$, the pair (p, σ) induces an *objective* posterior $p(\cdot | s) \in \Delta(\Omega)$. Denote the unconditional distribution of signals by $\Pr \in \Delta(S)$.

The subjective counterparts to the objects above need not coincide with the objective probabilities or be related to one another via Bayes' rule. Let $\hat{p} \in \Delta(\Omega)$ denote the agent's *subjective* prior and $\hat{\sigma}$ her *subjective* perception of the signal structure. I require that σ and $\hat{\sigma}$ share the same signal alphabet S , but nothing more. Because I do not require Bayesian updating, I define an operator $\phi : \Delta(\Omega) \times \Sigma \rightarrow \Delta(\Omega)^S$ that maps $(\hat{p}, \hat{\sigma})$ into the family of *subjective* posteriors $\{\hat{p}(\cdot | s)\}_{s \in S}$. I let Φ denote the class of operators ϕ and I denote the Bayesian operator by ϕ^* .

Henceforth, “objective uncertainty” refers to p , σ , and $\{p(\cdot | s)\}_{s \in S}$, while “subjective beliefs” refer to $\hat{p}$, $\hat{\sigma}$, and $\{\hat{p}(\cdot | s)\}_{s \in S}$.

Definition 2. A (primitive) ex-ante measure of meaningful information is a functional $M : \Delta(\Omega)^2 \times \Sigma^2 \times \Phi \rightarrow \mathbb{R}$ that assigns to each objective prior p , objective signal structure σ , subjective prior $\hat{p}$, perceived signal structure $\hat{\sigma}$, and belief-updating operator ϕ a real number summarizing how informative the signal structure is for an agent with the specified subjective interpretations.

As in the canonical case, two complementary approaches can be used to justify the properties that a measure of meaningful information should satisfy: an instrumental approach and an axiomatic approach. The *instrumental* approach still requires that M equal the value of information in some decision problem. However, when subjective and objective uncertainty might differ, the notion of value of information needs to be re-defined. The *axiomatic* approach takes as its starting point a set of intuitive properties a measure of meaningful information should satisfy. I elaborate on each approach in turn and, in the main theorem of this section, show once again that they are equivalent and that they admit a straightforward representation.

Instrumental Approach. When an agent holds misspecified beliefs or deviates from Bayesian updating, the canonical value of information remains a natural welfare benchmark from the agent's perspective. From the perspective of an analyst who knows the objective information environment, however, it is less natural.

In this paper, I adopt a welfare criterion under which the agent chooses actions based on her subjective beliefs, but the consequences of those actions are evaluated using the objective probabilities. This approach to welfare is common in related work on distorted beliefs and misspecification and is natural in many applications (Battigalli et al., 2022; Morris and Shin, 1997; Frick, Iijima and Ishii, 2024; Whitmeyer, 2026).¹⁴

Formally, fix a decision problem with action set A and payoff function $u : A \times \Omega \rightarrow \mathbb{R}$. Once again, I take a tie-breaking convention as part of the decision problem. For any subjective belief $\hat{p} \in \Delta(\Omega)$, let

$$a_{\hat{p}} \in \arg \max_{a \in A} \mathbb{E}_{\omega \sim \hat{p}}[u(a, \omega)]$$

be the element of the set of maximizers at $\hat{p}$ selected by the tie-breaking convention. Thus, the agent takes a subjectively optimal action at $\hat{p}$. Realized welfare, however, is evaluated under the objective distribution p :

$$W(p, \hat{p}) \equiv \mathbb{E}_{\omega \sim p}[u(a_{\hat{p}}, \omega)].$$

This welfare criterion leads naturally to a corresponding notion of the value of information: the expected change in realized welfare when the agent can condition her choice on the realization of a signal. Because this object reflects both the objective environment and the agent's subjective beliefs and updating behavior, I refer to it as the *behavioral* value of information.

Definition 3. The behavioral value of information is

$$V(p, \sigma, \hat{p}, \hat{\sigma}, \phi) \equiv \mathbb{E}_{s \sim \text{Pr}}[W(p(\cdot | s), \hat{p}(\cdot | s))] - W(p, \hat{p}) \quad (1)$$

where the expectation $\mathbb{E}_{s \sim \text{Pr}}$ is taken with respect to the objective unconditional distribution of signals.

In the previous section, I referred to a measure of information as valid if it corresponded to the value of information in some decision problem. In this section, I define a measure of *meaningful* information as *behaviorally valid* if it corresponds to the *behavioral value of information* in some decision problem.

¹⁴See Brunnermeier, Simsek and Xiong (2014) for a discussion of this welfare criterion and for a possible alternative.

Definition 4. A measure of meaningful information M is *behaviorally valid* if there exists a decision problem (A, u) such that

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = V(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$$

in that decision problem.

Axiomatic Approach. Because measures of meaningful information allow for misspecification and non-Bayesian updating, the canonical axioms alone do not identify them. Those axioms, however, are still normatively appealing. I therefore adopt natural counterparts of the previous section's axioms now stated in the joint objective/subjective environment.

In addition, I introduce two new axioms, which I call *invariance to null updating* and *dominance at correct posteriors*. Invariance to null updating requires that, if an agent does not update her beliefs at all in response to a signal, the amount of meaningful information transmitted to the agent is zero. Dominance at correct posteriors requires that any discrepancy between the objective posteriors and the subjective posteriors weakly decrease the amount of meaningful information transmitted to the agent. For an agent whose subjective prior equals the objective one, the requirement above is tantamount to requiring that any misinterpretation of signals or non-Bayesian updating weakly decrease the amount of meaningful information transmitted to the agent.¹⁵

I now present the axioms formally.

Axiom 4 (Positivity at correct beliefs). For any prior $p \in \Delta(\Omega)$ and signal structure $\sigma \in \Sigma$,

$$M(p, \sigma, p, \sigma, \phi^*) \geq 0.$$

Axiom 5 (Joint posterior dependence). M depends on $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$ only through $\hat{p}$ and the joint distribution of posterior pairs $\tau_{p, \sigma, \hat{p}, \hat{\sigma}, \phi} \in \Delta(\Delta(\Omega) \times \Delta(\Omega))$ induced by $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$, where

$$\tau_{p, \sigma, \hat{p}, \hat{\sigma}, \phi}(B) = \Pr(\{s \in S \mid (p(\cdot \mid s), \hat{p}(\cdot \mid s)) \in B\})$$

¹⁵Equivalently, dominance at correct posteriors has a “data processing” interpretation: subjective posteriors are functions of the realized signal, so any interpretation that does not use additional state-dependent information cannot increase what the agent can infer about the state relative to conditioning on the signal itself (i.e., relative to the objective posteriors).

for all Borel sets $B \subseteq \Delta(\Omega) \times \Delta(\Omega)$.

Axiom 6 (Joint path independence). For any compound signal structure that first reveals s_0 via σ_0 and, conditional on s_0 , reveals s_1 via $\sigma_1^{s_0}$ (with subjective counterparts $(\hat{\sigma}_0, \phi_0)$, and $(\hat{\sigma}_1^{s_0}, \phi_1^{s_0})$),

$$\begin{aligned} M(p, \sigma_1 \circ \sigma_0, \hat{p}, \hat{\sigma}_1 \circ \hat{\sigma}_0, \phi_1 \circ \phi_0) \\ = M(p, \sigma_0, \hat{p}, \hat{\sigma}_0, \phi_0) + \mathbb{E}_{s_0 \sim \text{Pr}_0} [M(p(\cdot | s_0), \sigma_1^{s_0}, \hat{p}(\cdot | s_0), \hat{\sigma}_1^{s_0}, \phi_1^{s_0})] \end{aligned}$$

where $\mathbb{E}_{s_0 \sim \text{Pr}_0}$ is the expectation with respect to the objective unconditional distribution of signals induced by (p, σ_0) .

A clarification about the scope of joint path independence might be useful. The axiom does not require the sequential belief-updating procedure to generate the same terminal beliefs as a one-shot updating rule that processes the compound signal (s_0, s_1) directly. It requires only that meaningful information be additive along the specified sequential process: the amount transmitted by the compound experiment equals the amount transmitted in the first stage plus the expected amount transmitted in the continuation. In other words, joint path independence is a requirement on M , not on the agent's potentially non-Bayesian updating procedure.

Axiom 7 (Invariance to null updating). For any objective prior $p \in \Delta(\Omega)$, any subjective prior $\hat{p} \in \Delta(\Omega)$, any objective signal structure $\sigma \in \Sigma$, and any subjective signal structure $\hat{\sigma}$

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi^{\hat{p}}) = 0$$

where $\phi^{\hat{p}}$ is the operator that sets $\hat{p}(\cdot | s) = \hat{p}$ after each signal realization.

Axiom 8 (Dominance at correct posteriors). For any objective prior $p \in \Delta(\Omega)$, any subjective prior $\hat{p} \in \Delta(\Omega)$, and any objective signal structure $\sigma \in \Sigma$,

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi^{\sigma}) \geq M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) \quad \forall \hat{\sigma} \in \Sigma \text{ and } \phi \in \Phi$$

where ϕ^{σ} is the operator that sets $\hat{p}(\cdot | s) = p(\cdot | s)$ for all $s \in S$.

Characterization of Ex-ante Measures of Meaningful Information. The main theoretical contribution of this section of the paper is to characterize ex-ante measures of meaningful information

transmission in terms of behavioral validity, the axioms introduced above, and a representation.

Theorem 2 (Characterization of ex-ante measures of meaningful information). *Let $M : \Delta(\Omega)^2 \times \Sigma^2 \times \Phi \rightarrow \mathbb{R}$ be a (primitive) ex-ante measure of meaningful information. The following are equivalent:*

(i) (Behavioral validity) *There exists a decision problem (A, u) such that, for all $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$,*

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = \mathbb{E}_{s \sim \text{Pr}} \left[W(p(\cdot | s), \hat{p}(\cdot | s)) \right] - W(p, \hat{p}),$$

(ii) (Axioms) *M satisfies Axioms 4, 5, 6, 7, and 8.*

(iii) (Representation) *There exists a convex potential $G : \Delta(\Omega) \rightarrow \mathbb{R}$ and a fixed selector $g(\cdot) \in \partial G(\cdot)$ such that, for all $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$,*

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = \underbrace{\mathbb{E}_{s \sim \text{Pr}} [G(p(\cdot | s))] - G(p)}_{\text{canonical ex-ante information}} + \underbrace{d^{(G, g)}(p, \hat{p})}_{\text{prior misspecification correction}} - \underbrace{\mathbb{E}_{s \sim \text{Pr}} [d^{(G, g)}(p(\cdot | s), \hat{p}(\cdot | s))]}_{\text{posterior misspecification correction}} \quad (2)$$

where $d^{(G, g)}(x, y) := G(x) - G(y) - g(y) \cdot (x - y)$. Moreover, the value of M is invariant to the choice of affine representative of G (with the corresponding affine shift of the selector g).

Equation (2) shows that meaningful information equals the signal's objective informational content plus two corrections: one for prior misspecification and one for posterior misspecification. The first term in the representation is a familiar canonical information measure: it depends only on the objective joint distribution of states of the world and signal realizations, and it captures how much the signal would improve the utility of a correctly specified Bayesian agent with value function G . The remaining two terms isolate what is new in environments with misspecification and non-Bayesian updating.

The term $d^{(G, g)}(p, \hat{p})$ is the welfare gain from replacing a misspecified prior with the objective prior before any signal is observed, measured in the geometry induced by G . A concrete setting in which an agent can reap that gain is if she begins with a misspecified but non-dogmatic prior and then correctly interprets a perfectly informative signal. In that case, sampling from the signal delivers a “double dividend”: it reveals the state of the world and it prevents the agent from acting on her misspecified prior. The latter benefit is captured by $d^{(G, g)}(p, \hat{p})$.

$\mathbb{E}_{s \sim \text{Pr}} \left[d^{(G,g)}(p(\cdot | s), \hat{p}(\cdot | s)) \right]$ captures the average penalty for holding incorrect posterior beliefs in the geometry induced by G . Subjective posteriors might deviate from objective ones because the agent starts off with an incorrect prior and samples from a signal structure that is less than fully informative, because she has a misspecified perception of the signal structure, or because of non-Bayesian updating.

I return to the interpretation of [Theorem 2](#) after introducing the main result of the next section, which makes precise the relationship between canonical ex-ante measures of information and measures of meaningful information. When useful, I write $M^{(G,g)}$ for the measure of meaningful information generated by the potential G and the selector g .

1.3 Relation to Canonical Ex-ante Measures

The decomposition in [\(2\)](#) already hints at how canonical measures of information and measures of meaningful information relate. This subsection formalizes that connection.

I begin by introducing two notions. The first ties a canonical information measure I and a meaningful information measure M to the same underlying decision problem. The second formalizes the idea that, since the domain of I lies on a lower-dimensional slice of the domain of M , we can view M as an extension of I from that slice to the full domain.

Definition 5 (Joint validity). A valid ex-ante measure of information I and a behaviorally valid ex-ante measure of meaningful information M are *jointly valid* if they arise in the context of the same decision problem (A, u) .

Definition 6 (Extension to subjective interpretations). Let $D \subset \Delta(\Omega)^2 \times \Sigma^2 \times \Phi$ denote the slice of the domain of M where I takes values.

$$D := \{ (p, \sigma, p, \sigma, \phi^*) : (p, \sigma) \in \Delta(\Omega) \times \Sigma \},$$

where ϕ^* is the Bayesian updating operator. Given $I : \Delta(\Omega) \times \Sigma \rightarrow \mathbb{R}$ and $M : \Delta(\Omega)^2 \times \Sigma^2 \times \Phi \rightarrow \mathbb{R}$, we say that M *extends* I to subjective interpretations if

$$M(p, \sigma, p, \sigma, \phi^*) = I(p, \sigma) \quad \text{for all } (p, \sigma) \in \Delta(\Omega) \times \Sigma.$$

The main theoretical contribution of this section is to leverage the definitions above to establish a formal connection between canonical ex-ante measures of information transmission and ex-ante measures of meaningful information transmission.

Theorem 3. *Let I be a valid ex-ante measure of information and M a behaviorally valid ex-ante measure of meaningful information. The following are equivalent:*

- (i) *I and M are jointly valid.*
- (ii) *If G is a convex potential representing I as in [Theorem 1](#), then there exists a fixed selector $g(\cdot) \in \partial G(\cdot)$ such that $M = M^{(G,g)}$, where*

$$M^{(G,g)}(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = I^G(p, \sigma) + d^{(G,g)}(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}} \left[d^{(G,g)}(p(\cdot | s), \hat{p}(\cdot | s)) \right]. \quad (3)$$

- (iii) *Conditional on the fixed selector $g(\cdot)$ in (ii), $M^{(G,g)}$ is the unique extension of I^G to subjective interpretations that satisfies Axioms [4](#), [5](#), [6](#), [7](#), and [8](#).*

[Theorem 3](#) ties together the two parallel structures I developed so far, one describing canonical ex-ante measures of information, the other describing measures of meaningful information. [Theorem 1](#) and [Theorem 2](#) show that each structure admits a representation that relies on a convex potential G . [Theorem 3](#) closes the loop by showing that these two representations are in fact the same object viewed from different vantage points.

[Theorem 3](#) can be unpacked as follows. Part (i) introduces joint validity as a conceptual bridge between canonical measures of information and measures of meaningful information. Part (ii) shows that, under joint validity, the meaningful information extension of a canonical measure is obtained by adding two Bregman-divergence adjustments to the canonical measure. Part (iii) shows that this extension is unique: once the axioms are imposed, there exists one and only one way to extend any valid canonical measure to environments with subjective interpretations.

[Theorem 3](#) delivers two key insights. First, it provides an operational recipe: given any canonical measure represented by a convex potential G , one immediately obtains the corresponding meaningful information measure by applying the simple adjustment in part (ii). Importantly, the invariance of both I and M to affine transformations of G ensures that the decomposition in [Theorem 3](#) is well defined and uniquely determined by the geometry of G itself.

Second, the theorem shows that meaningful information is not an ad-hoc behavioral modification but rather a structural continuation of the canonical framework. Since the Bregman divergences that appear in M are tied to the convex potential G used in the representation of I , meaningful information inherits its structure directly from canonical measures of information. When subjective and objective beliefs coincide, the Bregman terms vanish and M reduces exactly to I ; when they differ, the gap between M and I quantifies the informational consequences of misinterpretation and non-Bayesian updating.

The observation that each measure of meaningful information transmission inherits its structure directly from its jointly valid canonical counterpart reconnects with the earlier discussion of ex-post measures of information in [Frankel and Kamenica \(2019\)](#). There, I argued that the Bregman divergence $d^{(G,g)}(p(\cdot | s), p)$ can be interpreted as the ex-post penalty, in the geometry induced by the convex potential G and the fixed selector g , from acting on the prior p rather than on the correct posterior $p(\cdot | s)$ once signal realization s is observed. In that sense, the prior can be interpreted as an incorrect belief relative to the posterior as objective uncertainty is resolved, and $d^{(G,g)}$ measures the associated loss.

[Theorem 3](#) extends this logic. It shows that the same divergence $d^{(G,g)}$ remains the primitive object that quantifies the penalty from acting on incorrect beliefs, but now in a broader environment in which incorrect beliefs may arise not only because the agent has not yet conditioned on the realized signal, but also because she starts from a misspecified prior, misinterprets the signal structure, or updates in a non-Bayesian way. As a result, the geometry of G provides a single yardstick for measuring, in common units, both the gains from the resolution of objective uncertainty and the losses generated from misinterpretations.

1.3.1 Comparing Meaningful and Canonical Information

The representation in (3) makes it immediate to compare meaningful information to its jointly valid canonical counterpart.¹⁶

Corollary 1. Consider jointly valid I and M .

¹⁶Since affine transformations of the potential G leave both I (by [Theorem 1](#)) and the right-hand side of (2) (by [Theorem 2](#)) unchanged, the comparison $M \gtrless I$ and the conditions in [Corollary 1](#) are invariant to the choice of affine representative of G .

(i) $M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) > I(p, \sigma)$ if and only if

$$\underbrace{d^{(G,g)}(p, \hat{p})}_{\text{prior misspecification correction}} > \underbrace{\mathbb{E}_{s \sim \text{Pr}}[d^{(G,g)}(p(\cdot | s), \hat{p}(\cdot | s))]}_{\text{posterior misspecification correction}}$$

(ii) If $\hat{p}(\cdot | s) = p(\cdot | s)$ for all s , then

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = I(p, \sigma) + d^{(G,g)}(p, \hat{p}) \geq I(p, \sigma).$$

(iii) If $\hat{p} = p$, then

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = I(p, \sigma) - \mathbb{E}_{s \sim \text{Pr}}[d^{(G,g)}(p(\cdot | s), \hat{p}(\cdot | s))] \leq I(p, \sigma).$$

Part (i) of [Corollary 1](#) shows that a measure of meaningful information exceeds its jointly valid canonical counterpart if and only if the prior misspecification correction exceeds average posterior misspecification correction. Operationally, this occurs whenever the signal, on average, pushes the agent's posterior beliefs sufficiently toward the objective posteriors so that, in the geometry induced by G , the agent's expected divergence from the objective benchmark is smaller after updating than before.

Part (ii) illustrates a particularly transparent case in which measures of meaningful information are guaranteed to exceed their canonical counterparts. Whenever subjective posteriors are correct after every signal and the prior is misspecified, meaningful information exceeds the canonical measure. As discussed in the context of [Theorem 2](#), this can happen, for instance, when the agent has a misspecified but non-dogmatic prior and she samples from and correctly interprets a perfectly informative signal structure.

Part (iii) follows directly from [Axiom 8](#). For an agent with a correct prior, misinterpretations can never increase the amount of meaningful information extracted from a signal structure. The difference between a measure of meaningful information and its canonical counterpart is thus weakly negative and reflects the expected penalty from posterior distortion.

The discussion above helps bring to light two ways in which measures of meaningful information depart starkly from canonical ones.

Two Qualitative Departures from the Canonical Framework First, measures of meaningful information can be strictly negative. Axiom 4 requires measures of meaningful information to be non-negative when beliefs are correct, but neither this axiom nor the full set of axioms guarantees non-negativity elsewhere. By the tripartite decomposition in Theorem 2, meaningful information is negative exactly when, in the geometry induced by the convex potential G that underlies M , the expected penalty from posterior misspecification dominates the combined benefits from canonical ex-ante information and any prior correction.

An agent with a correct prior ($\hat{p} = p$) who samples from an uninformative signal structure ($p(\cdot | s) = p$ for all s) but mistakenly interprets it as informative ($\hat{p}(\cdot | s) \neq p$ for at least one s) provides the starkest example: in that case, Equation (2) reduces to

$$M(p, \sigma, p, \hat{\sigma}, \phi) = -\mathbb{E}_{s \sim \text{Pr}}[d^{(G, g)}(p, \hat{p}(\cdot | s))] \leq 0,$$

with strict inequality whenever G is strictly convex and $\hat{p}(\cdot | s) \neq p$ for some s with $\text{Pr}(s) > 0$.

Negative meaningful information is not a pathology; rather, it is a natural feature of environments in which signals may be misinterpreted. In the context of a given decision problem, an agent may form posterior beliefs that are sufficiently inaccurate that she would have been better off ignoring the signal altogether.

Second, measures of meaningful information need not be monotone in the Blackwell order. Recall that $\sigma_1 \succeq_B \sigma_2$ (“ σ_1 Blackwell-dominates σ_2 ”) if, for every decision problem, the value of information is weakly higher under σ_1 than under σ_2 (Blackwell, 1951, 1953). Since valid canonical ex-ante measures of information correspond to the value of information in some decision problem, $\sigma' \succeq_B \sigma$ implies $I(p, \sigma') \geq I(p, \sigma)$ for every valid canonical measure I .

No analogous implication holds for measures of meaningful information. If the agent does not update her subjective beliefs, Axiom 7 implies $M(p, \sigma, \hat{p}, \hat{\sigma}, \phi^{\hat{p}}) = 0$, for every σ , so meaningful information can be flat across Blackwell-ranked signal structures. More strikingly, reversals are possible. For instance, consider an environment with two states of the world and two possible signal realizations. Furthermore, suppose the agent systematically misinterprets which signal realization is diagnostic of which state. As the signal becomes objectively more informative, this mistake pushes the agent’s posterior beliefs further in the wrong direction. If the potential G under-

lying the meaningful information measure is strictly convex, meaningful information strictly falls. I formalize this example in [Appendix B](#).

1.4 Cardinal Measures: Takeaways

The main takeaway from [Section 1](#) is that measures of meaningful information transmission provide a natural extension of canonical measures to environments in which agents may misinterpret the information they receive. This extension is not imposed from outside the canonical framework: it is generated by the convex geometry that underlies standard cardinal measures of information. That geometry provides a common yardstick for both the gains from the resolution of objective uncertainty and the losses from subjective (mis)interpretations. Meaningful information therefore captures, in the same units as canonical measures of information, how much of a signal’s objective informational content is actually converted into welfare-relevant understanding by the agent who receives it.

The characterizations in [Section 1](#) are, by construction, decision-problem dependent. Indeed, this dependence is built into the notion of validity: each cardinal measure evaluates information through the welfare consequences generated by a particular decision problem. When an application leaves the downstream decision problem unspecified, however, [Section 1](#) does not provide guidance about which G to use, and different choices of G may rank interpreted experiments differently. This raises a natural question: which comparisons, if any, hold *robustly* across decision problems? The next section addresses this question.

2. A Blackwell-style Order for Interpreted Experiments

A natural way to address the dependence of cardinal measures of information (whether canonical or meaningful) on an underlying decision problem is to relinquish the idea of obtaining a complete cardinal ranking and, instead, seek a partial order that identifies the comparisons on which all decision problems agree. For canonical ex-ante measures of information, Blackwell’s partial order provides exactly such a comparison ([Blackwell, 1951, 1953](#)). In the language of this paper, Blackwell’s order characterizes when one signal structure yields weakly more information than another according to *all* canonical ex-ante measures characterized in [Theorem 1](#).

The corresponding robust-comparison problem for misinterpreted experiments was studied in Morris (1991). Morris asks when one interpreted experiment is uniformly preferred to another across all decision problems and provides a necessary-and-sufficient characterization in terms of a flow condition.¹⁷

In this section, I integrate the robust comparison of interpreted experiments with the framework developed in Section 1. Specifically, I express this robust comparison as unanimity across the behaviorally valid measures characterized in Theorem 2 and derive a new equivalent characterization of the resulting partial order on interpreted experiments in terms of the convex geometry from Section 1.

I begin the analysis by formalizing the notion of an interpreted experiment.

Definition 7. An *interpreted experiment* $\widehat{\mathcal{E}}_i = (p, \sigma_i, \hat{p}, \hat{\sigma}_i, \phi_i)$ consists of an objective prior $p \in \Delta(\Omega)$, an objective signal structure $\sigma_i : \Omega \rightarrow \Delta(S_i)$, a subjective prior $\hat{p} \in \Delta(\Omega)$, a perceived signal structure $\hat{\sigma}_i : \Omega \rightarrow \Delta(S_i)$, and an updating rule ϕ_i .

Throughout this section, I restrict attention to comparisons between interpreted experiments that share the same objective prior p and the same subjective prior $\hat{p}$. This is a natural restriction. As in the canonical Blackwell setting, holding the objective prior fixed ensures that the comparison isolates differences in the information structures rather than differences in the underlying distribution of states. Holding the subjective prior fixed plays an analogous role on the interpretation side: it ensures that differences in meaningful information arise from how the two experiments are interpreted, rather than from differences in the agent’s initial beliefs.

With the definition of an interpreted experiment at hand, I can simplify notation and write

$$M^{(G,g)}(\widehat{\mathcal{E}}_i) := M^{(G,g)}(p, \sigma_i, \hat{p}, \hat{\sigma}_i, \phi_i)$$

for the amount of meaningful information transmitted under interpreted experiment $\widehat{\mathcal{E}}_i$ according to the behaviorally valid measure generated by the convex potential G and the selector g .

I next introduce a binary relation that captures robust comparisons across measures of meaningful information.

¹⁷I present this condition in the notation of the current paper below.

Definition 8. Consider two interpreted experiments $\widehat{\mathcal{E}}_1$ and $\widehat{\mathcal{E}}_2$ with a common objective prior p and a common subjective prior $\hat{p}$. I write $\widehat{\mathcal{E}}_1 \succeq_{\text{MI}} \widehat{\mathcal{E}}_2$ if $M^{(G,g)}(\widehat{\mathcal{E}}_1) \geq M^{(G,g)}(\widehat{\mathcal{E}}_2)$ for every behaviorally valid measure $M^{(G,g)}$ characterized in [Theorem 2](#).

By [Theorem 2](#), the comparison $\widehat{\mathcal{E}}_1 \succeq_{\text{MI}} \widehat{\mathcal{E}}_2$ is equivalent to requiring that $\widehat{\mathcal{E}}_1$ yield a weakly higher behavioral value of information than $\widehat{\mathcal{E}}_2$ in every decision problem. Thus, when $\widehat{\mathcal{E}}_1 \succeq_{\text{MI}} \widehat{\mathcal{E}}_2$, I say that $\widehat{\mathcal{E}}_1$ is uniformly better than $\widehat{\mathcal{E}}_2$ in the Meaningful-Information order.¹⁸

I now introduce the objects used to characterize this order. First of all, a subscript i attached to the notation from [Section 1](#) indicates that the object relates to interpreted experiment $\widehat{\mathcal{E}}_i$. For instance, Pr_i indicates the objective unconditional distribution of signals for interpreted experiment $\widehat{\mathcal{E}}_i$, $\hat{p}_i(\cdot|s)$ indicates the agent's subjective posterior after observing signal realization s for interpreted experiment $\widehat{\mathcal{E}}_i$, etc.

Second, I denote by Q_i the set of posterior beliefs that the agent acts upon under interpreted experiment i :

$$Q_i = \{q \in \Delta(\Omega) : q = \hat{p}_i(\cdot | s) \text{ for some } s \in S_i \text{ with } \text{Pr}_i(s) > 0\}$$

and I let $Q := Q_1 \cup Q_2$. Note that Q is finite, because S_1 and S_2 are both assumed to be finite.

The next pieces of notation are necessary to state the condition in [Morris \(1991\)](#). For each $i \in \{1, 2\}$, $q \in Q$, and $\omega \in \Omega$, define

$$k_q^i(\omega) := \sum_{\substack{s \in S_i: \text{Pr}_i(s) > 0 \\ \hat{p}_i(\cdot|s) = q}} \text{Pr}_i(s) p_i(\omega | s).$$

Thus, $k_q^i(\omega)$ is the objective joint probability, under interpreted experiment i , that the state is ω and that the agent acts on posterior belief q . Let $k_q^i \in \mathbb{R}^\Omega$ denote the vector of $k_q^i(\omega)$.

Let

$$D_q := k_q^1 - k_q^2, \quad q \in Q.$$

The vector $D_q \in \mathbb{R}^\Omega$ summarizes how the objective joint probability of the state and the acted-upon belief q differs across the two interpreted experiments.

¹⁸As with Blackwell's relation on signal structures, $\succeq_{\text{MI}}$ is formally a preorder on interpreted experiments and induces a partial order once interpreted experiments that mutually dominate one another are identified.

In the notation of the current paper, Morris's condition requires the existence of non-negative weights $\lambda = (\lambda_{qq'})_{q,q' \in Q}$ satisfying the following system of vector equations:¹⁹

$$D_q = \sum_{q' \in Q} [\lambda_{qq'} q - \lambda_{q'q} q'], \quad q \in Q. \quad (\text{F})$$

Each equality in Condition (F) is an equality in $\mathbb{R}^\Omega$.

The last main theoretical contribution of the paper connects robust comparisons across measures of meaningful information, Morris's flow condition, and the convex geometry developed in Section 1.

Theorem 4. *Consider two interpreted experiments $\widehat{\mathcal{E}}_1$ and $\widehat{\mathcal{E}}_2$ with a common objective prior p and a common subjective prior $\hat{p}$. The following are equivalent:*

1. $\widehat{\mathcal{E}}_1 \succeq_{\text{MI}} \widehat{\mathcal{E}}_2$.
2. Condition (F) has a non-negative solution $\lambda = (\lambda_{qq'})_{q,q' \in Q}$.
3. There exist non-negative weights $\lambda = (\lambda_{qq'})_{q,q' \in Q}$ such that, for every convex potential G and every fixed selector $g(\cdot) \in \partial G(\cdot)$,

$$M^{(G,g)}(\widehat{\mathcal{E}}_1) - M^{(G,g)}(\widehat{\mathcal{E}}_2) = \sum_{q,q' \in Q} \lambda_{qq'} d^{(G,g)}(q, q'). \quad (4)$$

Any non-negative solution to Condition (F) provides weights satisfying Equation (4). Moreover, if Condition (F) is infeasible, there exists a finite decision problem in which the optimal action is unique at every acted-upon belief $q \in Q$ and in which the behavioral value of information is strictly lower under $\widehat{\mathcal{E}}_1$ than under $\widehat{\mathcal{E}}_2$.

The equivalence between parts (i) and (ii) recovers Morris's robust-comparison result in the language of interpreted experiments and measures of meaningful information. Specifically, the characterization in Theorem 2 translates unanimity across decision problems into unanimity across measures of meaningful information that satisfy Axioms 4, 5, 6, 7, and 8.

The additional contribution of Theorem 4 relative to Morris (1991) is the equivalence of parts (i) and (iii). This equivalence connects the robust comparison of interpreted experiments across

¹⁹See Morris (1991), Chapter III, Claim 2.

decision problems to the geometry from [Section 1](#). Specifically, it shows that the same convex geometry that determines each individual measure of meaningful information also determines the comparisons on which *all* such measures agree.

Part (iii) of [Theorem 4](#) can be interpreted as a portfolio representation of the meaningful-information order. The new representation has three components. The first is a collection of elementary assets, indexed by ordered pairs of beliefs (q, q') . Each asset corresponds to avoiding a pairwise mistake: taking an action that is optimal at q' when q is the relevant distribution of states. The second component is the portfolio's exposure to these assets. The non-negative weights $\lambda_{qq'}$, which depend only on the interpreted experiments themselves, summarize how the two interpreted experiments differ in their exposure to each pairwise mistake. The third component is the price of each asset. A particular cardinal measure of meaningful information, through its underlying convex geometry, assigns to each pairwise mistake (q, q') a penalty $d^{(G, g)}(q, q')$. Thus, the interpreted experiments determine the portfolio of exposures, while the decision problem determines how those exposures are priced. The economic interpretation is that a dominating experiment can be represented as avoiding a non-negative portfolio of pairwise mistakes relative to the dominated experiment.

The final statement of the theorem clarifies the meaning of a failure of robust dominance. If Condition (F) has no non-negative solution, the failure of $\hat{\mathcal{E}}_1 \succeq_{\text{MI}} \hat{\mathcal{E}}_2$ is not an artifact of an obscure tie-breaking rule or of an artificial decision problem with an infinite action space. Rather, there exists a finite decision problem, with a unique optimal action at each relevant belief, in which the ranking is strictly reversed.

2.1 Relation to Blackwell's Order and to No Information

[Section 1.3](#) recorded two qualitatively relevant differences between canonical measures of information and measures of meaningful information. The former respect the Blackwell order and are always non-negative. The latter need not respect the Blackwell order and can be strictly negative. [Theorem 4](#) sheds further light on both observations through the following two corollaries.

Corollary 2. Consider two interpreted experiments $\hat{\mathcal{E}}_1$ and $\hat{\mathcal{E}}_2$ with a common objective prior p and a common subjective prior $\hat{p}$. Suppose that, in both experiments, the agent acts on the objective

posterior after every signal realization with positive objective probability $\hat{p}_i(\cdot | s) = p_i(\cdot | s)$, for $i \in \{1, 2\}$. If p has full support, then

$$\hat{\mathcal{E}}_1 \succeq_{\text{MI}} \hat{\mathcal{E}}_2 \iff \sigma_1 \succeq_B \sigma_2.$$

Thus, although individual measures of meaningful information need not be monotone in the Blackwell order away from correct interpretation, the meaningful-information order reduces exactly to Blackwell's order when the agent acts on the objective posteriors in both experiments.

Corollary 3. Let $\hat{\mathcal{E}}^0 = (p, \sigma^0, \hat{p}, \hat{\sigma}^0, \phi^{\hat{p}})$ denote a null interpreted experiment, where σ^0 has a single signal realization and $\phi^{\hat{p}}$ leaves the agent's belief equal to $\hat{p}$. For any interpreted experiment $\hat{\mathcal{E}}$ with the same objective and subjective priors, $\hat{\mathcal{E}} \succeq_{\text{MI}} \hat{\mathcal{E}}^0$ if and only if $M^{(G,g)}(\hat{\mathcal{E}}) \geq 0$ for every meaningful-information measure $M^{(G,g)}$ or, equivalently, if and only if Condition (F) is feasible for the ordered pair $(\hat{\mathcal{E}}, \hat{\mathcal{E}}^0)$. An analogous characterization of robust non-positivity follows by swapping the roles of $\hat{\mathcal{E}}$ and $\hat{\mathcal{E}}^0$.

Corollary 3 recasts the robust sign of meaningful information as a comparison with a null interpreted experiment. In the correct-prior, Bayesian-updating specialization, this recovers the comparison with no information studied by Morris and Shin (1997). More generally, the result applies to arbitrary subjective priors, perceived signal structures, and updating rules. Thus, although meaningful information may be positive under some behaviorally valid measures and negative under others, the meaningful-information order identifies exactly when its sign is invariant across all such measures.

2.2 Cardinal Measures and Robust Orders: Takeaways

The overall picture that emerges from the theoretical analysis in Section 1 and Section 2 is as follows. Cardinal measurement and robust comparison provide two complementary ways of evaluating interpreted experiments. Theorem 3 identifies, for each canonical cardinal measure, its unique jointly valid extension to subjective interpretations. Theorem 4 identifies the comparisons that survive the choice of cardinal measure of meaningful information and connects the resulting partial order to the convex geometry that underpins those measures. Figure 1 summarizes this architecture.

The top row of [Figure 1](#) is axiomatic, cardinal, and decision-problem dependent. It answers how much information an experiment transmits in the context of a particular decision problem. Each decision problem, through its convex geometry, generates a canonical ex-ante measure of information characterized in [Theorem 1](#) and, through [Theorem 3](#), its unique jointly valid measure of meaningful information characterized in [Theorem 2](#).

The bottom row is ordinal and decision-problem independent. It asks which comparisons between experiments survive across all decision problems. Unanimity across all canonical ex-ante measures gives rise to Blackwell’s order, while unanimity across all behaviorally valid measures gives rise to the meaningful-information order characterized in [Morris \(1991\)](#) and in [Theorem 4](#). When the agent acts on the objective posteriors, the latter order reduces to the former by [Corollary 2](#).

The two rows are complementary. The meaningful-information order answers whether one interpreted experiment is uniformly better than another independently of the decision problem. When such a robust comparison exists, [Theorem 4](#) provides a decision-problem-independent portfolio of exposures to pairwise mistakes, summarized by the weights λ , while a particular cardinal measure prices those mistakes through the Bregman divergences generated by its convex potential G and selector g . Thus, the robust order identifies the sign of the comparison across all decision problems, whereas a cardinal measure determines its magnitude in the welfare units relevant for a particular decision problem. When two interpreted experiments are incomparable under $\succeq_{\text{MI}}$, no such robust ranking is available and their ordering genuinely depends on the decision problem; the choice of G is then substantive rather than merely a choice of units.

The distinction between cardinal measurement and robust comparison also provides a natural hierarchy for empirical work. A researcher can first compare an interpreted experiment with a null experiment in which the agent does not update. If the interpreted experiment dominates the null experiment, meaningful information is non-negative under every behaviorally valid measure. If the null experiment dominates, meaningful information is non-positive under every such measure. If neither holds, the sign of meaningful information genuinely depends on the downstream decision problem.

More generally, the flow condition can be used to identify pairwise comparisons between interpreted experiments that are robust to the choice of G , while cardinal measures can be used to

quantify comparisons left incomplete by the partial order.

3. Empirical Application to Political Correctness

In this section, I bring the theoretical framework to data using the two experiments on political correctness in [Braghieri \(2024\)](#). The first experiment (the encoding experiment) identifies the objective signal structures that map college students' political types, Democrats versus Independents/Republicans, into private and public speech. The second experiment (the decoding experiment) elicits, on the same college campus, other students' perceptions of those mappings.

At the level of the encoding experiment, the communication regimes can be shown to be unambiguously Blackwell ranked. Private speech is Blackwell more informative about political type than public speech because social-image concerns induce students to shift their public responses toward the position they perceive to be more socially acceptable. This partial pooling reduces the diagnosticity of public statements. Public speech nevertheless remains informative about political type and therefore Blackwell-dominates silence. The encoding experiment thus yields a clear Blackwell chain: private speech is more informative than public speech, which in turn is more informative than silence.

This structure makes it possible to ask whether the objective ranking survives receivers' actual interpretations. The analysis follows the four quadrants of [Figure 1](#): I first evaluate the different communication regimes using the decision-problem-independent meaningful information order and then, where that order is incomplete, using cardinal measures.²⁰

3.1 Setting and Measurement

The experiments in [Braghieri \(2024\)](#) examine a central claim in the debate over political correctness on college campuses: students may feel pressure to publicly espouse views they do not privately hold, and those distortions may make public discourse less informative. In the encoding exper-

²⁰The same empirical strategy applies whenever the objective information environment is observable and receivers' beliefs can be elicited. Potential applications include studies of discrimination ([Arrow, 1974](#); [Bohren et al., 2025](#)), biased memory ([Zimmermann, 2020](#); [Bordalo et al., 2023](#); [Gödker, Jiao and Smeets, 2025](#)), cursed behavior ([Eyster and Rabin, 2005](#); [Deversi, Ispano and Schwardmann, 2021](#)), lying ([Gneezy, 2005](#)), information cascades ([Banerjee, 1992](#); [Bikhchandani, Hirshleifer and Welch, 1992](#); [Cipriani and Guarino, 2005](#)), cheap-talk communication ([Blume, Lai and Lim, 2020](#)), motivated reasoning ([Kunda, 1990](#)), and complexity ([Oprea, 2024](#)).

iment, students are randomly assigned to a private treatment, which guarantees anonymity, or to a public treatment, in which students are given hints suggesting that their answers will be shared with other students at their college together with their names. All participants then report their agreement with a series of politically sensitive statements. I focus on the statement about illegal immigration because it is the statement that is also used in the decoding experiment. It reads: “People who immigrated to the U.S. illegally, when caught, should be deported and sent back to their countries of origin.” On the college campus where the experiment was conducted, disagreeing with the statement was perceived to be more socially acceptable than agreeing with it.

The encoding experiment supplies the objective side of the framework. The state is a randomly selected participant’s self-reported political ideology, $\Omega = \{\text{Democrat, Independent/Republican}\}$, with the objective prior given by the sample composition (38.5% Democrats).²¹ Each treatment generates a binary signal, $s \in \{\text{Agreement, Disagreement}\}$, whose objective likelihoods are the empirical agreement rates. Among Democrats, agreement rates are similar across treatments (10.2% in private and 12.3% in public). Among Independents and Republicans, agreement falls from 40.0% in private to 30.1% in public.

Taken together, these rates compress the partisan gap that makes agreement diagnostic of political ideology. In fact, the public signal structure is a garbling of the private signal structure. Together with the null structure σ^0 (“silence”), the environment therefore provides three Blackwell-ranked structures: $\sigma_{\text{private}} \succeq_B \sigma_{\text{public}} \succeq_B \sigma^0$.

The decoding experiment supplies the subjective side of the framework. A separate sample of students was asked to forecast, for each treatment, the fraction of Democrats and the fraction of Independents/Republicans who agreed with the statement. The pair of forecasts identifies each receiver’s perceived signal structure for that treatment. The aim of the decoding experiment was to isolate meaningful information loss arising from misperceptions of those signal structures. Thus, receivers were given the objective prior and their subjective posteriors were obtained by applying Bayes’ rule to that prior and to their perceived likelihoods. Accordingly, in the exercise below, the only possible departure from correct interpretation is a misspecified perception of the signal structure.

²¹Self-reported political ideology was elicited privately in the demographic section of the survey, before participants were asked to report their agreement with any of the controversial statements.

The data therefore associate each receiver with three interpreted experiments sharing a common objective and subjective prior: interpreted private speech, interpreted public speech, and silence, represented by the null experiment in [Corollary 3](#).²² I first compare these interpreted experiments using the meaningful-information order. Where that order is incomplete, I then turn to cardinal measures of meaningful information. This hierarchy allows me to determine which parts of the objective Blackwell chain survive independently of the specific measure of meaningful information selected and which depend on the specific measure employed. Furthermore, the cardinal measures allow me to isolate the amount of meaningful information that is lost because the public signal structure is less informative than the private one according to canonical measures, and the amount of meaningful information that is lost due to misinterpretations of public speech.

3.2 Does the Blackwell Chain Survive Interpretation?

For every receiver and every ordered pair of her three interpreted experiments, I test the feasibility of the flow condition in (F) in both directions. By [Theorem 4](#), these tests classify each pair into one of three categories. If the condition is feasible in only one direction, one interpreted experiment robustly dominates the other: it transmits weakly more meaningful information under every behaviorally valid measure of meaningful information and strictly more for at least one measure. If the condition is feasible in both directions, the experiments are universally equivalent. If it is feasible in neither direction, the experiments are incomparable, and there exist meaningful information measures ranking them in opposite directions. For comparisons with silence, the same classification determines the robust sign of meaningful information by [Corollary 3](#).

[Table 1](#) presents the central results in a structure that mirrors [Figure 1](#). The robust, decision-problem-independent order appears first, followed by two cardinal, decision-problem-dependent measures. For the cardinal comparisons, I use two common decision geometries: one based on Shannon’s entropy function and one based on a quadratic potential. The left-hand block of the table reports the counterfactual benchmark of correctly specified beliefs and Bayesian updating; the right-hand block uses receivers’ actual interpretations. Within each block, the table reports the fraction of receivers for whom the full private–public–silence chain holds, the fraction for whom

²²The decoding experiment did not include a condition with a null signal structure. In the analysis below, I simply assume that participants would not update their beliefs in response to a null signal structure. This provides a benchmark for when no information is preferable to misinterpreted information.

at least one comparison implied by that chain is reversed, and the fraction for whom the ranking is incomplete.

Under correct interpretation, every criterion preserves the objective ranking. The meaningful-information order reduces to the Blackwell order and every valid canonical measure is monotone in that order. Consequently, the full private–public–silence chain holds for every idealized Bayesian receiver with correctly specified beliefs in the left-hand block of the table.

The right-hand block looks very different. The full chain survives in the meaningful-information order for only 5% of receivers. For 19%, at least one comparison is robustly reversed: an objectively less informative communication regime transmits weakly *more* meaningful information than an objectively more informative regime under every measure of meaningful information, and strictly more for at least one measure. For the remaining 76% of receivers, no comparison is robustly reversed, but at least one comparison is inconclusive according to the meaningful-information partial order. For these receivers, the ranking genuinely depends on the measure of meaningful information employed.

Because the meaningful-information order is incomplete for most receivers, any complete ranking must select a decision geometry. I report the meaningful-information extensions of the two benchmark decision geometries introduced in [Section 1.1](#). In the present application, receivers are given the objective prior, so $\hat{p} = p$ and the prior-misspecification term $d^{(G,g)}(p, \hat{p})$ vanishes.

Under the Shannon geometry, I^{Sh} is mutual information and $d^{(G,g)}$ is the Kullback–Leibler divergence, so

$$M^{Sh} = I^{Sh} - \mathbb{E}_{s \sim \text{Pr}} [\text{KL}(p(\cdot | s) \| \hat{p}(\cdot | s))],$$

measured in bits. Under the quadratic geometry, $G(p) = \|p\|^2$ and $d^{(G,g)}(x, y) = \|x - y\|^2$. Hence,

$$M^Q = \mathbb{E}_{s \sim \text{Pr}} [\|p(\cdot | s) - p\|^2] - \mathbb{E}_{s \sim \text{Pr}} [\|p(\cdot | s) - \hat{p}(\cdot | s)\|^2].$$

These cardinal measures produce complete rankings by construction. As shown in [Table 1](#), the full private–public–silence chain holds for 37% of receivers under M^{Sh} and for 36% under M^Q .

The cardinal measures also allow me to distinguish two sources of information loss that are particularly natural in this application. Because receivers are given the objective prior, [Theorem 3](#) implies that meaningful information in each communication regime equals its objective informa-

tion content minus the loss generated by the receiver’s misinterpretation of that regime. Thus, for any decision geometry G , the shortfall between correctly interpreted private speech and actually interpreted public speech can be written as

$$I_{\text{private}}^G - M_{i,\text{public}}^{(G,g)} = \underbrace{I_{\text{private}}^G - I_{\text{public}}^G}_{\text{encoding loss}} + \underbrace{I_{\text{public}}^G - M_{i,\text{public}}^{(G,g)}}_{\text{decoding loss}}.$$

The first term is common across receivers and captures the objective loss in informativeness caused by moving from private to public communication. The second varies across receivers and captures the additional loss arising from their misinterpretation of public speech.

Under the Shannon geometry, the encoding loss is 0.050 bits, while the average decoding loss is 0.078 bits. Thus, of the average shortfall between correctly interpreted private speech and actually interpreted public speech, 39% is attributable to the change in what speakers say and 61% to how receivers interpret what they hear. The quadratic geometry yields a remarkably similar decomposition: 39% of the average shortfall is attributable to encoding and 61% to decoding. Thus, under both the Shannon and quadratic geometries, the larger share of the average loss in meaningful information transmission in [Braghieri \(2024\)](#) arises from receivers’ misinterpretation of public speech rather than from the objective loss in informativeness caused by moving from private to public communication.

The cardinal measures also complete particular comparisons that the robust order leaves unresolved. Among receivers whose public-versus-silence comparison is incomparable, M^{Sh} ranks public speech above silence for only 47%. Thus, even within the set of receivers for whom the meaningful-information order does not determine whether public speech is beneficial, the Shannon benchmark ranks silence above public speech for slightly more than half. Unlike the robust-harm classification, this conclusion is specific to the chosen decision geometry. The quadratic measure, however, yields a similar pattern.

Taken together, the results show that interpretation can sharply alter a clear objective informativeness ranking. The full Blackwell chain survives receivers’ interpretations robustly for only 5% of the sample, while at least one comparison is robustly reversed for 19%. For most remaining receivers, the meaningful-information order establishes that no decision-problem-independent ranking exists. Under either of the two cardinal geometries considered here, the objective chain

holds for barely more than one-third of receivers. Furthermore, erroneous interpretations of public speech (decoding) are responsible for a larger share of meaningful information loss than the objective loss in informativeness caused by moving from private to public communication (encoding).

4. Conclusion

How should we measure and rank information transmission when agents have a misspecified prior, misinterpret signals, or update their beliefs in a non-Bayesian way? Canonical measures of information and orders over experiments offer insufficient guidance, because they capture the objective features of a signal structure but not agents' interpretations of it. To address this gap, I developed and axiomatized measures of meaningful information transmission that account for both the objective information environment and the way agents interpret it. These measures preserve the decision-theoretic foundations of canonical measures, coincide with them under correct Bayesian beliefs, and otherwise extend them in a disciplined way using the same underlying decision geometry.

Measures of meaningful information connect naturally to the robust order over misinterpreted experiments developed by [Morris \(1991\)](#). Specifically, the robust order in [Morris \(1991\)](#) coincides with unanimity across all behaviorally valid measures of meaningful information and admits an equivalent portfolio representation: differences between interpreted experiments can be decomposed into a common set of exposures to pairwise belief mistakes, with different meaningful-information measures simply pricing those mistakes differently. The cardinal and robust approaches are complementary. Cardinal measures quantify how much information is transmitted for a particular decision problem; the meaningful-information order identifies comparisons that hold regardless of the decision problem.

The empirical application to political correctness illustrates the importance of taking into account both the objective features of a signal structure and the agent's interpretations. In [Braghieri \(2024\)](#), social-image concerns make private speech objectively more informative about political type than public speech, while public speech remains more informative than silence. Under correct interpretation, this Blackwell ranking would hold for every receiver. Once receivers' actual interpretations are taken into account, however, the ranking no longer holds. Cardinal measures

of meaningful information, furthermore, allow for an important decomposition: they quantify how much information is lost because social-image concerns alter the mapping from political type into public speech and how much additional information is lost because receivers misinterpret that public speech.

The broader implication is that evaluating information transmission in empirical applications requires more than asking whether signals are objectively informative. It also requires asking how those signals are interpreted and, when no robust ranking is available, what the resulting beliefs will be used for. The framework developed here separates these three ingredients. It distinguishes losses in the objective quality of information from losses due to interpretation, identifies which comparisons survive across all downstream uses, and makes explicit when a conclusion necessarily depends on a particular decision problem. In doing so, it provides a unified language to study information transmission in the real world.

References

- Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz.** 2025. “Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology.” *Working Paper*.
- Arrow, Kenneth J.** 1974. “The Theory of Discrimination.” In *Discrimination in Labor Markets*, ed. Orley Ashenfelter and Albert Rees. Princeton University Press.
- Athey, Susan, and Jonathan Levin.** 2018. “The Value of Information in Monotone Decision Problems.” *Research in Economics*, 72(1).
- Balashov, Vadim S., and Andrei Nikiforov.** 2019. “How Much Do Investors Trade Because of Name/Ticker Confusion?” *Journal of Financial Markets*, 46.
- Banerjee, Abhijit V.** 1992. “A Simple Model of Herd Behavior.” *The Quarterly Journal of Economics*, 107(3).
- Barberis, Nicholas, Andrei Shleifer, and Robert Vishny.** 1998. “A Model of Investor Sentiment.” *Journal of Financial Economics*, 49(3).
- Battigalli, Pierpaolo, Simone Cerreia-Vioglio, Fabio Maccheroni, Massimo Marinacci, and Thomas J. Sargent.** 2022. “A framework for the analysis of self-confirming policies.” *Theory and Decision*, 92.
- Benjamin, Daniel J.** 2019. “Errors in Probabilistic Reasoning and Judgment Biases.” In *Handbook of Behavioral Economics: Applications and Foundations I*. Vol. 2, Chapter 2. Elsevier.

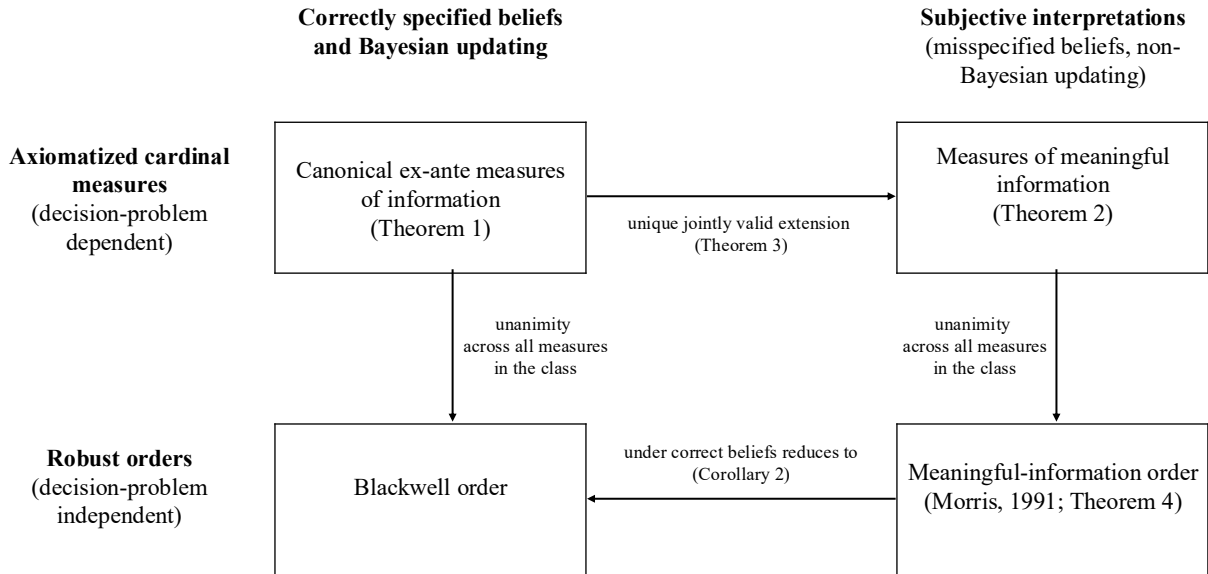
- Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch.** 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy*, 100(5).
- Blackwell, David.** 1951. "Comparison of Experiments." In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 2, , ed. Jerzy Neyman. Berkeley and Los Angeles:University of California Press.
- Blackwell, David.** 1953. "Equivalent Comparisons of Experiments." *The Annals of Mathematical Statistics*, 24(2).
- Blume, Andreas, Ernest K. Lai, and Wooyoung Lim.** 2020. "Strategic Information Transmission: A Survey of Experiments and Theoretical Foundations." In *Handbook of Experimental Game Theory*. , ed. C. Mónica Capra, Rachel T.A. Croson, Mary L. Rigdon and Tanya S. Rosenblat. Edward Elgar Publishing.
- Bohren, J. Aislinn, and Daniel N. Hauser.** 2021. "Learning with Heterogeneous Misspecified Models: Characterization and Robustness." *Econometrica*, 89(6).
- Bohren, J. Aislinn, and Daniel N. Hauser.** 2025. "Misspecified Models in Learning and Games." *Annual Review of Economics*, 17(1).
- Bohren, J. Aislinn, Kareem Haggag, Alex Imas, and Devin G. Pope.** 2025. "Inaccurate Statistical Discrimination: An Identification Problem." *The Review of Economics and Statistics*, 107(3).
- Bordalo, Pedro, John J. Conlon, Nicola Gennaioli, Spencer Y. Kwon, and Andrei Shleifer.** 2023. "Memory and Probability." *The Quarterly Journal of Economics*, 138(1).
- Braghieri, Luca.** 2024. "Political Correctness, Social Image, and Information Transmission." *American Economic Review*, 114(12).
- Brier, Glenn W.** 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review*, 78(1).
- Brunnermeier, Markus K., Alp Simsek, and Wei Xiong.** 2014. "A Welfare Criterion For Models With Distorted Beliefs." *The Quarterly Journal of Economics*, 129(4).
- Caplin, Andrew, and Mark Dean.** 2013. "Behavioral Implications of Rational Inattention with Shannon Entropy." National Bureau of Economic Research Working Paper 19318.
- Chan, David C., Matthew Gentzkow, and Chuan Yu.** 2022. "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." *The Quarterly Journal of Economics*, 137(2).
- Cipriani, Marco, and Antonio Guarino.** 2005. "Herd Behavior in a Laboratory Financial Market." *American Economic Review*, 95(5).
- Denti, Tommaso.** 2022. "Posterior Separable Cost of Information." *American Economic Review*, 112(10).
- Deversi, Marvin, Alessandro Ispano, and Peter Schwardmann.** 2021. "Spin Doctors: An Experiment on Vague Disclosure." *European Economic Review*, 139.

- Dreyfuss, Bnaya, and Ruru Hoong.** 2025. “Calibrated Coarsening: Designing Information for AI-Assisted Decisions.” *Working Paper*.
- Enke, Benjamin, and Florian Zimmermann.** 2019. “Correlation Neglect in Belief Formation.” *The Review of Economic Studies*, 86(1).
- Esponda, Ignacio, and Demian Pouzo.** 2016. “Berk–Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models.” *Econometrica*, 84(3).
- Frankel, Alex, and Emir Kamenica.** 2019. “Quantifying Information and Uncertainty.” *American Economic Review*, 109(10).
- Frick, Mira, Ryota Iijima, and Yuhta Ishii.** 2023. “Belief Convergence under Misspecified Learning: A Martingale Approach.” *The Review of Economic Studies*, 90(2).
- Frick, Mira, Ryota Iijima, and Yuhta Ishii.** 2024. “Welfare Comparisons for Biased Learning.” *American Economic Review*, 114(6).
- Fudenberg, Drew, Giacomo Lanzani, and Philipp Strack.** 2021. “Limit Points of Endogenous Misspecified Learning.” *Econometrica*, 89(3).
- Gneezy, Uri.** 2005. “Deception: The Role of Consequences.” *American Economic Review*, 95(1).
- Gödker, Katrin, Peiran Jiao, and Paul Smeets.** 2025. “Investor Memory.” *The Review of Financial Studies*, 38(6).
- Gossner, Olivier, and Jakub Steiner.** 2018. “On the cost of misperception: General results and behavioral applications.” *Journal of Economic Theory*, 177.
- Grether, David M.** 1980. “Bayes Rule as a Descriptive Model: The Representativeness Heuristic.” *The Quarterly Journal of Economics*, 95(3).
- Heidhues, Paul, Botond Kőszegi, and Philipp Strack.** 2021. “Convergence in Models of Misspecified Learning.” *Theoretical Economics*, 16(1).
- Jin, Ginger Zhe, Michael Luca, and Daniel Martin.** 2021. “Is No News (Perceived As) Bad News? An Experimental Investigation of Information Disclosure.” *American Economic Journal: Microeconomics*, 13(2).
- Kamenica, Emir, and Matthew Gentzkow.** 2011. “Bayesian Persuasion.” *American Economic Review*, 101(6).
- Kullback, Solomon, and Richard A. Leibler.** 1951. “On Information and Sufficiency.” *The Annals of Mathematical Statistics*, 22(1).
- Kunda, Ziva.** 1990. “The Case for Motivated Reasoning.” *Psychological Bulletin*, 108(3).
- Lehmann, E. L.** 1988. “Comparing Location Experiments.” *The Annals of Statistics*, 16(2).
- Lewis, David.** 1969. *Convention: A Philosophical Study*. Harvard University Press.

- Matějka, Filip, and Alisdair McKay.** 2015. “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model.” *American Economic Review*, 105(1).
- Morris, Stephen.** 1991. “The Role of Beliefs in Economic Theory.” PhD diss. Yale University.
- Morris, Stephen, and Hyun Song Shin.** 1997. “The Rationality and Efficacy of Decisions Under Uncertainty and the Value of an Experiment.” *Economic Theory*, 9(2): 309–324.
- Oprea, Ryan.** 2024. “Decisions under Risk Are Decisions under Complexity.” *American Economic Review*, 114(12).
- Ortoleva, Pietro.** 2024. “Alternatives to Bayesian Updating.” *Annual Review of Economics*, 16(1).
- Rabin, Matthew, and Joel L. Schrag.** 1999. “First Impressions Matter: A Model of Confirmatory Bias.” *The Quarterly Journal of Economics*, 114(1).
- Rashes, Michael S.** 2001. “Massively Confused Investors Making Conspicuously Ignorant Choices (MCI–MCIC).” *The Journal of Finance*, 56(5).
- Sims, Christopher A.** 2003. “Implications of Rational Inattention.” *Journal of Monetary Economics*, 50(3).
- Skyrms, Brian.** 2010. *Signals: Evolution, Learning, and Information*. Oxford University Press.
- Whitmeyer, Mark.** 2026. “Blackwell-Monotone Updating Rules.” *Journal of Political Economy*.
- Zhong, Weijie.** 2019. “Optimal Dynamic Information Acquisition.” *Working Paper*.
- Zimmermann, Florian.** 2020. “The Dynamics of Motivated Beliefs.” *American Economic Review*, 110(2).

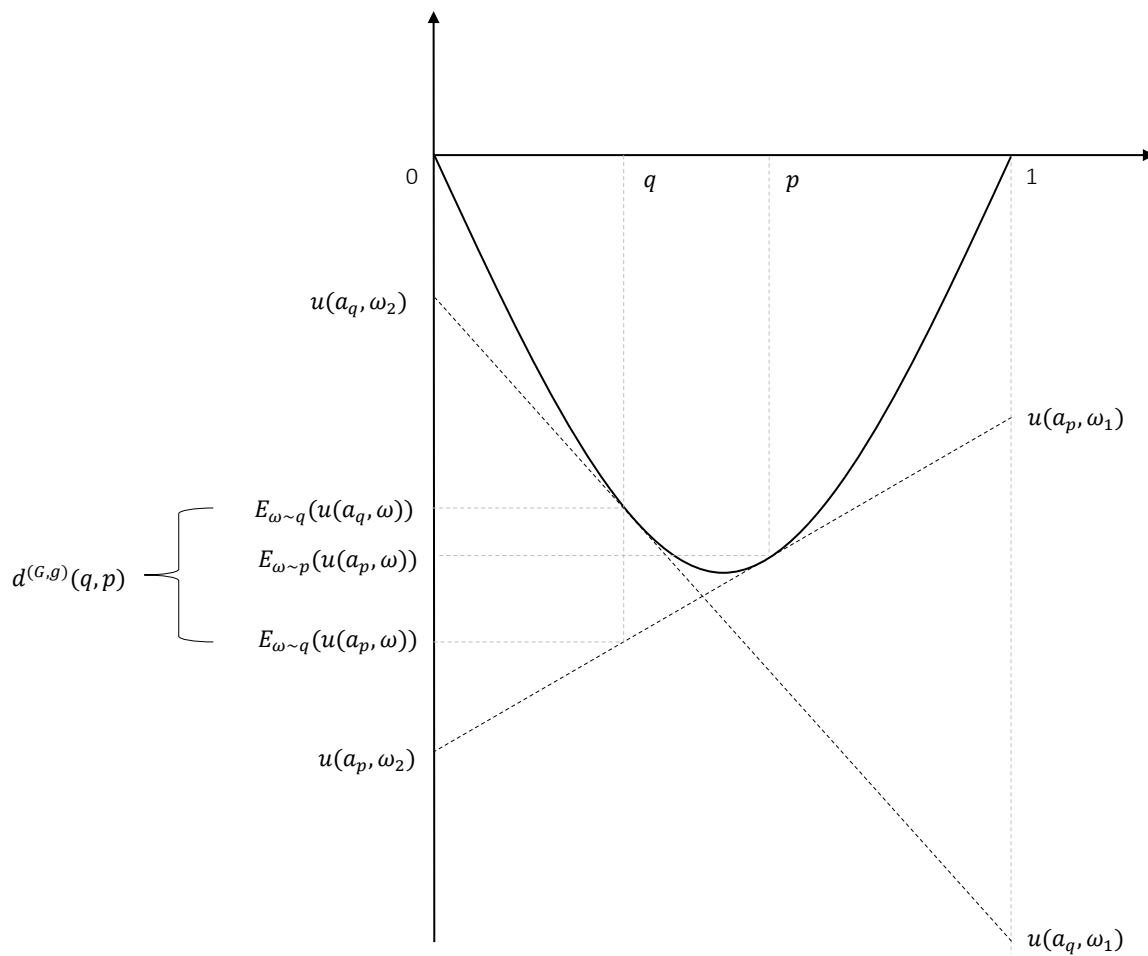
Tables and Figures

Figure 1: Cardinal Measures and Robust Orders



Notes: This figure summarizes the architecture of the theoretical results. The top row contains cardinal, decision-problem-dependent measures; the bottom row contains decision-problem-independent orders obtained through unanimity across the corresponding classes of measures. The first column relates to settings in which the agent has correctly specified beliefs and is Bayesian. The second column relates to settings in which the agent has potentially misspecified prior or posterior beliefs and is potentially non-Bayesian. [Morris \(1991\)](#) characterizes the robust comparison in the bottom-right cell through Condition (F) in [Section 2](#). [Theorem 4](#) connects that comparison to the class of behaviorally valid measures characterized in [Theorem 2](#) and provides its equivalent portfolio representation.

Figure 2: Bregman Divergence



Notes: This figure illustrates a Bregman divergence in the simple case of a binary state space. The solid convex function is the convex potential G . The dashed affine lines represent the expected utility from choosing actions a_p and a_q . The Bregman divergence corresponds to the vertical gap labeled $d^{(G,g)}(q,p)$.

Table 1: The Private–Public–Silence Ranking under Canonical and Meaningful Information

		Correctly specified beliefs and Bayesian updating			Subjective interpretations		
		Full chain	Reversal	Incomplete	Full chain	Reversal	Incomplete
Robust order (decision-problem independent)		100%	0%	0%	5%	19%	76%
Cardinal measures (decision-problem dependent)	Shannon decision geometry	100%	0%	0%	37%	63%	0%
	Quadratic decision geometry	100%	0%	0%	36%	64%	0%

Notes: This table reports the fraction of idealized and real participants in the decoding experiment in [Braghieri \(2024\)](#) for whom the benchmark ranking $\text{private} \succeq_B \text{public} \succeq_B \text{silence}$ is preserved (“Full chain”), the fraction for whom there is at least one reversal (“Reversal”), and the fraction for whom the order does not yield a conclusive comparison (“Incomplete”). Idealized participants, who are used as a benchmark, are hypothetical participants with correctly specified Bayesian beliefs. Real participants are actual students participating in the decoding experiment.

Appendix: For Online Publication Only

Table of Contents

A	Proofs	A.2
A.1	Proof of Theorem 1	A.2
A.2	Proof of Theorem 2	A.4
A.2.1	Representation $\Rightarrow$ Axioms	A.4
A.2.2	Axioms $\Rightarrow$ Representation	A.6
A.2.3	Representation $\Longleftrightarrow$ Valid	A.12
A.2.4	Invariance to affine reparametrization	A.14
A.3	Proof of Theorem 3	A.14
A.4	Proof of Theorem 4	A.17
B	Blackwell Reversal for Meaningful Information	A.20

A. Proofs

At various points throughout the proofs—and often without mentioning it—I will apply the standard result whereby any finite-support Bayes plausible distribution of posteriors can be generated by a finite signal structure.²³ Where I do mention the result, I refer to it as *implementability*.

A.1 Proof of Theorem 1

Proof. We prove (iii) $\Rightarrow$ (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) and the uniqueness claim.

Throughout the proof, I adopt a special boundary convention. The boundary convention is necessary because some convex potentials G do not have finite supporting hyperplanes at the boundary. When G has no finite supporting hyperplane at a boundary belief q , I let state-contingent payoffs take the value $-\infty$. Restricting G to the minimal face containing q , I choose a relative subgradient at q , and assign payoff $-\infty$ to states outside that face. This produces an extended-valued supporting action that is optimal at q . All arguments below therefore apply on the boundary with supporting hyperplanes understood in this extended sense.

(iii $\Rightarrow$ i). Assume there is a convex G with $I(p, \sigma) = \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p)$. Fix a selector $g(\cdot) \in \partial G(\cdot)$ and define the Bregman scoring decision problem with actions $A = \Delta(\Omega)$ and payoff

$$u(q, \omega) := G(q) + g(q) \cdot (\delta_\omega - q).$$

If the selector does not exist, adopt the special boundary condition described at the beginning of the proof.

For any belief r , $\mathbb{E}_{\omega \sim r}[u(q, \omega)] \leq G(r)$ by the subgradient inequality, with equality at $q = r$; hence $q = r$ is one of the subjectively optimal reports at belief r , and the ex-ante value function is $V(r) = G(r)$. Therefore,

$$\mathbb{E}_{s \sim \text{Pr}} \left[\max_{a \in A} \mathbb{E}_{\omega \sim p(\cdot | s)}[u(a, \omega)] \right] - \max_{a \in A} \mathbb{E}_{\omega \sim p}[u(a, \omega)] = \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p) = I(p, \sigma),$$

so I is valid.

(i $\Rightarrow$ ii). Assume validity: there exists a decision problem (A, u) such that

$$I(p, \sigma) = \mathbb{E}_{s \sim \text{Pr}} \left[\max_{a \in A} \mathbb{E}_{\omega \sim p(\cdot | s)}[u(a, \omega)] \right] - \max_{a \in A} \mathbb{E}_{\omega \sim p}[u(a, \omega)].$$

Let $V(r) := \max_{a \in A} \sum_{\omega} r(\omega) u(a, \omega)$.

Positivity (Axiom 1). Let $V(r) := \max_{a \in A} \sum_{\omega} r(\omega) u(a, \omega)$. Then V is convex (pointwise maximum of linear functions of r). By Bayes-plausibility $\mathbb{E}_{s \sim \text{Pr}}[p(\cdot | s)] = p$ and Jensen's inequality,

$$I(p, \sigma) = \mathbb{E}_{s \sim \text{Pr}}[V(p(\cdot | s))] - V(p) \geq V(\mathbb{E}_{s \sim \text{Pr}}[p(\cdot | s)]) - V(p) = V(p) - V(p) = 0.$$

Posterior dependence (Axiom 2). The integrand depends only on the posterior $p(\cdot | s)$; thus $I(p, \sigma) = \mathbb{E}_{s \sim \text{Pr}}[V(p(\cdot | s))] - V(p)$ depends on σ solely through the distribution of posteriors it induces.

²³See, for instance, [Kamenica and Gentzkow \(2011\)](#).

Path independence (Axiom 3). For a compound experiment that first draws s_0 via σ_0 and then s_1 via $\sigma_1^{s_0}$, we have

$$\begin{aligned} I(p, \sigma_1^{s_0} \circ \sigma_0) &= \mathbb{E}_{s_0 \sim \text{Pr}_0} [\mathbb{E}_{s_1 \sim \text{Pr}_1(\cdot | s_0)} [V(p(\cdot | s_0, s_1))] - V(p)] \\ &= (\mathbb{E}_{s_0 \sim \text{Pr}_0} [V(p(\cdot | s_0))] - V(p)) + \mathbb{E}_{s_0 \sim \text{Pr}_0} (\mathbb{E}_{s_1 \sim \text{Pr}_1(\cdot | s_0)} [V(p(\cdot | s_0, s_1))] - V(p(\cdot | s_0))) \\ &= I(p, \sigma_0) + \mathbb{E}_{s_0 \sim \text{Pr}_0} [I(p(\cdot | s_0), \sigma_1^{s_0})]. \end{aligned}$$

Thus all three axioms hold.

(ii $\Rightarrow$ iii). Assume I satisfies Axioms 1, 2, and 3. Let $\sigma^\dagger$ be a fully revealing experiment (signal alphabet $S = \Omega$, posteriors $\{\delta_\omega\}$ with weights $p(\omega)$). Define

$$G(p) := C - I(p, \sigma^\dagger),$$

for an arbitrary constant $C \in \mathbb{R}$ (the choice of C is immaterial).

Fix (p, σ) and write $p(\cdot | s)$ for the posteriors induced by σ . By posterior dependence, the final distribution of posteriors under the composite $\sigma^\dagger \circ \sigma$ coincides with that under $\sigma^\dagger$ (both yield $\{\delta_\omega\}$ with weights $p(\omega)$). Hence

$$I(p, \sigma^\dagger \circ \sigma) = I(p, \sigma^\dagger).$$

By path independence,

$$I(p, \sigma^\dagger \circ \sigma) = I(p, \sigma) + \mathbb{E}_{s \sim \text{Pr}} [I(p(\cdot | s), \sigma^\dagger)].$$

Combining the last two displays and rearranging,

$$I(p, \sigma) = I(p, \sigma^\dagger) - \mathbb{E}_{s \sim \text{Pr}} [I(p(\cdot | s), \sigma^\dagger)] = (C - G(p)) - \mathbb{E}_{s \sim \text{Pr}} [C - G(p(\cdot | s))] = \mathbb{E}_{s \sim \text{Pr}} [G(p(\cdot | s))] - G(p).$$

This proves the representation.

Finally, positivity implies that for every $p \in \Delta(\Omega)$ and every finite signal structure σ ,

$$0 \leq I(p, \sigma) = \mathbb{E}_{s \sim \text{Pr}} [G(p(\cdot | s))] - G(p),$$

hence

$$G(p) \leq \mathbb{E}_{s \sim \text{Pr}} [G(p(\cdot | s))]. \quad (5)$$

That is, G satisfies Jensen's inequality for every implementable finite Bayes-plausible distribution of posteriors.

By implementability, for any $q, q' \in \Delta(\Omega)$ and any $t \in (0, 1)$, letting $p := tq + (1-t)q'$, there exists a finite signal structure σ such that, under prior p , the induced posterior is q with probability t and q' with probability $1-t$. Applying (5) at prior p yields

$$G(p) \leq tG(q) + (1-t)G(q'),$$

i.e.,

$$G(tq + (1-t)q') \leq tG(q) + (1-t)G(q').$$

Therefore G is convex on $\Delta(\Omega)$.

Uniqueness of G up to affine transformations. Suppose both G_1 and G_2 represent I : $\mathbb{E}_{s \sim \text{Pr}}[G_1(p(\cdot | s))] - G_1(p) = \mathbb{E}_{s \sim \text{Pr}}[G_2(p(\cdot | s))] - G_2(p)$ for all (p, σ) . Let $H := G_1 - G_2$. Then for every Bayes-plausible posterior law,

$$\mathbb{E}_{s \sim \text{Pr}}[H(p(\cdot | s))] - H(p) = 0.$$

Hence H is affine on $\Delta(\Omega)$ (it preserves all finite martingale mixtures), i.e., $H(p) = \alpha + \beta \cdot p$. Thus $G_1 = G_2 + \alpha + \beta \cdot p$.

This completes the proof. $\square$

A.2 Proof of Theorem 2

We prove the theorem in four parts. First, we show that the representation implies the axioms. Second, we show that the axioms imply the representation. Third, we show that the representation is equivalent to behavioral validity. Fourth, we show that the representation is invariant to the choice of affine representative of G .

A.2.1 Representation $\Rightarrow$ Axioms

Proposition 1 (Representation $\Rightarrow$ Axioms). *Fix a convex $G : \Delta(\Omega) \rightarrow \mathbb{R}$ and a selector $g(\cdot) \in \partial G(\cdot)$. Define, for every $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$ with objective posteriors $\{p(\cdot | s)\}_{s \in \mathcal{S}}$, subjective posteriors $\{\hat{p}(\cdot | s)\}_{s \in \mathcal{S}}$, and objective signal law Pr ,*

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) := \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p) + d^{(G, g)}(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}}[d^{(G, g)}(p(\cdot | s), \hat{p}(\cdot | s))],$$

where $d^{(G, g)}(x, y) := G(x) - G(y) - g(y) \cdot (x - y)$. Then Axioms 4, 5, 6, 7, and 8 hold.

Proof. Fix a convex $G : \Delta(\Omega) \rightarrow \mathbb{R}$ and a selector $g(\cdot) \in \partial G(\cdot)$, and define

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) := \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p) + d^{(G, g)}(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}}[d^{(G, g)}(p(\cdot | s), \hat{p}(\cdot | s))],$$

where $d^{(G, g)}(x, y) := G(x) - G(y) - g(y) \cdot (x - y)$ and the expectation $\mathbb{E}_{s \sim \text{Pr}}$ is taken with respect to the objective signal law Pr . By Bayes plausibility, $\mathbb{E}_{s \sim \text{Pr}}[p(\cdot | s)] = p$.

Axiom 4 (Positivity at correct beliefs). At correct beliefs $(\hat{p}, \hat{\sigma}, \phi) = (p, \sigma, \phi^*)$ we have $d^{(G, g)}(p, p) = 0$ and $d^{(G, g)}(p(\cdot | s), p(\cdot | s)) = 0$ for all s , hence

$$\begin{aligned} M(p, \sigma, p, \sigma, \phi^*) &= \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p) \geq \\ &G(\mathbb{E}_{s \sim \text{Pr}}[p(\cdot | s)]) - G(p) = G(p) - G(p) = 0 \end{aligned}$$

where the second step follows by Jensen's inequality and the third by Bayes' plausibility.

Axiom 5 (Joint posterior dependence). In the displayed formula for M , the primitives $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$ enter only through the subjective prior $\hat{p}$ and the joint distribution of posterior pairs induced under the objective signal law. Indeed, the objective prior p is itself pinned down by that joint distribution,

since it equals the expectation of the first marginal:

$$p = \int_{\Delta(\Omega) \times \Delta(\Omega)} q d\tau_{p, \sigma, \hat{p}, \hat{\sigma}, \phi}(q, r).$$

Hence M depends on $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$ only through

$$(\hat{p}, \tau_{p, \sigma, \hat{p}, \hat{\sigma}, \phi}).$$

Relabeling signals preserves this pair, so the axiom holds.

Axiom 6 (Joint path independence). Consider a compound experiment that first draws s_0 from σ_0 and then, conditional on s_0 , draws s_1 from $\sigma_1^{s_0}$. Let $p(\cdot | s_0)$ and $p(\cdot | s_0, s_1)$ denote the corresponding objective posteriors, and define $\hat{p}(\cdot | s_0)$ and $\hat{p}(\cdot | s_0, s_1)$ analogously on the subjective side.

The meaningful information of the compound version of this procedure is:

$$\begin{aligned} M(p, \sigma_1^{s_0} \circ \sigma_0, \hat{p}, \hat{\sigma}_1^{s_0} \circ \hat{\sigma}_0, \phi_1^{s_0} \circ \phi_0) &= \\ &= \left(\mathbb{E}_{(s_0, s_1) \sim \text{Pr}_{0,1}} [G(p(\cdot | s_0, s_1))] - G(p) \right) + d^{(G,g)}(p, \hat{p}) - \mathbb{E}_{(s_0, s_1) \sim \text{Pr}_{0,1}} [d^{(G,g)}(p(\cdot | s_0, s_1), \hat{p}(\cdot | s_0, s_1))]. \end{aligned}$$

The meaningful information of stage 1 is:

$$\begin{aligned} M(p, \sigma_0, \hat{p}, \hat{\sigma}_0, \phi_0) &= \\ &= \mathbb{E}_{s_0 \sim \text{Pr}_0} [G(p(\cdot | s_0))] - G(p) + d^{(G,g)}(p, \hat{p}) - \mathbb{E}_{s_0 \sim \text{Pr}_0} [d^{(G,g)}(p(\cdot | s_0), \hat{p}(\cdot | s_0))]. \end{aligned}$$

The meaningful information of stage 2 at posterior $p(\cdot | s_0)$ is:

$$\begin{aligned} M(p(\cdot | s_0), \sigma_1^{s_0}, \hat{p}(\cdot | s_0), \hat{\sigma}_1^{s_0}, \phi_1^{s_0}) &= \\ \mathbb{E}_{s_1 \sim \text{Pr}_1(\cdot | s_0)} [G(p(\cdot | s_0, s_1))] - G(p(\cdot | s_0)) &+ d^{(G,g)}(p(\cdot | s_0), \hat{p}(\cdot | s_0)) - \mathbb{E}_{s_1 \sim \text{Pr}_1(\cdot | s_0)} [d^{(G,g)}(p(\cdot | s_0, s_1), \hat{p}(\cdot | s_0, s_1))] \end{aligned}$$

Adding $M(p, \sigma_0, \hat{p}, \hat{\sigma}_0, \phi_0)$ and taking the expectation of the previous expression with respect to s_0 , yields

$$\begin{aligned} &\mathbb{E}_{s_0 \sim \text{Pr}_0} [G(p(\cdot | s_0))] - G(p) + d^{(G,g)}(p, \hat{p}) - \mathbb{E}_{s_0 \sim \text{Pr}_0} [d^{(G,g)}(p(\cdot | s_0), \hat{p}(\cdot | s_0))] + \\ &\mathbb{E}_{s_0 \sim \text{Pr}_0} \left(\mathbb{E}_{s_1 \sim \text{Pr}_1(\cdot | s_0)} [G(p(\cdot | s_0, s_1))] - G(p(\cdot | s_0)) + d^{(G,g)}(p(\cdot | s_0), \hat{p}(\cdot | s_0)) - \right. \\ &\left. \mathbb{E}_{s_1 \sim \text{Pr}_1(\cdot | s_0)} [d^{(G,g)}(p(\cdot | s_0, s_1), \hat{p}(\cdot | s_0, s_1))] \right) \end{aligned}$$

which boils down to

$$\mathbb{E}_{s_0 \sim \text{Pr}_0} \left(\mathbb{E}_{s_1 \sim \text{Pr}_1(\cdot | s_0)} [G(p(\cdot | s_0, s_1))] \right) - G(p) + d^{(G,g)}(p, \hat{p}) - \mathbb{E}_{s_0 \sim \text{Pr}_0} \left(\mathbb{E}_{s_1 \sim \text{Pr}_1(\cdot | s_0)} [d^{(G,g)}(p(\cdot | s_0, s_1), \hat{p}(\cdot | s_0, s_1))] \right)$$

By the law of iterated expectation, the above equals $M(p, \sigma_1^{s_0} \circ \sigma_0, \hat{p}, \hat{\sigma}_1^{s_0} \circ \hat{\sigma}_0, \phi_1^{s_0} \circ \phi_0)$. Thus joint path independence holds.

Axiom 8 (Dominance at correct posteriors). Fix $p, \hat{p}, \sigma$ and consider arbitrary $\hat{\sigma} \in \Sigma$, and

$\phi \in \Phi$. Let ϕ^σ be the operator that sets $\hat{p}(\cdot | s) = p(\cdot | s)$ for all $s \in S$. I want to show

$$M(p, \sigma, \hat{p}, \sigma, \phi^\sigma) \geq M(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$$

By the tripartite representation, we can re-write the above inequality as

$$\begin{aligned} & \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p) + d^{(G,g)}(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}}[d^{(G,g)}(p(\cdot | s), p(\cdot | s))] \geq \\ & \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p) + d^{(G,g)}(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}}[d^{(G,g)}(p(\cdot | s), \hat{p}(\cdot | s))] \end{aligned}$$

which boils down to

$$\mathbb{E}_{s \sim \text{Pr}}[d^{(G,g)}(p(\cdot | s), \hat{p}(\cdot | s))] \geq 0$$

which is true because Bregman divergences are non-negative. Therefore, dominance at correct posteriors holds.

Axiom 7 (Invariance to null updating). Under $\phi^{\hat{p}}$ we have $\hat{p}(\cdot | s) \equiv \hat{p}$ for all signal realizations s . Hence

$$\mathbb{E}_{s \sim \text{Pr}}[d^{(G,g)}(p(\cdot | s), \hat{p}(\cdot | s))] = \mathbb{E}_{s \sim \text{Pr}}[d^{(G,g)}(p(\cdot | s), \hat{p})].$$

Expanding $d^{(G,g)}$ and using Bayes–plausibility $\mathbb{E}_{s \sim \text{Pr}}[p(\cdot | s)] = p$,

$$\begin{aligned} \mathbb{E}_{s \sim \text{Pr}}[d^{(G,g)}(p(\cdot | s), \hat{p})] &= \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(\hat{p}) - g(\hat{p}) \cdot (\mathbb{E}_{s \sim \text{Pr}}[p(\cdot | s)] - \hat{p}) \\ &= \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(\hat{p}) - g(\hat{p}) \cdot (p - \hat{p}). \end{aligned}$$

Similarly,

$$d^{(G,g)}(p, \hat{p}) = G(p) - G(\hat{p}) - g(\hat{p}) \cdot (p - \hat{p}).$$

Plugging these into the representation of M yields

$$\begin{aligned} M(p, \sigma, \hat{p}, \hat{\sigma}, \phi^{\hat{p}}) &= [\mathbb{E}_{s \sim \text{Pr}}G(p(\cdot | s)) - G(p)] + [G(p) - G(\hat{p}) - g(\hat{p}) \cdot (p - \hat{p})] \\ &\quad - [\mathbb{E}_{s \sim \text{Pr}}G(p(\cdot | s)) - G(\hat{p}) - g(\hat{p}) \cdot (p - \hat{p})] \\ &= 0. \end{aligned}$$

All five axioms therefore hold for M defined above. □

A.2.2 Axioms $\Rightarrow$ Representation

At a high level, the logical flow of the proof is as follows:

1. Identify the canonical potential G at correct beliefs;
2. Show any deviation from the canonical term must telescope via a two–argument “potential difference” Ψ ;
3. Show that, when the first component of Ψ is a vertex (atom), Ψ can be identified with the Bregman divergence of G ;
4. Show that Ψ can be identified with the Bregman divergence of G everywhere on the simplex.

5. Plug back into the residual to obtain the representation.

We proceed by proving a series of lemmas that help us articulate the proof along the steps described above. We assume Axioms 4, 5, 6, 7, and 8 throughout this section.

Lemma 1 (Canonical representation at correct beliefs). *Restrict M to correct beliefs and Bayesian updating. Then there exists a convex $G : \Delta(\Omega) \rightarrow \mathbb{R}$ such that, for all (p, σ) ,*

$$M(p, \sigma, p, \sigma, \phi^*) = \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p) =: I_G(p, \sigma).$$

Proof. At correct beliefs and Bayesian updating, Axioms 4, 5, and 6 reduce to the canonical axioms (Axioms 1, 2, and 3). By Theorem 1, there exists a convex potential G such that $M(p, \sigma, p, \sigma, \phi^*)$ coincides with I_G . $\square$

Lemma 2 (Residual telescopes). *Let*

$$R(p, \sigma, \hat{p}, \hat{\sigma}, \phi) := M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) - \left(\mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p) \right).$$

There exists a function $\Psi : \Delta(\Omega) \times \Delta(\Omega) \rightarrow \mathbb{R}$ with $\Psi(p, p) = 0$ such that, for every $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$ with induced posteriors $\{p(\cdot | s), \hat{p}(\cdot | s)\}_{s \in S}$ and objective weights $\text{Pr}(s)$,

$$R(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = \Psi(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}}[\Psi(p(\cdot | s), \hat{p}(\cdot | s))]. \quad (6)$$

Proof. Fix the potential G from Lemma 1 and define R as in the statement. We construct Ψ via a benchmark full-revelation signal structure.

Let $\sigma^\dagger$ be a fully revealing signal structure. From any objective prior $p \in \Delta(\Omega)$, $\sigma^\dagger$ induces the family of objective posteriors $\{\delta_\omega\}_{\omega \in \Omega}$ with $\text{Pr}(\delta_\omega) = p(\omega)$. Let $\phi^\dagger$ be a subjective-belief rule that, upon observing the signal in $\sigma^\dagger$, sets the final subjective posterior to the point mass on the realized state: $\hat{p}^\dagger(\cdot | s) = \delta_\omega$.

Define

$$\Psi(p, \hat{p}) := R(p, \sigma^\dagger, \hat{p}, \sigma^\dagger, \phi^\dagger).$$

By construction $\Psi(p, p) = 0$ for every p . Indeed,

$$\Psi(p, p) = R(p, \sigma^\dagger, p, \sigma^\dagger, \phi^\dagger),$$

and since $(p, \sigma^\dagger, p, \sigma^\dagger, \phi^\dagger)$ is a correct-beliefs, Bayesian-updating environment, Lemma 1 implies that this residual is zero.

Now fix any $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$. By joint path independence (Axiom 6) for M and the telescoping property of I_G , the residual $R = M - I_G$ is path independent as well. Hence, for the two-stage plan that runs $(\sigma, \hat{\sigma}, \phi)$ and then $(\sigma^\dagger, \sigma^\dagger, \phi^\dagger)$,

$$R(p, \sigma^\dagger \circ \sigma, \hat{p}, \sigma^\dagger \circ \hat{\sigma}, \phi^\dagger \circ \phi) = R(p, \sigma, \hat{p}, \hat{\sigma}, \phi) + \mathbb{E}_{s \sim \text{Pr}}[R(p(\cdot | s), \sigma^\dagger, \hat{p}(\cdot | s), \sigma^\dagger, \phi^\dagger)], \quad (7)$$

where the expectation is over first-stage signals s drawn according to Pr .

If we can show that

$$R(p, \sigma^\dagger \circ \sigma, \hat{p}, \sigma^\dagger \circ \hat{\sigma}, \phi^\dagger \circ \phi) = R(p, \sigma^\dagger, \hat{p}, \sigma^\dagger, \phi^\dagger).$$

then, we can rewrite (7) as

$$\Psi(p, \hat{p}) = R(p, \sigma, \hat{p}, \hat{\sigma}, \phi) + \mathbb{E}_{s \sim \text{Pr}}[\Psi(p(\cdot | s), \hat{p}(\cdot | s))].$$

which proves the lemma.

But

$$R(p, \sigma^\dagger \circ \sigma, \hat{p}, \sigma^\dagger \circ \hat{\sigma}, \phi^\dagger \circ \phi) = R(p, \sigma^\dagger, \hat{p}, \sigma^\dagger, \phi^\dagger).$$

follows directly from Axiom 5. Indeed, in both experiments the final objective posterior is δ_ω and the final subjective posterior is also δ_ω , with unconditional probability $p(\omega)$ for each $\omega \in \Omega$. Hence both induce the same joint distribution of posterior pairs, so

$$M(p, \sigma^\dagger \circ \sigma, \hat{p}, \sigma^\dagger \circ \hat{\sigma}, \phi^\dagger \circ \phi) = M(p, \sigma^\dagger, \hat{p}, \sigma^\dagger, \phi^\dagger).$$

Moreover, the canonical term

$$\mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p)$$

is the same in both cases, since the objective posterior is δ_ω with probability $p(\omega)$ under both experiments. Therefore the corresponding residuals are equal. $\square$

Lemma 3 (Constant–sum identity). *For arbitrary $p \in \Delta(\Omega)$, we have*

$$G(p) + \sum_{\omega} p(\omega) \Psi(\delta_\omega, p) = \sum_{\omega} p(\omega) G(\delta_\omega) \quad (8)$$

Proof. By Axiom 7, $M(p, \sigma^\dagger, p, \sigma^\dagger, \phi^p) = 0$. Using Lemma 1, the canonical term for $\sigma^\dagger$ is $\mathbb{E}_{\omega \sim p}[G(\delta_\omega)] - G(p)$. By Lemma 2, the residual is $-\mathbb{E}_{\omega \sim p}[\Psi(\delta_\omega, p)]$. Thus

$$0 = \sum_{\omega} p(\omega) G(\delta_\omega) - G(p) - \sum_{\omega} p(\omega) \Psi(\delta_\omega, p)$$

which rearranges to the desired identity. $\square$

Lemma 4 (Additive form on atoms). *For each $q \in \Delta(\Omega)$, we have*

$$\Psi(\delta_\omega, q) = G(\delta_\omega) - G(q) - g(q) \cdot (\delta_\omega - q) \quad (9)$$

where $g(q) \in \mathbb{R}^{|\Omega|}$ is defined componentwise by $g_\omega(q) := G(\delta_\omega) - G(q) - \Psi(\delta_\omega, q)$

Proof. Consider $g(q) \cdot q$:

$$\begin{aligned} g(q) \cdot q &= \sum_{\omega} q(\omega) [G(\delta_\omega) - G(q) - \Psi(\delta_\omega, q)] = \\ &= \sum_{\omega} q(\omega) G(\delta_\omega) - G(q) - \sum_{\omega} q(\omega) \Psi(\delta_\omega, q) \end{aligned}$$

By Lemma 3, the expression above equals 0.

But then, consider $\Psi(\delta_\omega, q)$

$$\begin{aligned}\Psi(\delta_\omega, q) &= G(\delta_\omega) - G(q) - g_\omega(q) = \\ &= G(\delta_\omega) - G(q) - (g_\omega(q) - g(q) \cdot q) = \\ &= G(\delta_\omega) - G(q) - g(q) \cdot (\delta_\omega - q)\end{aligned}$$

where the first equality follows from the definition of $g_\omega(q)$ and the second from the observation above that $g(q) \cdot q = 0$. $\square$

Lemma 5 (Calibration to the simplex using atoms). *For all $p, q \in \Delta(\Omega)$,*

$$\Psi(p, q) = \sum_{\omega} p(\omega) [\Psi(\delta_\omega, q) - \Psi(\delta_\omega, p)]. \quad (10)$$

Proof. Consider arbitrary $p, q \in \Delta(\Omega)$, the fully revealing experiment $\sigma^\dagger$, and interpretation ϕ^q that fixes posteriors to q independently of the signal realized. By Axiom 7, we have $M(p, \sigma^\dagger, q, \sigma^\dagger, \phi^q) = 0$. Using Lemmas 1 and 2, we can re-write the condition above as:

$$\sum_{\omega} p(\omega) G(\delta_\omega) - G(p) + \Psi(p, q) - \sum_{\omega} p(\omega) \Psi(\delta_\omega, q) = 0$$

By Equation (8), we can re-write the above as

$$G(p) + \sum_{\omega} p(\omega) \Psi(\delta_\omega, p) - G(p) + \Psi(p, q) - \sum_{\omega} p(\omega) \Psi(\delta_\omega, q) = 0$$

which boils down to

$$\Psi(p, q) = \sum_{\omega} p(\omega) \Psi(\delta_\omega, q) - \sum_{\omega} p(\omega) \Psi(\delta_\omega, p)$$

$\square$

Lemma 6 (Subgradient property). *For every $q \in \Delta(\Omega)$, $g(q) \in \partial G(q)$; i.e.*

$$G(p) \geq G(q) + g(q) \cdot (p - q) \quad \text{for all } p \in \Delta(\Omega).$$

Proof. Fix $q \in \Delta(\Omega)$ and an arbitrary $p \in \Delta(\Omega)$. Consider the null experiment σ_{null} (an experiment with a single signal realization works) and the interpretation ϕ^p that sets the subjective posterior to p (independently of the realized signal).

By Axiom 8, we know that for all other possible interpretations $(\hat{\sigma}, \phi) \in \Sigma \times \Phi$, we have

$$M(p, \sigma_{\text{null}}, q, \sigma_{\text{null}}, \phi^p) \geq M(p, \sigma_{\text{null}}, q, \hat{\sigma}, \phi)$$

In particular, the inequality above holds for interpretation $\sigma_{\text{null}}, \phi^q$ that sets the subjective posterior to q . I.e.,

$$M(p, \sigma_{\text{null}}, q, \sigma_{\text{null}}, \phi^p) \geq M(p, \sigma_{\text{null}}, q, \sigma_{\text{null}}, \phi^q)$$

By Axiom 7, the right-hand side equals zero. Furthermore, by Lemmas 1 and 2, the LHS equals $\Psi(p, q)$. Thus, we have

$$\Psi(p, q) \geq 0$$

By Lemma 5, we know that $\Psi(p, q) = \sum_{\omega} p(\omega) [\Psi(\delta_{\omega}, q) - \Psi(\delta_{\omega}, p)]$.

Thus, we have

$$\sum_{\omega} p(\omega) \Psi(\delta_{\omega}, q) - \sum_{\omega} p(\omega) \Psi(\delta_{\omega}, p) \geq 0$$

From Lemma 3, we know that $\sum_{\omega} p(\omega) \Psi(\delta_{\omega}, p) = \sum_{\omega} p(\omega) G(\delta_{\omega}) - G(p)$. From Lemma 4, we know $\Psi(\delta_{\omega}, q) = G(\delta_{\omega}) - G(q) - g(q) \cdot (\delta_{\omega} - q)$. Substituting these expressions in the inequality above, we have

$$\sum_{\omega} p(\omega) [G(\delta_{\omega}) - G(q) - g(q) \cdot (\delta_{\omega} - q)] - \sum_{\omega} p(\omega) G(\delta_{\omega}) + G(p) \geq 0$$

Simplifying, we obtain

$$\sum_{\omega} p(\omega) G(\delta_{\omega}) - G(q) - g(q) \cdot (p - q) - \sum_{\omega} p(\omega) G(\delta_{\omega}) + G(p) \geq 0$$

which boils down to

$$G(p) \geq G(q) + g(q) \cdot (p - q)$$

Thus, $g(q)$ is a subgradient of G at q . □

Lemma 7 (Bregman form on atoms). *For every $q \in \Delta(\Omega)$ and every state $\omega \in \Omega$,*

$$\Psi(\delta_{\omega}, q) = d^{(G, g)}(\delta_{\omega}, q) \quad \text{where} \quad d^{(G, g)}(x, y) := G(x) - G(y) - g(y) \cdot (x - y)$$

and $g(\cdot) \in \partial G(\cdot)$ is the selector from Lemma 6.

Proof. By Lemma 4, for each $q \in \Delta(\Omega)$ and $\omega \in \Omega$ we have

$$\Psi(\delta_{\omega}, q) = G(\delta_{\omega}) - G(q) - g(q) \cdot (\delta_{\omega} - q).$$

By Lemma 6, $g(q) \in \partial G(q)$, hence the right-hand side is exactly $d^{(G, g)}(\delta_{\omega}, q)$ by definition of the Bregman divergence generated by G with selector $g(\cdot)$. □

Lemma 8 (General property of Bregman divergences). *Let $G : \Delta(\Omega) \rightarrow \mathbb{R}$ be convex and fix a selector $g(\cdot) \in \partial G(\cdot)$. Define the Bregman divergence*

$$d^{(G, g)}(x, y) := G(x) - G(y) - g(y) \cdot (x - y) \quad (x, y \in \Delta(\Omega)).$$

Then for all $p, q \in \Delta(\Omega)$,

$$d^{(G, g)}(p, q) = \sum_{\omega \in \Omega} p(\omega) d^{(G, g)}(\delta_{\omega}, q) - \sum_{\omega \in \Omega} p(\omega) d^{(G, g)}(\delta_{\omega}, p). \quad (11)$$

Proof. Compute

$$\begin{aligned} \sum_{\omega} p(\omega) d^{(G, g)}(\delta_{\omega}, q) &= \sum_{\omega} p(\omega) \left(G(\delta_{\omega}) - G(q) - g(q) \cdot (\delta_{\omega} - q) \right) = \\ &= \sum_{\omega} p(\omega) G(\delta_{\omega}) - G(q) - g(q) \cdot (p - q) \end{aligned}$$

and

$$\begin{aligned}\sum_{\omega} p(\omega) d^{(G,g)}(\delta_{\omega}, p) &= \sum_{\omega} p(\omega) \left(G(\delta_{\omega}) - G(p) - g(p) \cdot (\delta_{\omega} - p) \right) = \\ &= \sum_{\omega} p(\omega) G(\delta_{\omega}) - G(p)\end{aligned}$$

since $\sum_{\omega} p(\omega) (\delta_{\omega} - p) = 0$. Subtracting the second expression from the first yields

$$\begin{aligned}\sum_{\omega} p(\omega) d^{(G,g)}(\delta_{\omega}, q) - \sum_{\omega} p(\omega) d^{(G,g)}(\delta_{\omega}, p) &= G(p) - G(q) - g(q) \cdot (p - q) = \\ &= d^{(G,g)}(p, q)\end{aligned}$$

which is (11). □

Lemma 9 (Bregman identification on the entire simplex). *For all $p, q \in \Delta(\Omega)$,*

$$\Psi(p, q) = d^{(G,g)}(p, q) \quad \text{where} \quad d^{(G,g)}(p, q) := G(p) - G(q) - g(q) \cdot (p - q)$$

and $g(\cdot) \in \partial G(\cdot)$ is the selector from Lemma 6.

Proof. From Lemma 5, we have

$$\Psi(p, q) = \sum_{\omega} p(\omega) \Psi(\delta_{\omega}, q) - \sum_{\omega} p(\omega) \Psi(\delta_{\omega}, p)$$

From Lemma 7, we have $\Psi(\delta_{\omega}, q) = d^{(G,g)}(\delta_{\omega}, q)$ using the selector $g(\cdot) \in \partial G(\cdot)$ from Lemma 6.

Thus,

$$\Psi(p, q) = \sum_{\omega} p(\omega) d^{(G,g)}(\delta_{\omega}, q) - \sum_{\omega} p(\omega) d^{(G,g)}(\delta_{\omega}, p)$$

By Lemma 8, we know

$$d^{(G,g)}(p, q) = \sum_{\omega} p(\omega) d^{(G,g)}(\delta_{\omega}, q) - \sum_{\omega} p(\omega) d^{(G,g)}(\delta_{\omega}, p)$$

Thus, $\Psi(p, q) = d^{(G,g)}(p, q)$ □

Proposition 2 (Axioms $\Rightarrow$ Representation). *A measure of meaningful information M that satisfies Axioms 4, 5, 6, 7, and 8 admits the representation*

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = \underbrace{\mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))]}_{\text{canonical ex-ante information}} - G(p) + \underbrace{d^{(G,g)}(p, \hat{p})}_{\text{prior misspecification}} - \underbrace{\mathbb{E}_{s \sim \text{Pr}}[d^{(G,g)}(p(\cdot | s), \hat{p}(\cdot | s))]}_{\text{posterior misspecification}}, \quad (12)$$

where $G : \Delta(\Omega) \rightarrow \mathbb{R}$ is a convex potential, and, for a fixed selector $g(\cdot) \in \partial G(\cdot)$, $d^{(G,g)}$ is the Bregman divergence generated by G ,

$$d^{(G,g)}(x, y) \equiv G(x) - G(y) - g(y) \cdot (x - y)$$

Proof. The proposition is proved in five steps.

Step 1: Canonical representation at correct beliefs.

By Lemma 1, there exists a convex $G : \Delta(\Omega) \rightarrow \mathbb{R}$ (unique up to an affine transformation) such that for correct beliefs and Bayesian updating,

$$M(p, \sigma, p, \sigma, \phi^*) = \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p) =: I_G(p, \sigma).$$

Step 2: The residual telescopes.

Define the residual

$$R(p, \sigma, \hat{p}, \hat{\sigma}, \phi) := M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) - \left(\mathbb{E}_{s \sim \text{Pr}} G(p(\cdot | s)) - G(p) \right).$$

By Lemma 2, there exists a function $\Psi : \Delta(\Omega) \times \Delta(\Omega) \rightarrow \mathbb{R}$ with $\Psi(p, p) = 0$ such that, for every $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$ with induced posteriors $\{p(\cdot | s), \hat{p}(\cdot | s)\}_{s \in \mathcal{S}}$ and objective weights $\text{Pr}(s)$,

$$R(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = \Psi(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}} [\Psi(p(\cdot | s), \hat{p}(\cdot | s))]. \quad (13)$$

Step 3: Bregman on Atoms

Lemma 7 shows that when the first component of Ψ is a vertex (atom), Ψ takes the form of the Bregman divergence of G identified in Step 1.

Step 4: Bregman on the Simplex.

By Lemma 9, we have $\Psi(p, q) = d^{(G, g)}(p, q)$. Ψ takes the form of the Bregman divergence of G identified in Step 1 everywhere on the simplex

Step 5: Conclude the representation.

Substitute $\Psi(p, q) = d^{(G, g)}(p, q)$ into (13):

$$R(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = d^{(G, g)}(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}} [d^{(G, g)}(p(\cdot | s), \hat{p}(\cdot | s))].$$

Adding back the canonical part $\mathbb{E}_{s \sim \text{Pr}} G(p(\cdot | s)) - G(p)$ from Lemma 1 gives

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = \underbrace{\mathbb{E}_{s \sim \text{Pr}} G(p(\cdot | s)) - G(p)}_{\text{canonical ex-ante information}} + \underbrace{d^{(G, g)}(p, \hat{p})}_{\text{prior misspecification}} - \underbrace{\mathbb{E}_{s \sim \text{Pr}} d^{(G, g)}(p(\cdot | s), \hat{p}(\cdot | s))}_{\text{posterior misspecification}}.$$

This is exactly (12). The selector $g(\cdot) \in \partial G(\cdot)$ is fixed by the construction above, completing the proof. $\square$

A.2.3 Representation $\iff$ Valid

Proposition 3 (Representation $\iff$ behavioral validity). *The tripartite representation in Theorem 2 is equivalent to behavioral validity.*

Proof. (Representation $\Rightarrow$ behavioral validity). Assume there exist a convex $G : \Delta(\Omega) \rightarrow \mathbb{R}$ and a fixed selector $g(\cdot) \in \partial G(\cdot)$ such that

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = \mathbb{E}_{s \sim \text{Pr}} [G(p(\cdot | s))] - G(p) + d^{(G, g)}(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}} [d^{(G, g)}(p(\cdot | s), \hat{p}(\cdot | s))],$$

with $d^{(G,g)}(x,y) := G(x) - G(y) - g(y) \cdot (x - y)$. Define a *Bregman scoring* decision problem with action set $A := \Delta(\Omega)$ and payoff

$$u(q, \omega) := G(q) + g(q) \cdot (\delta_\omega - q).$$

For any belief r , a (subjective) optimal report is $q = r$ and the (objective) welfare is

$$W(p, r) = \mathbb{E}_{\omega \sim p}[u(r, \omega)] = G(p) - d^{(G,g)}(p, r).$$

Hence the behavioral value of information equals

$$\mathbb{E}_{s \sim \text{Pr}} \left[G(p(\cdot | s)) - d^{(G,g)}(p(\cdot | s), \hat{p}(\cdot | s)) \right] - \left(G(p) - d^{(G,g)}(p, \hat{p}) \right),$$

which is exactly M . Thus M is behaviorally valid.

(*Behavioral validity $\Rightarrow$ representation*). Assume M is behaviorally valid: there exists a decision problem (A, u) and a fixed tie-breaking rule such that for any belief $r \in \Delta(\Omega)$ an action $a_r \in \arg \max_{a \in A} \sum_{\omega} r(\omega) u(a, \omega)$ is chosen, and

$$W(p, r) := \sum_{\omega} p(\omega) u(a_r, \omega).$$

Hence

$$M = \mathbb{E}_{s \sim \text{Pr}} [W(p(\cdot | s), \hat{p}(\cdot | s))] - W(p, \hat{p}).$$

Define the value function

$$G(q) := \max_{a \in A} \sum_{\omega} q(\omega) u(a, \omega),$$

which is convex (supremum of linear maps). Fix the selector $g(\cdot)$ by

$$g(r) := (u(a_r, \omega))_{\omega} \in \partial G(r) \quad \text{for each } r \in \Delta(\Omega),$$

(the subgradient characterization of a pointwise supremum). Then, for any p ,

$$\begin{aligned} W(p, r) &= \sum_{\omega} p(\omega) u(a_r, \omega) = g(r) \cdot p = \\ &= G(p) - (G(p) - G(r) - g(r) \cdot (p - r)) = G(p) - d^{(G,g)}(p, r) \end{aligned}$$

where $d^{(G,g)}(x,y) := G(x) - G(y) - g(y) \cdot (x - y)$.

Therefore,

$$M = \mathbb{E}_{s \sim \text{Pr}} \left[G(p(\cdot | s)) - d^{(G,g)}(p(\cdot | s), \hat{p}(\cdot | s)) \right] - \left(G(p) - d^{(G,g)}(p, \hat{p}) \right),$$

which is exactly the stated tripartite representation with potential G and selector $g(\cdot)$. □

A.2.4 Invariance to affine reparametrization

Proposition 4 (Affine invariance of the representation). *Suppose M admits the representation (12) with a convex potential $G : \Delta(\Omega) \rightarrow \mathbb{R}$ and a selector $g(\cdot) \in \partial G(\cdot)$. Then, for any affine reparametrization $\tilde{G}(p) = G(p) + \alpha + \beta \cdot p$ with $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^{|\Omega|}$, choosing $\tilde{g}(y) := g(y) + \beta \in \partial \tilde{G}(y)$ leaves the right-hand side of (12) unchanged. Consequently, the value of M in (12) is invariant to the choice of affine representative of G .*

Proof. By Bayes-plausibility, $\mathbb{E}_{s \sim \text{Pr}}[p(\cdot | s)] = p$, so

$$\mathbb{E}_{s \sim \text{Pr}}[\tilde{G}(p(\cdot | s))] - \tilde{G}(p) = \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s)) + \alpha + \beta \cdot p(\cdot | s)] - (G(p) + \alpha + \beta \cdot p) = \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p).$$

For the Bregman terms, using $\tilde{g}(y) = g(y) + \beta$,

$$d^{(\tilde{G}, \tilde{g})}(x, y) = \tilde{G}(x) - \tilde{G}(y) - \tilde{g}(y) \cdot (x - y) = G(x) - G(y) - g(y) \cdot (x - y) = d^{(G, g)}(x, y).$$

Hence each component of (12) is preserved under $(G, g) \mapsto (\tilde{G}, \tilde{g})$, and the value of M is unchanged. $\square$

A.3 Proof of Theorem 3

Proof. Throughout, fix a finite state space Ω . For any decision problem (A, u) with a well-defined maximum define the (ex-ante) indirect utility

$$G(q) := \max_{a \in A} \sum_{\omega \in \Omega} q(\omega) u(a, \omega), \quad q \in \Delta(\Omega).$$

Fix a selector $a(\cdot)$ with $a(q) \in \arg \max_{a \in A} \sum_{\omega} q(\omega) u(a, \omega)$ for each q . For each q , define a subgradient selector $g(q) \in \partial G(q)$ by

$$g(q)(\omega) := u(a(q), \omega), \quad \omega \in \Omega.$$

Define the associated Bregman divergence

$$d^{(G, g)}(x, y) := G(x) - G(y) - g(y) \cdot (x - y).$$

Lemma 10. *For every $x, y \in \Delta(\Omega)$,*

$$d^{(G, g)}(x, y) = G(x) - \sum_{\omega \in \Omega} x(\omega) u(a(y), \omega). \quad (14)$$

Proof. Since $g(y)(\omega) = u(a(y), \omega)$ and $G(y) = \sum_{\omega} y(\omega) u(a(y), \omega) = y \cdot g(y)$, we have

$$\begin{aligned} d^{(G, g)}(x, y) &= G(x) - G(y) - g(y) \cdot (x - y) \\ &= G(x) - y \cdot g(y) - g(y) \cdot x + g(y) \cdot y \\ &= G(x) - x \cdot g(y) = G(x) - \sum_{\omega} x(\omega) u(a(y), \omega), \end{aligned}$$

which is (14).

We now prove $(i) \Leftrightarrow (ii) \Leftrightarrow (iii)$.

(i) $\Rightarrow$ (ii). Assume I and M are jointly valid, i.e. they arise from the same decision problem (A, u) . By Theorem 1, I admits a representation via the same indirect utility G :

$$I(p, \sigma) = \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p), \quad (15)$$

where $p(\cdot | s)$ denotes the Bayesian posterior induced by (p, σ) .

Let $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$ be arbitrary and let $\hat{p}(\cdot | s)$ denote the (subjective) posterior induced by $(\hat{p}, \hat{\sigma}, \phi)$ upon observing s . Behavioral validity of M for the same decision problem (A, u) means that, ex-ante under the true signal law, M equals the gain in true expected utility from choosing the signal-contingent action that is optimal for the subjective posterior:

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = \mathbb{E}_{s \sim \text{Pr}} \left[\sum_{\omega} p(\omega | s) u(a(\hat{p}(\cdot | s)), \omega) \right] - \sum_{\omega} p(\omega) u(a(\hat{p}), \omega). \quad (16)$$

Apply Lemma 10 with $x = p(\cdot | s)$ and $y = \hat{p}(\cdot | s)$ to obtain, for each s ,

$$\sum_{\omega} p(\omega | s) u(a(\hat{p}(\cdot | s)), \omega) = G(p(\cdot | s)) - d^{(G, g)}(p(\cdot | s), \hat{p}(\cdot | s)).$$

Apply Lemma 10 again with $x = p$ and $y = \hat{p}$ to obtain

$$\sum_{\omega} p(\omega) u(a(\hat{p}), \omega) = G(p) - d^{(G, g)}(p, \hat{p}).$$

Substituting both equalities into (16) yields

$$\begin{aligned} M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) &= \mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - \mathbb{E}_{s \sim \text{Pr}}[d^{(G, g)}(p(\cdot | s), \hat{p}(\cdot | s))] - \left(G(p) - d^{(G, g)}(p, \hat{p}) \right) \\ &= \left(\mathbb{E}_{s \sim \text{Pr}}[G(p(\cdot | s))] - G(p) \right) + d^{(G, g)}(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}}[d^{(G, g)}(p(\cdot | s), \hat{p}(\cdot | s))]. \end{aligned}$$

Using (15) gives

$$M(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = I(p, \sigma) + d^{(G, g)}(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}}[d^{(G, g)}(p(\cdot | s), \hat{p}(\cdot | s))],$$

which is (3). □

(ii) $\Rightarrow$ (i). Assume (ii). Let G be a convex potential representing I as in Theorem 1 and fix the selector $g(\cdot) \in \partial G(\cdot)$ appearing in (3). We construct a decision problem (A, u) whose indirect utility is G (up to affine normalization).

Let $A := \Delta(\Omega)$ and define, for $q \in A$ and $\omega \in \Omega$,

$$u(q, \omega) := g(q)(\omega) + (G(q) - g(q) \cdot q). \quad (17)$$

Then the expected utility of action q under belief p is

$$\sum_{\omega} p(\omega) u(q, \omega) = p \cdot g(q) + G(q) - g(q) \cdot q = G(q) + g(q) \cdot (p - q).$$

Since $g(q) \in \partial G(q)$, the subgradient inequality gives

$$G(p) \geq G(q) + g(q) \cdot (p - q) \quad \forall p, q \in \Delta(\Omega),$$

with equality at $q = p$. Hence the indirect utility of the constructed decision problem equals G :

$$\max_{q \in A} \sum_{\omega} p(\omega) u(q, \omega) = G(p) \quad \forall p \in \Delta(\Omega).$$

Therefore, by [Theorem 1](#), the canonical ex-ante value of information generated by (A, u) is exactly I .

Moreover, [Lemma 10](#) applies to this same (A, u) and yields the same Bregman penalties, so the behaviorally valid meaningful measure generated by (A, u) coincides with the right-hand side of [\(3\)](#), i.e. with M . Thus I and M arise from the same decision problem, so they are jointly valid. $\square$

(ii) $\Rightarrow$ (iii). Let M_G denote the right-hand side of [\(3\)](#). First, M_G extends I to subjective interpretations: on the slice $D = \{(p, \sigma, p, \sigma, \phi^*)\}$ we have $\hat{p} = p$ and $\hat{p}(\cdot | s) = p(\cdot | s)$ for all s , hence both Bregman terms vanish and

$$M_G(p, \sigma, p, \sigma, \phi^*) = I(p, \sigma).$$

Second, M_G satisfies Axioms [4](#), [5](#), [6](#), [7](#), and [8](#). (These follow directly from the tripartite decomposition in [\(2\)](#) together with the properties of Bregman divergences induced by convex potentials.)

We next show uniqueness. Let $\tilde{M}$ be any extension of I to subjective interpretations that satisfies Axioms [4](#), [5](#), [6](#), [7](#), and [8](#). Let M_G denote the candidate extension on the right-hand side of [\(3\)](#).

Step 1: By the tripartite characterization (the decomposition in [\(2\)](#) together with Axioms [5](#) and [6](#)), there exist a convex potential $H : \Delta(\Omega) \rightarrow \mathbb{R}$ and a fixed selector $h(\cdot) \in \partial H(\cdot)$ such that, for all $(p, \sigma, \hat{p}, \hat{\sigma}, \phi)$,

$$\tilde{M}(p, \sigma, \hat{p}, \hat{\sigma}, \phi) = I_H(p, \sigma) + d^{(H, h)}(p, \hat{p}) - \mathbb{E}_{s \sim \text{Pr}}[d^{(H, h)}(p(\cdot | s), \hat{p}(\cdot | s))], \quad (18)$$

where

$$I_H(p, \sigma) := \mathbb{E}_{s \sim \text{Pr}}[H(p(\cdot | s))] - H(p), \quad d^{(H, h)}(x, y) := H(x) - H(y) - h(y) \cdot (x - y).$$

(Here $p(\cdot | s)$ is the Bayesian posterior under (p, σ) and $\hat{p}(\cdot | s)$ is the subjective posterior induced by $(\hat{p}, \hat{\sigma}, \phi)$.)

Step 2: Evaluate [\(18\)](#) on the slice D , i.e. set $(\hat{p}, \hat{\sigma}, \phi) = (p, \sigma, \phi^*)$. Then $\hat{p} = p$ and $\hat{p}(\cdot | s) = p(\cdot | s)$ for all s , so both Bregman terms vanish and

$$\tilde{M}(p, \sigma, p, \sigma, \phi^*) = I_H(p, \sigma) \quad \forall (p, \sigma).$$

Since $\tilde{M}$ extends I , the left-hand side equals $I(p, \sigma)$, hence

$$I_H(p, \sigma) = I(p, \sigma) \quad \forall (p, \sigma). \quad (19)$$

Step 3: Since H and G represent the same canonical measure, there exist $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^\Omega$ such that

$$H(q) = G(q) + \alpha + \beta \cdot q \quad \text{for all } q \in \Delta(\Omega).$$

Define

$$\bar{g}(q) := h(q) - \beta.$$

Then $\bar{g}(q) \in \partial G(q)$ for every $q \in \Delta(\Omega)$. Moreover, for every $x, y \in \Delta(\Omega)$,

$$\begin{aligned} d^{(H, h)}(x, y) &= H(x) - H(y) - h(y) \cdot (x - y) \\ &= G(x) - G(y) + \beta \cdot (x - y) - h(y) \cdot (x - y) \\ &= G(x) - G(y) - \bar{g}(y) \cdot (x - y) \\ &= d^{(G, \bar{g})}(x, y). \end{aligned}$$

Step 4: By (19), $I_H = I^G = I$. Substituting the conclusion of Step 3 into (18) therefore yields

$$\begin{aligned} \tilde{M}(p, \sigma, \hat{p}, \hat{\sigma}, \phi) &= I^G(p, \sigma) + d^{(G, \bar{g})}(p, \hat{p}) \\ &\quad - \mathbb{E}_{s \sim \text{Pr}} \left[d^{(G, \bar{g})}(p(\cdot | s), \hat{p}(\cdot | s)) \right] \\ &= M^{(G, \bar{g})}(p, \sigma, \hat{p}, \hat{\sigma}, \phi). \end{aligned}$$

Thus, every extension of I^G satisfying the axioms is generated by some fixed selector $\bar{g}(\cdot) \in \partial G(\cdot)$. Conversely, by Theorem 2, every fixed selector $g(\cdot) \in \partial G(\cdot)$ generates an extension $M^{(G, g)}$ satisfying the axioms. Hence, conditional on the fixed selector g in part (ii), any such extension must equal $M^{(G, g)}$. Therefore, $M^{(G, g)}$ is the unique extension of I^G associated with that selector. $\square$

(iii) $\Rightarrow$ (ii). Assume (iii). By Theorem 1, there exists a convex potential G (unique up to affine transformations) that represents I . By the tripartite decomposition (2), any extension of I to subjective interpretations that satisfies Axioms 4, 5, 6, 7, and 8 must coincide with the Bregman-penalty correction built from that same potential G (for some fixed selector $g(\cdot) \in \partial G(\cdot)$), namely the right-hand side of (3). Since M is the unique such extension by (iii), it follows that M satisfies (3). $\square$

Combining the implications proves $(i) \Leftrightarrow (ii) \Leftrightarrow (iii)$. $\square$

A.4 Proof of Theorem 4

Proof. We first establish the equivalence between (i) and (ii). By Theorem 2, requiring

$$\hat{\mathcal{E}}_1 \succeq_{\text{MI}} \hat{\mathcal{E}}_2$$

is equivalent to requiring that the behavioral value of information be weakly higher under $\hat{\mathcal{E}}_1$ than under $\hat{\mathcal{E}}_2$ in every decision problem and under every fixed tie-breaking rule. Because the two

interpreted experiments have the same objective prior p and subjective prior $\hat{p}$, their no-signal welfare benchmark is the same and cancels from the comparison. Theorem 1 and Claim 2 of Morris (1991, Chapter III) therefore imply that (i) holds if and only if there exist non-negative weights $\lambda = (\lambda_{qq'})_{q,q' \in Q}$ satisfying (F). Thus, Morris's theorem shows the equivalence between (i) and (ii).

It remains to connect the flow condition to the Bregman representation. Fix a convex potential G and a fixed selector $g(\cdot) \in \partial G(\cdot)$, and consider an associated decision problem implementing (G, g) . For each belief q , let a_q denote the selected action that is optimal at q . As discussed in Section 1, for any $p, q \in \Delta(\Omega)$,

$$p \cdot u(a_q, \cdot) = G(p) - d^{(G,g)}(p, q). \quad (20)$$

Using behavioral validity and grouping signal realizations according to the posterior belief on which the agent acts,

$$\begin{aligned} M^{(G,g)}(\hat{\mathcal{E}}_i) &= \sum_{\substack{s \in S_i \\ \Pr_i(s) > 0}} \Pr_i(s) W(p_i(\cdot | s), \hat{p}_i(\cdot | s)) - W(p, \hat{p}) \\ &= \sum_{\substack{s \in S_i \\ \Pr_i(s) > 0}} \Pr_i(s) [p_i(\cdot | s) \cdot u(a_{\hat{p}_i(\cdot | s)}, \cdot)] - p \cdot u(a_{\hat{p}}, \cdot) \\ &= \sum_{q \in Q} k_q^i \cdot u(a_q, \cdot) - p \cdot u(a_{\hat{p}}, \cdot). \end{aligned}$$

The last term is common across the two interpreted experiments. Hence,

$$M^{(G,g)}(\hat{\mathcal{E}}_1) - M^{(G,g)}(\hat{\mathcal{E}}_2) = \sum_{q \in Q} D_q \cdot u(a_q, \cdot). \quad (21)$$

Suppose now that $\lambda \geq 0$ solves (F). Substituting the flow equations into (21) and relabeling indices in the second term gives

$$\begin{aligned} M^{(G,g)}(\hat{\mathcal{E}}_1) - M^{(G,g)}(\hat{\mathcal{E}}_2) &= \sum_{q,q' \in Q} \lambda_{qq'} q \cdot [u(a_q, \cdot) - u(a_{q'}, \cdot)] \\ &= \sum_{q,q' \in Q} \lambda_{qq'} d^{(G,g)}(q, q'), \end{aligned}$$

where the last equality follows from (20), since

$$q \cdot [u(a_q, \cdot) - u(a_{q'}, \cdot)] = G(q) - [G(q) - d^{(G,g)}(q, q')] = d^{(G,g)}(q, q').$$

Thus, every non-negative solution to (F) provides weights satisfying (4), proving (ii) $\Rightarrow$ (iii). Since $d^{(G,g)}(q, q') \geq 0$ for every $q, q' \in Q$, condition (iii) immediately implies (i).

It remains to prove the final statement. Suppose that Condition (F) is infeasible. By the

equivalence between (i) and (ii), there exist a convex potential G and a fixed selector g such that

$$M^{(G,g)}(\widehat{\mathcal{E}}_1) - M^{(G,g)}(\widehat{\mathcal{E}}_2) = \sum_{q \in Q} D_q \cdot u(a_q, \cdot) < 0,$$

where a_q denotes the selected action that is optimal at belief q in an associated decision problem implementing (G, g) . Moreover, for every $q, q' \in Q$,

$$q \cdot [u(a_q, \cdot) - u(a_{q'}, \cdot)] = d^{(G,g)}(q, q') \geq 0.$$

Thus, consider the finite decision problem with one action a_q for each $q \in Q$, and with state-contingent payoff vector $u(a_q, \cdot)$ inherited from the associated decision problem above. In this finite decision problem, a_q is weakly optimal at belief q .

To make the optimum unique, define

$$B_q(\omega) := 2q(\omega) - \|q\|^2, \quad q \in Q, \omega \in \Omega,$$

and, for $\varepsilon > 0$, perturb the payoff of action a_q according to

$$u^\varepsilon(a_q, \omega) := u(a_q, \omega) + \varepsilon B_q(\omega).$$

For any distinct $q, q' \in Q$,

$$\begin{aligned} q \cdot u^\varepsilon(a_q, \cdot) - q \cdot u^\varepsilon(a_{q'}, \cdot) &= q \cdot [u(a_q, \cdot) - u(a_{q'}, \cdot)] + \varepsilon (q \cdot B_q - q \cdot B_{q'}) \\ &\geq \varepsilon \|q - q'\|^2 \\ &> 0. \end{aligned}$$

Hence a_q is uniquely optimal at belief q .

Because the two interpreted experiments have the same objective prior p and subjective prior $\hat{p}$, the no-signal welfare benchmark is common across them and cancels from the comparison. Therefore, the difference in behavioral value of information under the perturbed finite decision problem is

$$\sum_{q \in Q} D_q \cdot u(a_q, \cdot) + \varepsilon \sum_{q \in Q} D_q \cdot B_q.$$

The first term is strictly negative, while the second is finite. Therefore, for sufficiently small $\varepsilon > 0$, the expression remains strictly negative. This gives a finite decision problem with a unique optimal action at every acted-upon belief $q \in Q$ in which the behavioral value of information is strictly lower under $\widehat{\mathcal{E}}_1$ than under $\widehat{\mathcal{E}}_2$. $\square$

B. Blackwell Reversal for Meaningful Information

Example 1. Let $\Omega = \{0, 1\}$ with prior $p(0) = p(1) = \frac{1}{2}$. For $\eta \in [\frac{1}{2}, 1)$, consider the objective binary symmetric signal σ_η with accuracy η :

$$\Pr(s = \omega \mid \omega) = \eta, \quad \Pr(s \neq \omega \mid \omega) = 1 - \eta.$$

It is well known that $\sigma_{\eta'} \succeq_B \sigma_\eta$ whenever $\eta' \geq \eta \geq \frac{1}{2}$.

Suppose the agent's prior is correct, $\hat{p} = p$, but she misunderstands which signal is diagnostic of which state. In particular, she treats the messages as swapped, so that her subjective posterior when observing s equals the objective posterior conditional on the opposite message $s' \neq s$.

Under the meaningful information extension of mutual information, [Theorem 3](#) gives

$$M(p, \sigma_\eta, \hat{p}, \hat{\sigma}, \phi) = I(p, \sigma_\eta) - \mathbb{E}_{s \sim \Pr} \left[\text{KL}(p(\cdot \mid s) \parallel p(\cdot \mid s')) \right],$$

where $s' \neq s$. This expression can be shown to reduce to:

$$M(p, \sigma_\eta, \hat{p}, \hat{\sigma}, \phi) = \log 2 + \eta \log \eta + (1 - \eta) \log(1 - \eta) - (2\eta - 1) \log \frac{\eta}{1 - \eta}.$$

Taking a derivative with respect to η yields:

$$\frac{d}{d\eta} M(p, \sigma_\eta, \hat{p}, \hat{\sigma}, \phi) = -\log \frac{\eta}{1 - \eta} - \frac{2\eta - 1}{\eta(1 - \eta)} < 0$$

because $\eta \in (\frac{1}{2}, 1)$. Furthermore, since $M(p, \sigma_{1/2}, \hat{p}, \hat{\sigma}, \phi) = 0$ and the derivative is negative for $\eta > \frac{1}{2}$, it follows that

$$M(p, \sigma_\eta, \hat{p}, \hat{\sigma}, \phi) < 0 \quad \text{for all } \eta \in \left(\frac{1}{2}, 1\right)$$

Thus, within this family, the more informative the objective experiment is in the Blackwell order, the more negative the amount of meaningful information transmitted becomes. $\square$

The intuition behind the example is straightforward. When the agent systematically misinterprets which signal is diagnostic of which state of the world, more precise signals push her posterior beliefs further in the wrong direction.