

Learning from a Mixture of Information Sources*

[Draft; please do not distribute]

Nicole Immorlica[†] Brendan Lucier[‡] Clayton Thomas[§] Ruqing Xu[¶]

July 2025

Abstract

We often learn from multiple sources that convey information in different ways. How informative is it to know the source of a signal, and how is this informativeness shaped by the distribution of sources? We extend the standard (binary-state, binary-signal) Blackwell experiment model by introducing a commonly known distribution over signaling schemes, representing the distribution of information sources. We compare learning under two information models: *source-aware*, where decision makers observe a signaling scheme and its realization (e.g., raw reviews, search results), and *source-blind*, where only the signal realization is observed (e.g., aggregate ratings, generated summaries). We show that a mean-preserving spread in the distribution of signaling schemes translates into Blackwell dominance for source-aware decision makers, implying they are “risk-loving” in information sources. In contrast, it has no impact on source-blind decision makers. When learning from repeated draws of signaling schemes, source-blind decision makers learn more slowly. However, as long as the average signaling scheme is ε away from being completely uninformative, source-blind learning can match source-aware learning by using at most $O(1/\varepsilon)$ times more data.

*We thank Marco Battaglini, Takuma Habu, Rohit Lamba, Mira Frick, Navin Kartik, Ankur Mani, Krishnamurthy Iyer, Thodoris Lykouris, Yiding Feng, Liang Lyu, and participants at Cornell Microeconomic Theory reading group, Cornell Artificial Intelligence, Policy, and Practice (AIPP), and Marketplace Innovation Workshop (MIW) for helpful discussions and comments.

[†]Microsoft Research — *E-mail*: nicinm@gmail.com.

[‡]Microsoft Research — *E-mail*: brlucier@microsoft.com.

[§]Microsoft Research — *E-mail*: thomas.clay95@gmail.com.

[¶]Cornell University — *E-mail*: rx24@cornell.edu.

1 Introduction

In economic decision-making, individuals and institutions routinely rely on information from diverse sources. Whether a piece of information—such as a product review, a news article, or a medical recommendation—is informative depends not only on its content but critically on who produces it. A satirical news report, for instance, invites different interpretations compared to a reputable news outlet; medical advice from a layperson lacks the authority of a licensed professional; and the relevance of a restaurant review hinges on how the tastes of the reviewer and the reader compare. In these examples, knowing the source of a piece of information can be critical to evaluate its content and to inform decision making.

On the other hand, modern technology offers unprecedented opportunities for summarization and processing large amounts of data at low costs. While aggregating information opens up the possibility of harnessing large datasets, the source of each data point—reflecting its reliability, bias, or relevance—is often obscured in the process. In the age of big data, the common wisdom is that this loss of context can be overcome with ever-increasing amounts of data. This raises fundamental questions about the role of data provenance in learning. How is informativeness shaped by the distribution of sources? How does the value of knowing a signal’s source compare with the ability to process more samples?

Our paper proposes a framework for answering these questions. There is an unknown decision-relevant state of the world. A signaling scheme is a Blackwell experiment that maps the state to a probability distribution over finite signal realizations. We assume access to a commonly known *distribution* over signaling schemes, representing heterogeneity in information sources. We consider a learner who can observe multiple signal realizations, each generated by a separate signaling scheme independently drawn from this distribution. We compare two learning environments, representing different amounts of information available to the learner. In the *source-aware* setting, the learner observes the details of the signal-generating processes (e.g., reading raw reviews or examining the origin of search results) in addition to the realizations of the signals from those sources. In the *source-blind* setting, only the signal realizations are observed by the learner (e.g., seeing aggregated ratings or summary statistics) while the nature of the signal-generating processes is obscured. We can ask our main question in the language of the model:

How many more samples does a source-blind learner require to be as informed—and hence as effective at decision-making—as a source-aware learner?

As an initial foray into answering this question, we instantiate our framework in a simple binary setting. There is a state of the world $\theta \in \{0, 1\}$, initially unknown to the learner. Each potential signaling scheme generates a binary signal that can be correlated with θ . We assume “no fake data,” meaning that no signal realization is negatively correlated with the state, but we do allow perfectly uninformative signals that are not correlated with θ . While restrictive, this binary scenario is nevertheless sufficiently rich to illustrate the gap between source-blind and source-aware learning, while also capturing natural scenarios in which a decision-maker must select whether or not to take an action (such as purchasing a product) given a dataset that takes the form of action recommendations.

In our main result, we give an upper bound on the advantage of being source-aware when learning from many different sources drawn from the distribution. We prove that if the *average* of all signaling schemes is at least an ε -distance (in an appropriate metric) from being completely uninformative, then a source-blind learner needs no more than $O(1/\varepsilon)$ times as much data to be at least as informed as a source-aware learner in the sense of Blackwell dominance. This means,

in particular, that for any decision problem that depends on the state θ , a source-aware decision maker will be outperformed (in expectation) by a source-blind decision maker with access to $O(1/\varepsilon)$ times as many signals.

Our bound depends on the informativeness of an average signaling scheme with respect to the distribution from which they are drawn. To see why, imagine a distribution in which a very small fraction of signals are fully informative of the state, while the rest are completely uninformative. A source-aware learner would then learn the state perfectly if even a single informative signal is observed, whereas for a source-blind learner a single informative signal is likely to become lost in the noise. However, if sufficiently many informative signals are present in the dataset, then the state could be inferred even from aggregate statistics. Our analysis essentially shows that this type of example is the worst-case scenario for a source-blind learner, and the resulting bound applies to all potential distributions over data sources.

Along the way to establishing our main result, we analyze how changes in the underlying distribution over signal-generating processes affect the rate at which a source-aware or source-blind decision maker becomes informed. We show that as the distribution over signal-generating processes becomes “more spread out,” in the sense of a mean-preserving spread, a source-aware decision maker becomes strictly more informed in the sense of Blackwell informativeness. Again, this implies that the source-aware decision-maker gets weakly higher expected utility, under any decision problem, when learning from more heterogeneous data sources. In contrast, we show that the source-blind decision maker cares only about the mean signal, and hence her level of informativeness is unchanged when the source become more spread out in the same sense as above. These results, which may be of independent interest, apply to any finite signal and state space (not necessarily binary).

Related Work. There is a *vast* literature on learning from signaling schemes, i.e., Blackwell experiments [Blackwell et al., 1951, Blackwell, 1953]. This literature considers a host of important topics and results, including (but not limited to) how different information experiments can feasibly correlate [e.g., Brooks et al., 2022, 2024], how agents learn in social environments [e.g., Banerjee, 1992, Huang, 2024], how costs can be assigned to experiments [e.g., Pomatto et al., 2023, Bloedel and Zhong, 2020], and how experiments can be ranked [e.g., Moscarini and Smith, 2002, Azrieli, 2014, Kosenko, 2024]. Our main result in Section 7.1 relies on a related concept from Mu et al. [2021], the dominance ratio, which gives a quantitative comparison between two experiments based on the information content of many repeated independent samples from the experiments.

There is relatively little work that, like our model, considers learning in a *distribution* over signaling schemes. Acemoglu et al. [2016] consider agents with different subjective priors over signaling schemes. They consider how agents learn from repeated signal realizations generated by a single signaling scheme drawn once from their prior. In contrast, our main result in Section 7.1 consider learning from many repeatedly-drawn signaling schemes from a commonly-known distribution of such schemes, so there is no learning about signal-generating processes overtime. Cripps et al. [2008] considers agents that learn from correlated signals and studies conditions under which joint learning occurs and is common knowledge. In our scenario, a source-aware learner has only more information than the source-blind learner, and hence joint learning is assured at the rate at which the source-blind agent learns; our focus is on comparing the rates at which they learn. Our result in Theorem 5.1 concerning the mean-preserving spread of signaling schemes generalizes a part of a result in Whitmeyer [2024], which finds that a Bayesian decision maker weakly prefers randomization over two experiments to their convex combination. They define this as “convexity,” and show that Bayes’ rule is the only updating rule satisfying convexity and certain other requirements.

Beyond the context of learning, recent works on communication games have incorporated ambiguity into signaling schemes, similar to how we introduce heterogeneous signaling schemes. In ambiguous persuasion [Beauchêne et al., 2019, Cheng, 2023, Cheng et al., 2024], the sender is allowed to use ambiguous communication devices comprising a finite set of signaling schemes, while the receiver is assumed to be ambiguity-averse and to maximize their maxmin expected utility. Similarly, Kellner and Le Quement [2017] studies cheap talk communication with purely exogenous ambiguity and Kellner and Le Quement [2018] studies when sender can strategically choose to communicate ambiguously.

Other papers study conceptually similar questions to ours in different settings. Motivated by investigating the “value of context,” Iakovlev and Liang [2024] ask whether an agent prefers to be evaluated by a human who observes a smaller but customized set of covariates, or an algorithm which observes a bigger set of non-customizable covariates. They find that the expected value of customization vanishes to zero as the number of covariates grows. In a more classic setting, several papers have considered the value of context when learning from signals generated by individuals connected in a social network. For example, if individuals receive binary signals about an unknown state of the world and repeatedly report their best guess of this state to their neighbors, the “context” (i.e., the structure of the network) impacts whether the society can learn the state. If individuals are Bayesian and correctly incorporate the context in their updating process, learning will be asymptotically efficient for a wide range of dynamics Gale and Kariv [2003], Mossel et al. [2015, 2016]. However, if individuals simply update using a majority dynamic, then the society might fail to learn the state, and multiple beliefs might persist indefinitely unless the social network and dynamics satisfy certain properties Feldman et al. [2014], Mossel et al. [2014]. Another domain related to the value of context is the problem of learning mixture models, whereby data is known to be generated from multiple sources but the source label (i.e., context) is hidden. Depending on the underlying data distribution from each individual source, learning the mixture model may be substantially more difficult than if the source labels were revealed Dasgupta [1999], Dasgupta et al. [2005].

Our main result, Theorem 7.3, shows that if the average signal is not too uninformative, then for any decision problem, a source-blind learner with access to only $O(1/\epsilon)$ times more signals will outperform a source-aware learner. This has a similar flavor to Bulow and Klemperer [1994]’s seminal result that in independent private value auctions, an auctioneer can generate more profit by adding one more bidder than by switching to the optimal auction mechanism. Similarly, the seminal paper in algorithmic game theory, Roughgarden and Tardos [2002], shows that a network suffers less total latency with Nash routing and twice the edge capacity than that of the optimal routing. As such, our result joins the common refrain in the Economics and Computation literature that the optimal mechanism can often be bested by a simple mechanism plus a bit more resources.

Paper Organization. The rest of the paper proceeds as follows. We define our model of source-aware and source-blind learning in Section 2. We give background on the main concept needed to express our results—the Blackwell order—in Section 3, and makes basic observations comparing source-aware and -blind learning in Section 4. We study source-aware learners in Section 5, and prove they are “risk-loving” in information sources. Section 6 introduces a tractable special-case of our general model: that of binary states and signals under a “no-fake-data” assumption. In Section 7, we consider this special case and give our main result, which bounds the advantage of knowing the source when learning from many sources. Section 8 concludes.

2 Model and Preliminaries

2.1 Signaling Schemes

Let $\Theta = \{0, 1, \dots, k-1\}$ be a finite set of $k \geq 2$ states. For concreteness, we assume that the decision maker initially has a uniform prior belief over Θ ; since our results all concern Blackwell dominance and other prior-free notions, standard arguments imply that our results hold for any other fixed prior.

A *signaling scheme* is a function $\pi : \Theta \rightarrow \Delta(\Omega)$ for some finite realization space Ω . We refer to an $\omega \in \Omega$ as a *signal realization*. We can equivalently view a signaling scheme π as a set of probability measures $(P_\theta^\pi)_{\theta \in \Theta}$ over Ω , where P_θ^π is the measure of signals realized when the state is θ . We use $p_{\theta\omega}^\pi = \pi(\omega \mid \theta)$ for $\omega \in \Omega$ and $\theta \in \Theta$ to denote the specific entries in the probability measure P_θ^π .¹ Whenever it is clear which signaling scheme we refer to, we may omit the superscript and write P_θ and $p_{\theta\omega}$. We let \mathcal{P} denote the set of feasible signaling schemes, which we refer to as the *domain*.

Given a signaling scheme π , let $\tau_\pi \in \Delta(\Delta(\Theta))$ denote the distribution over posterior beliefs which arises from the signaling scheme π . Since $|\Theta| = k$, we typically represent τ_π as an element of $\Delta(S_k)$, where $S_k = \{(q_0, \dots, q_{k-1}) \in [0, 1]^k \mid \sum_{i=0}^{k-1} q_i = 1\}$ denotes the k -simplex and the event $\tau_\pi = (q_0, \dots, q_{k-1})$ represents the posterior belief that the state is i with probability q_i .

For an concrete example of a signaling scheme and the induced distribution of posterior beliefs, we refer the reader to [Section 6](#) for a discussion under binary state and signal spaces.

2.2 Learning and Information Models

We focus on a decision maker who wishes to learn about the state $\theta \in \Theta$ from one or more signal realizations. Informally, these are realized according to signal processes that are themselves repeatedly drawn from a *distribution* Π over the domain of signaling schemes \mathcal{P} , i.e., $\Pi \in \Delta(\mathcal{P})$. The standard Blackwell experiment framework with a single signaling scheme $\pi \in \mathcal{P}$ thus corresponds to the distribution Π which is a single point mass at π .

We compare learning under two information models: source-aware and source-blind learning.

First, we introduce source-aware learning. Intuitively, we say a decision maker is *source-aware* if she learns how a piece of information is generated, together with the content of the information. This learning regime may represent environments where one reads raw reviews on a platform or read sources from a search engine.

Formally, we define the source-aware model as follows. Nature draws (i) a state $\theta \in \Theta$ according to the uniform prior, (ii) a signal scheme $\pi \sim \Pi$, and (iii) a signal realization $\omega \sim P_\theta^\pi$. The decision maker learns the tuple (π, ω) , i.e., the signal scheme and a realization from it. This process is equivalently a single signal scheme with an enriched signal space $\mathcal{P} \times \Omega$. Note that the draw of π from the distribution Π is independent of the state—only the realization ω is state-dependent. We denote this signaling scheme by $A(\Pi)$ where A is a mnemonic for “aware.”

Second, we introduce source-blind learning. Intuitively, we say the decision maker is *source-blind* if she sees a piece of information without learning its source. This learning regime may represent environments where one reads aggregated ratings or generated summaries where information about sources are often suppressed.

Formally, in source-blind learning, nature draws a random θ, π , and ω exactly as before, but the

¹The standard notation $\pi(\omega \mid \theta)$ represents $\mathbb{P}[\pi(\theta) = \omega]$, the probability of observing signal realization ω given that the state is θ . Additionally, when no confusion can arise, we write a distribution (e.g., $\pi(\theta)$ or τ) with the same notation as a random variable following that distribution (e.g., we consider events $\pi(\theta) = \omega$ or $\tau = q$).

decision maker only sees ω and not π . Thus, this constitutes its own signaling scheme with signal space Ω . We refer to this signaling scheme as $B(\Pi)$ where B is a mnemonic for “blind.”

3 Background: The Blackwell Order

We are interested in how useful different signaling schemes are to decision makers. To that end, in this section we review a classic tool for comparing different information sources.

The canonical way to capture and compare the usefulness of signaling schemes is via the Blackwell informativeness ordering [Blackwell, 1953], which is a partial order \succeq over signaling schemes such that $P \succeq Q$ if and only if *every* decision maker prefers P over Q . We use several classical equivalent definitions for the Blackwell order, particularly using the concepts of garblings of signaling schemes and mean-preserving spreads of posterior distributions.

Definition 3.1 (Garblings of signaling schemes). Let $P : \Theta \rightarrow \Delta(\Omega)$ and $Q : \Theta \rightarrow \Delta(\Xi)$ be two signaling schemes with state space Θ and realization spaces Ω and Ξ . We say that a *garbling function* from P to Q is a function $T : \Omega \rightarrow \Delta(\Xi)$ such that for all $\theta \in \Theta$, if we sample $\omega \sim P(\theta)$ and $\xi_g \sim g(\omega)$, then ξ_g is equal in distribution to $Q(\theta)$. We say Q is a *garbling* of P if there exists a garbling function from P to Q .

Definition 3.2 (Mean-preserving spread in \mathbb{R}^m). For random variables X_1 and X_2 in $\Delta(\mathbb{R}^m)$, we say that a *spread function* from X_1 to X_2 is a function $s : \mathbb{R}^m \rightarrow \Delta(\mathbb{R}^m)$ such that:

- (1) For all t in the support of X_1 , we have $\mathbb{E}[s(t)] = t$.
- (2) If we draw $z \sim X_1$ and then $y \sim s(z)$, then y is equal in distribution to X_2 .

We say X_2 is a *mean-preserving spread* of X_1 if there exists a spread function from X_1 to X_2 .²

Formally, the equivalences we use can now be states as follows.

Theorem 3.3 (Blackwell [1953]). Let $P : \Theta \rightarrow \Delta(\Omega)$ and $Q : \Theta \rightarrow \Delta(\Xi)$ be two signaling schemes with state space Θ and realization spaces Ω and Ξ . The following statements are equivalent:

- (i) For every decision problem with state space Θ , any Bayesian decision maker can achieve weakly higher expected utility under P than under Q .³
- (ii) Q is a garbling of P .
- (iii) The posterior distribution τ_P is a mean-preserving spread of τ_Q .

In this case, we say that P **Blackwell dominates** Q , denoted by $P \succeq Q$.

²Equivalently, X_2 is a mean-preserving spread of X_1 if there exists a random variable Z such that for all $x_1 \in \text{supp}(X_1)$, we have (1) $\mathbb{E}[Z \mid X_1 = x_1] = 0$, and (2) $X_1 + Z$ is equal in distribution to X_2 . Note that we define mean-preserving spreads over the slightly more-general space of \mathbb{R}^m , and not just the space of posterior distributions (as is needed to state Theorem 3.3); this generality will be useful in Section 5 when we discuss mean-preserving spreads over the space of signaling schemes \mathcal{P} .

³Formally, a decision problem specifies a prior over Θ , a set of actions A , and a utility function $u : \Theta \times A \rightarrow \mathbb{R}$. Under the signaling scheme P , the model is as follows. Nature draws the state θ , the decision maker sees a signal realization drawn from $P(\theta)$, and then takes an action that maximizes her utility based on the signal realization she receives.

3.1 The Dominance Ratio

A Blackwell ordering relation $P \succeq Q$ captures a canonical and strong notion of P being more informative than Q ; however, it remains an ordinal ranking. In other words, the Blackwell order can tell us which signaling scheme is better, but cannot tell us *by how much*.

To address the above, a line of work has considered quantitative comparisons between signaling schemes P, Q when learning from a large number of repeated independent samples from $P(\theta)$ vs. $Q(\theta)$. We follow the approach given by the *dominance ratio* in [Mu et al. \[2021\]](#).⁴ The dominance ratio bounds the comparative number of samples from one signaling scheme needed to Blackwell dominate another signaling scheme. For example, if observing 25 independent samples from a signaling scheme P is Blackwell more informative than 100 independent samples from a signaling scheme Q , then intuitively, we say P is (at least) 4 times as informative as Q for large enough samples, and the dominance ratio is (at least) 4. This notion is formalized as follows.

Definition 3.4 ([Mu et al. \[2021\]](#)). The dominance ratio of two signaling schemes P and Q is defined as:

$$P/Q = \sup \left\{ \frac{m}{n} : P^{\otimes n} \succeq Q^{\otimes m} \right\}$$

where $P^{\otimes n}$ denote the n -fold signaling scheme where n independent observations are generated according to the signaling scheme P .

Intuitively, a dominance ratio of r suggests that P will be at least r times as informative as Q in large samples. We use the characterization from [Mu et al. \[2021\]](#) of the dominance ratio in terms of the statistical notion of Rényi divergences.

Definition 3.5 ([Rényi \[1961\]](#)). Given a signaling scheme P with binary state space $\Theta = \{0, 1\}$, signal space Ω and probability measures P_0 and P_1 . The Rényi t -divergence of P under θ is

$$R_P^\theta(t) = R_t(P_\theta \parallel P_{1-\theta}) = \frac{1}{t-1} \log \left(\int_{\Omega} \left(\frac{dP_\theta}{dP_{1-\theta}}(\omega) \right)^{t-1} dP_\theta \right)$$

When $t = 1$, it is the Kullback-Leibler divergence.

$$R_P^\theta(1) = R_1(P_\theta \parallel P_{1-\theta}) = \int_{\Omega} \log \left(\frac{dP_\theta}{dP_{1-\theta}}(\omega) \right) dP_\theta$$

Theorem 3.6 ([Mu et al. \[2021\]](#)). Under binary states $\Theta = \{0, 1\}$, we have

$$P/Q = \inf_{\theta \in \{0,1\}, t > 0} \frac{R_P^\theta(t)}{R_Q^\theta(t)}$$

⁴The main result of [Mu et al. \[2021\]](#) characterizes the *large sample* ordering of signaling schemes, which P dominating Q in large samples means that there is an n large enough so that n repeated draws from P Blackwell dominates n repeated draws from Q . This ordering was first introduced by [Azrieli \[2014\]](#) as a refinement of the ordering proposed by [Moscarini and Smith \[2002\]](#) (where the above n is allowed to depend on the particular decision problem). [Mu et al. \[2021\]](#) prove [Theorem 3.6](#) as a corollary of their main characterization. In addition to providing a quantification of the advantage of a signal over another, this notion is able to compare all pairs of signaling schemes, and not only those that are Blackwell ranked.

4 Source-Aware and Source-Blind Learning: Preliminaries

We now return to our model of learning from a mixture of information sources, both with and without the knowledge of the source, and provide some preliminary analysis.

To begin our analysis, we consider the source-blind signaling scheme. While one may worry that a source-blind decision maker may need to think in a sophisticated way about the distribution of signaling schemes, it turns out that there is a simple way to view source-blind learning. Namely, the blind signaling scheme $B(\Pi)$ exactly corresponds to the average of all signaling schemes in Π , where the average is calculated by treating π as an element of $\mathbb{R}^{|\Theta| \times |\Omega|}$, given by $(p_{\theta\omega})_{\theta \in \Theta, \omega \in \Omega}$.

Proposition 4.1. *For any distribution of signaling schemes $\Pi \in \Delta(\mathcal{P})$, we have that $B(\Pi)$ is equivalent to the “mean signal” $\bar{\pi} = \mathbb{E}[\Pi]$.*

Proof. Observe that the distribution of signal realizations under $B(\Pi)$ exactly equals the distribution of signal realizations under $\bar{\pi}$ for each state. Formally, in state θ , signal realization ω occurs under $B(\Pi)$ with probability exactly $\mathbb{E}_{\pi \sim \Pi} [p_{\theta\omega}^{\pi}] = p_{\theta\omega}^{\bar{\pi}}$. \square

Next, we show that for a given Π , source-aware decision makers always learn more information than source-blind ones. This holds because the source-aware decision maker has the option to simply *ignore the source*. Formally:

Proposition 4.2. *For any distribution of signaling schemes $\Pi \in \Delta(\mathcal{P})$, we have $A(\Pi) \succeq B(\Pi)$.*

Proof. Recall that $A(\Pi)$ is a signaling scheme over $\mathcal{P} \times \Omega$, and $B(\Pi)$ is a signaling scheme over Ω . Consider the function $\gamma : \mathcal{P} \times \Omega \rightarrow \Delta(\Omega)$ such that $\gamma(\pi, \omega) = \omega$. It holds directly by the definitions that γ is a garbling from $A(\Pi)$ to $B(\Pi)$, as defined in [Theorem 3.3](#), and hence $A(\Pi)$ Blackwell dominates $B(\Pi)$. \square

We end this section by introducing some notation. In our analysis, it is sometimes useful to consider the distribution of Π conditional on a given posterior $q \in S_k$ resulting from $A(\Pi)$. We denote this by $\Pi|q$. In other words, $\Pi|q$ denotes a source-aware decision maker’s posterior distribution over Π , conditional on having received a signal realization such that their posterior distribution over Θ is q .

[Section 6.1](#) provides an example illustrating [Proposition 4.1](#) and [Proposition 4.2](#) in the binary state and signal setting. Under this setting, we also provide a visualization for the distribution $\Pi|q$ in [Figure 4](#).

5 Source-Aware Learners Are Risk-loving in Information

In this section we derive a result about how source-aware decision makers compare different signaling scheme distributions. We show that source-aware decision makers always prefer “more spread out” distributions over signaling schemes. This is formalized with the standard notion of a mean-preserving spread in the space of distributions over \mathcal{P} , where \mathcal{P} is treated as a subset of multi-dimensional Euclidean space, via precisely the same definition as in [Section 3](#). This will hold due to much the same intuition as in [Proposition 4.1](#): a source-aware learner will always prefer “more spread out” signaling schemes in the same way that a learner would rather *be* source-aware, instead of learning based on only the mean of the distribution of signaling schemes, i.e., being source-blind.

The main result of this section is the following.

Theorem 5.1. *For $\Pi_s, \Pi \in \Delta(\mathcal{P})$, suppose Π_s is a mean-preserving spread of Π , as in [Definition 3.2](#). Then, $A(\Pi_s)$ Blackwell dominates $A(\Pi)$.*

Proof. Let Π_s be a mean preserving spread of Π , and let $s : \mathbb{R}^m \rightarrow \Delta(\mathbb{R}^m)$ be the spread function in the space of signaling schemes as in [Definition 3.2](#). Recall that this means that (1) for all $\pi \in \text{supp}(\Pi)$, we have $\mathbb{E}[s(\pi)] = \pi$, and (2) if we sample $\pi \sim \Pi$ and then $\pi_s \sim s(\pi)$, then π_s is equal in distribution to Π_s . By the equivalences in [Theorem 3.3](#), it suffices to show that $\tau_{A(\Pi_s)}$ is *also* a mean-preserving spread of $\tau_{A(\Pi)}$, in the space of posteriors S_k (where recall that S_k denotes the k -simplex). We use $s(\cdot)$ to construct a spread function in the posterior space $t : S_k \rightarrow \Delta(S_k)$ from $\tau_{A(\Pi)}$ to $\tau_{A(\Pi_s)}$.

Fix a certain π , consider $s(\pi)$, which is itself a distribution over signaling schemes. Observe that for all π in the support of Π , we have that $A(s(\pi))$ Blackwell dominates π , by the same logic we used to establish that $A(\Pi)$ always Blackwell dominates $B(\Pi)$ in [Proposition 4.2](#).⁵ By the equivalences in [Theorem 3.3](#), $\tau_{A(s(\pi))}$ is a mean-preserving spread of τ_π . Thus, there is a spread function $t_\pi : S_k \rightarrow \Delta(S_k)$ from τ_π to $\tau_{A(s(\pi))}$. Now, for any posterior $q \in \text{supp}(\tau_{A(\Pi)}) \subseteq S_k$, define $t(q)$ as follows: First, draw π from $\Pi|q$; Second, draw q_s from $t_\pi(q)$; Finally $t(q)$ is the distribution of q_s that arises from the above process.⁶ We will show that $t(\cdot)$ is a valid mean-preserving spread by showing: (1) it is mean-preserving at each point; and (2) the distribution after the spread equals the distribution $\tau_{A(\Pi_s)}$.

(1) To see that $\mathbb{E}[t(q)] = q$ for each q , we use the law of iterated expectation, and observe that for all π in the support of $\Pi|q$, we know $t_\pi(q)$ is (by the definition of a spread function) a distribution with mean q . Therefore we can write:

$$\mathbb{E}[t(q)] = \mathbb{E}_{\substack{q_s \sim t_\pi(q) \\ \pi \sim \Pi|q}} [q_s] = \mathbb{E}_{\pi \sim \Pi|q} \left[\mathbb{E}_{q_s \sim t_\pi(q)} [q_s | \pi] \right] = \mathbb{E}_{\pi \sim \Pi|q} [q] = q$$

(2) Now we must show that if q_s is distributed by drawing q from $\tau_{A(\Pi)}$ and then q_s from $t(q)$, then q_s is equal in distribution to $\tau_{A(\Pi_s)}$. To see this, observe that by the definition of $t(q)$, this is equivalent to: First, draw q from $\tau_{A(\Pi)}$; Second, draw π from $\Pi|q$; Third, draw q_s from $t_\pi(q)$. In this process, π is distributed according to Π . Additionally, since t_π is a spread function from τ_π to $\tau_{A(s(\pi))}$, we know that for any fixed π , the distribution of q_s conditioned on this π equals the distribution of $\tau_{A(s(\pi))}$. Since s is a spread function from Π to Π_s as in [Definition 3.2](#), this means that if we draw $\pi \sim \Pi$ and then $\pi_s \sim s(\pi)$, then π_s is equal in distribution to Π_s . Therefore, the overall distribution of q_s follows $\tau_{A(\Pi_s)}$, as desired. This finishes the proof. \square

6 A Special Case: Binary States and Signals

All of our results to this point have applied to a general framework with finitely many states and signal realizations. We now instantiate this general framework in the special case with binary states and binary signaling schemes. I.e., suppose we have $\Theta = \{0, 1\}$ and $\Omega = \{0, 1\}$. In this case, we denote a signaling scheme π as (p_{00}, p_{11}) , where p_{00} and p_{11} are the probability of the signal realization matching the state when the state is 0 and 1, respectively. [Table 6](#) illustrates such a signaling scheme.

We often further restrict attention to the domain of signaling schemes such that signal realization 0 (resp., 1) is always evidence that the state is 0 (resp., 1). Formally, we denote the domain as

⁵The crucial reliance of this step on [Proposition 4.1](#) illustrates why mean-preserving spread over \mathcal{P} are useful. It also shows that our parametrization of the space of signaling schemes \mathcal{P} , and its notion of averages, may be the “correct” way to represent learning from a mixture of information sources.

⁶Note that when π is drawn from $\Pi|q$, posterior τ_π must include q in the support and t_π must include q in its domain. That’s why we can always evaluate $t_\pi(q)$.

$\mathcal{P}_{\triangleright} = \{(p_{00}, p_{11}) \mid p_{00} + p_{11} \geq 1\}$.⁷ We interpret this to mean that there is “no fake data.” Hence, we call this restriction the *two-signal, no-fake-data* setting. This domain of signaling schemes is illustrated in [Figure 1](#).

Given a signaling scheme π , recall that $\tau_{\pi} \in \Delta(\Delta(\Theta))$ denotes the distribution over posterior beliefs which arises from the signaling scheme π . Since we consider binary states, we typically represent τ_{π} as an element of $\Delta([0, 1])$, where the event $\tau_{\pi} = q$ represents the posterior belief that the state is 1 with probability q . By Bayes theorem, the decision maker with a uniform prior who knows π and observes ω forms the posterior $p_{1\omega}/(p_{0\omega} + p_{1\omega})$.

		Signal space Ω	
		0	1
State space Θ	0	p_{00}	p_{01} $= 1 - p_{00}$
	1	p_{10} $= 1 - p_{11}$	p_{11}

Table 1: Binary signaling scheme specified by (p_{00}, p_{11})

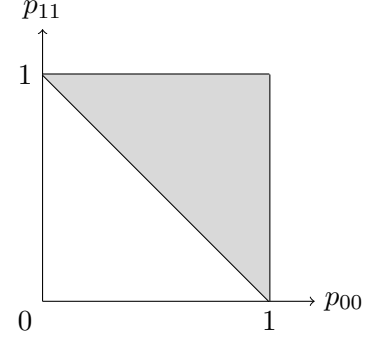


Figure 1: Two-signal, “no fake data” domain

[Table 2](#) shows an example of a (no-fake-data) binary signaling scheme π_z with $p_{00} = 1/4$ and $p_{11} = 7/8$. The top mini-figure in [Figure 2](#) illustrates of the prior: with probability (abbreviated “w.p.”) equal to 1, the decision maker begins with the belief that the state is 1 with probability equal to $1/2$ (with this belief represented as the horizontal axis). The bottom mini-figure illustrates the distribution of posteriors the decision maker will have after observing a signal realization from π_z : when the realization is 0 (which occurs w.p. $3/16$), her posterior is $3/16$; when the realization is 1 (which occurs w.p. $13/16$), her posterior is $7/13$.

		Signal space Ω	
		0	1
State space Θ	0	$1/4$	$3/4$
	1	$1/8$	$7/8$

Table 2: Binary signaling scheme specified by $(1/4, 7/8)$

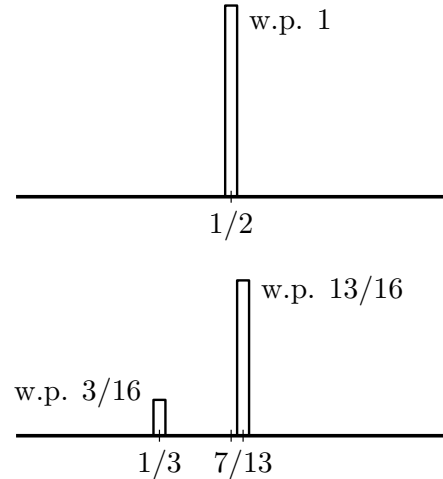


Figure 2: Illustration of the prior, and of the distribution of posteriors generated by the signaling scheme in [Table 2](#)

⁷A straightforward application of Bayes rule show that a signaling scheme π increases the posterior belief that the state is 1 when the signal realization is 1 if and only if $\pi = (p_{00}, p_{11})$ has $p_{00} + p_{11} \geq 1$

6.1 Source-Aware and Source-Blind Learning in Binary Settings

We next present an example illustrating [Proposition 4.1](#) and [Proposition 4.2](#) in the context of the binary setting. Consider the case of binary signals under the “no fake data” assumption, where we represent signaling scheme as the pair (p_{00}, p_{11}) , and let Π be a uniform distribution over $\{(1, 0), (0, 1), (1, 1)\}$. [Proposition 4.1](#) says that $B(\Pi)$ is equivalent to the single signaling scheme $(2/3, 2/3)$; hence, $\tau_{B(\Pi)}$ is the uniform distribution on the posteriors $\{1/3, 2/3\}$. On the other hand, $A(\Pi)$ is equivalent to the signalling scheme that reveals the state with probability $1/3$ (and hence yielding posterior 0 or 1 with total probability $1/6$ each), and is completely uninformative with probability $2/3$. [Proposition 4.2](#) says that $A(\Pi) \succeq B(\Pi)$. Thus, $\tau_{A(\Pi)}$ is a mean-preserving spread of $\tau_{B(\Pi)}$; to see this directly from the distribution of posteriors, consider the random variable Z such that:

$$\begin{aligned} \mathbb{P}[Z = 0 \mid \tau_{B(\Pi)} = 1/3] &= 1/3 & \mathbb{P}[Z = 1/2 \mid \tau_{B(\Pi)} = 2/3] &= 2/3 \\ \mathbb{P}[Z = 1/2 \mid \tau_{B(\Pi)} = 1/3] &= 2/3 & \mathbb{P}[Z = 1 \mid \tau_{B(\Pi)} = 2/3] &= 1/3. \end{aligned}$$

Then, one can check that Z satisfies the conditions of witnessing a mean-preserving spread as in [Theorem 3.3](#). [Figure 3](#) illustrates this example of $B(\Pi)$ and $A(\Pi)$.

Posterior Distribution over Signal Sources: The Binary Case The binary model likewise makes it easier to visualize the distribution $\Pi|q$ over signal sources conditional on $A(\Pi)$. Consider a full support Π in the case of binary states and signals under the “no fake data” assumption, and recall that a posterior q in this setting can be thought of as lying in $[0, 1]$. For $q > 1/2$, the support of $\Pi|q$ consists of those points (p_{00}, p_{11}) that satisfy $\frac{p_{11}}{p_{11} + 1 - p_{00}} = q$; one can show that this is the line through $(1, 0)$ and (q, q) (not including the end point at $(1, 0)$). [Figure 4](#) illustrates this. Note, however, that $\Pi|q$ is not the distribution of Π conditioned on lying in this line segment. Instead, $\Pi|q$ must weight each π in this line segment by the probability that this π induces posterior q for the source-aware learner; for example, when $q > 1/2$, distribution $\Pi|q$ must weight each (p_{00}, p_{11}) by $\frac{1}{2}(1 - p_{00}) + \frac{1}{2}p_{11}$.

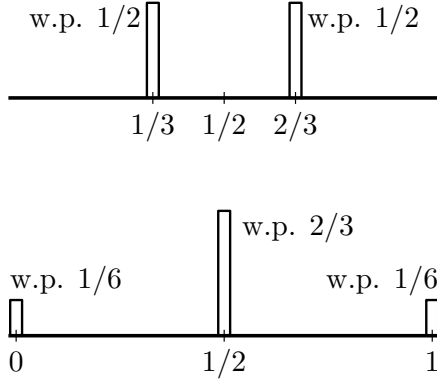


Figure 3: Example of distribution of posteriors induced by $B(\Pi)$ and $A(\Pi)$

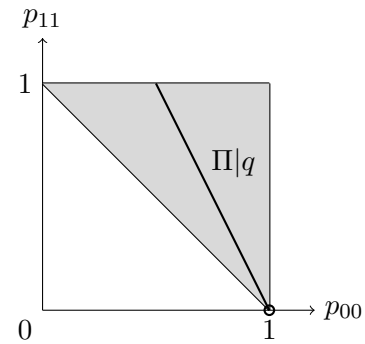


Figure 4: $\Pi|q$ in binary signaling schemes

7 The Value of Many Sources in Binary Settings

7.1 Over-provisioning Theorem

Proposition 4.2 formalizes the fact that, for a fixed distribution of signaling schemes Π , source-aware learning, $A(\Pi)$, is more informative than source-blind learning, $B(\Pi)$. However, our primary interest lies in quantifying *how much* more informative $A(\Pi)$ is than $B(\Pi)$. To formalize this, we make use of the dominance ratio P/Q of [Mu et al. \[2021\]](#) (recalled in [Section 3.1](#)). We study $A(\Pi)/B(\Pi)$, representing “how many times” $A(\Pi)$ is more informative than $B(\Pi)$ in large samples. We put an analytical upper bound on the value of $A(\Pi)/B(\Pi)$ in terms of the mean signal $\bar{\pi} = \mathbb{E}[\Pi]$.

In this section, we focus on binary states and the *two-signal, no-fake-data* domain of signaling schemes. [Mu et al. \[2021\]](#) restricts attention to binary states for technical reasons and defines dominance ratio in this environment. Moreover, restricting to the domain $\mathcal{P}_{\triangleleft}$ gives us a natural approach to putting a meaningful upper bound on the value of sources. In [Section 7.2](#), we discuss the challenges to upper bounding the value of sources with finite signal realizations or without the “no fake data” assumption.

Our approach is to use [Theorem 5.1](#) to find the distribution of signaling schemes Π on which source-aware learning, when compared to source-blind learning, has the largest dominance ratio. As stated in [Theorem 5.1](#), $A(\Pi')$ Blackwell dominates $A(\Pi)$ as long as Π' is a mean-preserving spread of Π . Since the domain of signaling schemes $\mathcal{P}_{\triangleleft}$ is convex, the distribution supported only on the vertices of the triangle is a mean-preserving spread to any distribution over $\mathcal{P}_{\triangleleft}$. This is formalized in the following lemma:

Lemma 7.1. *For any $\Pi \in \Delta(\mathcal{P}_{\triangleleft})$, there exists a unique Π^* with support only on the vertices of $\mathcal{P}_{\triangleleft}$ such that Π^* is a mean-preserving spread of Π .*

Proof. Consider a spread function s such that, for each point $(x, y) \in \mathcal{P}_{\triangleleft} \subseteq \mathbb{R}^2$, the distribution $s(x, y)$ is supported only on the vertices $(1, 1)$, $(0, 1)$ and $(1, 0)$. For any (x, y) , it is straightforward to see that there exists a unique distribution $s(x, y)$ on these vertices such that $\mathbb{E}[s(x, y)] = (x, y)$. In particular, $s(x, y)$ is the distribution over $(1, 1)$, $(0, 1)$ and $(1, 0)$ with probability $(x + y - 1)$, $(1 - x)$, and $(1 - y)$, respectively. Hence, distribution Π^* arising from applying $s(\cdot)$ is unique as well. \square

[Lemma 7.1](#) establishes that for any Π , there is a unique Π^* supported on the vertices that is a mean-preserving spread of Π . Observe that a Π^* is uniquely identified by its mean, which gives the weights on the vertices. Therefore, any distribution Π with the same mean signal $\mathbb{E}[\Pi] = \bar{\pi}$ will have the same Π^* with $\mathbb{E}[\Pi^*] = \bar{\pi}$ as their mean-preserving spread. With this observation, we provide an upper bound on the dominance ratio of any distribution Π in terms of its mean signal $\bar{\pi} = \mathbb{E}[\Pi]$.

Lemma 7.2. *For any $\bar{\pi} \in \mathcal{P}_{\triangleleft}$, let Π^* denote the distribution supported on the vertices of $\mathcal{P}_{\triangleleft}$ with $\mathbb{E}[\Pi^*] = \bar{\pi}$. Then, for any other distribution Π over $\mathcal{P}_{\triangleleft}$, we have $A(\Pi)/B(\Pi) \leq A(\Pi^*)/B(\Pi^*)$.*

Proof. Recall the definition of the dominance ratio of $A(\Pi)$ over $B(\Pi)$:

$$A(\Pi)/B(\Pi) = \sup \left\{ \frac{m}{n} : A(\Pi)^{\otimes n} \succeq B(\Pi)^{\otimes m} \right\}$$

Since [Lemma 7.1](#) establishes that Π^* is the unique mean-preserving spread of any Π with the same mean, [Theorem 5.1](#) implies that $A(\Pi^*)$ Blackwell dominates $A(\Pi)$. This clearly implies that $A(\Pi^*)^{\otimes n}$ Blackwell dominates $A(\Pi)^{\otimes n}$ when repeatedly drawing from the same experiment n times.

Furthermore, we know $B(\Pi) = B(\Pi^*)$ from [Proposition 4.1](#) since the distributions have the same mean. Now, for any m, n such that $A(\Pi)^{\otimes n} \succeq B(\Pi)^{\otimes m}$ holds, we have that

$$A(\Pi^*)^{\otimes n} \succeq A(\Pi)^{\otimes n} \succeq B(\Pi)^{\otimes m} = B(\Pi^*)^{\otimes m}$$

i.e., $A(\Pi^*)^{\otimes n} \succeq B(\Pi^*)^{\otimes m}$ also holds. We thus have that $A(\Pi^*)/B(\Pi^*) \geq A(\Pi)/B(\Pi)$ by definition. \square

Now we are ready to state our main theorem, which provides an upper bound on the value of sources by analyzing the worst-case ratio obtained by Π^* .

Theorem 7.3 (Over-provisioning). *Consider two-signal, no-fake-data signaling schemes. For any distribution of signaling schemes $\Pi \in \Delta(\mathcal{P}_{\mathfrak{A}})$ with the average signal $\mathbb{E}[\Pi] = (x, y)$, $A(\Pi)/B(\Pi)$ is at most $\frac{\log(2-x-y)}{\log(\sqrt{x(1-y)} + \sqrt{y(1-x)})}$.*

Furthermore, if the average signal is ε -away from being completely uninformative, i.e., $x + y \geq 1 + \varepsilon$ for some $\varepsilon > 0$, then $A(\Pi)/B(\Pi)$ is at most $\frac{2\log(1-\varepsilon)}{\log(1-\varepsilon^2)} = O(1/\varepsilon)$.

We use numerical simulations to check the approximate tightness of the derived bound. Recall in [Theorem 3.6](#), the dominance ratio is the infimum of a possibly non-convex function. Therefore, in [Figure 5](#), we numerically approximate the ratio for each (x, y) pair by discretizing the domain into a grid, evaluating the function at those grid points, and taking the minimum value as an approximation of the infimum. On the other hand, in [Figure 6](#), we directly evaluate the function $\frac{\log(2-x-y)}{\log(\sqrt{x(1-y)} + \sqrt{y(1-x)})}$ as in the theorem. We plot the contour lines of values obtained by both methods. The analytical bound is reasonably close to the simulated infimum, providing informal evidence that this bound approximates the true value.

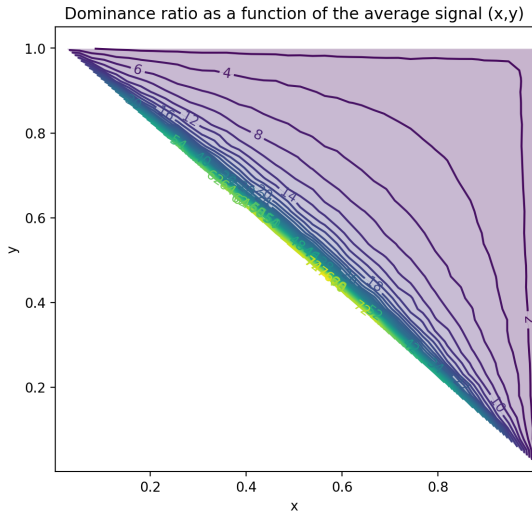


Figure 5: Simulated dominance ratio

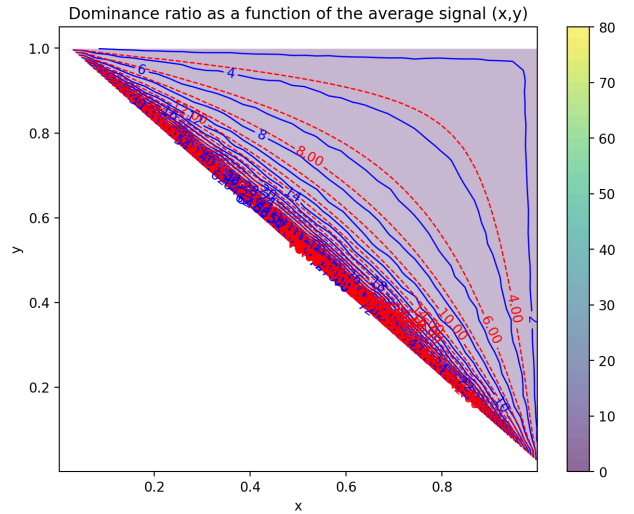


Figure 6: Analytical upper bound

Remark 7.4. The average signal quality is directly related to the maximal value one can obtain from knowing the sources. For distributions with an average signal that is relatively away from being uninformative, e.g. $(0.75, 0.75)$, the maximal dominance ratio is 8. On the other hand, for distributions with an average signal of $(0.51, 0.51)$, the maximal dominance ratio is 200. The intuition is that suppose an user only cares about the review made by one other person in the

world, then it makes more sense to look for the source as opposed to read summarizations of more data. This corresponds to the situation where the average review is close to being uninformative to this user.

Proof of Theorem 7.3. We use Lemma 7.2 and consider the dominance ratio for distributions supported on the vertices Π^* as an upper bound. We also use the fact that the infimum of the Rényi ratio over $t > 0$ is weakly smaller than the ratio evaluated at any t . Specifically, we use $t = \frac{1}{2}$ as it simplifies the expression while providing a reasonably tight upper bound.

$$\frac{A(\Pi)}{B(\Pi)} \leq \frac{A(\Pi^*)}{B(\Pi^*)} = \inf_{\theta \in \{0,1\}, t > 0} \frac{R_{A(\Pi^*)}^\theta(t)}{R_{B(\Pi^*)}^\theta(t)} \leq \frac{R_{A(\Pi^*)}^\theta(\frac{1}{2})}{R_{B(\Pi^*)}^\theta(\frac{1}{2})}$$

By Proposition 4.1, $B(\Pi^*)$ is equivalent to the (binary) mean signal (x, y) . In binary signaling scheme, the Rényi divergences simplifies to:

$$R_P^\theta(t) = \frac{1}{t-1} \log \left[\sum_{\omega=0}^1 \left(\frac{p_{\theta\omega}}{p_{(1-\theta)\omega}} \right)^{t-1} \cdot p_{\theta\omega} \right]$$

When $t = \frac{1}{2}$, it further simplifies to:

$$R_P^\theta(1/2) = -2 \log \left[\sum_{\omega=0}^1 \sqrt{p_{\theta\omega} p_{(1-\theta)\omega}} \right]$$

Substituting $p_{00} = x$ and $p_{11} = y$:

$$R_{(x,y)}^{\theta=0}(1/2) = R_{(x,y)}^{\theta=1}(1/2) = -2 \log \left[\sqrt{x(1-y)} + \sqrt{(1-x)y} \right]$$

For $A(\Pi^*)$, on the other hand, the enriched signal space of $\mathcal{P}_\mathbb{N} \times \Omega$ is supported on $\{(1,0), (0,1), (1,1)\} \times \{0,1\}$. The probability measure over the enriched signal space under state θ is

$$P_\theta(p, q, \omega) = f_{\Pi^*}(p, q) \cdot p_{\theta\omega}$$

for $(p, q) \in \{(1,0), (0,1), (1,1)\}$ and $\theta, \omega \in \{0,1\}$, where f_{Π^*} is the probability mass function supported on the vertices $(1,0)$, $(0,1)$, and $(1,1)$. The Rényi divergence for $A(\Pi^*)$ is calculated as follows, where we sum over $(p, q) \in \{(1,0), (0,1), (1,1)\}$.

$$\begin{aligned} R_{A(\Pi^*)}^{\theta=0}(t) &= \frac{1}{t-1} \log \left[\sum_{p,q} \left(\frac{p^t}{(1-q)^{t-1}} + \frac{(1-p)^t}{q^{t-1}} \right) f_{\Pi^*}(p, q) \right] \\ R_{A(\Pi^*)}^{\theta=1}(t) &= \frac{1}{t-1} \log \left[\sum_{p,q} \left(\frac{(1-q)^t}{p^{t-1}} + \frac{q^t}{(1-p)^{t-1}} \right) f_{\Pi^*}(p, q) \right] \end{aligned}$$

When $t = \frac{1}{2}$, this simplifies to:

$$R_{A(\Pi^*)}^{\theta=0}(1/2) = R_{A(\Pi^*)}^{\theta=1}(1/2) = -2 \log \left[\sum_{p,q} \left(\sqrt{p(1-q)} + \sqrt{(1-p)q} \right) f_{\Pi^*}(p, q) \right]$$

When Π^* has mean (x, y) , the weight on the supporting signaling schemes for $A(\Pi^*)$ are:

$f_{\Pi^*}(1, 1) = x + y - 1$, $f_{\Pi^*}(0, 1) = 1 - x$, and $f_{\Pi^*}(1, 0) = 1 - y$. Substituting to the expression, we obtain:

$$R_{A(\Pi^*)}^\theta(1/2) = -2 \log \left[\sqrt{0} f_{\Pi}(1, 1) + \sqrt{1} f_{\Pi}(1, 0) + \sqrt{1} f_{\Pi}(0, 1) \right] = -2 \log(2 - x - y)$$

Therefore, for any signal distribution Π with mean signal $(x, y) \in \mathcal{P}_{\nabla}$, $A(\Pi)/B(\Pi)$ can be upper bounded by the ratio of Π^* with the same mean:

$$\frac{A(\Pi)}{B(\Pi)} \leq \frac{R_{A(\Pi^*)}^\theta(1/2)}{R_{B(\Pi^*)}^\theta(1/2)} = \frac{\log(2 - x - y)}{\log(\sqrt{x(1-y)} + \sqrt{y(1-x)})} \quad (1)$$

Furthermore, if we put restrictions on the mean of Π , e.g., requiring that $x + y \geq 1 + \varepsilon$, we can get more interpretable bounds in terms of ε .

First, let us consider for the (x, y) pairs on the line $x + y = 1 + \varepsilon'$ for any ε' . We want to find which (x, y) pair on the line attains the highest dominance ratio. Substituting $y = 1 + \varepsilon' - x$ to [Equation 1](#), we obtain

$$\frac{A(\Pi)}{B(\Pi)} \leq \frac{\log(1 - \varepsilon')}{\log(\sqrt{x(x - \varepsilon')} + \sqrt{(1 - x + \varepsilon')(1 - x)})}$$

To bound the right hand side, we show that when $\varepsilon' \in (0, 1)$ and $x \in [\varepsilon', 1]$, the logarithm attains its maximum at $x = \frac{1+\varepsilon'}{2}$. We start by showing that the argument inside the logarithm is log-concave in x . Since every concave function that is nonnegative on its domain is log-concave, we need to verify the concavity of argument. The second order derivative of $\sqrt{x(x - \varepsilon')}$ is $\frac{-(\varepsilon')^2}{4(x^2 - x\varepsilon')^{3/2}} < 0$ when $\varepsilon' \in (0, 1)$ and $x \in [\varepsilon', 1]$. By a symmetric argument, the second order derivative of $\sqrt{(1 - x + \varepsilon')(1 - x)}$ is also strictly negative in the domain. The sum of two strictly concave functions is strictly concave. Therefore, the denominator is strictly concave in x when $x \in [\varepsilon', 1]$ and attains the maximum at $x = \frac{1+\varepsilon'}{2}$. Since the numerator $\log(1 - \varepsilon') < 0$, the fraction also attains maximum at $x = \frac{1+\varepsilon'}{2}$. This means that the dominance ratio for (x, y) along the line $x + y = 1 + \varepsilon'$ is upper bounded by the dominance ratio at $x = y = \frac{1+\varepsilon'}{2}$.

Substituting in $x = y = \frac{1+\varepsilon'}{2}$ to [Equation 1](#) we obtain an upper bound in terms of ε' :

$$\frac{A(\Pi)}{B(\Pi)} \leq \frac{R_{A(\Pi^*)}^\theta(1/2)}{R_{B(\Pi^*)}^\theta(1/2)} = \frac{2 \log(1 - \varepsilon')}{\log(1 - (\varepsilon')^2)} \quad (2)$$

We can rewrite the function as $\frac{2}{1+g(\varepsilon')}$ where $g(\varepsilon') = \frac{\log(1+\varepsilon')}{\log(1-\varepsilon')}$. By quotient rule, the numerator of the first-order derivative of g is $\frac{\log(1-\varepsilon')}{1+\varepsilon'} + \frac{\log(1+\varepsilon')}{1-\varepsilon'}$, which is strictly positive when $\varepsilon' \in (0, 1)$. Since $\frac{2}{1+g(\varepsilon')}$ is decreasing in $g(\varepsilon')$, we have proved that [Equation 2](#) is decreasing in ε' when $\varepsilon' \in (0, 1)$.

Therefore, among the (x, y) pairs such that $x + y \geq 1 + \varepsilon$, those with $x + y = 1 + \varepsilon$ attains the largest upper bound. Replacing ε' with ε , we have that for any distribution of signaling schemes Π with mean signal (x, y) such that $x + y \geq 1 + \varepsilon$,

$$\frac{A(\Pi)}{B(\Pi)} \leq \frac{2 \log(1 - \varepsilon)}{\log(1 - \varepsilon^2)} \approx \frac{2}{\varepsilon} \quad (3)$$

Using the approximation $\log(1 + x) \approx x$ when the absolute value of x is small, this fraction is approximately $\frac{2}{\varepsilon}$ when ε is close to zero. \square

7.2 Challenges in Extending to General Signaling Schemes

In the previous section, our restriction to the *two-signal, no-fake-data* signaling schemes provided a particularly natural path forward. Under both of these conditions, there is a unique mean-preserving spread of any arbitrary distribution $\Pi \in \Delta(\mathcal{P}_{\mathcal{N}})$ to the vertices of the space $\mathcal{P}_{\mathcal{N}}$. This property let us define an Π^* as the unique mean-preserving spread of Π that is supported on the vertices. The Rényi divergences of $A(\Pi^*)$ and $B(\Pi^*)$ have good properties and can be upper bounded by a simple function form.

However, when we drop either one of the assumption, i.e., if we have signaling schemes with more than two realizations, or if the domain of binary signaling schemes has no restriction, our approach runs into a problem: we lose the uniqueness of the mean-preserving spread supported on the vertices. Hence, it is unclear how to approach getting a natural upper bound as [Section 7.1](#) above. We illustrate the challenge in a few examples.

First, consider the space of binary signaling schemes without the “no-fake data” assumption. Observe that in this domain, there are distributions Π such that $A(\Pi)$ is informative, but $B(\Pi)$ is completely uninformative. For example, when Π is a uniform distribution over the bottom left and top right corners $(0, 0)$ and $(1, 1)$, then $A(\Pi)$ always reveals the state, but $B(\Pi)$ is equivalent to $(1/2, 1/2)$ and is completely uninformative. In particular, the dominance ratio $A(\Pi)/B(\Pi)$ is unbounded.

Moreover, even when we restrict attention to when the mean signal $\bar{\pi} = \mathbb{E}[\Pi]$ is bounded away from uninformativeness, there is no longer a unique mean-preserving spread from Π to the vertices of the domain. For example, the point $(3/4, 3/4)$ could be spread to a distribution Π_a^* supported on vertices of $\mathcal{P}_{\mathcal{N}}$, as in [Section 7.1](#), or it could be spread to the distribution Π_b^* that puts $1/4$ weight on $(0, 0)$ and $3/4$ weight on $(1, 1)$; the latter distribution Π_b^* would (similar to the previous paragraph) have infinite dominance ratio and thus give no useful upper bound on the dominance ratio of Π .

Analogous problems arise when considering our problem for a larger number of states or signals. To illustrate, consider the space of signaling schemes with two states and three signal realizations. Since $|\Omega| = 3$ and $\mathcal{P} = \Delta(\Omega) \times \Delta(\Omega)$ is naturally viewed as a subspace of \mathbb{R}^6 (with dimension 4). There are 9 vertices of this polytope, specifically:

$$\begin{aligned} \pi_1 &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} & \pi_2 &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} & \pi_3 &= \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \\ \pi_4 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} & \pi_5 &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} & \pi_6 &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \\ \pi_7 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \pi_8 &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \pi_9 &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

Each vertices represents a signaling scheme where the columns denotes the space of signal realizations, the first row denotes $P_0 \in \Delta(\Omega)$, and the second row denotes $P_1 \in \Delta(\Omega)$. These vertices are either perfectly uninformative for all realizations (when $P_0 = P_1$), or perfectly informative (when $P_0 \neq P_1$). Consider a completely uninformative signaling scheme

$$\bar{\pi} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

This signaling scheme can be written as $\frac{1}{3}\pi_1 + \frac{1}{3}\pi_5 + \frac{1}{3}\pi_9$, which are three uninformative signaling schemes. However, it can also be written as a $\frac{1}{9}$ equal weighting of *all* the vertices, which contains some informative signals. More broadly, beyond two signals and states, it is unclear how to impose any sort of “no-fake-data” restriction which makes the above representations unique, as the signal

realizations need not have any interpretation with regard to the states.

The above discussion shows that, without any restrictions on the space of signaling schemes, it is unclear how to get useful bounds on the dominance ratio due to non-uniqueness and possibly unbounded ratios. We leave further investigations in the space of general signaling schemes to future work.

8 Conclusion

In today’s information-rich environment, economic decision-making increasingly relies on aggregating large volumes of data from diverse sources. Whether evaluating product reviews, news articles, or medical recommendations, understanding who generates the information is crucial to judging its reliability and relevance. Yet, modern technologies that facilitate rapid data aggregation often obscure the origins of each data point, raising important questions about the trade-offs involved.

To address these concerns, we develop a formal framework that contrasts source-aware with source-blind learning. We extend the classic Blackwell model of learning by repeatedly drawing from a distribution of signaling schemes and either reveal (source-aware) or conceal (source-blind) the realized signaling scheme to the learner.

Our analysis reveals source-aware learners benefit from increased heterogeneity among data sources whereas source-blind learning is mean-based. Focusing on a tractable binary model, our main result demonstrates that if the average signaling scheme is at least an ε -distance from being completely uninformative, then a source-blind learner requires at most $O(1/\varepsilon)$ times as much data to match the informativeness of a source-aware learner, as measured by Blackwell dominance in large samples.

Our findings invite further exploration into more complex settings beyond the binary framework, as well as empirical investigations into real-world applications where the trade-off between data volume and source information is paramount.

References

- D. Acemoglu, V. Chernozhukov, and M. Wold. Fragility of asymptotic agreement under bayesian learning. *Theoretical Economics*, 11(1):187–225, 2016. 2
- Y. Azrieli. Comment on “the law of large demand for information”. *Econometrica*, 82(1):415–423, 2014. 2, 6
- A. V. Banerjee. A simple model of herd behavior. *The quarterly journal of economics*, 107(3):797–817, 1992. 2
- D. Beauchêne, J. Li, and M. Li. Ambiguous persuasion. *Journal of Economic Theory*, 179:312–365, 2019. 3
- D. Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272, 1953. 2, 5
- D. Blackwell et al. Comparison of experiments. In *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, volume 1, page 26, 1951. 2
- A. W. Bloedel and W. Zhong. The cost of optimally-acquired information. *Unpublished Manuscript*, November, 2020. 2

- B. Brooks, A. Frankel, and E. Kamenica. Information hierarchies. *Econometrica*, 90(5):2187–2214, 2022. 2
- B. Brooks, A. Frankel, and E. Kamenica. Comparisons of signals. *American Economic Review*, 114(9):2981–3006, 2024. 2
- J. I. Bulow and P. D. Klemperer. Auctions vs. negotiations, 1994. 3
- X. Cheng. Ambiguous persuasion: An ex-ante formulation, 2023. URL <https://arxiv.org/abs/2010.05376>. 3
- X. Cheng, P. Klibanoff, S. Mukerji, and L. Renou. Persuasion with ambiguous communication. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 1201–1202, 2024. 3
- M. W. Cripps, J. C. Ely, G. J. Mailath, and L. Samuelson. Common learning. *Econometrica*, 76(4):909–933, 2008. 2
- A. Dasgupta, J. Hopcroft, J. Kleinberg, and M. Sandler. On learning mixtures of heavy-tailed distributions. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05)*, pages 491–500. IEEE, 2005. 3
- S. Dasgupta. Learning mixtures of gaussians. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 634–644. IEEE, 1999. 3
- M. Feldman, N. Immorlica, B. Lucier, and S. M. Weinberg. Reaching consensus via non-bayesian asynchronous learning in social networks. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, page 192, 2014. 3
- D. Gale and S. Kariv. Bayesian learning in social networks. *Games and economic behavior*, 45(2):329–346, 2003. 3
- W. Huang. Learning about informativeness. Working Paper. Abstract in EC’24, 2024. 2
- A. Iakovlev and A. Liang. The value of context: Human versus black box evaluators. *arXiv preprint arXiv:2402.11157*, 2024. 3
- C. Kellner and M. T. Le Quement. Modes of ambiguous communication. *Games and Economic Behavior*, 104:271–292, 2017. 3
- C. Kellner and M. T. Le Quement. Endogenous ambiguity in cheap talk. *Journal of Economic Theory*, 173:1–17, 2018. 3
- A. Kosenko. Beyond the blackwell order in dichotomies. *Available at SSRN 4649851*, 2024. 2
- G. Moscarini and L. Smith. The law of large demand for information. *Econometrica*, 70(6):2351–2366, 2002. 2, 6
- E. Mossel, J. Neeman, and O. Tamuz. Majority dynamics and aggregation of information in social networks. *Autonomous Agents and Multi-Agent Systems*, 28:408–429, 2014. 3
- E. Mossel, A. Sly, and O. Tamuz. Strategic learning and the topology of social networks. *Econometrica*, 83(5):1755–1794, 2015. 3

- E. Mossel, N. Olsman, and O. Tamuz. Efficient bayesian learning in social networks with gaussian estimators. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 425–432. IEEE, 2016. [3](#)
- X. Mu, L. Pomatto, P. Strack, and O. Tamuz. From blackwell dominance in large samples to rényi divergences and back again. *Econometrica*, 89(1):475–506, 2021. [2](#), [6](#), [11](#)
- L. Pomatto, P. Strack, and O. Tamuz. The cost of information: The case of constant marginal costs. *American Economic Review*, 113(5):1360–1393, 2023. [2](#)
- A. Rényi. On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics*, volume 4, pages 547–562. University of California Press, 1961. [6](#)
- T. Roughgarden and É. Tardos. How bad is selfish routing? *Journal of the ACM (JACM)*, 49(2): 236–259, 2002. [3](#)
- M. Whitmeyer. Blackwell-monotone updating rules, 2024. URL <https://arxiv.org/abs/2302.13956>. [2](#)