Robust Aggregation of Correlated Information*

Henrique de Oliveira† Yuhta Ishii‡ Xiao Lin§

August 16, 2025

Abstract

Decisions are typically made by drawing on multiple sources of information, yet the correlation among these sources is often hard to assess. We study robustly optimal strategies — those that maximize the worst-case payoff guarantee across all correlation structures. With two states and two actions, the robustly optimal strategy is strikingly simple: ignore all but the best information source. With more actions, the analysis can be extended by decomposing the problem into binary-action subproblems. With more states, we show a more sophisticated characterization of robustly optimal strategies via concavification. Our results offer a new rationale for why ignoring information can be optimal.

1 Introduction

Out of 21 U.S. general election polls showcased by the website RealClearPolitics on Nov. 7, 2016, only two predicted a Trump victory. Poll aggregators used a variety of models to predict Trump's probability of losing, with estimates ranging from 70% to 99%. The lowest estimate of 70% was reached by Nate Silver's FiveThirtyEight, whose model emphasized the possibility of correlation between state polls. The highest estimate of 99% was reached by Sam Wang's Princeton Election Consortium, assuming that errors were unlikely to be correlated. The extreme predictions led to widespread belief that the election was a done deal, setting the stage for a stunning political upset that few saw coming.[1]

*This paper has previously been circulated under the title "Robust Merging of Information." We are grateful to Nageeb Ali, Alex Bloedel, Kfir Eliaz, Marcos Fernandes, Alex Frankel, Marc Henry, Nicole Immorlica, Asen Kochov, Jiangtao Li, Elliot Lipnowski, Ce Liu, George Mailath, Pietro Ortoleva, Collin Raymond, Fedor Sandomirskiy, Shamim Sinnar, Ran Spiegler, Rui Tang, Rakesh Vohra, Leeat Yariv, and participants at various conferences and seminars for valuable comments. We also thank Tiago Botelho for his excellent research assistance.

†São Paulo School of Economics - FGV. Email: henrique.oliveira@fgv.br

‡Pennsylvania State University. Email: yxi5014@psu.edu

§University of Pennsylvania. Email: xiaolin7@sas.upenn.edu

[1] Poll count from Time (2017); probability range summarized by Westwood, Messing, and Lelkes (2020); methods underlying the two estimates are described in Silver (2016) and Wang (2016).

Most decisions in life, from the mundane to the important, rely on multiple correlated information sources. For example, treatment decisions can be made by consulting multiple doctors, who base their recommendations on the same test results; investment plans are often informed by the advice of many financial analysts, who may have incentives to provide contrarian opinions. Yet, assessing how information sources are correlated is hard, as the complexity increases exponentially with their number — a phenomenon known as the curse of dimensionality. Misinterpreting correlations can lead to flawed inferences and suboptimal decisions. Therefore, an agent may look to make decisions that are robust against potential misspecification of correlation.

In this paper, we study agents who understand each information source individually but lack knowledge about their correlations, and look for a strategy that yields the best payoff guarantee against all possible correlation structures. Our main results characterize how agents optimally disregard certain sources of information to hedge against this uncertainty.

As an example, consider the following scenario: The Centers for Disease Control (CDC) is setting guidelines for administering a new Covid treatment to a patient population, who has equal prior probabilities of having either Covid or the Flu. The treatment is designed for Covid, so it is beneficial for Covid patients, but causes only side effects for those with the Flu. The payoff matrix is given in Table 1, where the payoffs from no treatment are normalized to zero.

	Treatment	No Treatment
Covid	30	0
Flu	-20	0

Table 1: Payoffs from the Treatment

Since patients with different diseases may develop different symptoms with different probabilities, these symptoms can serve as informative signals to guide treatment decisions. Suppose there are two well-understood studies: one describes the relationship between the diseases (Covid/Flu) and the Cough symptom; the other describes the relationship between the diseases and the Fever symptom. These relationships, represented as Blackwell experiments, are shown in Table 2, where "+" denotes the presence of a symptom and "−" denotes its absence.

	+	−
Covid	0.9	0.1
Flu	0.5	0.5

Cough

	+	−
Covid	0.5	0.5
Flu	0.1	0.9

Fever

Table 2: Known Relationships between Diseases and Symptoms

However, no studies have jointly examined both Cough and Fever symptoms, so the CDC does not know their joint probabilities. Without this knowledge, a strategy that best responds to a misspecified correlation structure can perform poorly, so the goal is to design a treatment guideline that utilizes the available information in a way that is robust to all possible correlations.

A simple strategy that guards against the unknown correlation is to base the treatment decision on only one symptom. If using only the Cough symptom, the treatment should be administered if and only if the patient has a positive Cough symptom. This strategy guarantees a value of $\frac{1}{2}[0.9 \times 30 + 0.5 \times (-20)] = 8.5$ regardless of the correlation. Similarly, the CDC could also base the treatment decision solely on the Fever symptom, which guarantees a value of $\frac{1}{2}[0.5 \times 30 + 0.1 \times (-20)] = 6.5$. Since the strategy using the Cough symptom guarantees a higher value, we call it a *best-source strategy*, which selects a *single* information source — the best one when considered individually — and best responds to it.

While the best-source strategy has the virtue of being simple, it completely forfeits the potential benefits from observing multiple information sources. Could the CDC do better by using a more sophisticated treatment strategy that incorporates both symptoms? Theorem 1 shows that the answer is no. In fact, it shows that a best-source strategy is always robustly optimal in any decision problem involving two states and two actions. Moreover, whenever the best information source is unique, e.g. the Cough symptom in this example, the best-source strategy is the unique robustly optimal strategy.

With more than two actions, best-source strategies are no longer always optimal, and robustly optimal strategies will typically use multiple information sources. Nevertheless, we will see that best-source strategies serve as the basic building blocks from which robustly optimal strategies are constructed. To illustrate, let us revise the example and suppose now there are two treatments: one is the previous treatment, designed for Covid, and the other is an additional treatment, designed for the Flu. The payoff from each treatment is given in Table 3; if the patient receives both treatments, the total payoff is the sum of the two.

	T_1	N_1
Covid	30	0
Flu	-20	0

Treatment 1

	T_2	N_2
Covid	-20	0
Flu	30	0

Treatment 2

Table 3: Payoffs from Two Treatments

The CDC now chooses among four actions, in the form of $\{T_1, N_1\} \times \{T_2, N_2\}$, specifying whether to administer each of the treatments. Again, a simple strategy that is not vulnerable to misspecified correlations is to base the treatment decision on only one symptom. It can be easily checked that using either the Cough or the Fever symptom alone guarantees a value of $8.5 + 6.5 = 15$. However, the CDC can do better by basing the decision of Treatment 1 on the Cough symptom and Treatment 2 on the Fever symptom, as described in Table 4. This strategy, which uses the respective best information source for each treatment, guarantees a value of $8.5 + 8.5 = 17$ regardless of the actual correlations between the information sources.

	Fever_+	Fever_-
Cough_+	$T_1 + T_2$	T_1
Cough_-	T_2	No Treatment

Table 4: Using Information from Both Symptoms

A key property of the decision problem above is the additive separability of payoffs across the two treatments. Indeed, for any decision problem consisting of a collection of binary-state binary-action subproblems whose utilities are summed, which we call a *separable problem*, we show that a robustly optimal strategy is to use the best-source strategy for each subproblem separately.

The separability property may seem rather restrictive, but in fact, *every* binary-state decision problem can be written as a separable problem via what we call the binary decomposition of a decision problem. Building on this idea, Theorem 2 provides a general construction of robustly optimal strategies for all finite-action, binary-state decision problems.

With three or more states, we adopt a different approach, analyzing the interim value function of the decision problem. The space of interim value functions is generally infinite dimensional, but it can be reduced to a finite-dimensional value vector by focusing on the interim values at a finite set of beliefs, which we call extremal beliefs. The robustly optimal value, as a function of these value vectors, satisfies two properties. First, it must be concave. Second, it must be greater than the value derived from the best-source strategy. Theorem 3 establishes that it is, in fact, the smallest function satisfying these two properties — the concavification of the value derived from the best source strategy.

A few behavioral implications follow immediately from our characterizations. First, we derive a bound on the number of sources needed for robustly optimal strategies, which equals the number of non-degenerate extremal beliefs of the decision problem.[2] The bound depends only on the decision problem, not on the number or specifics of the information sources. This means that when the number of information sources grows large, the fraction of information sources a decision maker pays attention to converges to zero. Second, we characterize which information sources might be used and which can be safely ignored. Specifically, the sources that may be used are those that serve as the best standalone source for some decision problems. This highlights the advantage of gathering information from specialized information sources. Lastly, when the state is binary, our results also imply an intuitive property of robust aggregation of action recommendations — the unanimity rule; that is, to always take an action that is unanimously recommended by all information sources. Although unanimity may seem natural, a well-known puzzle is that it fails to hold under conditional independence (see, e.g., Sobel, 2014).

[2]This bound simplifies to $|A| - 1$ with two states.

Our results show that the agent tends to ignore some freely available information when facing ambiguity on correlations. Information neglect is well-documented, with existing explanations often attributing this behavior to hidden costs or psychological distortions (see Handel and Schwartzstein (2018) for a detailed discussion). Our results offer a different rationale: ignoring certain information can lead to more robust decisions when there is ambiguity about the correlations among various information sources. This explanation has distinct counterfactual implications. For instance, an agent who finds it costly to acquire or process information would become more informed as stakes are raised, but one who is concerned with correlation robustness would not react to such an incentive.

The rest of the paper is organized as follows: Section 2 introduces the formal model. Section 3 establishes preliminary results that will be useful throughout the paper. Sections 4 and 5 consider the binary-state and general-state environments, respectively. Section 6 summarizes the behavioral implications of robustly optimal strategies. Section 7 discusses extensions. Section 8 concludes. The remainder of this introduction situates our contribution within the broader literature.

Related Literature: Our paper studies robust decision making under uncertain correlations among information sources. The practice of finding robust strategies traces back at least to Wald (1950). The worst-case approach we adopt is in line with the literature on ambiguity aversion (Gilboa and Schmeidler, 1989). In particular, a recent experiment by Epstein and Halevy (2019) documents aversion to ambiguity on correlation structures.

Our approach to modeling information aggregation is closely related to the robust forecast aggregation literature, which seeks to combine multiple forecasts into a single prediction without detailed knowledge of the underlying information structure.[3] Arieli, Babichenko, and Smorodinsky (2018) first proposed an adversarial framework for combining forecasts, and considered various types of ambiguity, such as when one information source is Blackwell more informative than the other, but the agent does not know which. They study a specific decision problem where the agent aims to minimize the quadratic loss to the true state. By contrast, we focus solely on ambiguity in the correlation structure and consider general decision problems. Our ambiguity set is also closely related to that in Levy and Razin (2020), who consider both the correlation among signals and the correlation across different dimensions of the a multi-dimensional state space. They adopt an interim approach, where ambiguity arises after the signals have been realized. By contrast, our approach is ex-ante, where the worst-case correlation does not vary with signal realizations.[4]

[3]This literature often assumes that only forecasts — experts' beliefs about the state — are observable, instead of the raw information informing those beliefs, as in our model. See Section 7.3 for further discussion of this assumption.

[4]The ex-ante approach may be more natural when the decision maker designs a guideline that must be established upfront and apply broadly to an organization or a population, while the interim approach may be

The agent in our model has a maxmin objective — evaluating each strategy by its worst-case payoff across all correlation structures. Arieli, Babichenko, Talgam-Cohen, and Zabarnyi (2023) adopts a complementary approach, minmax regret, where the agent concerns the largest opportunity loss relative to what she could have achieved if she knew the correlation and best responded accordingly. They show that when the marginal experiments are symmetric, following a single random information source is robustly optimal under both robustness paradigms.

Our analysis involves understanding and characterizing the least informative joint experiments with given marginals. These least informative joint experiments capture an extreme form of substitution between information sources, a notion introduced by Börgers, Hernando-Veciana, and Krähmer (2013). Cheng and Börgers (2024) further explore the relationship between the joint informativeness of experts' recommendations and their chance of disagreement.

Several studies have investigated learning from multiple information sources with known correlations. Liang and Mu (2020) examine a social learning setting where agents' information is complementary. Ichihashi (2021) looks at how a firm purchases data from consumers with potentially correlated information sources. Liang, Mu, and Syrgkanis (2022) study an agent's optimal dynamic allocation of attention to multiple correlated information sources. Finally, Brooks, Frankel, and Kamenica (2024) explores the comparison of experiments with known correlations and characterize their ranking that is robust to any additional information.

Robustness to correlations has also been studied in other contexts, such as mechanism design. Carroll (2017) studies a multi-dimensional screening problem, where the principal knows only the marginals of the agent's type distribution, and designs a mechanism that is robust to all possible correlation structures. He and Li (2020) and Zhang (2021) study an auctioneer's robust design problem when selling an indivisible good, concerning the correlation of values among different agents.

2 Model

An agent faces a decision problem $\Gamma = (\Theta, \mu_0, A, \rho)$, with a finite state space Θ, a prior $\mu_0 \in \Delta\Theta$, a finite action space A, and a utility function $\rho : \Theta \times A \to \mathbb{R}$. To simplify notation, we define a prior-weighted utility function $u(\theta, a) = \mu_0(\theta)\rho(\theta, a)$, and will refer to the decision problem simply as (A, u). With a slight abuse of notation, for a mixed action $\alpha \in \Delta(A)$, we let $u(\theta, \alpha) = \sum_{a\in A} u(\theta, a)\alpha(a)$.

The agent has access to m information sources, denoted by $\{P_j\}_{j=1}^m$. Each source is a **marginal experiment**, $P_j : \Theta \to \Delta Y_j$, mapping each state to a distribution over some finite signal set Y_j. Let $\mathbf{Y} = Y_1 \times \cdots \times Y_m$ denote the set of all possible profiles of signal realizations, with typical element $\mathbf{y} = (y_1, \ldots, y_m)$. The agent can observe the signals from all marginal

better suited for a decision maker facing a specific, individualized problem where signals have already realized.

experiments, $\{P_j\}_{j=1}^m$, but does not know the joint. Thus, the agent conceives of the following set of **joint experiments**:

$$\mathcal{J}(P_1, ..., P_m) = \left\{ P : \Theta \to \Delta(\mathbf{Y}) : \sum_{y_{-j}} P(y_1, \ldots, y_m|\theta) = P_j(y_j|\theta) \text{ for all } \theta, j, y_j \right\}.$$

A strategy for the agent is a mapping, $\sigma : \mathbf{Y} \to \Delta(A)$, and the set of all strategies is denoted by Σ. The agent's problem is to maximize her expected payoff considering the worst possible joint experiment:

$$V(P_1, \ldots, P_m; (A, u)) := \max_{\sigma \in \Sigma} \min_{P \in \mathcal{J}(P_1, \ldots, P_m)} \sum_{\theta \in \Theta} \sum_{\mathbf{y} \in \mathbf{Y}} P(\mathbf{y}|\theta) u(\theta, \sigma(\mathbf{y})).$$

When there is no confusion about the relevant decision problem, we omit (A, u) from the argument of V. We call a solution to the problem a **robustly optimal** strategy.

If $m = 1$, the agent observes only a single experiment $P : \Theta \to \Delta(Y)$ and $V(P)$ is the classical value of a Blackwell experiment. In this case, a robustly optimal strategy is just an optimal strategy for a Bayesian agent.

3 Preliminaries

We begin with some groundwork for our main results. Section 3.1 represents decision problems as payoff polyhedrons that capture all feasible payoff vectors. We also introduce the dominance and equivalence relationship between decision problems — tools that will be instrumental in characterizing robustly optimal strategies in Theorems 2 and 3. Section 3.2 reviews the Blackwell order, drawing on a convenient Zonotope representation of Blackwell experiments and the existence of the Blackwell supremum in binary-state environments. Finally, Section 3.3 solves the dual problem — Nature's minmax problem, which characterizes the value function and plays a central role in our analysis.

3.1 Payoff Polyhedron

For any decision problem (A, u), the utility for a given action can be seen as a vector in $\mathbb{R}^\Theta$, denoted by $u(\cdot, a)$. Let

$$\mathcal{H}(A, u) = \text{co}\{u(\cdot, a) : a \in A\} - \mathbb{R}_+^\Theta$$

be the associated polyhedron containing all payoff vectors that are achievable or weakly dominated by some mixed action.[5] An example of $\mathcal{H}(A, u)$ when $|\Theta| = 2$ is depicted in Figure 1.

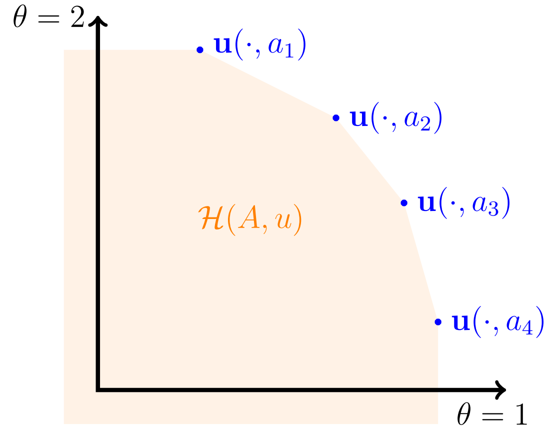


Figure 1: The shaded area represents $\mathcal{H}(A, u)$

Whenever $\mathcal{H}(\tilde{A}, \tilde{u}) \supseteq \mathcal{H}(A, u)$, it is immediate that

$$V(P_1, \ldots, P_m; (\tilde{A}, \tilde{u})) \geq V(P_1, \ldots, P_m; (A, u))$$

for all Blackwell experiments $P_1, \ldots, P_m$.

Definition 1. *We say that (A, u)* ***dominates*** *$(\tilde{A}, \tilde{u})$ if $\mathcal{H}(A, u) \supseteq \mathcal{H}(\tilde{A}, \tilde{u})$. We say that (A, u) is* ***equivalent*** *to $(\tilde{A}, \tilde{u})$ if $\mathcal{H}(A, u) = \mathcal{H}(\tilde{A}, \tilde{u})$.*

For instance, if an action is weakly dominated by a mixture of other actions, adding or removing this action leaves us with an equivalent decision problem.

We can also think of the dominance relationship in terms of a direct map between actions that lead to a higher utility across all states.

Definition 2. *A* ***dominating map*** *from $(\tilde{A}, \tilde{u})$ to (A, u) is a function $f : \tilde{A} \to \Delta(A)$ such that $u(\theta, f(\tilde{a})) \geq \tilde{u}(\theta, \tilde{a})$ for every $\tilde{a} \in \tilde{A}$ and every $\theta \in \Theta$.*

Clearly, (A, u) dominates $(\tilde{A}, \tilde{u})$ if and only if there exists a dominating map from $(\tilde{A}, \tilde{u})$ to (A, u).

3.2 The Blackwell Order and the Blackwell Supremum

We will use the Blackwell order of experiments throughout the paper. For the sake of completeness, we briefly review it in this subsection.[6]

[5]Here and in what follows, whenever $+$ and $-$ are used as operations between sets, they denote the Minkowski sum and difference.

[6]The zonotope approach to the Blackwell supremum presented in this section is due to Bertschinger and Rauh (2014). The lattice structure of the binary-state Blackwell order can also be derived from Kertz and Rösler (1992),

Definition 3. *$P : \Theta \to \Delta(Y)$ is more informative than $Q : \Theta \to \Delta(Z)$ if, for every decision problem, we have the inequality $V(P) \geq V(Q)$. We also say that P Blackwell dominates Q.*

We say that two experiments are **Blackwell equivalent** if they Blackwell-dominate each other. We also say that P is **strictly more informative** than Q (or P strictly Blackwell dominates Q) if P is more informative than Q and they are not Blackwell equivalent.

There are two other natural ways of ranking experiments. The first uses the notion of a *garbling*.

Definition 4. *$Q : \Theta \to \Delta(Z)$ is a garbling of $P : \Theta \to \Delta(Y)$ if there exists a function $g : Y \to \Delta(Z)$ (the "garbling") such that $Q(z|\theta) = \sum_y g(z|y)P(y|\theta)$.*

Thus Q is a garbling of P when one can replicate Q by "adding noise" to the signal generated from P. The second ranking uses the feasible state-action distributions.

Definition 5. *Given a set of actions A and an experiment $P : \Theta \to \Delta(Y)$, the feasible set of P is*

$$\Lambda_P(A) = \left\{ \lambda : \Theta \to \Delta A \;\middle|\; \lambda(a|\theta) = \sum_y \sigma(a|y)P(y|\theta) \text{ for some } \sigma : Y \to \Delta(A) \right\}.$$

The feasible set of an experiment specifies what conditional action distributions can be obtained by some choice of strategy σ. One might then say that more information allows for a larger feasible set.

Blackwell's Theorem states that these rankings of informativeness are equivalent.[7]

Blackwell's Theorem. *The following statements are equivalent*

1. *P is more informative than Q;*
2. *Q is a garbling of P;*
3. *For all sets A, $\Lambda_Q(A) \subseteq \Lambda_P(A)$.*

In addition, when $|\Theta| = 2$, Theorem 10 in Blackwell (1953) shows that the above statements are also equivalent to

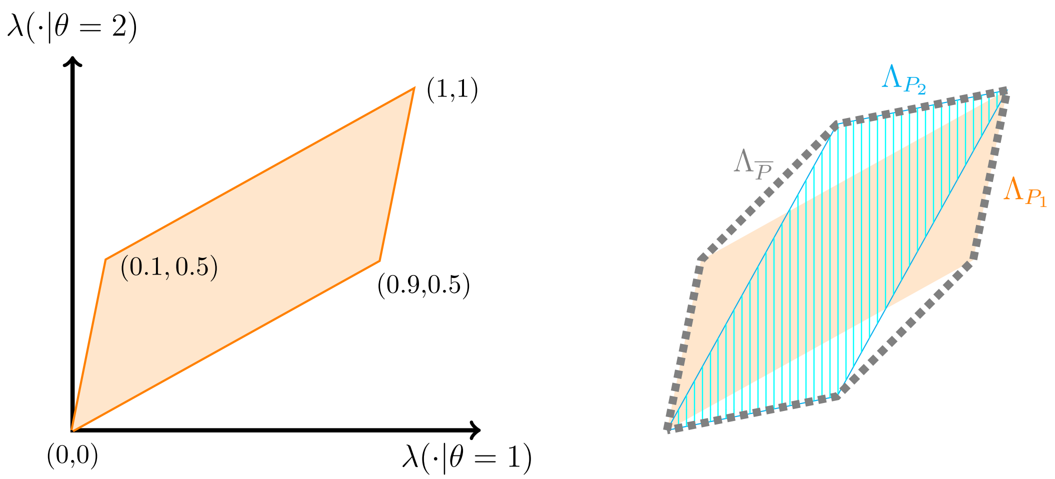
4. *For all sets A with $|A| = 2$, $\Lambda_Q(A) \subseteq \Lambda_P(A)$.*

Note that all sets A with the same cardinality give essentially the same set $\Lambda_P(A)$, so condition (3) could equivalently be stated as follows: for every $n \in \mathbb{N}$, we have $\Lambda_Q(\{1, \ldots, n\}) \subseteq \Lambda_P(\{1, \ldots, n\})$. Similarly, condition (4) can be stated as $\Lambda_Q(\{1, 2\}) \subseteq \Lambda_P(\{1, 2\})$. To simplify notation, when $|A| = 2$, we will omit A in the notation, simply writing Λ_P.

who establish the lattice structure of the univariate convex order.

[7] For a proof, see e.g. Blackwell (1953) or de Oliveira (2018).

Condition (4) is particularly useful as it offers a simple graphical representation of Blackwell experiments when $|\Theta| = 2$. Figure 2(a) illustrates this using the cough symptom from the introduction (see Table 2). To characterize Λ_P, it suffices to specify the probability of taking one of the two actions, as the probability of taking the other action is the complementary probability. The x-axis denotes the probability of taking this action in state 1, and the y-axis denotes the probability in state 2. This way, Λ_P is depicted as a subset of $[0,1]^2$. Clearly $(0,0),(1,1) \in \Lambda_P$ for all P, because these two points represent taking a constant action regardless of the signal realization. With the information obtained from the Blackwell experiment, additional points can be obtained. For example, the point $(0.1, 0.5)$ in Figure 2(a) can be achieved if the decision-maker chooses this action precisely when the patient does not have a cough symptom. Symmetrically, the decision-maker could choose the same action precisely when the agent has a cough symptom, which yields the point $(0.9, 0.5)$. Such pure strategies give us the extreme points of the polytope Λ_P and the possibility of randomization convexifies the set. Thus, Λ_P is a convex and symmetric[8] polytope in $[0,1]^2$, corresponding to the entire shaded area. Conversely, as shown in Bertschinger and Rauh (2014), any convex and symmetric polytope in $[0,1]^2$ correspond to Λ_P for some P.



(a) An example of $\Lambda_P(A)$ with $|\Theta| = |A| = 2$ (b) $\Lambda_{\overline{P}}$ as the convex hull of $\Lambda_{P_1} \cup \Lambda_{P_2}$

Figure 2

Having reviewed the Blackwell order, we now turn our attention to a concept that will be used extensively in our analysis — the Blackwell supremum.

Definition 6. *Let $P_1, P_2, \ldots, P_m$ be arbitrary Blackwell experiments. We say that $\overline{P}$ is the* ***Blackwell supremum*** *of $P_1, P_2, \ldots, P_m$ if*

1. *$\overline{P}$ is more informative than $P_1, P_2, \ldots, P_m$;*

2. *If Q is more informative than $P_1, P_2, \ldots, P_m$, then Q is also more informative than $\overline{P}$.*

[8]By symmetric we mean if $\lambda \in \Lambda_P$, $(1,1) - \lambda \in \Lambda_P$.

By definition, if there are two Blackwell suprema, they must Blackwell dominate each other. This means that any two Blackwell suprema must be Blackwell equivalent and so Blackwell suprema, if they exist, are unique up to Blackwell equivalence.

Furthermore, when the state space is binary, the Blackwell supremum always exists and can be characterized using the feasible set, as illustrated in Figure 2(b). From Blackwell's theorem, for any Q that is more informative than $P_1, ..., P_m$, the corresponding feasible set Λ_Q must contain $\Lambda_{P_1}, \ldots, \Lambda_{P_m}$. Since the feasible set is always convex, Λ_Q must also contain $\text{co}(\Lambda_{P_1} \cup \Lambda_{P_2} \cdots \cup \Lambda_{P_m})$. Moreover, the set $\text{co}(\Lambda_{P_1} \cup \Lambda_{P_2} \cdots \cup \Lambda_{P_m})$ is convex and symmetric, and so it corresponds to some Blackwell experiment $\overline{P}$, which is thus the least informative Blackwell experiment that dominates $P_1, ..., P_m$ — the Blackwell supremum. This observation yields the following lemma:

Lemma 1. *When $|\Theta| = 2$, the Blackwell supremum always exists. An experiment $\overline{P}$ is the Blackwell supremum of $P_1, P_2, \ldots, P_m$ if and only if $\Lambda_{\overline{P}} = co(\Lambda_{P_1} \cup \Lambda_{P_2} \cdots \cup \Lambda_{P_m})$.*

Proof. See Proposition 16 in Bertschinger and Rauh (2014). □

It is useful to note that the above lemma holds specifically for $|\Theta| = 2$. When $|\Theta| \geq 3$, a Blackwell supremum may not exist, as illustrated in Example 18 of Bertschinger and Rauh (2014).

3.3 Nature's MinMax Problem

Most of our focus will be on the robustly optimal strategies for the agent, but it will be helpful to first understand Nature's MinMax problem. Since the objective function is linear in both σ and P, and the choice sets of σ and P are both convex and compact, the minimax theorem (Sion, 1958) implies that

$$\begin{aligned} V(P_1, \ldots, P_m) &= \min_{P \in \mathcal{J}(P_1,\ldots,P_m)} \max_{\sigma \in \Sigma} \sum_{\theta \in \Theta} \sum_{(y_1,\ldots,y_m) \in \mathbf{Y}} P(y_1, \ldots, y_m|\theta) u(\theta, \sigma(y_1, \ldots, y_m)) \\ &= \min_{P \in \mathcal{J}(P_1,\ldots,P_m)} V(P). \end{aligned} \tag{1}$$

That is, the value of the agent's maxmin problem equals the value of a minmax problem where Nature chooses an experiment in the set $\mathcal{J}(P_1, \ldots, P_m)$ to minimize a Bayesian agent's value in the decision problem.

Let $\mathcal{D}(P_1, \ldots, P_m)$ denote the set of Blackwell experiments that dominates $P_1, ..., P_m$.[9] Every experiment in $\mathcal{J}(P_1, \ldots, P_m)$ must be more informative than each P_j, since the projection onto the jth coordinate is a garbling, so $\mathcal{J}(P_1, \ldots, P_m) \subseteq \mathcal{D}(P_1, \ldots, P_m)$. The set $\mathcal{D}(P_1, \ldots, P_m)$ is in

[9]Technically, if we allow any finite set to be a signal space, $\mathcal{D}$ is not a set in the strict set-theoretical sense. We can resolve this issue by fixing a large enough universe U of signals. For our purposes, it is enough if every finite subset of $\mathbb{N}$ are in U and all cartesian products of sets in U are also in U.

general a larger set, because not every experiment that dominates $P_1, ..., P_m$ can be represented as a joint experiments with marginals $P_1, ..., P_m$.[10] However, the next lemma shows that relaxing Nature's problem to choosing an experiment from the set $\mathcal{D}(P_1, \ldots, P_m)$ does not change the value of the problem.

Lemma 2.

$$V(P_1, \ldots, P_m) = \min_{P \in \mathcal{D}(P_1, \ldots, P_m)} V(P) \tag{2}$$

Proof. See Appendix A.1. □

The idea underlying Lemma 2 is that in the relaxed problem above, Nature could restrict attention to the experiments that are Blackwell minimal — those that do not strictly Blackwell dominate any other experiment in $\mathcal{D}(P_1, \ldots, P_m)$. Additionally, any Blackwell minimal element in this set can be represented as a joint experiment in $\mathcal{J}(P_1, \ldots, P_m)$, as shown in Appendix A.1.

Lemma 2 is particularly useful when the state is binary. Under binary states, the Blackwell supremum $\overline{P}$ of $P_1, ..., P_m$ exists, and it is the unique (up to Blackwell equivalence) Blackwell minimal element in $\mathcal{D}(P_1, \ldots, P_m)$. Therefore, $\overline{P}$ solves (2) regardless of the decision problem, yielding the following corollary.

Corollary 1. *When* $|\Theta| = 2$,

$$V(P_1, ..., P_m) = V(\overline{P}(P_1, ..., P_m))$$

where $\overline{P}(P_1, ..., P_m)$ *is the Blackwell supremum of experiments* $\{P_1, ..., P_m\}$.

Thus, in binary-state decision problems, the agent's value from using a robust strategy is the same as the value she would obtain if she faced a single experiment — the Blackwell supremum of all marginal experiments. Moreover, the Blackwell supremum depends only on the marginal experiments, and not on the particular decision problem.

4 Binary State Environment

4.1 Binary-State Binary-Action Problems

As seen in the introductory example, one simple strategy that generates a robust value independent of the correlations among the marginal information sources is to choose exactly one marginal experiment from $\{P_1, \ldots, P_m\}$ and play the optimal strategy that uses that information alone, ignoring the signal realizations of all other experiments. By choosing the marginal experiment

[10] For a simple example, consider two experiments P_1 and P_2 whose signal spaces Y_1 and Y_2 are singletons. Then $\mathcal{J}(P_1, P_2)$ contains only the completely uninformative experiment while $\mathcal{D}(P_1, P_2)$ contains all Blackwell experiments.

optimally, the agent achieves an expected payoff of $\max_{j=1,\ldots,n} V(P_j)$, regardless of the actual joint experiment $P \in \mathcal{J}(P_1, \ldots, P_m)$. We call such a strategy a *best-source strategy*.

In some cases, it is clear that a best-source strategy is robustly optimal. Suppose, for example, that we have two information sources, P_1 and P_2, and that P_1 is more informative than P_2. We can then consider a correlation structure where the signal of P_2 is generated by garbling the signal of P_1. Consequently, after observing the signal from P_1, observing signals from P_2 provides no additional information. Therefore, the agent loses nothing by ignoring P_2, and the best-source strategy that uses only P_1 is optimal.

The interesting cases are when information sources are not Blackwell-ranked (which will often be the case, given the demanding nature of the Blackwell order). In such cases, the Blackwell supremum is strictly more informative than any single information source, so one may hope to do better than a best-source strategy by combining different sources. Surprisingly, Theorem 1 shows that, in decision problems with binary states and binary actions, the agent can never do better than a best-source strategy. Moreover, if the information sources satisfy full support and we exclude cases where multiple information sources are optimal, then any strategy that uses more than one source is strictly suboptimal.

Theorem 1. *For all (A, u) with $|A| = |\Theta| = 2$, any best-source strategy is robustly optimal. In other words,*

$$V(P_1, \ldots, P_m) = \max_{j=1,\ldots,m} V(P_j).$$

In addition, if each marginal experiment has full support, i.e., $P_j(y_j|\theta) > 0$ for all j, y_j, θ, and $\operatorname{argmax}_{j=1,\ldots,m} V(P_j)$ is a singleton, then all robustly optimal strategies are best-source strategies.

Proof. To simplify notation, we write $\overline{P}$ to refer to the Blackwell supremum, $\overline{P}(P_1, \ldots, P_m)$. By Corollary 1, it suffices to show that $V(\overline{P}) = \max_{j=1,\ldots,m} V(P_j)$. By Lemma 1,

$$\Lambda_{\overline{P}} = \operatorname{co}\left(\Lambda_{P_1} \cup \cdots \cup \Lambda_{P_m}\right). \tag{3}$$

Now, the maximum utility achievable given $\overline{P}$ is $V(\overline{P}) = \max_{\lambda \in \Lambda_{\overline{P}}} \sum_{a,\theta} u(\theta, a)\lambda(a|\theta)$. Since the maximand is linear in λ, the fundamental theorem of linear programming states that the maximum is achieved at an extreme point of $\Lambda_{\overline{P}}$. By (3), an extreme point of $\Lambda_{\overline{P}}$ must belong to some Λ_{P_j}. Hence, we have

$$V(\overline{P}) = \max_{\lambda \in \Lambda_{P_j}} \sum_{a,\theta} u(\theta, a)\lambda(a|\theta) = V(P_j) \leq \max_{j'=1,\ldots,m} V(P_{j'}).$$

Since $\overline{P}$ is more informative than every P_j, we also have $V(\overline{P}) \geq \max_{j'=1,\ldots,m} V(P_{j'})$, which concludes the proof of the theorem's first statement. The proof of the second part (uniqueness) requires different arguments, so we defer it to Appendix B.6. □

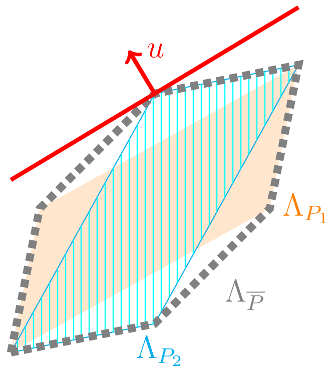


Figure 3: The maximum is achieved at an extreme point that belongs to Λ_{P_2}

The idea of Theorem 1 can be visualized in Figure 3 for two marginal experiments. Each marginal Blackwell experiment P_1, P_2 can be represented by $\Lambda_{P_1}, \Lambda_{P_2}$, the set of feasible state-action distributions generated by the experiment. The corresponding $\Lambda_{\overline{P}}$ for Blackwell supremum $\overline{P}$ is the convex hull of $\Lambda_{P_1} \cup \Lambda_{P_2}$. Since the utility is linear with respect to $\lambda \in \Lambda_{\overline{P}}$, the maximum is achieved at an extreme point, which belongs to either Λ_{P_1} or Λ_{P_2}, and thus can be achieved by using a single marginal experiment.

4.2 Separable Problems

While best-source strategies are sufficient in binary-state, binary-action decision problems, more complicated problems often require the agent to use more sophisticated strategies to robustly aggregate information from multiple sources. However, as we will see, best-source strategies form the building blocks from which the robustly optimal strategies are constructed. For example, in the Covid example with two treatments presented in the introduction, we saw that a simple yet robust strategy that uses multiple information sources is to consider the two treatments separately, using the corresponding best-source strategy for each treatment decision.

As a first step toward the analysis of robustly optimal strategies in general binary-state decision problems, we generalize this idea in the example to a class of decision problems, which we call *separable*.

Definition 7. *A binary-state decision problem (A, u) is a* ***separable problem*** *if A can be written as a product $A_1 \times \cdots \times A_k$ where $|A_\ell| = 2$ for all $\ell = 1, ..., k$, and*

$$u(\theta, a) = u_1(\theta, a_1) + \cdots + u_k(\theta, a_k)$$

for some $\{u_\ell : \Theta \times A_\ell \to \mathbb{R}\}_{\ell=1}^{k}$.

We will use $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$ to refer to a separable problem and we refer to each of the binary-action decision problems, (A_ℓ, u_ℓ), as a *subproblem*.

As highlighted in the Covid example, a simple strategy in separable problems is to use the best-source strategy in each of the subproblems, which guarantees a payoff of $\sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j; (A_\ell, u_\ell))$ regardless of the correlations between the information sources. Hence, this value provides a lower bound on the robustly optimal value:

$$V\left(P_1, \ldots, P_m; \bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)\right) \geq \sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j; (A_\ell, u_\ell)).$$

The following lemma further shows that the above inequality is indeed an equality. This follows from a special property highlighted in Corollary 1 — that in binary-state environments, there exists a single $\overline{P}(P_1, \ldots, P_m)$ that uniformly minimizes the agent's value across all decision problems.[11]

Lemma 3. *For any separable problem* $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$,

$$V\left(P_1, \ldots, P_m; \bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)\right) = \sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j; (A_\ell, u_\ell)).$$

Moreover, let $\sigma_\ell : \mathbf{Y} \to A_\ell$ *be a pure best-source strategy for subproblem* (A_ℓ, u_ℓ). *Then* $\sigma : \mathbf{Y} \to A_1 \times \cdots \times A_k$ *defined by*

$$\sigma(y_1, ..., y_m) = \Big(\sigma_\ell(y_1, ..., y_m)\Big)_{\ell=1}^{k} \qquad \text{for all } y_1, ..., y_m \tag{4}$$

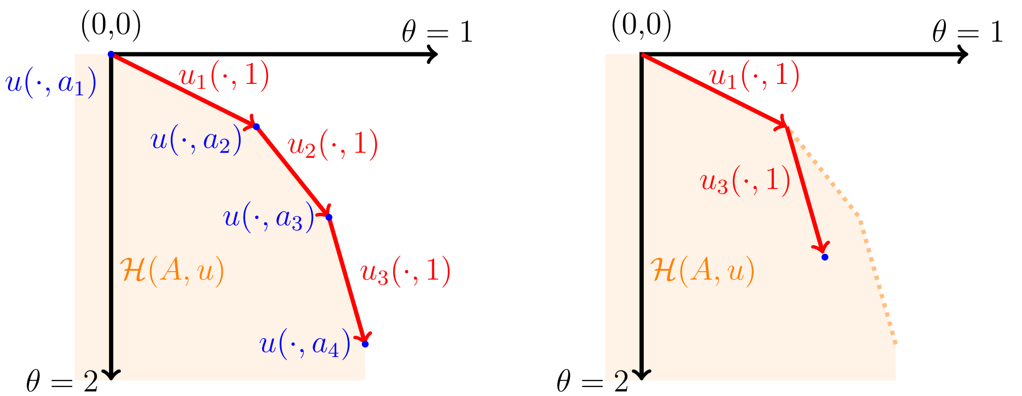
is a robustly optimal strategy for decision problem $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$.

Proof. See Appendix A.2. □

4.3 General Decision Problems and Decompositions

The special structure of separable problems yields simple robustly optimal strategies. To what extent can this structure be applied in tackling more general decision problems? We show in this section that *every* binary-state decision problem is equivalent to a separable problem in the sense of Definition 1. The idea is to decompose an n-action decision problem into $n-1$ binary-action decision problems and use these subproblems to construct the separable problem that is equivalent to the original problem. We call the resulting separable problem the *binary decomposition*.

[11] In contrast, with three or more states, Nature's worst-case joint experiment in Eq. (1) typically depends on the decision problem. Therefore, $\min_{P \in \mathcal{J}} V(P; \bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)) \geq \sum_{\ell=1}^{k} \min_{P \in \mathcal{J}} V(P; (A_\ell, u_\ell))$, which in general is not an equality.



(a) Binary decomposition

(b) A nonconsecutive sum of $u_\ell(\cdot, 1)$ lies in the interior of $\mathcal{H}(A, u)$

Figure 4

Before we define the binary decomposition, we make a normalization to simplify exposition. First we remove all weakly-dominated actions,[12] so that actions can be ordered such that a higher action best responds to a higher belief on θ_1:

$$u(\theta_1, a_1) < u(\theta_1, a_2) < \cdots < u(\theta_1, a_n),$$
$$u(\theta_2, a_1) > u(\theta_2, a_2) > \cdots > u(\theta_2, a_n).$$

Moreover, by adding a constant vector, we can normalize $u(\cdot, a_1) = (0, 0)$.

Definition 8. *Given a decision problem (A, u), the **binary decomposition** of (A, u) is a separable problem $\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)$ where*

$$A_\ell := \{0, 1\}, u_\ell(\cdot, 0) = (0, 0), u_\ell(\cdot, 1) = u(\cdot, a_{\ell+1}) - u(\cdot, a_\ell).$$

The key idea underlying the binary decomposition is to decompose the original problem into binary-action decision problems that compare each pair of consecutive actions. This can be visualized in Figure 4(a). The four-action decision problem is decomposed into three binary-action decision problems, by examining the difference vectors $u(\cdot, a_{\ell+1}) - u(\cdot, a_\ell)$. Each subproblem of the decomposition can be interpreted as a "local" comparison between two consecutive actions.

Notice that every feasible payoff vector in the original problem can be replicated in the binary decomposition, due to the fact that $u(\cdot, a_i) = \sum_{\ell=1}^{i-1} u_\ell(\cdot, 1) + \sum_{\ell=i}^{n-1} u_\ell(\cdot, 0)$ for all $i = 1, ..., n$. So $\mathcal{H}(A, u) \subseteq \mathcal{H}\big(\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)\big)$. Of course, the binary decomposition $\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)$ could introduce additional feasible payoff vectors. For example, in the example in Figure 4(b), the strategy $(1, 0, 1)$ in the binary decomposition yields a payoff vector that is infeasible in the

[12] An action $a \in A$ is weakly-dominated if there exists $\alpha \in \Delta A$ such that $u(\cdot, a) \leq u(\cdot, \alpha)$ and $u(\cdot, a) \neq u(\cdot, \alpha)$. If there are duplicated actions, we remove all but keep one copy.

original problem. However, this additional payoff vector lies in the interior of $\mathcal{H}(A, u)$, and thus it is dominated by one of the original (possibly mixed) actions. The next lemma shows that this is generally true, so we have $\mathcal{H}(A, u) = \mathcal{H}\big(\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)\big)$ — any decision problem is equivalent to its binary decomposition in the sense of Definition 1.

Lemma 4. (A, u) *is equivalent to its binary decomposition.*

Proof. See Appendix A.3. □

Lemma 4 implies that the set of dominating maps (see Definition 2) from $\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)$ to (A, u) is non-empty. Together with Lemma 3, it allows us to derive a robustly optimal strategy for any decision problem (A, u) through its binary decomposition.

Theorem 2. *Let* $\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)$ *be the binary decomposition of* (A, u)*, and* σ_ℓ *be a pure best-source strategy for* (A_ℓ, u_ℓ)*. Then*

$$V(P_1, \ldots, P_m; (A, u)) = \sum_{\ell=1}^{n-1} \max_{j=1,\ldots,m} V(P_j; (A_\ell, u_\ell)).$$

Moreover, for any dominating map f from $\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)$ *to* (A, u)*,* $\sigma_f^*(\mathbf{y}) := f(\sigma_1(\mathbf{y}), \ldots, \sigma_{n-1}(\mathbf{y}))$ *is robustly optimal.*

Proof. See Appendix A.4. □

Theorem 2 allows the construction of a robustly optimal strategy for any decision problem (A, u) according to a two-step procedure:

1. For each subproblem, (A_ℓ, u_ℓ), find a best-source strategy σ_ℓ (which we know is robustly optimal by Theorem 1).

2. For each realization $\mathbf{y}$, pick a (mixed) action $\sigma^*(\mathbf{y}) \in \Delta(A)$ such that $u(\cdot, \sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{n-1} u_\ell(\cdot, \sigma_\ell(\mathbf{y}))$.[13]

Notably, once $\sigma_\ell(\cdot)$ has been determined in Step 1, the marginal experiments, $P_1, \ldots, P_m$, play no role in Step 2. In other words, the marginal experiments only influence the ultimate choice of action in Step 1, and more specifically through its effect on the choice of $\sigma_\ell(\mathbf{y})$ in each of the subproblems.

In contrast to Theorem 1, Theorem 2 also highlights the non-uniqueness of robustly optimal strategies when there are three or more actions. This is because there could be multiple σ^* satisfying $u(\cdot, \sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{n-1} u_\ell(\cdot, \sigma_\ell(\mathbf{y}))$ for all $\mathbf{y}$. For example, in the Covid example with two

[13]Note that in Theorem 1, the agent does not need to randomize. In contrast, the construction of a robustly optimal strategy in Theorem 2 may require randomization. We provide an example in Appendix B.1 where all robustly optimal strategies require randomization.

treatments in the introduction, the robustly optimal strategy we derived in Table 4 recommends no treatment when a patient has no symptoms. However, note that giving neither treatment is dominated by giving both treatments, so replacing No Treatment with $T_1 + T_2$ does not decrease the guaranteed value, thus yielding another robustly optimal strategy. Why can a robustly optimal strategy play a dominated action? Because the worst-case correlation structure against that strategy will put probability 0 on the symptom realization $(Cough_-, Fever_-)$.

Theorem 2 delivers two immediate corollaries.

Corollary 2. *For any decision problem (A, u) and any collection of experiments $\{P_j\}_{j=1}^m$, there exists a subset of marginal experiments $\{P_j\}_{j\in J\subseteq\{1,\dots,m\}}$ with $|J| \le |A| - 1$, such that*

$$V(P_1, \cdots, P_m; (A, u)) = V(\{P_j\}_{j\in J}; (A, u)).$$

Corollary 2 establishes a bound, $|A| - 1$, on the number of information sources needed for a robustly optimal strategy. Note that this bound is *independent* of the fine details of the decision problem, such as the exact cardinal utilities of the agent, and the details of the marginal information sources available to the agent.

Corollary 3. *Suppose $\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)$ is the binary decomposition of (A, u). For any j,*

$$V(P_1, ..., P_m; (A, u)) > V(P_1, \dots, P_{j-1}, P_{j+1}, \dots, P_m; (A, u))$$

if and only if $V(P_j; (A_\ell, u_\ell)) > \max_{j'\neq j} V(P_{j'}; (A_\ell, u_\ell))$ for some $\ell = 1, ..., n-1$.

Corollary 3 shows that an additional marginal experiment robustly improves the agent's value if and only if it outperforms all other marginal experiments in at least one of the subproblems of the decomposition. In particular, an experiment that performs reasonably well across all subproblems can be completely ignored if, for each subproblem, there is some other, more specialized experiment that is the best. The next example further illustrates that even the best standalone information source may be ignored.

Example 1. *We revisit the Covid example with two treatments in the introduction. Suppose in addition to the Cough and Fever, now we have a third informative symptom, Headache, whose relationship to the diseases is given in Table 5.*

	+	−
Covid	0.72	0.28
Flu	0.28	0.72

Table 5: Headache symptom

Note that in either treatment 1 or treatment 2, when using the Headache symptom alone, the agent can achieve a value of $\frac{1}{2}(0.72 \times 30 + 0.28 \times (-20)) = 8$. *This means that Headache is the best standalone information source, because it yields a value of* $8 + 8 = 16$, *which is greater than* 15, *the value of using either the Cough or Fever symptom alone. However, this symptom is never the best information source for either treatment 1 or treatment 2, as the value it yields is lower than* 8.5, *the value achieved by using the Cough symptom for treatment 1 or using the Fever symptom for treatment 2. Thus, despite being the best standalone information source for the overall decision problem, Headache can be ignored because it fails to be the best information source for any single subproblem.*

5 General-State Decision Problems

In this section, we turn to general-state decision problems and characterize the robustly optimal value using concavification. This characterization offers an explicit formula for the robustly optimal value and implies several corollaries that speak to the structure of robustly optimal strategies.

It is worth noting why our previous approach for binary-state decision problems would not extend to general-state decision problems: First, with more states, it is not immediately clear how to decompose a general decision problem into more "elementary" ones. Second, the non-existence of the Blackwell supremum implies that in Nature's minmax problem Eq. (1), there may no longer be a single experiment that uniformly minimizes the agent's value across all decision problems, exacerbating the complexity of the analysis (see Footnote 11). Lastly, an agent may want to use multiple information sources even in simple binary-action decision problems, as illustrated in Example 2.

Example 2. *Suppose that there are three states* $\theta_1, \theta_2, \theta_3$. *The marginal experiments are both binary with respective signals* x_1, x_2 *and* y_1, y_2, *as given by Table 6.*

P_X

$P_X(x\|\theta)$	x_1	x_2
θ_1	1	0
θ_2	1	0
θ_3	0	1

P_Y

$P_X(y\|\theta)$	y_1	y_2
θ_1	1	0
θ_2	0	1
θ_3	0	1

Table 6

Intuitively, experiment P_X *indicates whether the state is* θ_3 *or not and experiment* P_Y *indicates whether the state is* θ_1 *or not. Note that upon observing both experiments, the agent obtains perfect information, and so in any decision problem, the agent achieves the perfect information payoff.*

Let $A = \{0, 1\}$ and suppose that the utilities are as follows:[14]

$$u(\theta, a = 1) = \mathbf{1}(\theta \in \{\theta_1, \theta_3\}) - 0.9 \cdot \mathbf{1}(\theta = \theta_2),$$
$$u(\theta, a = 0) = 0.$$

By using only one information source (either P_X or P_Y), $a = 1$ is the unique optimal action for any signal realization. Therefore, the agent's expected payoff is $1 - 0.9 + 1 = 1.1$. By contrast, when using both information sources, the full information payoff is $1 + 0 + 1 = 2$.

5.1 Value Functions and Extremal Beliefs

Before proceeding to our characterization theorem, we first introduce useful terminology that will be important for our general approach in this section. Rather than viewing a decision problem as a collection of actions and a utility function, we instead describe a decision problem via its (interim) value function $v : \Delta(\Theta) \to \mathbb{R}$:

$$v(\mu) := \max_{a \in A} \sum_{\theta \in \Theta} \mu(\theta)\rho(\theta, a).$$

If the value functions of two decision problems differ by an affine function, $g(\mu)$, then these two decision problems are equivalent, in the sense that their robustly optimal values will differ by $g(\mu_0)$ for any marginal experiments. Hence, without loss of generality, we normalize $v(\delta_\theta) = 0$ for all $\theta \in \Theta$.

The value function of a finite decision problem is convex and piecewise linear. We denote the epigraph of v by $\text{epi}(v) = \{(\mu, w) : w \geq v(\mu), \mu \in \Delta(\Theta)\}$. The set of extreme points of the epigraph, denoted by $\text{ext}(\text{epi}(v))$, is finite and contains $\{(\delta_\theta, v(\delta_\theta))\}_{\theta \in \Theta}$, where δ_θ denotes the degenerate belief on θ. Let E_v denote the projection of $\text{ext}(\text{epi}(v))$ on $\Delta(\Theta)$. We call the elements of E_v the **extremal beliefs**.[15] Let K_v denote the set of **non-degenerate extremal beliefs**, that is, those extremal beliefs that do not belong to $\{\delta_\theta\}_{\theta \in \Theta}$. See Figure 5 for an illustration when $|\Theta| = 2$ and $|A| = 3$.

As we will mainly work with value functions throughout this section, we rewrite the robustly optimal values in terms of v. Note that each Blackwell experiment P induces a posterior distribution $\tau^P \in \Delta(\Delta\Theta))$.[16] From Eq. (1), the robustly optimal value of a value function v when

[14]Recall that the payoffs here have been weighted by the prior: $u(\theta, a) = \mu_0(\theta)\rho(\theta, a)$.

[15]Similar approaches have been used in Bergemann, Brooks, and Morris (2015) and Lipnowski and Mathevet (2017), where these objects are called "extremal markets" or "outer points."

[16]More precisely, the induced posterior distribution $\tau^P \in \Delta(\Delta(\Theta))$ associated with a Blackwell experiment $P : \Theta \to \Delta Y$ is defined as

$$\tau^P(E) = \sum_{y \in Y_E} \sum_{\theta} \mu_0(\theta)P(y|\theta),$$

facing information sources, $P_1, ..., P_m$, is then

$$V(P_1, ..., P_m; v) = \min_{P \in \mathcal{J}(P_1,...,P_m)} \int v(\mu)\tau^P(d\mu). \tag{5}$$

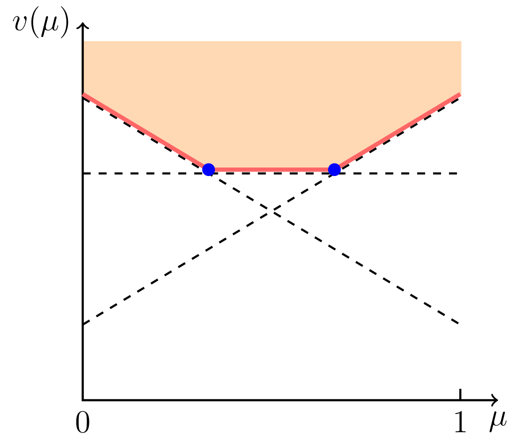


Figure 5: Interim value function and extremal beliefs when $|\Theta| = 2$ and $|A| = 3$. Each dashed line denotes the agent's interim payoff from an action, and their upper envelope (in red) is the interim value function. The shaded area represents the epigraph, and the blue dots are its non-degenerate extreme points, whose projections are non-degenerate extremal beliefs.

5.2 Convexification and Concavification

The concepts of convex and concave envelope of a function will play an important role in our analysis, so we briefly review them.

Let X be a subset of a vector space, and $f : X \to \mathbb{R}$. The **convexification** of f, denoted by $conv(f)$, is the largest convex function defined on $co(X)$ below f. That is, for any $z \in co(X)$,

$$conv(f)(z) = \sup\{g(z)|g \in \mathbb{R}^{co(X)} \text{ is convex and } g(x) \leq f(x), \forall x \in X\}.$$

Similarly, we define the **concavification** of f as

$$conc(f)(z) = \inf\{g(z)|g \in \mathbb{R}^{co(X)} \text{ is concave and } g(x) \geq f(x), \forall x \in X\}.$$

The convexification and concavification are well-defined real functions, as long as the sets in their

where $E \subseteq \Delta(\Theta)$ is a Borel set and

$$Y_E = \left\{ y \in Y \middle| \frac{\mu_0(\theta)P(y|\theta)}{\sum_\theta \mu_0(\theta)P(y|\theta)} \in E, \forall\theta \right\}.$$

In our case, the posterior distribution always has finite support, so it is enough to consider $E = \{\mu\}$ for some $\mu \in \Delta(\Theta)$.

definitions are non-empty, which will be assumed whenever these concepts are used. Note that f need not be defined on a convex set, but its convexification or concavification will be extended to the convex hull of f's domain.

An equivalent definition of concavification that will be useful in our analysis is $conc(f)(z) = \sup\{y|(z, y) \in co(G(f))\}$, where $G(f) = \{(x, f(x))|x \in X\}$ is the graph of f.[17]

5.3 Robustly Optimal Value

As shown by Example 2, with general state spaces, best-source strategies are not robustly optimal even in very simple decision problems involving two actions. Nevertheless, they form the backbone of robustly optimal strategies, similarly to Theorem 2. To see the idea, for any decision problem with value function v, the best source strategy provides a lower bound on the robustly optimal value:

$$V(P_1, \dots, P_m; v) \geq \max_{j=1,\dots,m} V(P_j, v).$$

Now suppose that (A_ℓ, u_ℓ) are decisions problems with corresponding value functions $\lambda_\ell v_\ell$, where $\lambda_1, \dots, \lambda_k \in [0, 1]$ with $\sum_{\ell=1}^k \lambda_\ell = 1$ and $\sum_{\ell=1}^k \lambda_\ell v_\ell = v$. Then, (A, u) is equivalent to $\bigoplus_{\ell=1}^k (A_\ell, u_\ell)$, and by the same argument used in Section 4.2, $V(P_1, \dots, P_m; v) \geq \sum_{\ell=1}^k \lambda_\ell \max_{j=1,\dots,m} V(P_j, v_\ell)$. This proves that

$$V(P_1, \dots, P_m; v) \geq conc\left(\max_{j=1,\dots,m} V(P_j, \cdot)\right)(v).$$

Theorem 3 shows two things: First, it simplifies the above concavification problem by reducing the dimension of the relevant value functions. Second, it shows that the above inequality holds as an equality in this reduced dimension.

To reduce the dimension, we summarize a value function by its values on extremal beliefs $\mu \in E_v$. Formally, for each (normalized) value function v, evaluating it on E_v obtains $\boldsymbol{v}_v \in \mathbb{R}_-^{E_v}$ where $\boldsymbol{v}_v(\mu) = v(\mu)$ for all $\mu \in E_v$. Following our normalization of decision problems, we call a vector $\boldsymbol{v} \in \mathbb{R}_-^{E_v}$ a **value vector** if $\boldsymbol{v}(\delta_\theta) = 0$ for all $\theta \in \Theta$. Thus, the set of all value vectors is equivalent to $\mathbb{R}_-^{K_v}$. Note that some value vector $\boldsymbol{v} \in \mathbb{R}_-^{K_v}$ may not correspond to a value function, due to the convexity constraint of the value functions. However, the convexification $conv(\boldsymbol{v})$ is always a well-defined value function. In particular, convexifying the value vector obtained from a value function yields the value function itself, i.e., $conv(\boldsymbol{v}_v) = v$.

For any experiment P, define

$$\begin{aligned} W(P; \cdot) : \mathbb{R}_-^{K_v} &\to \mathbb{R} \\ \boldsymbol{v} &\mapsto V(P; conv(\boldsymbol{v})) \end{aligned}$$

[17]See Chapter IV, Proposition 2.5.1 in Hiriart-Urruty and Lemaréchal (1996).

as the value of the experiment P given a value vector $\boldsymbol{v}$. We are now ready to formally state Theorem 3.

Theorem 3. *For any value function* v *with a corresponding value vector* $\boldsymbol{v}_v$*,*

$$V(P_1, ..., P_m; v) = conc(\max\{W(P_1; \cdot), ..., W(P_m; \cdot)\})(\boldsymbol{v}_v).$$

Proof. See Appendix A.5. □

Theorem 3 offers a geometric characterization for the robustly optimal value through concavification. To illustrate the connection to Theorem 2, suppose that $(\lambda_\ell, \boldsymbol{v}_\ell)_{\ell=1}^m$ constitutes a solution to the concavification:

$$\lambda_\ell \geq 0, \sum_{\ell=1}^m \lambda_\ell = 1, \sum_{\ell=1}^m \lambda_\ell \boldsymbol{v}_\ell = \boldsymbol{v}_v, \sum_{\ell=1}^m \lambda_\ell W(P_\ell, \boldsymbol{v}_\ell) = conc(\max\{W(P_1; \cdot), ..., W(P_m; \cdot)\})(\boldsymbol{v}_v).$$

Now let (A_ℓ, u_ℓ) be a finite action decision problem whose interim value function is $\lambda_\ell conv(\boldsymbol{v}_\ell)$. It can be shown that $\sum_{\ell=1}^m \lambda_\ell conv(\boldsymbol{v}_\ell) \leq v$ and so using the language of Section 3, (A, u) is a decision problem that dominates $\bigoplus_{\ell=1}^m (A_\ell, u_\ell)$. The above equality essentially shows that

$$V(P_1, \ldots, P_m; (A, u)) = V\left(P_1, \ldots, P_m; \bigoplus_{\ell=1}^m (A_\ell, u_\ell)\right) = \sum_{\ell=1}^m V(P_\ell, (A_\ell, u_\ell)).$$

Given the above analogy to the idea of decomposition underlying Theorem 2, the solution to the concavification also generates robustly optimal strategies in the same manner here.

Corollary 4. *Suppose that* $\lambda_\ell^* \in \mathbb{R}_+, \boldsymbol{v}_\ell^* \in \mathbb{R}_-^{K_v}$*,* $j = 1, ..., m$ *are such that*

$$\sum_{\ell=1}^m \lambda_\ell^* = 1, \sum_{\ell=1}^m \lambda_\ell^* \boldsymbol{v}_\ell^* = \boldsymbol{v}_v, \sum_{\ell=1}^m \lambda_\ell^* W(P_j; \boldsymbol{v}_\ell^*) = V(P_1, \ldots, P_m; \boldsymbol{v}_v).$$

Let (A_ℓ, u_ℓ) *be a finite-action decision problem whose induced value function is* $conv(\lambda_\ell \boldsymbol{v}_\ell^*)$ *and* $\sigma_\ell^* : Y_\ell \to \Delta(A_\ell)$ *be an optimal strategy given Blackwell experiment* P_ℓ*. Then* (A, u) *dominates* $\bigoplus_{\ell=1}^m (A_\ell, u_\ell)$ *and for any dominating map,* $f : \bigoplus_{\ell=1}^m (A_\ell, u_\ell) \to \Delta(A)$*,*

$$\sigma_f^*(y_1, \ldots, y_m) = f(\sigma_1^*(y_1), \ldots, \sigma_m^*(y_m))$$

is a robustly optimal strategy.

Proof. See Appendix A.6. □

6 Behavioral Implications of Robustly Optimal Strategies

6.1 Ignoring Information

Theorem 3 offers two immediate corollaries that speak to which information sources can be ignored and how many of them are needed.

Corollary 5. *There exists a subset of information sources* $\{P_j\}_{j\in J\subseteq\{1,...,m\}}$ *with* $|J| \leq |K_v|$*, such that*

$$V(P_1, ..., P_m; v) = V(\{P_j\}_{j\in J}; v).$$

Proof. See Appendix A.7. □

Corollary 5 establishes that a decision maker would never need to use more information sources than the number of nondegenerate extremal beliefs. The proof follows from the familiar idea that, in a k-dimensional concavification problem, the concavification value can be achieved by a convex combination of at most $k + 1$ points. One subtlety here is that such an argument only gives us a bound of $|K_v| + 1$. To derive the tighter bound $|K_v|$, we utilize the positive homogeneous property of the $W(P; \cdot)$ function to further reduce dimensions.

We show in Appendix B.2 that $|K_v|$ can be further bounded by a function of $|\Theta|$ and $|A|$:

$$|K_v| \leq \binom{|\Theta| + |A| + 1 - \left\lfloor \frac{|\Theta|+1}{2} \right\rfloor}{|A| + 1} + \binom{|\Theta| + |A| + 1 - \left\lfloor \frac{|\Theta|+2}{2} \right\rfloor}{|A| + 1} - 2|\Theta|.$$

When $|\Theta| = 2$, this bound reduces to $|A| - 1$, the bound given in Corollary 2.

Corollary 6. *Suppose that* $V(P_\ell; v) \leq \max_{j\neq\ell} V(P_j; v)$ *for all* v*. Then* $V(P_1, ..., P_m; v) = V(P_1, ..., P_{\ell-1}, P_{\ell+1}, ..., P_m; v)$ *for all* v*.*

Proof. See Appendix A.8. □

Corollary 6 tells us that if an information source P_ℓ is never the unique best information source among $\{P_j\}_{j=1}^m$ for any decision problem, then it can be safely ignored. Note that being never the best information source is weaker than P_ℓ being Blackwell dominated by one of the other experiments, because the experiment that outperforms P_ℓ may depend on the particular decision problem.[18]

6.2 Aggregating Action Recommendations and Unanimity

Consider the leading question from Piccione and Rubinstein (2024):

[18] Cheng and Börgers (2024) show that this condition is equivalent to P_ℓ being dominated by a convex combination of the rest of the experiments.

> "The proportion of newborns with a specific genetic trait is 1%. Two conditionally independent screening tests, A and B, are used to identify this trait in all newborns. However, the tests are not precise. Specifically, it has been found that:
>
> 1. 70% of the newborns who are found to be positive according to test A have the trait.
> 2. 20% of the newborns who are found to be positive according to test B have the trait.
>
> Suppose that a newborn is found to be positive according to both tests. What is your estimate of the probability that this newborn has the trait?"

A correct understanding that the signals are conditionally independent would result in an updated posterior of 98% that the newborn has the trait! However, Piccione and Rubinstein (2024) report that many responses lie between 20% and 70%. Our model provides a possible explanation for this pattern: If subjects do not believe or understand the conditional independence of the tests and instead are cautious as to what may be their correlation, they would indeed act as if their beliefs are between 20% and 70%.

To formalize this statement, consider a decision problem (A, u) with a binary state space, $\{\theta_1, \theta_2\}$, normalized as in Section 4.3. Recall that, after normalization, actions can be ordered according to their optimality with respect to beliefs. To extend this order to mixed strategies, let $B(\alpha) \subset [0, 1]$ denote the set of beliefs that are consistent with interim optimality of action $\alpha \in \Delta(A)$:

$$B(\alpha) := \left\{ \mu(\theta_1) : \sum_{\theta} \mu(\theta) u(\alpha, \theta) = \max_{a \in A} \sum_{\theta} \mu(\theta) u(a, \theta) \right\}.$$

We can interpret $B(\alpha)$ as the revealed posterior beliefs that rationalize an agent's choice of α. Now, define the **belief order** as $\alpha \succeq \beta$ if $B(\alpha)$ dominates $B(\beta)$ in the strong set order[19] — higher actions correspond to higher revealed beliefs. For $a \in A$, this order corresponds to the labeling in Section 4.3.

Now, suppose that the agent sees only a single signal y_j from a single information source P_j. Let $a_j^*(y_j; A, u)$ denote the optimal action given the observation of y_j in isolation. We will interpret these actions as the recommended actions from individual information sources.

The following corollary of Theorem 2 shows that there exists some robustly optimal strategy such that the realized actions lie between the recommended actions from the individual information sources.

[19] Given $S, T \subseteq [0, 1]$, S dominates T in the strong set order if, for any $s \in S$ and $t \in T$, $\max\{s, t\} \in S$ and $\min\{s, t\} \in T$.

Corollary 7. *Let (A, u) be a binary-state decision problem. There exists some robustly optimal strategy σ, such that for all realizations of signals $\boldsymbol{y}$,*

$$\min_{j=1,\dots,m} a_j^*(y_j; A, u) \preceq \sigma(y_1, \dots, y_m) \preceq \max_{j=1,\dots,m} a_j^*(y_j; A, u),$$

where the minimum and maximum is with respect to the belief order. In other words, if $\underline{a} = \min_{j=1,\dots,m} a_j^*(y_j; A, u)$ and $\overline{a} = \max_{j=1,\dots,m} a_j^*(y_j; A, u)$,

$$B(\underline{a}) \le B(\sigma(y_1, \dots, y_m)) \le B(\overline{a}),$$

where the inequalities are in the strong set order.

Proof. See Appendix A.9. □

The above corollary says that the revealed beliefs of the robustly optimal strategy always lie between the most extreme beliefs from the individual signals. In particular, considering the worst-case correlation in the question posed by Piccione and Rubinstein (2024) will always lead to an answer between 20% and 70%, in contrast to the 98% if we assume conditional independence.

Another implication of the above corollary is that the robustly optimal strategy satisfies *unanimity*: if under a signal realization $\mathbf{y}$ all marginal information sources agree on the optimal action, i.e.

$$a = a_1^*(y_1) = \cdots = a_m^*(y_m),$$

then the robustly optimal strategy will prescribe that action. While unanimity seems like an appealing and natural property, it does not necessarily hold when the joint information structure is perfectly known to the decision maker. For instance, in the example above, if both tests led to 70% of newborns having the trait, the updated posterior would be even higher than 98%.

7 Discussion

This section discusses some extensions of our model. Section 7.1 discusses the implications of additional knowledge about the correlation structure. Section 7.2 shows that Theorem 1 extends to scenarios where the agent has even less knowledge about the information sources — introducing an additional layer of ambiguity regarding the marginal experiments. Section 7.3 considers the case where the information sources available to the agent have already been processed by experts. Section 7.4 studies the scenario where marginal experiments are perfect news signals.

7.1 Knowledge of Correlation

7.1.1 Common Origin

A natural reason for the correlation among multiple information sources is a shared, common origin. For instance, financial consultants may base their recommendations on the same dataset, leading to correlations among their recommendations. If we know that a common origin is the *only* possible channel generating the correlation among information sources, does this additional knowledge help restrict the presumed set of correlations? In other words, what types of correlation structures can be rationalized by sharing a common origin?

Formally, we say a joint experiment $P \in \mathcal{J}(P_1, ..., P_m)$ is *rationalizable by a common origin* if there exists $Q : \Theta \to \Delta X$ and a collection, $\{\gamma_j : X \to \Delta(Y_j)\}_j$, such that

$$P(y_1, ..., y_m|\theta) = \sum_x \prod_{j=1}^m \gamma_j(y_j|x)Q(x|\theta).$$

The interpretation is that Q is the common but unknown origin, and the experiments $P_1, ..., P_j$ are generated by *independent* garblings of signals from Q.

We have the following straightforward observation.

Observation. *Every $P \in \mathcal{J}(P_1, ..., P_m)$ is rationalizable by a common origin.*

To see why, note that we can let the common source Q be P itself, and the garblings γ_j be the deterministic functions that project each vector $y_1, ..., y_m$ onto y_j. Therefore, sharing a common origin does not exclude any possible correlation.

7.1.2 Partial Knowledge of Correlations

In certain situations, an agent may understand the correlation among some information sources, even if she does not comprehend all of them. For example, in medical diagnoses, older technologies such as X-rays and MRI have well-understood correlations. On the other hand, newer technologies, such as genetic sequencing, may have correlations with these traditional tests that are not yet fully understood.

In the context of our model, such knowledge can be modeled as imposing additional constraints on the set of conceived joint experiments $\mathcal{J}(P_1, ..., P_m)$. A simple case in which our results extend in a straightforward manner is the following: Suppose that there is a partition, $\Pi = \{S_1, \dots, S_k\}$, of $\{1, 2, \dots, m\}$ such that for all $S \in \Pi$, the agent knows that joint distribution over signals in S is given by:

$$\sum_{y_{-S}} P(y_S, y_{-S}|\theta) = P_S(y_S|\theta).$$

Then the set of conceived information structures is given by:

$$\left\{ P \in \mathcal{J}(P_1, \ldots, P_m) : \sum_{y_{-S}} P(y_S, y_{-S}|\theta) = P_S(y_S|\theta), \forall \theta, S \in \Pi, y_S \in Y_S \right\}.$$

But note that we could treat each joint experiment, $P_{S_1}, \ldots, P_{S_k}$, as a separate marginal experiment and use our previous analysis.

However, our analysis does not immediately extend to other, more complex situations. In particular, when the knowledge on the correlations spans across non-disjoint subsets, the set of possible joint experiments cannot be treated by replacing a subset of experiments with a single experiment, and our existing results no longer apply. To illustrate, suppose there are three information sources, $\{P_1, P_2, P_3\}$, and that the agent knows that P_1 and P_2 are correlated according to $P_{12} : \Theta \to \Delta(Y_1 \times Y_2)$, and that P_2 and P_3 are correlated according to $P_{23} : \Theta \to \Delta(Y_2 \times Y_3)$. The set of feasible joint experiments would be

$$\left\{ P : \Theta \to \Delta(Y_1 \times Y_2 \times Y_3) \;\middle|\; \begin{array}{l} \sum_{y_3} P(y_1, y_2, y_3|\theta) = P_{12}(y_1, y_2|\theta), \forall \theta, y_1, y_2 \\ \sum_{y_1} P(y_1, y_2, y_3|\theta) = P_{23}(y_2, y_3|\theta), \forall \theta, y_2, y_3 \end{array} \right\}.$$

An interesting direction for future research would be to consider general restrictions on the set of correlation structures derived from causality diagrams (see Pearl (2009) and Spiegler (2016)).

7.2 Ambiguity about Marginals

Our model so far assumes that the agent understands each information source precisely; that is, she knows P_j for $j = 1, ..., m$. In this section, we extend our model to allow for additional ambiguity about the marginal information sources.

Let $\mathcal{P}_j$ denote the set of possible marginal experiments for information source $j = 1, ..., m$. Let all $P_j \in \mathcal{P}_j$ have the same finite signal space Y_j. In addition, each $\mathcal{P}_j$ is assumed to be convex. That is, if $P_j : \Theta \to \Delta(Y_j)$ and $P'_j : \Theta \to \Delta(Y_j)$ are both in $\mathcal{P}_j$, then for any $\lambda \in (0, 1)$, $Q_\lambda : \Theta \to \Delta(Y_j)$ defined as $\theta \mapsto \lambda P_j(\cdot|\theta) + (1 - \lambda) P'_j(\cdot|\theta)$ is also in $\mathcal{P}_j$.

The agent conceives of the following set of joint experiments:

$$\mathcal{J}(\mathcal{P}_1, ..., \mathcal{P}_m) = \left\{ P : \Theta \to \Delta(\mathbf{Y}) : \exists P_j \in \mathcal{P}_j, \sum_{-j} P(y_1, \ldots, y_m|\theta) = P_j(y_j|\theta) \text{ for all } \theta, j, y_j \right\}.$$

The agent's decision problem is similarly defined:

$$V(\mathcal{P}_1, \ldots, \mathcal{P}_m) := \max_{\sigma:\mathbf{Y}\to\Delta(A)} \min_{P\in\mathcal{J}(\mathcal{P}_1,\ldots,\mathcal{P}_m)} \sum_{\theta\in\Theta} \sum_{\mathbf{y}\in\mathbf{Y}} P(\mathbf{y}|\theta) u(\theta, \sigma(\mathbf{y})).$$

We show that the prediction in Theorem 1 is robust to this additional layer of ambiguity.

Proposition 1. *For all* (A, u) *with* $|A| = |\Theta| = 2$*,*

$$V(\mathcal{P}_1, \ldots, \mathcal{P}_m) = \max_{j=1,\ldots,m} V(\mathcal{P}_j).$$

Proof. See Appendix B.3. □

7.3 Aggregating Experts' opinions

In certain instances, an agent may not have the expertise to process raw information sources. Instead, she may rely on experts who understand the information sources to offer their opinions, such as in the form of beliefs (e.g., doctors offering beliefs on the likelihood of a successful surgery) or action recommendations (e.g., financial consultants providing investment recommendations).

Reporting beliefs and offering action recommendations can both be viewed as garblings of the original, raw information sources. For any given information source $P_j : \Theta \to \Delta(Y_j)$, we call the induced *belief information structure*, denoted by $B_{P_j} : \Theta \to \Delta(\Theta)$, as the information structure derived by garbling each signal into the corresponding induced beliefs. In addition, we call the induced *recommendation information structure*, denoted by $R_{P_j} : \Theta \to \Delta A$, as the information structure derived by a garbling σ_j^*, given by an optimal strategy:

$$\sigma_j^* \in \operatorname*{argmax}_{\sigma_j : Y_j \to A} \sum_{\theta, y_j} P_j(y_j|\theta) u(\theta, \sigma_j(y_j)).$$

Note that, in contrast to the belief information structure, the recommendation information structure depends on the decision problem.

When the agent has access to only a single source of information, garbling information through reporting beliefs or action recommendations does not hurt the agent, that is, $V(P_j) = V(B_{P_j}) = V(R_{P_j})$ for any j. This is because beliefs and action recommendations already contain all the information needed to make an optimal decision.

When multiple information sources are available, garbling information by reporting only beliefs or recommendations could potentially hurt payoffs because some of the lost information, which is not useful on its own, could become valuable when combined with other sources. This begs the question of whether the agent could still achieve the same value as if she had access to the raw information sources. In other words, does

$$V(P_1, ..., P_m) = V(B_{P_1}, ..., B_{P_m}) = V(R_{P_1}, ..., R_{P_m})$$

hold when $m > 1$?

First, it is indeed the case that $V(P_1, ..., P_m) = V(B_{P_1}, ..., B_{P_m})$: since P_j is Blackwell equivalent to B_{P_j} for all j, Lemma 2 implies the values $V(P_1, ..., P_m)$ and $V(B_{P_1}, ..., B_{P_m})$ must be equal. The relationship between $V(R_{P_1}, ..., R_{P_m})$ and $V(P_1, ..., P_m)$ is more interesting: when $|\Theta| = 2$, these values coincide, but in general, we could have $V(R_{P_1}, ..., R_{P_m}) < V(P_1, ..., P_m)$.

Proposition 2. *When $|\Theta| = 2$, for any (A, u),*

$$V(P_1, ..., P_m) = V(R_{P_1}, ..., R_{P_m}).$$

Proof. See Appendix B.4. □

When there are three or more states, the recommendation information structure could generate a strictly lower value than the raw information structure. This can be seen by revisiting Example 2. Recall that in the example, under both P_X and P_Y, $a = 1$ is the unique optimal action to any signal realization. Therefore, both R_{P_X} and R_{P_Y} are uninformative experiments, and so $V(R_{P_X}, R_{P_Y}) = 1 - 0.9 + 1 = 1.1$. By contrast, the agent obtains perfect information when observing the raw information structures, and thus $V(P_X, P_Y) = 1 + 0 + 1 = 2 > V(R_{P_X}, R_{P_Y})$.

7.4 Perfect News

Thus far, we have made no assumptions about the marginal experiments. In this section, we study a specific parametric class: perfect news signals. We say that P with signal space $Y = \{y_\theta : \theta \in \Theta\} \cup \{y_\emptyset\}$ is a **perfect news signal** if for all $\theta \in \Theta$, $P(y_{\theta'} \mid \theta) = 0$ for all $\theta' \neq \theta$. In words, a perfect news signal either discloses a state perfectly or sends a null message. When all information sources are perfect news signals, then the characterization of the robustly optimal value simplifies substantially.

Proposition 3. *Suppose that $P_1, \dots, P_m$ are all perfect news signal structures. For each θ, let*

$$P^*_\theta \in \arg\max \{P^*_1(y_\theta \mid \theta), \dots, P^*_m(y_\theta \mid \theta)\}.$$

Let $\hat{\mu}$ be such that

$$\sum_{\theta' \in \Theta} \mu_0(\theta') P^*_{\theta'}(y_{\theta'} \mid \theta')\delta_{\theta'} + \left(1 - \sum_{\theta' \in \Theta} \mu_0(\theta') P^*_{\theta'}(y_{\theta'} \mid \theta')\right) \hat{\mu} = \mu_0.$$

Then

$$\begin{aligned} V(P_1, \dots, P_m) &= V\left(\{P^*_\theta\}_{\theta \in \Theta}\right) \\ &= \left(1 - \sum_{\theta' \in \Theta} \mu_0(\theta') P^*_{\theta'}(y_{\theta'} \mid \theta')\right) v\left(\hat{\mu}\right). \end{aligned}$$

The above proposition tells us that only those information sources that maximize the arrival rate of perfect news in each state are used under the robustly optimal strategy. Notice that, unlike in Sections 4 and 5, this result holds independently of the decision problem.

Proof. See Appendix B.5. □

8 Conclusion

Our findings have both normative and positive implications.

In a normative sense, there are settings in which decisions have to be made in highly uncertain environments, where correlations between information sources are hard to ascertain, and worst-case performance is of primary concern. Examples include the design of artificial intelligence and robotic agents, as well as the formulation of public guidelines. Our results can help guide such decision-makers on how to construct robustly optimal strategies from best-source strategies.

In a positive sense, our work offers an alternative rationale for information neglect, with implications that differ from existing explanations. For example, in models of rational inattention (see Maćkowiak, Matějka, and Wiederholt (2023) for a survey), higher stakes lead agents to acquire and use more information. In contrast, in our model, multiplying the utility function by any constant does not alter the set of information sources attended to. This distinction helps account for why information is ignored even in high-stakes decision problems.[20]

[20]For example, Olver et al. (2020) found that only 16.1% of patients sought a second opinion about their cancer treatment.

References

Itai Arieli, Yakov Babichenko, and Rann Smorodinsky. Robust forecast aggregation. *Proceedings of the National Academy of Sciences*, 115(52):E12135–E12143, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1813934115. URL https://www.pnas.org/content/115/52/E12135.

Itai Arieli, Yakov Babichenko, Inbal Talgam-Cohen, and Konstantin Zabarnyi. A random dictator is all you need. *arXiv preprint arXiv:2302.03667*, 2023.

Dirk Bergemann, Benjamin Brooks, and Stephen Morris. The limits of price discrimination. *American Economic Review*, 105(3):921–957, 2015.

Nils Bertschinger and Johannes Rauh. The blackwell relation defines no lattice. *2014 IEEE International Symposium on Information Theory*, pages 2479–2483, 2014.

David Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272, 1953.

Tilman Börgers, Angel Hernando-Veciana, and Daniel Krähmer. When are signals complements or substitutes? *Journal of Economic Theory*, 148(1):165–195, 2013.

Benjamin Brooks, Alexander Frankel, and Emir Kamenica. Comparisons of signals. *American Economic Review (Forthcoming)*, 2024.

Gabriel Carroll. Robustness and separation in multidimensional screening. *Econometrica*, 85(2): 453–488, 2017.

Xienan Cheng and Tilman Börgers. Dominance and optimality. *Working paper*, 2024.

Xienan Cheng and Tilman Börgers. Diversity, disagreement, and information aggregation. *Working Paper*, 2024.

Henrique de Oliveira. Blackwell's informativeness theorem using diagrams. *Games and Economic Behavior*, 109:126–131, 2018.

Larry G Epstein and Yoram Halevy. Ambiguous correlation. *The Review of Economic Studies*, 86(2):668–693, 2019.

Itzhak Gilboa and David Schmeidler. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

Branko Grünbaum. *Convex Polytopes*, volume 221. Springer Science & Business Media, 2003.

Benjamin Handel and Joshua Schwartzstein. Frictions or mental gaps: what's behind the information we (don't) use and when do we care? *Journal of Economic Perspectives*, 32(1):155–78, 2018.

Wei He and Jiangtao Li. Correlation-robust auction design. *Working paper*, 2020.

Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer science & business media, 1996.

Shota Ichihashi. The economics of data externalities. *Journal of Economic Theory*, 196:105316, 2021.

Robert P Kertz and Uwe Rösler. Stochastic and convex orders and lattices of probability measures, with a martingale interpretation. *Israel Journal of Mathematics*, 77:129–164, 1992.

Gilat Levy and Ronny Razin. Combining forecasts in the presence of ambiguity over correlation structures. *Journal of Economic Theory*, page 105075, 2020.

Annie Liang and Xiaosheng Mu. Complementary information and learning traps. *The Quarterly Journal of Economics*, 135(1):389–448, 2020.

Annie Liang, Xiaosheng Mu, and Vasilis Syrgkanis. Dynamically aggregating diverse information. *Econometrica*, 90(1):47–80, 2022.

Elliot Lipnowski and Laurent Mathevet. Simplifying bayesian persuasion. *Unpublished Paper, Columbia University.[642]*, 2017.

Bartosz Maćkowiak, Filip Matějka, and Mirko Wiederholt. Rational inattention: A review. *Journal of Economic Literature*, 61(1):226–273, 2023.

Ian Olver, Mariko Carey, Jamie Bryant, Allison Boyes, Tiffany Evans, and Rob Sanson-Fisher. Second opinions in medical oncology. *BMC Palliative Care*, 19:1–6, 2020.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Michele Piccione and Ariel Rubinstein. Failing to correctly aggregate signals. *Available at SSRN 4795431*, 2024.

R Tyrrell Rockafellar. *Convex analysis*, volume 36. Princeton university press, 1970.

Nate Silver. Election update: Why our model is more bullish than others on trump, 2016. URL https://web.archive.org/web/20161105210912/http://fivethirtyeight.com/features/election-update-why-our-model-is-more-bullish-than-others-on-trump/.

Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematic*, 8:171–176, 1958.

Joel Sobel. On the relationship between individual and group decisions. *Theoretical Economics*, 9(1):163–185, 2014.

Ran Spiegler. Bayesian networks and boundedly rational expectations. *The Quarterly Journal of Economics*, 131(3):1243–1290, 2016.

Time. President trump just said this poll was the 'most inaccurate' around the election. it wasn't, 2017. URL https://time.com/4860080/donald-trump-approval-ratings-poll-twitter-response/.

Abraham Wald. *Statistical decision functions.* Wiley, 1950.

Sam Wang. Is 99% a reasonable probability?, 2016. URL https://web.archive.org/web/20161110051827/http://election.princeton.edu/2016/11/06/is-99-a-reasonable-probability/comment-page-1#comment-107218.

Sean Jeremy Westwood, Solomon Messing, and Yphtach Lelkes. Projecting confidence: How the probabilistic horse race confuses and demobilizes the public. *The Journal of politics*, 82(4): 1530–1544, 2020.

Wanchang Zhang. Correlation-robust optimal auctions. *arXiv preprint arXiv:2105.04697*, 2021.

Günter M Ziegler. *Lectures on polytopes*, volume 152. Springer Science & Business Media, 2012.

A Appendix

A.1 Proof of Lemma 2

Proof. It suffices to show that for any $Q \in \mathcal{D}(P_1, ..., P_m)$, there exists $P \in \mathcal{J}(P_1, ..., P_m)$ such that P is Blackwell dominated by Q.

Take any $Q \in \mathcal{D}(P_1, ..., P_m)$ and let X be the signal space of Q. By Blackwell's Theorem, there exist $\gamma_j : X \to \Delta Y_j$ such that for each j,

$$P_j(y_j|\theta) = \sum_x \gamma_j(y_j|x)Q(x|\theta).$$

Define the following joint Blackwell experiment $P : \Theta \to \Delta(Y_1 \times ... \times Y_m)$:

$$P(y_1, ..., y_m|\theta) = \sum_x \prod_{j=1}^{m} \gamma_j(y_j|x)Q(x|\theta).$$

Clearly, $P \in \mathcal{J}(P_1, ..., P_m)$ because $\sum_{y_{-j}} P(y_1, ..., y_m|\theta) = \sum_x \gamma_j(y_j|x)Q(x|\theta) = P_j(y_j|\theta)$. Moreover, $\prod_{j=1}^{m} \gamma_j(y_j|x)$ defines a garbling, so P is Blackwell Dominated by Q. □

A.2 Proof of Lemma 3

Proof. To reduce notation, let's write $\mathbb{E}_P[u_\ell(\theta, \sigma_\ell)] = \sum_{\theta,\mathbf{y}} u_\ell(\theta, \sigma_\ell(\mathbf{y}))P(\mathbf{y}|\theta)$. Since $\sigma = (\sigma_\ell)_{\ell=1}^{k}$ is a feasible strategy,

$$\begin{aligned} V\left(P_1, \ldots, P_m; \bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)\right) &\geq \min_{P \in \mathcal{J}(P_1,\ldots,P_m)} \sum_{\ell=1}^{k} \mathbb{E}_P[u_\ell(\theta, \sigma_\ell)] \\ &\geq \sum_{\ell=1}^{k} \min_{P \in \mathcal{J}(P_1,\ldots,P_m)} \mathbb{E}_P[u_\ell(\theta, \sigma_\ell)] \\ &= \sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j; (A_\ell, u_\ell)). \end{aligned}$$

Moreover, by Theorem 1 and Corollary 1,

$$\sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j;(A_\ell,u_\ell)) = \sum_{\ell=1}^{k} V(\overline{P}(P_1,...,P_m);(A_\ell,u_\ell))$$

$$= V\left(\overline{P}(P_1,\ldots,P_m); \bigoplus_{\ell=1}^{k}(A_\ell,u_\ell)\right)$$

$$\geq V\left(P_1,\ldots,P_m; \bigoplus_{\ell=1}^{k}(A_\ell,u_\ell)\right).$$

Together, these inequalities prove our claim that

$$V\left(P_1,\ldots,P_m; \bigoplus_{\ell=1}^{k}(A_\ell,u_\ell)\right) = \sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j;(A_\ell,u_\ell))$$

and that σ is a robustly optimal strategy.

□

A.3 Proof of Lemma 4

Proof. Consider the binary decomposition $\bigoplus_{\ell=1}^{n-1}(A_\ell,u_\ell)$. We prove that for any $\boldsymbol{\delta} \in \{0,1\}^{n-1}$, $\sum_{\ell=1}^{n-1} \boldsymbol{\delta}_\ell u_\ell(\cdot,1) \in \mathcal{H}(A,u)$.

Suppose, by way of contradiction, that there exists $\boldsymbol{\delta} \in \{0,1\}^{n-1}$ for which $u^* := \sum_{\ell=1}^{n-1} \boldsymbol{\delta}_\ell u_\ell(\cdot,1) \notin \mathcal{H}(A,u)$. Since $\mathcal{H}(A,u)$ is convex and closed, we can strictly separate it from the singleton u^*(Corollary 11.4.2 of Rockafellar (1970)), i.e. there exists $\lambda \in \mathbb{R}^2 \setminus \{(0,0)\}$ such that

$$\lambda \cdot u^* > \sup_{v\in\mathcal{H}(A,u)} \lambda \cdot v. \tag{6}$$

Note that $\lambda \geq 0$ since otherwise $\sup_{v\in\mathcal{H}(A,u)} \lambda \cdot v = +\infty$.

From the ordering of the actions and the binary decomposition, $u_\ell(\theta_2,1)/u_\ell(\theta_1,1)$ is decreasing in ℓ. Therefore, for any $\ell' > \ell$,

$$\lambda \cdot u_\ell(\cdot,1) \leq 0 \Longrightarrow \lambda \cdot u_{\ell'}(\cdot,1) \leq 0.$$

So there exists ℓ^* such that $\lambda \cdot u_\ell(\cdot,1) > 0$ for $\ell < \ell^*$ and $\lambda \cdot u_\ell(\cdot,1) \leq 0$ for $\ell \geq \ell^*$.

Thus

$$\max_{\boldsymbol{\delta}'\in\{0,1\}^{n-1}} \sum_{\ell=1}^{n-1} \lambda \cdot \boldsymbol{\delta}'_\ell u_\ell(\cdot,1)$$

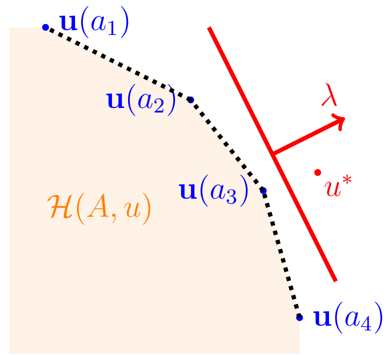


Figure 6

is solved by choosing $\boldsymbol{\delta}'_\ell = 1$ for $\ell < \ell^*$ and $\boldsymbol{\delta}'_\ell = 0$ for $\ell \geq \ell^*$. Hence

$$\lambda \cdot u(\cdot, a_{\ell^*}) = \lambda \cdot \sum_{\ell=1}^{\ell^*-1} u_\ell(\cdot, 1) \geq \lambda \cdot \sum_{\ell=1}^{n-1} \boldsymbol{\delta}_\ell u_\ell(\cdot, 1) = \lambda \cdot u^*.$$

But $u(\cdot, a_{\ell^*}) \in \mathcal{H}(A, u)$, contradicting (6). □

A.4 Proof of Theorem 2

Proof. From Lemma 4, (A, u) is equivalent to $\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)$, so

$$V(P_1, ..., P_m; (A, u)) = V\left(P_1, ..., P_m; \bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)\right) = \sum_{\ell=1}^{n-1} \max_{j=1,...,m} V(P_j; (A_\ell, u_\ell)),$$

where the second equality follows from Lemma 3. This establishes the first statement of the theorem.

By the definition of dominating maps, $u(\cdot, \sigma_f^*(\mathbf{y})) \geq \sum_{\ell=1}^{n-1} u_\ell(\cdot, \sigma_\ell(\mathbf{y}))$ for each $\mathbf{y}$. For any $P \in \mathcal{J}(P_1, ..., P_m)$,

$$\begin{aligned}
\mathbb{E}_P\left[u(\theta, \sigma_f^*(\mathbf{y}))\right] &\geq \mathbb{E}_P\left[\sum_{\ell=1}^{n-1} u_\ell(\theta, \sigma_\ell(\mathbf{y}))\right] \\
&= V\left(P_1, ..., P_m; \bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)\right) \\
&= V\left(P_1, ..., P_m; (A, u)\right)
\end{aligned}$$

where the second line follows from Lemma 3 and the third line follows from Lemma 4. So σ_f^* is a robustly optimal strategy. □

A.5 Proof of Theorem 3

A.5.1 Posterior Polytopes

Let $\succeq_{MPS}$ denote the mean-preserving spread order on posterior distributions. For any experiment P, define $\mathcal{E}_v(P) = \{\tau \in \Delta(\Delta\Theta) \mid \tau \succeq_{\text{MPS}} \tau^P \text{ and } \text{supp}(\tau) \subseteq E_v\}$ as the set of posterior distributions that are mean-preserving spreads of τ^P and whose supports are contained in the set of extremal beliefs E_v. This set is non-empty because it always contains the posterior distribution induced by a perfectly informative experiment.

Note that $\mathcal{E}_v(P)$ resides in a $|K_v|$-dimensional space, since by the martingale constraint, the probabilities that a posterior distribution assigns to the $|K_v|$ nondegenerate extremal beliefs pin down the probabilities assigned to the degenerate extremal beliefs.[21] In addition, because the mean-preserving spread constraints are linear, $\mathcal{E}_v(P)$ is a $|K_v|$-dimensional polytope. Hence, we call $\mathcal{E}_v(P)$ the *posterior polytope* induced by experiment P. We will sometimes use the vector notation $\boldsymbol{\tau} \in \mathbb{R}_+^{|K_v|}$ to denote a typical element in $\mathcal{E}_v(P)$, whose components are probabilities assigned to the nondegenerate extremal beliefs.

We first present a lemma that allows us to simplify Nature's problem into choosing elements in the intersection of the posterior polytopes induced by the marginal experiments.

Lemma 5.

$$V(P_1, ..., P_m; v) = \min_{\tau \succeq_{MPS} \tau_{P_1}, ..., \tau_{P_m}} \int v(\mu)\tau(d\mu) = \min_{\tau \in \cap_{j=1}^m \mathcal{E}_v(P_j)} \sum_{\mu \in E_v} \tau(\mu)v(\mu).$$

Proof. From Eq. (5), we have $V(P_1, ..., P_m; v) = \min_{P \in \mathcal{J}(P_1,...,P_m)} \int v(\mu)\tau^P(d\mu)$. By Lemma 2, Nature's problem can be relaxed to choosing an experiment that Blackwell dominates all marginal experiments; that is, $V(P_1, ..., P_m; v) = \min_{P \in D(P_1,...,P_m)} \int v(\mu)\tau^P(d\mu)$. A posterior distribution can be induced by a Blackwell experiment that dominates all marginals if and only if it is a mean-preserving spread of the posterior distributions induced by each marginal experiment (Blackwell, 1953). This yields the first equality in the lemma.

To see the second equality, clearly

$$\min_{\tau \succeq_{\text{MPS}} \tau_{P_1}, ..., \tau_{P_m}} \int v(\mu)\tau(d\mu) \leq \min_{\tau \in \cap_{j=1}^m \mathcal{E}_v(P_j)} \sum_{\mu \in E_v} \tau(\mu)v(\mu).$$

To establish the reverse inequality, consider any $\tau \succsim_{MPS} \tau_{P_1}, \ldots, \tau_{P_m}$. Lemma 7 implies that any posterior $\mu \notin E_v$ can be split into nearby extremal beliefs without changing the value. Hence,

[21] Formally, for any posterior distribution τ supported on E_v with a mean μ_0, we must have $\sum_{\mu \in K_v} \tau(\mu) \leq 1$ and $\tau(\delta_\theta) = \mu_0(\theta) - \sum_{\mu \in K_v} \tau(\mu)\mu(\theta)$ for all θ. Conversely, for any $t \in [0, 1]^{K_v}$ such that $\sum_{\mu \in K_v} t(\mu) \leq 1$, by letting $\tau(\mu) = t(\mu)$ for $\mu \in K_v$ and $\tau(\delta_\theta) = \mu_0(\theta) - \sum_{\mu \in K_v} t(\mu)\mu(\theta)$ for all θ, τ is a well-defined posterior distribution.

there exists $\tilde{\tau} \succsim_{MPS} \tau$ with $\tilde{\tau} \in \Delta(E_v)$ such that

$$\sum_{\mu \in E_v} v(\mu)\tilde{\tau}(\mu) = \int v(\mu)\tau(d\mu).$$

Moreover by transitivity, $\tilde{\tau} \succsim_{MPS} \tau_{P_1}, \dots, \tau_{P_m}$ and so, $\tilde{\tau} \in \bigcap_{j=1}^m \mathcal{E}_v(P_j)$. Thus, the reverse inequality follows. □

Lemma 6. *Let* $V = \{(\mu, w) \in epi(v) : w \leq 0, \mu \in \Delta(\Theta)\}$. *Then for every* $(\mu, w) \in ext(V)$, $w = v(\mu)$ *and* $\mu \in E_v$.

Proof. The first claim is obvious. For the second claim, if $\mu \in \delta_\theta$ for some $\theta \in \Theta$, we are done. So let us assume that $\mu \neq \delta_\theta$.

In this case, note that $w < 0$. Otherwise, $(\mu, 0)$ can be represented as a convex combination of $\{(\delta_\theta, v(\delta_\theta))\}_{\theta \in \Theta}$, which contradicts the assumption that $(\mu, w) \in ext(V)$. Now consider $(\mu_1, w_1), (\mu_2, w_2) \in epi(v)$ and some $\alpha \in (0, 1)$ such that

$$\alpha(\mu_1, w_1) + (1 - \alpha)(\mu_2, w_2) = (\mu, w).$$

Letting $\mu_i(\beta) = \beta\mu_i + (1 - \beta)\mu$ and $w_i(\beta) = \beta w_i + (1 - \beta)w$, we have for any $\beta \in (0, 1)$, $(\mu_i(\beta), w_i(\beta)) \in epi(v)$ and

$$\alpha\big(\mu_1(\beta), w_1(\beta)\big) + (1 - \alpha)\big(\mu_2(\beta), w_2(\beta)\big) = (\mu, w).$$

Moreover, for β sufficiently small, $w_i(\beta) < 0$, and so $(\mu_i(\beta, w_i(\beta)) \in V$ for each $i = 1, 2$. Because $(\mu, w) \in ext(V)$, $(\mu_1(\beta), w_1(\beta)) = (\mu_2(\beta), w_2(\beta))$. This however, implies that $(\mu_1, w_1) = (\mu_2, w_2)$. Hence (μ, w) is an extreme point of $epi(v)$ and therefore, $\mu \in E_v$. □

Lemma 7 (Splitting Lemma)**.** *For every* τ*, there exists some* $\tilde{\tau} \succsim_{MPS} \tau$ *such that* $\tilde{\tau} \in \Delta(E_v)$ *with*

$$\sum_{\mu \in E_v} v(\mu)\tilde{\tau}(\mu) = \int v(\mu)\tau(d\mu).$$

Proof. Consider any $\mu \in \Delta(\Theta)$. Since V is convex and compact, there exists a finite collection, $(\mu_1, w_1), \dots, (\mu_k, w_k) \in ext(V)$ and some $\lambda_1, \dots, \lambda_k \in [0, 1]$ with $\sum_{\ell=1}^k \lambda_\ell = 1$ such that

$$(\mu, v(\mu)) = \sum_{\ell=1}^k \lambda_\ell(\mu_\ell, w_\ell).$$

By Lemma 6, $\mu_\ell \in E_v$ and $w_\ell = v(\mu_\ell)$. Hence, for every $\mu \in \Delta(\Theta)$, we have shown the existence of $\lambda(\cdot \mid \mu) \in \Delta(E_v)$ such that $(\mu, v(\mu)) = \sum_{\nu \in E_v} \gamma(\nu \mid \mu)(\nu, v(\nu))$.

Now define $\tilde{\tau} \in \Delta(E_v)$ as $\tilde{\tau}(\nu) = \int \gamma(\nu \mid \mu)\tau(d\mu)$. Clearly, $\tilde{\tau} \succsim_{MPS} \tau$ and

$$\int v(\nu)\tilde{\tau}(d\nu) = \int \left(\sum_{\nu \in E_v} v(\nu)\gamma(\nu \mid \mu) \right) \tau(d\mu) = \int v(\mu)\tau(d\mu).$$

□

A.5.2 Support Functions

For any convex compact set $S \subseteq \mathbb{R}_+^{K_v}$, we define

$$\begin{aligned} h_S : \mathbb{R}_-^{K_v} &\to \mathbb{R} \\ \boldsymbol{v} &\mapsto \min_{\boldsymbol{\tau} \in S} \boldsymbol{\tau} \cdot \boldsymbol{v} \end{aligned} \tag{7}$$

as the support function of S.[22]

The next lemma shows that for any collection of convex sets $\{S_j\}_{j=1}^m$, the support function of their intersection, $h_{\cap_{j=1}^m S_j}(\lambda)$, can be characterized by the concavification of the pointwise maximum of the support functions of S_j.

Lemma 8. *Suppose $\{S_j\}_{j=1}^m$ is a collection of non-empty, compact, convex sets in $\mathbb{R}_+^{K_v}$ such that for any $s_j \in S_j$ (for $j = 1, ..., m$), their meet $\wedge_{j=1}^m s_j$ lies in $\cap_j S_j$. Then for any $\boldsymbol{v} \in \mathbb{R}_-^{K_v}$,*

$$h_{\cap_{j=1}^m S_j}(\boldsymbol{v}) = conc(\max\{h_{S_1}, ..., h_{S_m}\})(\boldsymbol{v}).$$

Proof. To simplify notation, let us denote $H(\cdot) = \max\{h_{S_1}(\cdot), \dots, h_{S_m}(\cdot)\}$. Clearly, $h_{\cap_{j=1}^m S_j}(\cdot)$ is concave as it is the minimum of linear functions. In addition, $h_{\cap_{j=1}^m S_j}(\cdot) \geq H(\cdot)$, because any solution to $\min_{\boldsymbol{\tau} \in \cap_{j=1}^m S_j} \boldsymbol{\tau} \cdot \boldsymbol{v}$ is feasible in $\min_{\boldsymbol{\tau} \in S_j} \boldsymbol{\tau} \cdot \boldsymbol{v}$. Therefore, $h_{\cap_{j=1}^m S_j}(\cdot) \geq conc(H)(\cdot)$.

For the reverse inequality, recall that $conc(H)(\cdot)$ is the point-wise infimum of all affine functions, g, that dominate H pointwise, i.e. $g(\boldsymbol{v}) \geq H(\boldsymbol{v})$ for all $\boldsymbol{v} \in \mathbb{R}_-^{K_v}$. Thus, it suffices to show that for any such function, we also have $g(\cdot) \geq h_{\cap_{j=1}^m S_j}(\cdot)$.

Consider any affine function $g(\boldsymbol{v}) = \lambda \cdot \boldsymbol{v} + c$ that dominates H pointwise. Note that for any $\beta > 0$, $\boldsymbol{v} \in \mathbb{R}_-^{K_v}$ and j, $0 \leq \frac{g(\beta \boldsymbol{v}) - h_{S_j}(\beta \boldsymbol{v})}{\beta} = (\lambda \cdot \boldsymbol{v} - h_{S_j}(\boldsymbol{v})) + \frac{c}{\beta}$. Taking $\beta \to 0$, we first note that $c \geq 0$. Moreover, taking $\beta \to \infty$, we have $\lambda \cdot \boldsymbol{v} \geq h_{S_j}(\boldsymbol{v})$. To summarize, we have shown that for all $\boldsymbol{v} \in \mathbb{R}_-^{K_v}$:

$$g(\boldsymbol{v}) \geq \lambda \cdot \boldsymbol{v} \geq h_{S_j}(\boldsymbol{v}) = h_{S_j^-}(\boldsymbol{v}), \tag{8}$$

where $S_j^- \equiv S_j - \mathbb{R}_+^{K_v}$. Since $\lambda \cdot \boldsymbol{v} \geq h_{S_j^-}(\boldsymbol{v})$ for all $\boldsymbol{v} \in \mathbb{R}_-^{K_v}$, we must have $\lambda \in S_j^-$. To see why, suppose toward a contradiction that $\lambda \notin S_j^-$. By the separating hyperplane theorem, there must

[22] Recall that we have normalized the value functions to be non-positive, so the domain here is restricted to the negative quadrant.

exist $r \in \mathbb{R}_-^{K_v}$ and $\alpha \in \mathbb{R}$ such that $\lambda \cdot r < \alpha < x \cdot r$ for all $x \in S_j^-$, which implies $\lambda \cdot r < h_{S_j^-}(r)$, a contradiction.

Hence, we have

$$g(\boldsymbol{v}) \geq \lambda \cdot \boldsymbol{v} \geq \inf_{s \in \cap_{j=1}^m S_j^-} s \cdot \boldsymbol{v} = h_{\cap_{j=1}^m S_j}(\boldsymbol{v})$$

where the last equality holds because, by assumption, for any $s \in \cap_{j=1}^m S_j^-$, there exists $s_j \in S_j$ such that $s \leq s_j$ for $j = 1, ..., m$; therefore, $s \leq \wedge_j s_j \in \cap_j S_j$, which implies that $\inf_{s \in \cap_{j=1}^m S_j^-} s \cdot \boldsymbol{v} = \min_{s \in \cap_{j=1}^m S_j} s \cdot \boldsymbol{v} = h_{\cap_{j=1}^m S_j}(\boldsymbol{v})$. This concludes the proof. □

We will apply Lemma 8 to $\cap_{j=1}^m \mathcal{E}_v(P_j)$, and the following lemma ensures that $\{\mathcal{E}_v(P_j)\}_{j=1}^m$ satisfies the desired properties.

Lemma 9. *For any $\boldsymbol{\tau}_j \in \mathcal{E}_v(P_j)$ (for $j = 1, ..., m$), their meet $\wedge_{j=1}^m \boldsymbol{\tau}_j$ lies in $\cap_{j=1}^m \mathcal{E}_v(P_j)$.*

Proof. Note that if $\boldsymbol{\tau}_j \in \mathcal{E}_v(P_j)$, and $\boldsymbol{\tau}'$ satisfies $0 \leq \boldsymbol{\tau}' \leq \boldsymbol{\tau}_j$, then $\boldsymbol{\tau}' \in \mathcal{E}_v(P_j)$. This is because any probability mass on nondegenerate extremal beliefs can be split into degenerate extremal beliefs, which yields a mean-preserving spread that retains its support on E_v. Since $0 \leq \wedge_{j=1}^m \boldsymbol{\tau}_j \leq \boldsymbol{\tau}_j$ for all j, it follows that $\wedge_{j=1}^m \boldsymbol{\tau}_j \in \cap_{j=1}^m \mathcal{E}_v(P_j)$. □

The next lemma characterizes $W(P; \boldsymbol{v})$ using support functions.

Lemma 10. *For any P and $\boldsymbol{v} \in \mathbb{R}_-^{K_v}$,*

$$W(P; \boldsymbol{v}) = h_{\mathcal{E}_v(P)}(\boldsymbol{v}).$$

Proof. Recall that $W(P; \boldsymbol{v}) = V(P; conv(\boldsymbol{v}))$. By Lemma 5,

$$W(P; \boldsymbol{v}) = V(P; conv(\boldsymbol{v})) = \min_{\tau \in \mathcal{E}_v(P)} \sum_{\mu \in E_v} conv(\boldsymbol{v})(\mu)\boldsymbol{\tau}(\mu).$$

Now let $\hat{E}_v = \{\mu \in E_v | conv(\boldsymbol{v})(\mu) = \boldsymbol{v}(\mu)\}$, which is non-empty as it always contains $\{\delta_\theta\}_{\theta \in \Theta}$. Let $\hat{\mathcal{E}}_v(P) = \{\tau \in \Delta(\Delta\Theta) | \tau \succeq_{\text{MPS}} \tau^P \text{ and } \text{supp}(\tau) \subseteq \hat{E}_v\}$.

We will establish the following:

$$\begin{aligned}
\min_{\boldsymbol{\tau} \in \mathcal{E}_v(P)} \sum_{\mu \in E_v} conv(\boldsymbol{v})(\mu)\boldsymbol{\tau}(\mu) &= \min_{\boldsymbol{\tau} \in \hat{\mathcal{E}}_v(P)} \sum_{\mu \in \hat{E}_v} conv(\boldsymbol{v})(\mu)\boldsymbol{\tau}(\mu) \\
&= \min_{\boldsymbol{\tau} \in \hat{\mathcal{E}}_v(P)} \sum_{\mu \in \hat{E}_v} \boldsymbol{v}(\mu)\boldsymbol{\tau}(\mu) \\
&\geq \min_{\boldsymbol{\tau} \in \mathcal{E}_v(P)} \sum_{\mu \in E_v} \boldsymbol{v}(\mu)\boldsymbol{\tau}(\mu) \\
&\geq \min_{\boldsymbol{\tau} \in \mathcal{E}_v(P)} \sum_{\mu \in E_v} conv(\boldsymbol{v})(\mu)\boldsymbol{\tau}(\mu);
\end{aligned}$$

hence, all hold with equality. The first equality follows from the fact that any μ such that $\boldsymbol{v}(\mu) > conv(\boldsymbol{v})(\mu)$ can be split into $\mu_i \in \hat{E}_v$ without changing the value. The second equality is due to $conv(\boldsymbol{v})(\mu) = \boldsymbol{v}(\mu)$ for all $\mu \in \hat{E}_v$. The second to last inequality holds because $\hat{\mathcal{E}}_v(P) \subseteq \mathcal{E}_v(P)$, and the last inequality holds because $conv(\boldsymbol{v}) \leq \boldsymbol{v}$.

Therefore, $W(P;\boldsymbol{v}) = \min_{\boldsymbol{\tau}\in\mathcal{E}_v(P)} \sum_{\mu\in E_v} \boldsymbol{v}(\mu)\boldsymbol{\tau}(\mu) = h_{\mathcal{E}_v(P)}(\boldsymbol{v})$, which concludes the proof.

□

A.5.3 Robustly Optimal Value

By Lemma 5, Lemma 8, Lemma 9, and Lemma 10, we have

$$\begin{aligned} V(P_1,...,P_m;v) &= \min_{\boldsymbol{\tau}\in\cap_{j=1}^m \mathcal{E}_v(P_j)} \boldsymbol{\tau}\cdot\boldsymbol{v}_v \\ &= h_{\cap_{j=1}^m \mathcal{E}_v(P_j)}(\boldsymbol{v}_v) \\ &= conc(\max\{h_{\mathcal{E}_v(P_1)},...,h_{\mathcal{E}_v(P_m)}\})(\boldsymbol{v}_v) \\ &= conc(\max\{W(P_1;\cdot),...,W(P_m;\cdot)\})(\boldsymbol{v}_v) \end{aligned}$$

and Theorem 3 follows.

A.6 Proof of Corollary 4

We first establish that (A,u) dominates $\bigoplus_{j=1}^m (A_j,u_j)$. Suppose, toward a contradiction, that $\mathcal{H}(\bigoplus_{j=1}^m (A_j,u_j)) \nsubseteq \mathcal{H}(A,u)$. Then there is some $r \in \mathcal{H}(\bigoplus_{j=1}^m (A_j,u_j))$ with $r \notin \mathcal{H}(A,u)$. By the separating hyperplane theorem, there exists $h \in \mathbb{R}_+^{|\Theta|}/\{\mathbf{0}\}$ such that $h\cdot r > h\cdot x$ for all $x \in \mathcal{H}(A,u)$. Normalizing h so that $h \in \Delta(\Theta)$, it follows that there exists $\alpha_j \in A_j$, $j = 1,...,m$ such that

$$\sum_\theta h(\theta)\sum_j u_j(\theta,\alpha_j) > \sum_\theta h(\theta)u(\theta,\alpha) \quad \text{for all } \alpha\in\Delta A.$$

This means that $\sum_{j=1}^m \lambda_j conv(\boldsymbol{v}_j)(h) > v(h) = conv(\boldsymbol{v}_v)(h) = conv(\sum_{j=1}^m \lambda_j^*\boldsymbol{v}_j^*)(h) \geq \sum_{j=1}^m \lambda_j conv(\boldsymbol{v}_j^*)(h)$, a contradiction.

Next, we show that for any dominating map, $f : \bigoplus_{j=1}^m (A_j,u_j) \to \Delta(A)$, the corresponding strategy $\sigma_f^*(y_1,\ldots,y_m) = f(\sigma_1^*(y_1),\ldots,\sigma_m^*(y_m))$ is a robustly optimal strategy. For any $P \in$

$\mathcal{J}(P_1, ..., P_m)$, the agent's expected payoff when playing σ_f^* is

$$\begin{aligned}
\sum_{\theta} \sum_{y_1,...,y_m} P(y_1, ..., y_m|\theta) u(\theta, \sigma_f^*(y_1, ..., y_m)) &\geq \sum_{\theta} \sum_{y_1,...,y_m} P(y_1, ..., y_m|\theta) \left[\sum_{j=1}^{m} u_j(\theta, \sigma_j^*(y_j)) \right] \\
&= \sum_{j=1}^{m} \sum_{y_j} P_j(y_j|\theta) u_j(\theta, \sigma_j^*(y_j)) \\
&= \sum_{j=1}^{m} \lambda_j^* W(P_j; \boldsymbol{v}_j^*) \\
&= V(P_1, \ldots, P_m; \boldsymbol{v}_v).
\end{aligned}$$

Therefore, σ_f^* is a robustly optimal strategy.

A.7 Proof of Corollary 5

Let $f(\cdot) := \max\{h_{\mathcal{E}_v(P_1)}, ..., h_{\mathcal{E}_v(P_m)}\}(\cdot)$. By Theorem 3,

$$V(P_1, ..., P_m; v) = conc(f)(\boldsymbol{v}_v) = \max\{r|(\boldsymbol{v}_v, r) \in co(G(f))\},$$

where $G(f) = \{(\boldsymbol{v}, f(\boldsymbol{v}))|\boldsymbol{v} \in \mathbb{R}_-^{K_v}\}$ is the graph of f.[23] Therefore, there exists $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$, and $\boldsymbol{v}_i \in \mathbb{R}_-^{K_v}$ such that $\sum_i \lambda_i \boldsymbol{v}_i = \boldsymbol{v}_v$ and $V(P_1, ..., P_m; v) = \sum_i \lambda_i f(\boldsymbol{v}_i)$.

Let $||\boldsymbol{v}|| = |\boldsymbol{v}_v \cdot \mathbf{1}| = -\boldsymbol{v}_v \cdot \mathbf{1}$ denote the L1 norm of a vector $\boldsymbol{v} \in \mathbb{R}_-^{K_v}$, and $\boldsymbol{V} = \{\boldsymbol{v} \in \mathbb{R}_-^{K_v} \mid ||\boldsymbol{v}|| = ||\boldsymbol{v}_v||\}$ denote the set of vectors that have the same norm as $\boldsymbol{v}_v$. Let $\hat{\lambda}_i := \lambda_i \frac{||\boldsymbol{v}_i||}{||\boldsymbol{v}_v||}$ and $\hat{\boldsymbol{v}}_i := \frac{||\boldsymbol{v}_v||}{||\boldsymbol{v}_i||} \boldsymbol{v}_i$. Note that f is a positive homogeneous function; that is, $f(\alpha \boldsymbol{v}) = \alpha f(\boldsymbol{v})$ for any $\alpha \geq 0$ and $\boldsymbol{v} \in \mathbb{R}_-^{K_v}$. Therefore, $conc(f)(\boldsymbol{v}_v) = \sum_i \lambda_i f(\boldsymbol{v}_i) = \sum_i \lambda_i \frac{||\boldsymbol{v}_i||}{||\boldsymbol{v}_v||} f(\frac{||\boldsymbol{v}_v||}{||\boldsymbol{v}_i||} \boldsymbol{v}_i) = \sum_i \hat{\lambda}_i f(\hat{\boldsymbol{v}}_i)$. Since $\hat{\boldsymbol{v}}_i \in \boldsymbol{V}$ and $\sum_i \hat{\lambda}_i = 1$, it follows that $(\boldsymbol{v}_v, conc(f)(\boldsymbol{v}_v))$ is in the convex hull of $G_{\boldsymbol{V}}(f) := \{(\boldsymbol{v}, f(\boldsymbol{v}))|\boldsymbol{v} \in \boldsymbol{V}\}$. Therefore, the concavification with domain on $\boldsymbol{V}$ obtains the same value; that is, $conc(f)(\boldsymbol{v}_v) = \max\{r|(\boldsymbol{v}_v, r) \in co(G_{\boldsymbol{V}}(f))\}$. Since $(\boldsymbol{v}_v, conc(f)(\boldsymbol{v}_v))$ lies on the boundary of $co(G_{\boldsymbol{V}}(f))$, there exists a supporting hyperplane crossing $(\boldsymbol{v}_v, conc(f)(\boldsymbol{v}_v))$. In addition, $(\hat{\boldsymbol{v}}_i, f(\hat{\boldsymbol{v}}_i))$ must all lie on this supporting hyperplane. Since the supporting hyperplane is of dimension $|K_v|-1$, by Carathéodory's theorem, there exists an index set I and some $\tilde{\lambda}_i \geq 0$, $\sum_{i \in I} \tilde{\lambda}_i$ such that $|I| \leq |K_v|$ and $(\hat{\boldsymbol{v}}_v, conc(f)(\hat{\boldsymbol{v}}_v)) = \sum_{i \in I} \tilde{\lambda}_i (\hat{\boldsymbol{v}}_i, f(\hat{\boldsymbol{v}}_i))$.

For each $\hat{\boldsymbol{v}}_i$, there exists a j such that $f(\hat{\boldsymbol{v}}_i) = h_{\mathcal{E}_v(P_j)}(\hat{\boldsymbol{v}}_i)$; we select such a j and denote it by $j(i)$. Let $J = \{j|j = j(i) \text{ for some } i\}$. Note that $|J| \leq |K_v|$, and that $V(\{P_j\}_{j \in J}; v) = conc(\max\{h_{\mathcal{E}_v(P_j)}; j \in J\}) \geq \sum_{i \in I} \tilde{\lambda}_i h_{\mathcal{E}_v(P_{j(i)})}(\hat{\boldsymbol{v}}_i) = \sum_{i \in I} \tilde{\lambda}_i f(\hat{\boldsymbol{v}}_i) = conc(f)(\boldsymbol{v}_v) = V(P_1, ..., P_m; v)$, and the corollary follows.

[23] We can take max instead of sup in the definition because f has a closed graph.

A.8 Proof of Corollary 6

If $V(P_\ell; v) \leq \max_{j\neq \ell; v} V(P_j; v)$ for all v, we have

$$\max\{W(P_1;\cdot),...,W(P_m;\cdot))\} = \max\{W(P_1;\cdot),...,W(P_{\ell-1};\cdot),W(P_{\ell+1};\cdot)...,W(P_m;\cdot))\}.$$

The corollary then follows immediately from Theorem 3.

A.9 Proof of Corollary 7

Before proving the result, we prove a lemma connecting the best-response to a signal in the original problem with the best responses in the decomposed problems.

Lemma 11. *Suppose that $a_i = a_j^*(y_j; A, u)$. Then, using only information source P_j, a best response to signal y_j for the decomposed decision problem (A_ℓ, u_ℓ) is $a_\ell = 1$ for $\ell = 1, \ldots, i-1$ and $a_\ell = 0$ for $\ell = i, \ldots, n-1$.*

Proof. Let $\mu \in \Delta(\Theta)$ be the belief that comes from observing y_j from P_j, so $a_i \in \operatorname{argmax}_{a\in A} \sum_\theta \mu(\theta)u(\theta, a)$. By Lemma 13 in Appendix B.4, for $\ell = 1, \ldots, i-1$, we have $\sum_\theta \mu(\theta)u(\theta, a_{\ell+1}) \geq \sum_\theta \mu(\theta)u(\theta, a_\ell)$, whereas for $\ell = i, \ldots, n-1$, we have $\sum_\theta \mu(\theta)u(\theta, a_\ell) \geq \sum_\theta \mu(\theta)u(\theta, a_{\ell+1})$. This gives the result by the definition of (A_ℓ, u_ℓ). □

In particular, using the belief order, it follows that if $a_j^*(y_j; A, u) \succeq a_i$, then $a_j^*(y_j; A_\ell, u_\ell) = 1$ for $\ell = 0, \ldots, i-1$; likewise, if $a_k \succeq a_j^*(y_j; A, u)$, then $a_j^*(y_j; A_\ell, u_\ell) = 0$ for $\ell = k, \ldots, n-1$.

Now let $\mathbf{y}$ be a signal realization for all information sources and

$$a_i = \min_j a_j^*(y_j; A, u) \qquad a_k = \max_j a_j^*(y_j; A, u),$$

where the minimum and maximum are with respect to the belief order. Then it follows that, for every j,

$$a_j^*(y_j; A_\ell, u_\ell) = \begin{cases} 1 & \text{if } \ell \leq i-1 \\ 0 \text{ or } 1 & \text{if } i \leq \ell \leq k-1 \\ 0 & \text{if } k \leq \ell. \end{cases}$$

Moreover, if $a_j^*(y_j; A_\ell, u_\ell) = a_\ell \in \{0, 1\}$ for all j, then there must be a pure best-source strategy σ_ℓ for (A_ℓ, u_ℓ) with $\sigma_\ell(\mathbf{y}) = a_\ell$. Hence there are best-source strategies σ_ℓ with $\sigma_\ell(\mathbf{y}) = 1$ for $\ell \leq i-1$ and $\sigma_\ell(\mathbf{y}) = 0$ for $k \leq \ell$.

Now let $f : \prod_{\ell=1}^{n-1} A_\ell \to \Delta(A)$ be a dominating map that always assigns strategies that are not weakly dominated by any other mixed strategy. By Theorem 2, $\sigma_f^*(\mathbf{y}) := f(\sigma_1(\mathbf{y}), \ldots, \sigma_{n-1}(\mathbf{y}))$

is a robustly optimal strategy for (A, u). Letting $a_\ell = \sigma_\ell(\mathbf{y})$, we have

$$\begin{aligned} u\left(\theta, f[\sigma_1(\mathbf{y}), \ldots, \sigma_{n-1}(\mathbf{y})]\right) &= \sum_{\ell=1}^{i-1} u_\ell(\theta, 1) + \sum_{\ell=i}^{k-1} u_\ell(\theta, \sigma_\ell(\mathbf{y})) + \sum_{\ell=k}^{n-1} u_\ell(\theta, 0) \\ &= \sum_{\ell=1}^{i-1} [u(\theta, a_{\ell+1}) - u(\theta, a_\ell)] + \sum_{\ell=i}^{k-1} u_\ell(\theta, \sigma_\ell(\mathbf{y})) + 0 \\ &= u(\theta, a_i) + \sum_{\ell=i}^{k-1} u_\ell(\theta, \sigma_\ell(\mathbf{y})). \end{aligned}$$

Notice that, given our normalization, $u_\ell(\theta_1, 1) > 0$ and $u_\ell(\theta_2, 1) < 0$ for every ℓ,. Therefore, the equality above implies

$$u\left(\theta_1, f\left(\sigma_1(\mathbf{y}), \ldots, \sigma_{n-1}(\mathbf{y})\right)\right) \geq u(\theta_1, a_i)$$

and

$$u\left(\theta_2, f\left(\sigma_1(\mathbf{y}), \ldots, \sigma_{n-1}(\mathbf{y})\right)\right) \geq u(\theta_2, a_i) + \sum_{\ell=i}^{k-1} u_\ell(\theta_2, 1) = u(\theta_2, a_k).$$

We now prove another lemma.

Lemma 12. *Let $\alpha \in \Delta(A)$ be not weakly dominated by any other mixed strategy. If $u(\theta_1, \alpha) > u(\theta_1, a_i)$ then $\alpha \succeq a_i$. If $u(\theta_2, \alpha) > u(\theta_2, a_k)$, then $\alpha \preceq a_k$.*

Proof. By definition of the belief order, showing that $\alpha \succeq a_i$ is equivalent to showing that, whenever α is optimal for a belief η, a_i is optimal for a belief μ and $\mu(\theta_1) > \eta(\theta_1)$ then α is optimal for μ and a_i is optimal for η. So let a_i be optimal for μ and α be optimal for η (since both a_i and α are not weakly dominated, they must be a best response to some belief). Then we have

$$\begin{aligned} \mu(\theta_1)[u(\theta_1, a_i) - u(\theta_1, \alpha)] + (1 - \mu(\theta_1))[u(\theta_2, a_i) - u(\theta_2, \alpha)] \geq 0 \\ \eta(\theta_1)[u(\theta_1, \alpha) - u(\theta_1, a_i)] + (1 - \eta(\theta_1))[u(\theta_2, \alpha) - u(\theta_2, a_i)] \geq 0. \end{aligned}$$

Combining these inequalities, we get

$$(\eta(\theta_1) - \mu(\theta_1))[u(\theta_1, \alpha) - u(\theta_1, a_i)] + (\mu(\theta_1) - \eta(\theta_1))[u(\theta_2, \alpha) - u(\theta_2, a_i)] \geq 0.$$

Since a_i is not weakly dominated and $u(\theta_1, \alpha) > u(\theta_1, a_i)$, we have that $u(\theta_2, \alpha) < u(\theta_2, a_i)$. For the inequality above to hold, it is necessary that $\mu(\theta_1) \leq \eta(\theta_1)$. This means that the condition $\mu(\theta_1) > \eta(\theta_1)$ never holds when a_i is optimal for μ and α is optimal for η, so $\alpha \succeq a_i$.

The proof that "if $u(\theta_2, \alpha) > u(\theta_2, a_k)$, then $\alpha \preceq a_k$" is analogous. □

Now fix a signal profile $\mathbf{y}$ and let $\alpha = f\big(\sigma_1(\mathbf{y}), \ldots, \sigma_{n-1}(\mathbf{y})\big)$. Then α is not weakly dominated and $u\,(\theta_1, \alpha) \geq u(\theta_1, a_i)$. If this inequality holds with equality, it must hold with equality for θ_2 as well, otherwise either a_i or α would be weakly dominated (in that case, $\alpha \succeq a_i$ holds trivially). In the case where the inequality holds strictly, we can use Lemma 12 to conclude that $\alpha \succeq a_i$. Similarly, since $u(\theta_2, \alpha) \geq u(\theta_2, a_k)$, we can conclude that $\alpha \preceq a_k$. Since we can do this for every $\mathbf{y}$, Corollary 7 follows.

B Online Appendix

B.1 Necessity of Randomization

We provide an example that highlights the necessity of mixed actions for robustly optimal strategies. There are two states, Covid ($\theta = 1$) and Flu ($\theta = 2$), and two information sources. These two information sources are the same as the ones in the introductory example, reproduced below for the ease of reference.

	+	−
Covid	0.9	0.1
Flu	0.5	0.5

Cough

	+	−
Covid	0.5	0.5
Flu	0.1	0.9

Fever

We consider the following three-action decision problem: $u(\cdot, a_1) = (0, 0)$, $u(\cdot, a_2) = (3, -2)$, $u(\cdot, a_3) = (7, -8)$. The corresponding binary decomposition consists of two subproblems with payoffs $u_1(\cdot, 1) = (3, -2)$, $u_2(\cdot, 1) = (4, -6)$, as illustrated in Fig. 7.

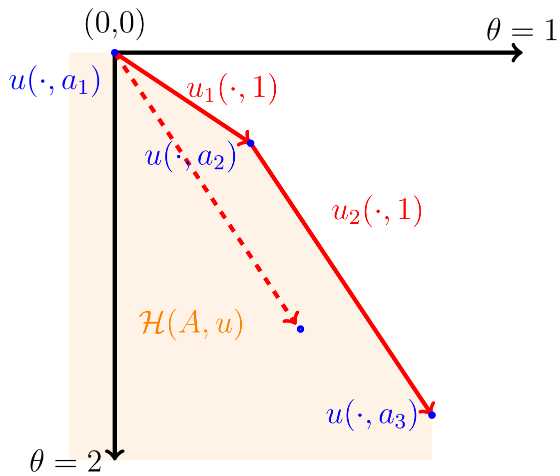


Figure 7: Binary decomposition of the three-action decision problem

In the binary decomposition, the action $(0, 1)$ corresponds to a payoff vector $(4, -6)$, which is not dominated by any pure action, but is dominated by a convex combination of a_2 and a_3. This means that the construction of robustly optimal strategies in Theorem 2 requires the decision maker to mix under some signal realization. We will further establish that *every* robustly optimal strategy must involve randomization in this example.

In light of Corollary 1, we can derive the robustly optimal value through the Blackwell supremum of $\{P_{Cough}, P_{Fever}\}$. Fig. 8 geometrically characterizes the induced feasible set of the Blackwell supremum, $\overline{P}(P_{Cough}, P_{Fever})$. It can be verified that the joint experiment P^* given in Table 7 induces this feasible set, which means P^* is the Blackwell supremum.

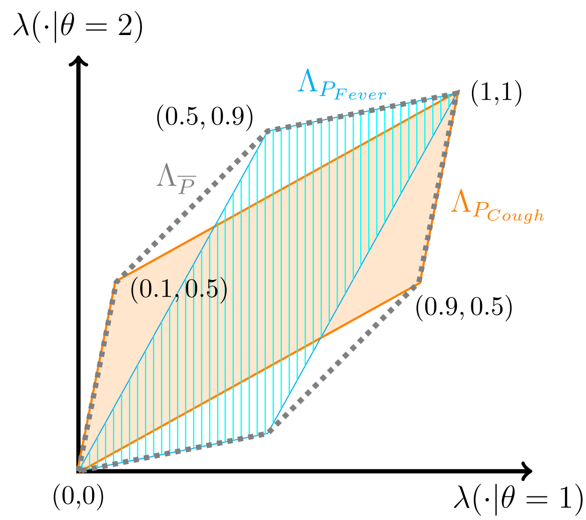


Figure 8: Blackwell Supremum of $\{P_{Cough}, P_{Fever}\}$

Covid	Fever_+	Fever_-	
Cough_+	0.5	0.4	0.9
Cough_-	0	0.1	0.1
	0.5	0.5	

Flu	Fever_+	Fever_-	
Cough_+	0.1	0.4	0.5
Cough_-	0	0.5	0.5
	0.1	0.9	

Table 7: Blackwell Supremum as Joint Experiment

From Eq. (1), any robustly optimal strategy σ^* must be a best response to P^*. Therefore, $\sigma^*(C_+, F_+) = a_3$, $\sigma^*(C_+, F_-) = a_2$, and $\sigma^*(C_-, F_-) = a_1$. Note that $\sigma^*(C_-, F_+)$ is not determined by the best response requirement because it is not in the support of P^*. A simple calculation gives us the robustly optimal value $V(P^*) = 1.55$.

We will show that $\sigma^*(C_-, F_+)$ must be mixed to guarantee such a value. We do so by showing that if the decision maker plays any of the three pure actions under (C_-, F_+), there exists a joint experiment that leads to a value strictly less than 1.55.

If $\sigma(C_-, F_+) = a_1$, consider the joint experiment given in Table 8, which yields a value of $\frac{1}{2}[0.4 * 7 + 0.5 * 3 - 0.1 * 8 - 0.4 * 2] = 1.35 < 1.55$. If $\sigma(C_-, F_+) = a_2$, the same joint experiment yields a value of $\frac{1}{2}[0.4 * 7 + 0.5 * 3 + 0.1 * 3 - 0.1 * 8 - 0.4 * 2] = 1.5 < 1.55$.

Covid	Fever_+	Fever_-	
Cough_+	0.4	0.5	0.9
Cough_-	0.1	0	0.1
	0.5	0.5	

Flu	Fever_+	Fever_-	
Cough_+	0.1	0.4	0.5
Cough_-	0	0.5	0.5
	0.1	0.9	

Table 8: Nature's alternative choice of joint experiment

If $\sigma(C_-, F_+) = a_3$, consider the joint experiment given in Table 9, which yields a value of $\frac{1}{2}[0.5 * 7 + 0.4 * 3 - 0.5 * 2 - 0.1 * 8] = 1.45 < 1.55$.

Covid	Fever$_+$	Fever$_-$	
Cough$_+$	0.5	0.4	0.9
Cough$_-$	0	0.1	0.1
	0.5	0.5	

Flu	Fever$_+$	Fever$_-$	
Cough$_+$	0	0.5	0.5
Cough$_-$	0.1	0.4	0.5
	0.1	0.9	

Table 9: Nature's alternative choice of joint experiment 2

B.2 Bounding $|K_v|$ by $|\Theta|$ and $|A|$

Proof. Recall that $\text{epi}(v) = \{(\mu, w) \in \Delta(\Theta) \times \mathbb{R} \mid w \geq v(\mu)\}$. Let $n = |\Theta|$. We can represent $\text{epi}(v)$ as a polyhedron in $\mathbb{R}^n$ that is the intersection of $|A| + |\Theta|$ halfspaces, as follows:

$$\left\{ (\mu_1, \ldots, \mu_{n-1}, w) \in \mathbb{R}^n \;\middle|\; \begin{array}{l} w \geq \sum_{i=1}^{n-1} \mu_i \rho(\theta_i, a) + (1 - \sum_{i=1}^{n-1} \mu_i) \rho(\theta_n, a) \quad \forall a \in A \\ \mu_i \geq 0 \quad i = 1, \ldots, n-1 \\ \mu_1 + \cdots + \mu_{n-1} \leqslant 1 \end{array} \right\}.$$

Here, we simply replaced the set $\Delta(\Theta)$ by its first $n-1$ coordinates; the original element $\mu \in \Delta(\Theta)$ can be recovered by $\mu_n = 1 - \mu_1 - \cdots - \mu_{n-1}$, so this change is inconsequential. In this representation, we have $|A|$ halfspaces corresponding to the constraints $w \geq \sum_{\theta \in \Theta} \mu(\theta) \rho(\theta, a)$ and $|\Theta| = n$ constraints corresponding to the description of $\Delta(\Theta)$.

This polyhedron is unbounded. To bound it, we also intersect $\text{epi}(v)$ with an additional halfspace, creating a bounded polytope $B = \text{epi}(v) \cap \{(\mu, w) : w \leq \max_{\theta, a} u(\theta, a) + 1\}$, which has at most $|A| + |\Theta| + 1$ facets.

The Upper Bound Theorem (see Theorem 8.23 in Ziegler (2012)) gives an upper bound on the number of facets that a polytope with a given number of vertices can have. Every polytope has a dual polytope (see Section 3.4 in Grünbaum (2003)), where each vertex corresponds to a facet and each facet corresponds to a vertex. Thus, we can apply the Upper Bound Theorem to the dual of B, which implies B can have at most

$$\binom{|\Theta| + |A| + 1 - \left\lfloor \frac{|\Theta|+1}{2} \right\rfloor}{|A| + 1} + \binom{|\Theta| + |A| + 1 - \left\lfloor \frac{|\Theta|+2}{2} \right\rfloor}{|A| + 1}$$

number of vertices.

These vertices include $\{(\delta_i, v(\delta_i))\}_{i=1}^n$ and $\{(\delta_i, \max_{\theta, a} u(\theta, a) + 1)\}_{i=1}^n$, which means $|K_v|$ can be no more than

$$\binom{|\Theta| + |A| + 1 - \left\lfloor \frac{|\Theta|+1}{2} \right\rfloor}{|A| + 1} + \binom{|\Theta| + |A| + 1 - \left\lfloor \frac{|\Theta|+2}{2} \right\rfloor}{|A| + 1} - 2|\Theta|.$$

□

B.3 Proof of Proposition 1

Proof. First observe that the agent's maxmin value is no more than her minmax value:

$$V(\mathcal{P}_1, \ldots, \mathcal{P}_m) \le \min_{P \in \mathcal{J}(\mathcal{P}_1, \ldots, \mathcal{P}_m)} \max_{\sigma: \mathbf{Y} \to \Delta(A)} \sum_{\theta} \sum_{\mathbf{y}} P(\mathbf{y}|\theta) u(\theta, \sigma(\mathbf{y}))$$

Now in the minmax problem, Nature's choice can be split into first choosing each marginal experiment $P_j \in \mathcal{P}_j$, and then choosing a joint experiment $P \in \mathcal{J}(P_1, ..., P_m)$:

$$= \min_{\substack{P_j \in \mathcal{P}_j \\ j=1,\ldots,m}} \min_{P \in \mathcal{J}(P_1, \ldots, P_m)} \max_{\sigma: \mathbf{Y} \to \Delta(A)} \sum_{\theta} \sum_{\mathbf{y}} P(\mathbf{y}|\theta) u(\theta, \sigma(\mathbf{y}))$$

And the value of the inner minmax problem is exactly $V(P_1, ..., P_m)$, which equals $\max_j V(P_j)$ from Theorem 1:

$$\begin{aligned} &= \min_{\substack{P_j \in \mathcal{P}_j \\ j=1,\ldots,m}} \max_{j=1,\ldots,m} V(P_j) \\ &= \max_{j=1,\ldots,m} V(\underline{P_j}) \end{aligned}$$

where $\underline{P}_j \in \operatorname{argmin}_{P_j \in \mathcal{P}_j} V(P_j)$ is a worst experiment among the set $\mathcal{P}_j$ if the agent faces this information source solely. Let $j^* \in \operatorname{argmax}_j V(\underline{P_j})$, and consider the problem where the decision maker faces only a single set of marginal experiments $\mathcal{P}_{j^*}$:

$$V(\mathcal{P}_{j^*}) = \max_{\sigma: Y_{j^*} \to \Delta(A)} \min_{P_{j^*} \in \mathcal{P}_{j^*}} \sum_{\theta} \sum_{y_{j^*} \in Y_{j^*}} P_{j^*}(y_{j^*}|\theta) u(\theta, \sigma(y_j^*)).$$

Since $\mathcal{P}_{j^*}$ is convex, from the minmax theorem, the value of the problem equals

$$V(\mathcal{P}_{j^*}) = \min_{P_{j^*} \in \mathcal{P}_{j^*}} \max_{\sigma: Y_{j^*} \to \Delta(A)} \sum_{\theta} \sum_{y_{j^*} \in Y_{j^*}} P_{j^*}(y_{j^*}|\theta) u(\theta, \sigma(y_j^*)) = V(\underline{P_{j^*}}).$$

So there exists a best-source strategy, using only signals from the experiment P_{j^*}, that guarantees the robustly optimal value $V(\underline{P_{j^*}}) = \max_j V(\underline{P_j}) \ge V(\mathcal{P}_1, \ldots, \mathcal{P}_m)$. □

B.4 Proof of Proposition 2

Lemma 13 (Single-Peaked Property)**.** *Suppose in a binary-state decision problem* (A, u)*, every action is a unique best response to some belief, and actions are ordered as follows*

$$u(\theta_1, a_1) < u(\theta_1, a_2) < \cdots < u(\theta_1, a_n),$$
$$u(\theta_2, a_1) > u(\theta_2, a_2) > \cdots > u(\theta_2, a_n).$$

Then, for any belief $\mu \in \Delta(\Theta)$*,*

$$a_i \in \operatorname*{argmax}_{a \in A} \sum_\theta \mu(\theta) u(\theta, a)$$

implies that for $k > j \geq i$*,*

$$\sum_\theta \mu(\theta) u(\theta, a_j) \geq \sum_\theta \mu(\theta) u(\theta, a_k)$$

and for $k < j \leq i$*,*

$$\sum_\theta \mu(\theta) u(\theta, a_j) \geq \sum_\theta \mu(\theta) u(\theta, a_k).$$

Proof. Suppose by contradiction that there exists $k > j \geq i$, such that

$$\mu(\theta_1) u(\theta_1, a_j) + \mu(\theta_2) u(\theta_2, a_j) < \mu(\theta_1) u(\theta_1, a_k) + \mu(\theta_2) u(\theta_2, a_k).$$

Rearranging, we obtain

$$\mu(\theta_2)[u(\theta_2, a_j) - u(\theta_2, a_k)] < \mu(\theta_1)[u(\theta_1, a_k) - u(\theta_1, a_j)].$$

Given that $u(\theta_2, a_j) - u(\theta_2, a_k) > 0$ and $u(\theta_1, a_k) - u(\theta_1, a_j) > 0$, the inequality above still holds if we raise $\mu(\theta_1)$ (and consequently lower $\mu(\theta_2)$). That is, for any $\mu' \in \Delta(\Theta)$ such that $\mu'(\theta_1) \geq \mu(\theta_1)$, we have

$$\mu'(\theta_1) u(\theta_1, a_j) + \mu'(\theta_2) u(\theta_2, a_j) < \mu'(\theta_1) u(\theta_1, a_k) + \mu'(\theta_2) u(\theta_2, a_k). \tag{9}$$

Since a_i is, by definition, a best response for μ,

$$\mu(\theta_1) u(\theta_1, a_j) + \mu(\theta_2) u(\theta_2, a_j) \leq \mu(\theta_1) u(\theta_1, a_i) + \mu(\theta_2) u(\theta_2, a_i).$$

Since $u(\theta_1, a_j) \geq u(\theta_1, a_i)$ and $u(\theta_2, a_j) \leq u(\theta_2, a_i)$, for any $\mu' \in \Delta(\Theta)$ such that $\mu'(\theta_1) \leq \mu(\theta_1)$, we have

$$\mu'(\theta_1) u(\theta_1, a_j) + \mu'(\theta_2) u(\theta_2, a_j) \leq \mu'(\theta_1) u(\theta_1, a_i) + \mu'(\theta_2) u(\theta_2, a_i) \tag{10}$$

The inequalities (9) and (10) together imply that a_j is never a unique best response to any belief, contradicting our assumption.

The case where $k < j \leq i$ follows from a similar argument. □

Lemma 14. *Let (A_ℓ, u_ℓ) be a subproblem in a binary decomposition of (A, u) and let R_{P_j} be a recommendation information structure with respect to (A, u). Then*

$$V(P_j; (A_\ell, u_\ell)) = V(R_{P_j}; (A_\ell, u_\ell)).$$

Proof. Recall that P_j Blackwell dominates R_{P_j}, so $V(P_j; (A_\ell, u_\ell)) \geq V(R_{P_j}; (A_\ell, u_\ell))$. We prove the result by constructing a recommendation information structure $R^\ell_{P_j}$ for (A_ℓ, u_ℓ) and showing that $V(R_{P_j}; (A_\ell, u_\ell)) \geq V(R^\ell_{P_j}; (A_\ell, u_\ell)) = V(P_j; (A_\ell, u_\ell))$.

Recall that R_{P_j} is defined using a garbling of P_j given by $\sigma^* : Y_j \to A$ that satisfies, for each y_j in the support,

$$\sigma^*(y_j) \in \operatorname*{argmax}_{a \in A} \sum_\theta P_j(y_j|\theta) u(\theta, a).$$

From Lemma 13, if $a_i \in \operatorname{argmax}_{a \in A} \sum_\theta P_j(y_j|\theta) u(\theta, a)$, for all $i \leq \ell \leq n-1$, $\sum_\theta P_j(y_j|\theta) u(\theta, a_\ell) \geq \sum_\theta P_j(y_j|\theta) u(\theta, a_{\ell+1})$, and for all $2 \leq \ell \leq i$, $\sum_\theta P_j(y_j|\theta) u(\theta, a_\ell) \geq \sum_\theta P_j(y_j|\theta) u(\theta, a_{\ell-1})$. This means that, if R_{P_j} recommends action a_i, then $1 \in A_\ell$ is optimal for the subproblems with $i \leqslant \ell$ and $0 \in A_\ell$ is optimal for the subproblems with $i > \ell$. Now let $\gamma_\ell : A \to \{0, 1\}$ be the garbling defined by

$$\gamma_\ell(a_i) = \begin{cases} 0 & \text{if } i \leq \ell \\ 1 & \text{if } i > \ell. \end{cases}$$

By construction, for each y_i in the support,

$$\gamma_\ell(\sigma^*(y_j)) \in \operatorname*{argmax}_{a \in A_\ell} \sum_{\theta, y_j} P_j(y_j|\theta) u_\ell(\theta, a),$$

so the experiment $R^\ell_{P_j}$, induced by garbling P_j according to $\gamma_\ell \circ \sigma^* : Y_j \to A$, is a recommendation information structure for the decision problem (A_ℓ, u_ℓ), so $V(R^\ell_{P_j}; (A_\ell, u_\ell)) = V(P_j; (A_\ell, u_\ell))$. Moreover, by construction, R_{P_j} Blackwell dominates $R^\ell_{P_j}$, so $V(R_{P_j}; (A_\ell, u_\ell)) \geq V(R^\ell_{P_j}; (A_\ell, u_\ell))$. □

Proof of Proposition 2. Let $\bigoplus_{\ell=1}^k (A_\ell, u_\ell)$ be a binary decomposition of (A, u). From Theorem 2 and Lemma 14,

$$\begin{aligned}V(P_1,\dots,P_m;(A,u)) &= \sum_{l=1}^{k} \max_{j=1,\dots,m} V(P_j;(A_\ell,u_\ell)) \\ &= \sum_{l=1}^{k} \max_{j=1,\dots,m} V(R_{P_j};(A_\ell,u_\ell)) \\ &= V(R_{P_j},\dots,R_{P_j};(A,u)).\end{aligned}$$

□

B.5 Proof of Proposition 3

Consider the following Blackwell experiment:

$$P^*(y \mid \theta) = \begin{cases} P^*(y_\theta \mid \theta) & \text{if } y = y_\theta, \\ 1 - P^*(y_\theta \mid \theta) & \text{if } y = \emptyset. \end{cases}$$

First, we will show that P^* Blackwell dominates $P_1,\dots,P_m$. To see this, first note that by construction, for each θ,

$$P^*(y_\theta \mid \theta) \geq P_\ell(y_\theta \mid \theta).$$

Thus, consider the following garbling matrix: for each θ, $Q(y_\theta \mid y_\theta) = \frac{P_\ell(y_\theta|\theta)}{P^*(y_\theta|\theta)}$ and $Q(\emptyset \mid y_\theta) = 1 - Q(y_\theta \mid y_\theta)$. Then clearly, P_ℓ is a Q-garbling of P^*. Thus, by Blackwell's theorem, P^* Blackwell dominates P_ℓ. Hence, we have:

$$V(P_1,\dots,P_m) \leq V(P^*).$$

Second, it remains to show that for any $P \in \mathcal{P}(P_1,\dots,P_m)$, $V(P) \geq V(P^*)$. To show this, it suffices to show that $\tau_P(\delta_\theta) \geq \tau_{P^*}(\delta_\theta)$ for all θ. By construction, $\tau_{P^*}(\delta_\theta) = \max_{\ell=1,\dots,m} \tau_{P_\ell}(\delta_\theta)$. So it suffices to show that $\tau_P(\delta_\theta) \geq \tau_{P_\ell}(\delta_\theta)$ for all ℓ.

First if $\tau_{P_\ell}(\delta_\theta) = 0$, then the inequality is trivial. So assume that $\tau_{P_\ell}(\delta_\theta) > 0$. Since P Blackwell dominates P_ℓ, there exists some garbling Q_ℓ such that P_ℓ is a Q_ℓ-garbling of P. Therefore, for every θ'', θ',

$$P_\ell(y_{\theta''} \mid \theta') = \sum_{y \in Y} Q(y_{\theta''} \mid y) P(y \mid \theta').$$

First note that if $Q(y_\theta \mid y) > 0$ then $P(y \mid \theta') = 0$ for all $\theta' \neq \theta$. Hence, the interim belief after observing y is δ_θ. Then letting $Y_\theta := \{y : Q(y_\theta \mid y) > 0\}$, we have:

$$\sum_{y \in Y_\theta} P(y \mid \theta)\mu_0(\theta) \geq \sum_{y \in Y_\theta} Q(y_\theta \mid y) P(y \mid \theta)\mu_0(\theta) = P_\ell(y_\theta \mid \theta)\mu_0(\theta).$$

So, $\tau_P(\delta_\theta) \geq \tau_{P_\ell}(\delta_\theta)$.

B.6 Proof of Uniqueness for Theorem 1

Consider any binary-state binary-action decision problem, denoted by (A^{bi}, u^{bi}). Without loss of generality, suppose P_1 is the unique best marginal information source: $V(P_1; (A^{bi}, u^{bi})) > V(P_j; (A^{bi}, u^{bi}))$ for $j \neq 1$.

B.6.1 Payoff Sets

Recall that as in Section 4.3, any binary-state decision problem (A, u) induces a payoff polyhedron:

$$\mathcal{H}(A, u) = co\{u(\cdot, a) : a \in A\} - \mathbb{R}^2_+,$$

which captures the feasible payoff vectors that can be achieved by the decision maker when allowing for free disposal of utils. Such a polyhedron is upper bounded, convex, closed, and has a finite number of extreme points.

Definition 9. *A non-empty subset $D \subseteq \mathbb{R}^{|\Theta|}$ is a **payoff set** if D is upper bounded, convex, closed, and has a finite number of extreme points.*

For any payoff set D, we define the robustly optimal value in a manner similar to that for decision problems:

$$W(P_1, ..., P_m; D) = \max_{t:\mathbf{Y}\to D} \min_{P \in \mathcal{J}(P_1,...,P_m)} \sum_{\mathbf{y}} \mathbf{P}(\mathbf{y}) \cdot t(\mathbf{y})$$

where $\mathbf{P}(\mathbf{y}) = P(\mathbf{y}|\cdot) \in \mathbb{R}^{|\Theta|}$ denotes the vector corresponding to the probability of $\mathbf{y}$ in each state.

If only a single experiment $P : \Theta \to \Delta(Y)$ is considered ($m = 1$),

$$W(P; D) = \max_{t:Y\to D} \sum_{y} \mathbf{P}(y) \cdot t(y).$$

Note that the value for a payoff set is tightly connected to the value of the decision problem that induces it. Specifically, we have $V(P_1, ..., P_m; (A, u)) = W(P_1, ..., P_m; \mathcal{H}(A, u))$.

Similar to V, W also has the property that having access to more experiments can be no worse than having access to just one experiment.

Lemma 15. *For any decision problem D,*

$$W(P_1, ..., P_m; D) \geq W(P_1; D)$$

Proof. Suppose $t_1 : Y_1 \to D$ is the solution to $W(P_1; D)$. Define $\tilde{t} : Y_1 \times \cdots \times Y_m \to D$ as $\tilde{t}(y_1, ..., y_m) = t_1(y_1)$, and we have

$$W(P_1, ..., P_m; D) \geq \min_{P \in \mathcal{P}(P_1,...,P_m)} \sum_{\mathbf{y}} \mathbf{P}(\mathbf{y}) \cdot \tilde{t}(\mathbf{y}) = \sum_{y_1} P_1(y_1) \cdot t_1(y) = W(P_1; D).$$

□

Another useful property of W is its separability with respect to payoff sets, analogous to the separability of V with respect to separable decision problems.

Lemma 16. *Let $C, D \subseteq \mathbb{R}^2$ be two payoff sets, and $C + D$ denote their Minkowski sum. Then*

$$W(P; C + D) = W(P; C) + W(P; D).$$

Proof. Let t^*_C and t^*_D be solutions to $W(P; C)$ and $W(P; C)$, respectively. Define $t : Y \to C + D$ to be $t(y) = t^*_C(y) + t^*_D(y)$. Then

$$\begin{aligned} W(P; C + D) &\geq \sum_y \mathbf{P}(y) \cdot t(y) \\ &= \sum_y \mathbf{P}(y) \cdot (t^*_C(y) + t^*_D(y)) \\ &= \sum_y \mathbf{P}(y) \cdot t^*_C(y) + \sum_y \mathbf{P}(y) \cdot t^*_D(y) \\ &= W(P; C) + W(P; D). \end{aligned}$$

Conversely, let t^* be a solution to $W(P; C + D)$. Then for any y, there exists $c_y \in C$ and $d_y \in D$ such that $t^*(y) = c_y + d_y$. Define $t_C(y) = c_y$ and $t_D(y) = d_y$, then

$$\begin{aligned} W(P; C) + W(P; D) &\geq \sum_y \mathbf{P}(y) \cdot t_C(y) + \sum_y \mathbf{P}(y) \cdot t_D(y) \\ &= \sum_y \mathbf{P}(y) \cdot t^*(y) \\ &= W(P; C + D). \end{aligned}$$

□

B.6.2 Binary-Action Decision Problems

Now we return to the binary action decision problem (A^{bi}, u^{bi}). The payoff polyhedron corresponding to (A^{bi}, u^{bi}) can be represented as intersection of three halfspaces:

$$\mathcal{H}(A^{bi}, u^{bi}) = \bigcap_{\beta \in \mathcal{B}_{(A^{bi}, u^{bi})}} \{v \in \mathbb{R}^2 : \beta \cdot v \leq k_\beta\}$$

where $\mathcal{B}_{(A^{bi},u^{bi})} = \{e_1, e_2, \beta^*\}$ with $e_1 = (1, 0)$, $e_2 = (0, 1)$, and $\beta^* \in \mathbb{R}^2_{++}$ denote the set of normal vectors, and $k_{e_1} = \max_{a \in A} u(\theta = 1, a)$, $k_{e_2} = \max_{a \in A} u(\theta = 2, a)$, and $k_{\beta^*} \in \mathbb{R}$. This is visualized in Fig. 9.

The set of normal vectors, $\mathcal{B}_{(A^{bi},u^{bi})}$, depends on the binary action decision problem, where β^* is proportional to the belief at which the decision maker is indifferent between the two actions. Since the decision problem (A^{bi}, u^{bi}) is fixed, for notational simplicity, we will henceforth omit the dependence of $\mathcal{B}$ on (A^{bi}, u^{bi}).

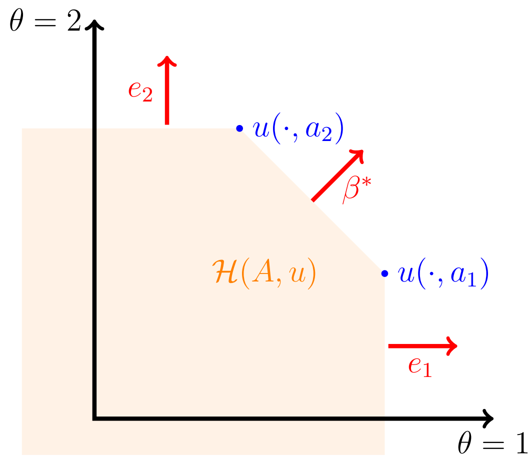


Figure 9: Payoff polyhedron for a binary-state binary-action problem

We next define payoff sets that have the same shape as the $\mathcal{H}(A^{bi}, u^{bi})$.

Definition 10. *A payoff set $D \subset \mathbb{R}^2$ is a $\mathcal{B}$-shape polyhedron if*

$$D = \bigcap_{\beta \in \mathcal{B}} \{v \in \mathbb{R}^2 : \beta \cdot v \leq k_\beta\}$$

for some constants $\{k_\beta\}_{\beta \in \mathcal{B}} \in \mathbb{R}$.

Note that the constraint $\beta^* \cdot v \leq k_{\beta^*}$ may be redundant in a $\mathcal{B}$-shape polyhedron, in which case the polyhedron is an unbounded rectangle. Such a polyhedron can be represented as $\{v : v \leq v^*\}$ for some $v^* \in \mathbb{R}^2$ and corresponds to a single-action decision problem. We call such a $\mathcal{B}$-shape polyhedron *trivial*.

Clearly, if D is a trivial $\mathcal{B}$-shape polyhedron, $W(P; D) = W(P'; D)$ for any P, P'. The next lemma shows that for any non-trivial $\mathcal{B}$-shape polyhedron, the relative value of experiments under (A^{bi}, u^{bi}) is preserved.

Lemma 17. *If D is a non-trivial $\mathcal{B}-$shape polyhedron, then $W(P_1; D) > \max_{j\neq 1} W(P_j; D)$.*

Proof. Any non-trivial $\mathcal{B}$-shape polyhedron D has two extreme points, denoted by $ex(D)_1$ and $ex(D)_2$. See Fig. 10 for an illustration.

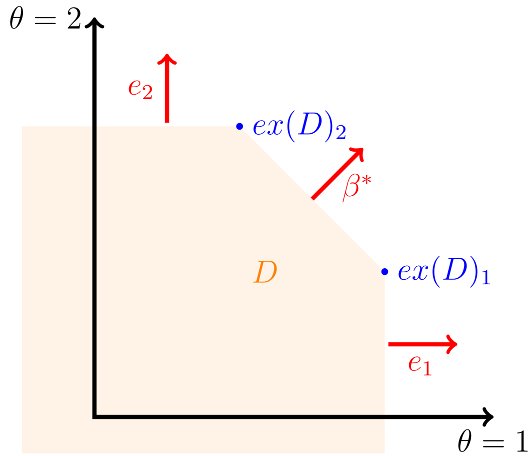


Figure 10: Extreme points of a non-trivial $\mathcal{B}$-polyhedron

The two extreme points are defined by two linear equations:

$$\begin{pmatrix} e_1 \\ \beta^* \end{pmatrix} v = \begin{pmatrix} k_{e_1} \\ k_{\beta^*} \end{pmatrix} \qquad \begin{pmatrix} e_2 \\ \beta^* \end{pmatrix} v = \begin{pmatrix} k_{e_2} \\ k_{\beta^*} \end{pmatrix},$$

with the closed-form solutions $ex(D)_1 = \begin{pmatrix} k_{e_1} \\ \frac{k_{\beta^*} - \beta_1^* k_{e_1}}{\beta_2^*} \end{pmatrix}$ and $ex(D)_2 = \begin{pmatrix} \frac{k_{\beta^*} - \beta_2^* k_{e_2}}{\beta_1^*} \\ k_{e_2} \end{pmatrix}$. A useful observation is that $(ex(D)_2 - ex(D)_1) = (k_{\beta^*} - k_{e_1}\beta_1^* - k_{e_2}\beta_2^*) \begin{pmatrix} -\frac{1}{\beta_1^*} \\ \frac{1}{\beta_2^*} \end{pmatrix}$. That is, β^* determines the direction of the vector $(ex(D)_2 - ex(D)_1)$, and the constant terms k_β only affect the scalar multiplier. Moreover, the multiplier $(k_{\beta^*} - k_{e_1}\beta_1^* - k_{e_2}\beta_2^*) > 0$, because $(k_{e_1}, k_{e_2}) \in int(D)$ and $k_{\beta^*} = \max_{v\in D} \beta^* \cdot v$.

For any $\mathcal{B}$-shape polyhedron D, and any P_j,

$$W(P_j; D) = \max_{t_j : Y_j \to D} \sum_{y_j \in Y_j} \mathbf{P}_j(y_j) \cdot t_j(y_j)$$

Since the objective function is linear and the extreme points of D are $ex(D)_1$ and $ex(D)_2$, a solution to the problem is

$$t_j^*(y_j) = \begin{cases} ex(D)_1 & \text{if } \mathbf{P}_j(y_j) \cdot \begin{pmatrix} -\frac{1}{\beta_1^*} \\ \frac{1}{\beta_2^*} \end{pmatrix} \leq 0 \\ ex(D)_2 & \text{if } \mathbf{P}_j(y_j) \cdot \begin{pmatrix} -\frac{1}{\beta_1^*} \\ \frac{1}{\beta_2^*} \end{pmatrix} > 0. \end{cases}$$

For each P_j, let $\tilde{Y}_j = \{y \in Y_j : \mathbf{P}_j(y_j) \cdot \begin{pmatrix} -\frac{1}{\beta_1^*} \\ \frac{1}{\beta_2^*} \end{pmatrix} \leq 0\}$, and we can rewrite:

$$W(P_j; D) = \sum_{y_j \in \tilde{Y}_j} \mathbf{P}_j(y_j) \cdot ex(D)_1 + \sum_{y_j \in Y_j / \tilde{Y}_j} \mathbf{P}_j(y_j) \cdot ex(D)_2.$$

Let $\mathbf{x}_{P_j} = \sum_{y_j \in \tilde{Y}_j} \mathbf{P}_j(y_j)$, then

$$\begin{aligned} W(P_j; D) &= \mathbf{x}_{P_j} \cdot ex(D)_1 + (\mathbf{1} - \mathbf{x}_{P_j}) \cdot ex(D)_2 \\ &= \mathbf{1} \cdot ex(D)_2 + \mathbf{x}_{P_j} \cdot (ex(D)_1 - ex(D)_2). \end{aligned}$$

Now consider any $j \neq 1$, we have

$$\begin{aligned} W(P_1; D) - W(P_j; D) &= (\mathbf{x}_{P_j} - \mathbf{x}_{P_1}) \cdot (ex(D)_2 - ex(D)_1) \\ &= (k_{\beta^*} - k_{e_1}\beta_1^* - k_{e_2}\beta_2^*)(\mathbf{x}_{P_j} - \mathbf{x}_{P_1}) \cdot \begin{pmatrix} -\frac{1}{\beta_1^*} \\ \frac{1}{\beta_2^*} \end{pmatrix} \end{aligned}$$

Note that for different non-trivial $\mathcal{B}$-shape polyhedra D (i.e., different parameters $k_{e_1}, k_{e_2}, k_{\beta^*}$), the above value differs only by a positive constant factor. This implies that if $W(P_1; D) - W(P_j; D) > 0$ for one non-trivial $\mathcal{B}$-shape polyhedron, the value is also strictly positive for any non-trivial $\mathcal{B}$-shape polyhedron.

Recall that

$$W(P_1; \mathcal{H}(A^{bi}, u^{bi})) - W(P_j; \mathcal{H}(A^{bi}, u^{bi})) = V(P_1; (A^{bi}, u^{bi})) - V(P_j; (A^{bi}, u^{bi})) > 0$$

where $\mathcal{H}(A^{bi}, u^{bi})$ is a $\mathcal{B}$-shape polyhedron. Therefore,

$$W(P_1; D) - W(P_j; D) > 0,$$

for any non-trivial $\mathcal{B}$-shape polyhedron.

□

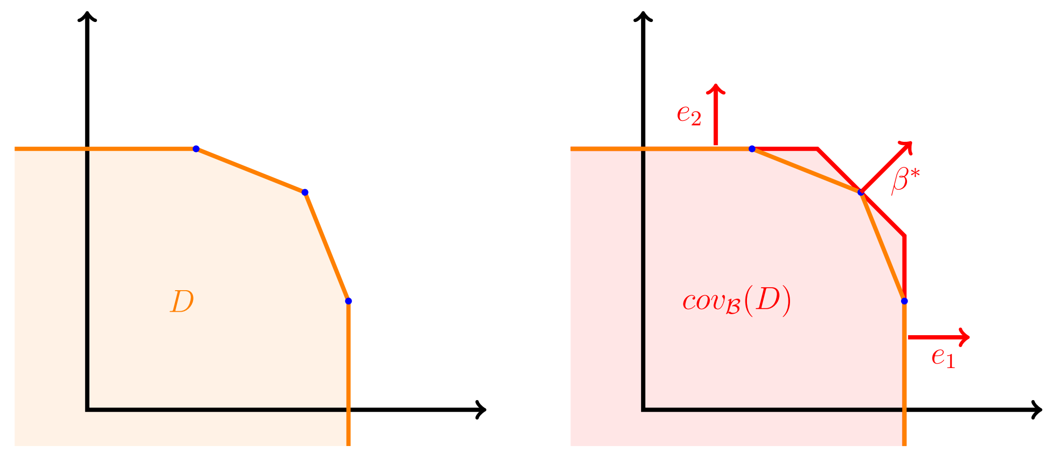
B.6.3 $\mathcal{B}$-cover

For any payoff set D, we define the smallest $\mathcal{B}$-shape polyhedron that covers D as its $\mathcal{B}$-*cover*. See Fig. 11 for an illustration.

Definition 11. *For any payoff set D, its **$\mathcal{B}$-cover** is defined as*

$$cov_{\mathcal{B}}(D) \doteq \bigcap_{\beta \in \mathcal{B}} \{v : \beta \cdot v \leq \rho_D(\beta)\},$$

where $\rho_D(\beta) = \sup_{v \in D} \beta \cdot v$ is the support function of D.



(a) A payoff set D derived from some three-action decision problem

(b) The corresponding $\mathcal{B}$-cover $cov_{\mathcal{B}}(D)$

Figure 11

We state a few properties of $\mathcal{B}$-cover that will be useful in our analysis.

Lemma 18. *1. (Monotonicity) If $D \subseteq D'$, $cov_{\mathcal{B}}(D) \subseteq cov_{\mathcal{B}}(D')$.*

2. (Reflexive) If D is a $\mathcal{B}$-shape polyhedron, $cov_{\mathcal{B}}(D) = D$.

3. (Superadditivity) $cov_{\mathcal{B}}(D + D') \supseteq cov_{\mathcal{B}}(D) + cov_{\mathcal{B}}(D')$

4. (Preserving Triviality) If $cov_{\mathcal{B}}(D)$ is trivial, then there exists a maximum in D. That is, $\exists \bar{v} \in D$ such that $v \leq \bar{v}$ for all $v \in D$.

Proof. 1. Since $D \subseteq D'$, $\rho_D(\beta) \leq \rho_{D'}(\beta)$ for all $\beta \in \mathcal{B}$. Therefore,

$$\bigcap_{\beta \in \mathcal{B}} \{v : \beta \cdot v \leq \rho_D(\beta)\} \subseteq \bigcap_{\beta \in \mathcal{B}} \{v : \beta \cdot v \leq \rho_{D'}(\beta)\}.$$

2. Clearly $D \subseteq cov_{\mathcal{B}}(D)$, because for every $v \in D$ and every $\beta \in \mathcal{B}$, $\beta \cdot v \leq \rho_D(\beta)$.

   Now consider any $\mathcal{B}$-shape polyhedron, represented by

$$D = \bigcap_{\beta \in \mathcal{B}} \{v \in \mathbb{R}^2 : \beta \cdot v \leq k_\beta\}$$

   for some $\{k_\beta\}_{\beta \in \mathcal{B}} \in \mathbb{R}^2$. Note that for all $\beta \in \mathcal{B}$ and $v \in D$, $\beta \cdot v \leq k_\beta$, so we have $\rho_D(\beta) = \max_{v \in D} \beta \cdot v \leq k_\beta$. Therefore,

$$cov_{\mathcal{B}}(D) = \bigcap_{\beta \in \mathcal{B}} \{v : \beta \cdot v \leq \rho_D(\beta)\} \subseteq \bigcap_{\beta \in \mathcal{B}} \{v : \beta \cdot v \leq k_\beta\} = D,$$

   which implies $cov_{\mathcal{B}}(D) = D$.

3. For any $\tilde{v} \in cov_{\mathcal{B}}(D) + cov_{\mathcal{B}}(D')$, there exists $v \in cov_{\mathcal{B}}(D)$ and $v' \in cov_{\mathcal{B}}(D')$ such that $\tilde{v} = v + v'$. Since $v \in cov_{\mathcal{B}}(D)$ and $v' \in cov_{\mathcal{B}}(D')$, we have $\beta \cdot v \leq \rho_D(\beta)$ and $\beta \cdot v' \leq \rho_{D'}(\beta)$ for all $\beta \in \mathcal{B}$. Therefore, for every $\beta \in \mathcal{B}$, $\beta \cdot \tilde{v} = \beta \cdot (v + v') \leq \rho_D(\beta) + \rho_{D'}(\beta) = \rho_{D+D'}(\beta)$, which implies $\tilde{v} \in cov_{\mathcal{B}}(D + D')$.

4. If $cov_{\mathcal{B}}(D)$ is trivial, the constraint $\beta^* \cdot v \leq \rho_D(\beta^*)$ is redundant. That is $\{v : \beta^* \cdot v \leq \rho_D(\beta^*)\} \supseteq \{v : e_1 \cdot v \leq \rho_D(e_1)\} \cap \{v : e_2 \cdot v \leq \rho_D(e_2)\}$.

   Let $\bar{v}_1 = \max_{v \in D} e_1 \cdot v$ and $\bar{v}_2 = \max_{v \in D} e_2 \cdot v$. We claim that $\bar{v} = (\bar{v}_1, \bar{v}_2) \in D$. Suppose not, then we have $\max_{v \in D} \beta^* \cdot v < \beta^* \cdot \bar{v}$. However, $\bar{v} \in \{v : e_1 \cdot v \leq \rho_D(e_1)\} \cap \{v : e_2 \cdot v \leq \rho_D(e_2)\}$ but $\bar{v} \notin \{v : \beta^* \cdot v \leq \rho_D(\beta^*)\}$, contradicting to the constraint $\beta^* \cdot v \leq \rho_D(\beta^*)$ being redundant. Thus, $\bar{v} \in D$ and for all $v \in D$, $v \leq \bar{v}$, which concludes the proof.

□

B.6.4 Dominance

We say a collection of payoff sets $D_1, ..., D_k \subseteq \mathbb{R}^{|\Theta|}$ is *dominated* by D if

$$D_1 + \cdots + D_k \subseteq D.$$

The following observation is immediate:

Lemma 19. *If $\{D_\ell\}_{\ell=1}^k$ is dominated by D,*

$$W(P_1, ..., P_m; D) \geq \sum_{\ell=1}^{k} W(P_1, ..., P_m; D_\ell).$$

Proof. Let t_ℓ be a maxmin strategy to $W(P_1, ..., P_m; D_\ell)$. Construct

$$\begin{aligned} t : \mathbf{Y} &\to D \\ y &\mapsto \sum_{\ell=1}^{k} t_\ell(y). \end{aligned}$$

Then

$$\begin{aligned} W(P_1, ..., P_m; D) &\geq \min_{P \in \mathcal{J}} \sum_y \mathbf{P}(y) \cdot t(y) \\ &= \min_{P \in \mathcal{J}} \sum_y \mathbf{P}(y) \cdot \sum_{\ell=1}^{k} t_\ell(y) \\ &= \min_{P \in \mathcal{J}} \sum_{\ell=1}^{k} \sum_y \mathbf{P}(y) \cdot t_\ell(y) \\ &\geq \sum_{\ell=1}^{k} \min_{P \in \mathcal{J}} \sum_y \mathbf{P}(y) \cdot t_\ell(y) \\ &= \sum_{\ell=1}^{k} W(P_1, ..., P_m; D_\ell). \end{aligned}$$

□

Next, we present the key lemma underlying our uniqueness theorem.

Lemma 20. *Suppose a collection of decision problems $D_1, ..., D_m$ is dominated by a $\mathcal{B}$-shape polyhedron D, and satisfies*

$$\sum_{j=1}^{m} W(P_j; D_j) \geq W(P_1, ..., P_m; D).$$

Then $cov_{\mathcal{B}}(D_j)$ must be trivial for all $j \neq 1$.

Proof. Since $D_1 + \cdots + D_m \subseteq D$, from properties 1 and 2 in Lemma 18,

$$cov_{\mathcal{B}}(D_1 + \cdots + D_m) \subseteq cov_{\mathcal{B}}(D) = D.$$

From property 3 in Lemma 18,

$$cov_{\mathcal{B}}(D_1) + \cdots + cov_{\mathcal{B}}(D_m) \subseteq cov_{\mathcal{B}}(D_1 + \cdots + D_m),$$

so $cov_{\mathcal{B}}(D_1), \cdots, cov_{\mathcal{B}}(D_m)$ is also dominated by D.

Now suppose by contradiction that $cov_{\mathcal{B}}(D_j)$ is not trivial for some $j \neq 1$. Then

$$\begin{aligned} W(P_1, ..., P_m; D) &\geq \sum_{j=1}^{m} W(P_1, ..., P_m; cov_{\mathcal{B}}(D_j)) \\ &\geq \sum_{j=1}^{m} W(P_1; cov_{\mathcal{B}}(D_j)) \\ &> \sum_{j=1}^{m} W(P_j; cov_{\mathcal{B}}(D_j)) \\ &\geq \sum_{j=1}^{m} W(P_j; D_j) \end{aligned}$$

where the first inequality follows from Lemma 19, second inequality follows from Lemma 15, the third inequality follows from Lemma 17, and the last inequality follows from $cov(D_j) \supseteq D_j$. Therefore, it contradicts to $\sum_{j=1}^{m} W(P_j; D_j) \geq W(P_1, ..., P_m; D)$, and D_j must be trivial for all $j \neq 1$. □

B.6.5 Common Support of the Blackwell Supremum

Lemma 21. *Suppose $P_j(y_j|\theta) > 0$ for all j, y_j, θ, and $P^* \in \mathcal{J}(P_1, ..., P_m)$ is a Blackwell supremum of $P_1, ..., P_m$. Then, $P^*(\cdot|\theta_1)$ and $P^*(\cdot|\theta_2)$ have common support; that is, for any $y_1, ..., y_m$, $P^*(y_1, ..., y_m|\theta_1) > 0$ if and only if $P^*(y_1, ..., y_m|\theta_2) > 0$.*

Proof. If $P^*(\cdot|\theta_1)$ and $P^*(\cdot|\theta_2)$ have different supports, then there exists $\mathbf{y}$ that induces a point-mass belief either on state θ_1 or θ_2. So the corresponding Zonotope Λ_{P^*} will include either a point $(x, 0)$ or $(0, x)$ for some $x > 0$. Since $P_j(y_j|\theta) > 0$ for all j, y_j, θ, none of the Zonotopes Λ_{P_j} contains such points. From Lemma 1, $\Lambda_{P^*} = \text{co}(\Lambda_{P_1} \cup \cdots \cup \Lambda_{P_m})$, which also should not contain such points, leading to a contradiction. □

B.6.6 Proof of the Theorem

Proof of Uniqueness for Theorem 1. Let σ^* be a robustly optimal strategy in the decision problem (A^{bi}, u^{bi}). We have

$$V(P_1, ..., P_m; (A^{bi}, u^{bi})) = \min_{P \in \mathcal{J}(P_1, ..., P_m)} \sum_{\theta} P(\mathbf{y}|\theta) u^{bi}(\theta, \sigma^*(\mathbf{y})).$$

This is a state-by-state optimal transport problem, and so the corresponding dual problem is

$$\max_{\phi_j : \Theta \times Y_j \to \mathbb{R},\, j=1,...,m} \sum_{\theta} \sum_{j} \sum_{y_j} \phi_j(\theta, y_j) P_j(y_j|\theta)$$

$$s.t. \quad \sum_{j=1}^{m} \phi_j(\theta, y_j) \leq u^{bi}(\theta, \sigma^*(\mathbf{y})) \quad \forall \theta, \mathbf{y}.$$

Or in vector form:

$$\max_{\phi_j : Y_j \to \mathbb{R}^{|\Theta|}, j=1,\ldots,m} \sum_j \sum_{y_j} \phi_j(y_j) \cdot \mathbf{P}_j(y_j)$$

$$s.t. \quad \sum_{j=1}^{m} \phi_j(y_j) \leq u^{bi}(\cdot, \sigma^*(\mathbf{y})) \quad \forall \mathbf{y}.$$

Let $\{\phi_j^*\}_{j=1}^m$ be a solution to the dual problem. Define $D_j = co(\{\phi_j^*(y_j)|y_j \in Y_j\}) - \mathbb{R}_+^2$ for $j = 1, ..., m$. Note that $D_1 + \cdots + D_m \subseteq \mathcal{H}(A^{bi}, u^{bi})$, so $\{D_j\}_{j=1}^m$ is dominated by $\mathcal{H}(A^{bi}, u^{bi})$, and satisfies

$$\begin{aligned}
\sum_{j=1}^{m} W(P_j; D_j) &\geq \sum_{j=1}^{m} \sum_{y_j} \phi_j^*(\cdot, y_j) \cdot \mathbf{P_j}(y_j) \\
&= V(P_1, ..., P_m; (A^{bi}, u^{bi})) \\
&= W(P_1, ..., P_m; \mathcal{H}(A^{bi}, u^{bi})).
\end{aligned}$$

From Lemma 20, $cov(D_2), ..., cov(D_m)$ must be trivial, and property 4 of Lemma 18 implies that for each $j \neq 1$, there exists y_j^* such that $\phi_j^*(y_j^*) \geq \phi_j^*(y_j)$ for all y_j. Now we define $\tilde{\phi}_j(y_j) = \phi_j^*(y_j^*)$ for all y_j as a constant function. Since $\tilde{\phi}_j(y_j) \geq \phi_j^*(y_j)$ and $\phi_1^*, \tilde{\phi}_2, ..., \tilde{\phi}_m$ is feasible in the dual problem, $\phi_1^*, \tilde{\phi}_2, ..., \tilde{\phi}_m$ is also a solution to the dual problem.

From Lemma 2 and Corollary 1, a Blackwell supremum $P^* \in \mathcal{J}(P_1, ..., P_m)$ solves Nature's MinMax Problem. From the minmax theorem, P^* is a solution to

$$\min_{P \in \mathcal{J}(P_1, ..., P_m)} \sum_{\theta} P(\mathbf{y}|\theta) u^{bi}(\theta, \sigma^*(\mathbf{y})).$$

Lemma 21 implies that $P^*(\cdot|\theta_1)$ and $P^*(\cdot|\theta_2)$ have a common support, which we denote by $\bar{\mathbf{Y}} = \{\mathbf{y} \in \mathbf{Y}, \overline{\mathbf{P}}(\mathbf{y}) > 0\}$.

Now for any $(y_1, \bar{y}_{-1}) \in \bar{Y}$, complementary slackness implies

$$\phi_1^*(\cdot, y_1) + \sum_{j=2}^{m} \tilde{\phi}_j(\cdot, \bar{y}_j) = u^{bi}(\cdot, \sigma^*(y_1, \bar{y}_{-1})).$$

For any $(y_1, y_{-1}) \in Y$, the dual constraint says

$$\phi_1^*(\cdot, y_1) + \sum_{j=2}^{m} \tilde{\phi}_j(\cdot, y_j) \leq u^{bi}(\cdot, \sigma^*(y_1, y_{-1})).$$

Since $\tilde{\phi}_j$ is constant for $j \geq 2$, the left-hand side of the two equations above are the same, which implies $u(\cdot, \sigma^*(y_1, \bar{y}_{-1})) \leq u(\cdot, \sigma^*(y_1, y_{-1}))$. Since (A^{bi}, u^{bi}) is a non-trivial binary-action decision problem, any two (mixed) actions are either identical or induce payoff vectors that are not ordered. Therefore, $u^{bi}(\cdot, \sigma^*(y_1, \bar{y}_{-1})) \leq u^{bi}(\cdot, \sigma^*(y_1, y_{-1}))$ implies $\sigma^*(y_1, \bar{y}_{-1}) = \sigma^*(y_1, y_{-1})$. So we have derived that for any $y_1 \in Y_1$ and $y_{-1}, y'_{-1} \in Y_{-1}$, $\sigma^*(y_1, y_{-1}) = \sigma^*(y_1, y'_{-1})$, which concludes the proof.

□