Robust Aggregation of Correlated Information[*]

Henrique de Oliveira[†]   Yuhta Ishii[‡]   Xiao Lin[§]

August 29, 2024

Abstract

An agent makes decisions based on multiple sources of information. In isolation, each source is well understood, but their correlation is unknown. We study the agent's robustly optimal strategies — those that give the best possible guaranteed payoff, even under the worst possible correlation. With two states and two actions, we show that a robustly optimal strategy uses a single information source, ignoring all others. In general decision problems, robustly optimal strategies combine multiple sources of information, but the number of information sources that are needed has a bound that only depends on the decision problem. These findings provide a new rationale for why information is ignored.

Contents

1   Introduction

From the mundane to the important, most decisions are made with the aid of many readily available information sources. Treatment decisions can be made by consulting multiple doctors. Retirement plans can follow the advice of numerous financial experts. These different information sources are often correlated, as doctors may base their recommendations on the same study and financial analysts may have incentives to echo each other.

Understanding how multiple sources are correlated is hard. In a scientific study, for example, determining the correlation among variables requires an exponentially increasing sample size — known as the "curse of dimensionality." This difficulty brings the risk of misinterpreting correlations, which can lead to flawed inferences and, consequently, suboptimal decisions. Therefore, an agent may look to make decisions that do not leave them vulnerable to misspecifications in correlation.

In this paper, we assume that an agent fully understands each information source in isolation, but has no knowledge about the correlations among different information sources. To guard against this lack of knowledge, the agent chooses a strategy that performs well even under the worst possible correlation structure. As we will see, this robustness concern could lead the agent to ignore information.

As an example, consider the following hypothetical scenario: The Centers for Disease Control (CDC) is setting guidelines for administrating a new Covid treatment to a patient population that 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 only causes 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 of 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. For instance, we do not have data on how likely it is that a Covid patient will simultaneously develop both cough and fever symptoms. Lacking knowledge on correlation, the goal is to design a treatment guideline that utilizes the available information while being robust to all possible correlations.

A simple strategy that protects against the hidden 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 only on the Fever symptom, which guarantees a value of $\frac{1}{2}[0.5 \times 30 + 0.1 \times (-20)] = 7$. 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 makes use of both symptoms? Theorem 1 says the answer is *no*: *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. 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,

2

designed for the Flu. The payoff from each treatment is given in Table 3; the total payoff is the sum of the payoffs from the two treatments.

	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 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 $8.5 + 7 = 15.5$. However, the CDC can do better by basing the Treatment 1 decision on the Cough symptom and the Treatment 2 decision on the Fever symptom, as described in Table 4. This strategy, which uses the 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 in the decision problem in the above example is the additive separability of payoffs across the two treatments. Indeed, for any decision problem consisting of a collection of binary-state binary-action decision 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 surprisingly, when the state is binary, *every decision problem can be written as a separable problem*. Based on this observation, Theorem 2 provides a general construction of robustly optimal strategies for every binary-state decision problem.

The construction in Theorem 2 starts by removing all dominated actions and ordering the remaining n actions according to how much utility they generate in the

first state. The decision problem is then decomposed into $n-1$ binary-action decision problem, each specifying a local comparison between a pair of consecutive actions. These $n-1$ decomposed problems are combined into one separable problem, which proves to be equivalent to the original decision problem. In this separable problem, each local comparison uses a best-source strategy to determine the recommended action. Finally, a robustly optimal strategy can be constructed by mapping the profile of recommendations in each local comparison into a (possibly mixed) action in the original problem. Such a robustly optimal strategy uses an information source if and only if it is the best source in one of the local comparisons, and so it uses multiple information sources precisely when the best sources differ across local comparisons. Thus, the number of information sources used cannot exceed $n-1$, the number of decomposed problems.

With three or more states, we do not have a full characterization of the robustly optimal strategy. However, as in Theorems 1 and 2, Theorem 3 establishes a bound, N, such that there is always a robustly optimal strategy that uses at most N information sources. Again, this bound depends only on the decision problem, meaning that as the number of information sources grows large, the fraction of information sources used converges to 0.

Together, Theorems 1, 2, and 3 reveal a common theme: the agent tends to ignore some freely available information. The phenomenon of ignoring readily accessible information is well-documented. Existing explanations often attribute this behavior to costs or psychological distortions (see, e.g., Handel and Schwartzstein (2018) for a detailed discussion). Our results offer a different and less recognized rationale: ignorance of information can lead to more robust decisions when there is uncertainty 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.

To isolate the effect of ambiguity about the correlation among information sources, our model makes two assumptions: First, the agent has no knowledge about the correlation of the different information sources. Second, the agent possesses perfect knowledge of each of the information sources in isolation. In reality, the situation might be less extreme—there might be some knowledge about the correlations or some uncertainty about the marginal sources. Section 6.1 shows that our results ex-

tend straightforwardly to certain additional knowledge about the correlations among information sources. Section 6.2 extends Theorem 1 to a setting in which the agent also faces ambiguity about the marginal information sources.

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 discusses extensions. Section 7 concludes. The remainder of this introduction places our contribution within the context of 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.[1] 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. Moreover, 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.

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

---

[1]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 6.3 for further discussion of this assumption.

5

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.

A crucial step in our analysis involves understanding the joint informativeness of correlated information sources. Börgers, Hernando-Veciana, and Krähmer (2013) first introduce the notions of complementarity and substitutability between two information structures and provide an explicit characterization. Cheng and Börgers (2024) further explore the relationship between joint informativeness of experts' recommendations and their chance of disagreement. In contrast, our analysis specifically focuses on those least informative correlation structures, where the information sources can be viewed as extreme substitutes for each other.

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 source. 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 later simplify notation, we will later refer to decision problems as (A, u), where $u(\theta, a) = \mu_0(\theta)\rho(\theta, a)$ is the prior-weighted utility function.

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. The agent can observe the signals from all marginal experiments, $\{P_j\}_{j=1}^m$, but does not have detailed knowledge of 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:

$$V(P_1, \ldots, P_m; (A, u)) := \max_{\sigma \in \Sigma} \min_{P \in \mathcal{J}(P_1, \ldots, P_m)} \sum_{\theta \in \Theta} \sum_{(y_1, \ldots, y_m) \in \mathbf{Y}} P(y_1, \ldots, y_m | \theta) u(\theta, \sigma(y_1, \ldots, y_m)).$$

Whenever 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

3.1   The Blackwell Order

We will use the Blackwell order of experiments throughout the paper. For the sake of completeness, we briefly review it in this subsection. Readers familiar with the Blackwell order may choose to skip this subsection.

Definition 1. $P : \Theta \to \Delta(Y)$ *is* more informative *than* $Q : \Theta \to \Delta(Z)$ *if, for every decision problem, we have the inequality* $V(P) \geqslant V(Q)$. *We also say that* P *Blackwell dominates* Q.

We say two experiments are *Blackwell equivalent* if they dominate each other.

7

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

Definition 2. $Q : \Theta \rightarrow \Delta(Z)$ *is a garbling of* $P : \Theta \rightarrow \Delta(Y)$ *if there exists a function* $g : Y \rightarrow \Delta(Z)$ *(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 3. *Given a set of actions* A *and an experiment* $P : \Theta \rightarrow \Delta(Y)$, *the feasible set of* P *is*

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

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 (for a proof, see e.g. Blackwell (1953) or de Oliveira (2018)).

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 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.

Condition (4) is particularly useful as it offers a simple graphical representation of Blackwell experiments when $|\Theta| = 2$. Figure 1(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. The x-axis denotes the probability of taking this action in state 1, and the y-axis denotes the probability in state 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 1(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, thus Λ_P is a convex and symmetric[2] 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 $\Lambda_{P_1} \cup \Lambda_{P_2}$
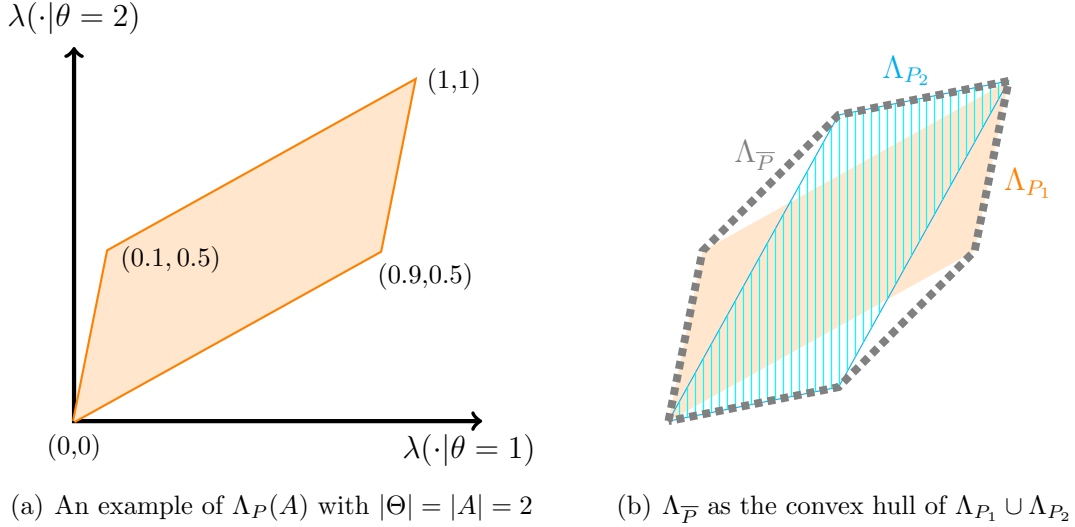
Figure 1

---

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

3.2 The Blackwell Supremum

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

Definition 4. *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 $\overline{P}$.*

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 1(b). From Blackwell's theorem, for any P' that is more informative than $P_1, ..., P_m$, the corresponding feasible set $\Lambda_{P'}$ must contain $\Lambda_{P_1}, \ldots, \Lambda_{P_m}$. Since the feasible set is always convex, $\Lambda_{P'}$ must also contain $co(\Lambda_{P_1} \cup \Lambda_{P_2} \cdots \cup \Lambda_{P_m})$. Moreover, the set $co(\Lambda_{P_1} \cup \Lambda_{P_2} \cdots \cup \Lambda_{P_m})$ is convex and symmetric, and so it corresponds to some Blackwell experiment $\overline{P}$, which is thus the least information Blackwell experiment that dominates $P_1, ..., P_m$ — the Blackwell supremum. This observation yields the following lemma:[3]

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 $\Lambda_{\overline{P}} = co(\Lambda_{P_1} \cup \Lambda_{P_2} \cdots \cup \Lambda_{P_m})$.*

It is useful to note that the above lemma holds specifically for when $|\Theta| = 2$. 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

---

[3]For a proof, see Proposition 16 in Bertschinger and Rauh (2014).

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

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

$$= \min_{P \in \mathcal{J}(P_1, \ldots, P_m)} V(P) \tag{1}$$

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.

Observe that every experiment in $\mathcal{J}(P_1, \ldots, P_m)$ must be more informative than every P_j, since the projection into the jth coordinate is a garbling. So if we let $\mathcal{D}(P_1, \ldots, P_m)$ denote the set of Blackwell experiments that dominates $P_1, \ldots, P_j$, then $\mathcal{J}(P_1, \ldots, P_m) \subseteq \mathcal{D}(P_1, \ldots, P_m)$.[4] The set $\mathcal{D}(P_1, \ldots, P_m)$ is in general a larger set, because not every experiment that dominates $P_1, \ldots, P_m$ can be represented as a joint experiments with marginals $P_1, \ldots, P_m$.[5] However, the next lemma shows that relaxing the 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 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 $P_1, \ldots, P_m$ exists, and it is the unique (up to Blackwell

---

[4]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, such that all sets $Y_j \subset U$. For our purposes $U = \mathbb{N}$ is large enough.

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

equivalence) Blackwell minimal element in $\mathcal{D}(P_1, \ldots, P_m)$. Therefore, $\overline{P}$ solves (2) regardless of the decision problem, which yields the following corollary.

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

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

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

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

In this section, we present our results for the special case in which $|\Theta| = 2$. We characterize both the robustly optimal strategies and values in this 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 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*.

It is sometimes 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 the agent 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. 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. We prove the first part of the theorem. The proof of the second part requires different arguments and so we defer it to Appendix B.5.

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,

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

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 $V(\overline{P}) \geq \max_{j'=1,\ldots,m} V(P_{j'})$, concluding the proof. $\square$

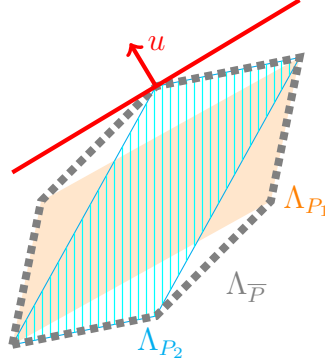The idea of Theorem 1 can be visualized in Figure 2 for two marginal experiments.

Figure 2: The maximum is achieved at an extreme point that belongs to Λ_{P_2}

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 function 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 using a single information source is 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. 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 best corresponding information source to determine whether to administer each treatment.

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

Definition 5. *A 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*

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

14

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 decision problems, (A_ℓ, u_ℓ), as a *subproblem.*

The next result provides a simple solution to separable problems: for each binary-action subproblem, by Theorem 1, one can derive a robustly optimal strategy by paying attention to the best marginal experiment and best responding to it. Assembling these strategies then yields a robustly optimal strategy for the original problem.

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 $\sigma_\ell : \mathbf{Y} \to \Delta A_\ell$ *be a robustly optimal strategy for subproblem* (A_ℓ, u_ℓ). *Then* $\sigma : \mathbf{Y} \to \Delta A_1 \times \cdots \times \Delta A_k$ *defined by*

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

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

Proof. See Appendix A.2. □

Remark. In any separable decision problem, it is immediate that

$$V\left(P_1, \ldots, P_m; \bigoplus_{\ell=1}^k (A_\ell, u_\ell)\right) \geq \sum_{\ell=1}^k \max_{j=1,\ldots,m} V(P_j; (A_\ell, u_\ell)). \tag{5}$$

The equality in Lemma 3 follows as a result of the 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.[6]

---

[6]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.

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 a sense to be made precise. The central idea involves decomposing an n-action decision problem into $n-1$ binary-action decision problems, and using these subproblems to construct the corresponding separable problem that is equivalent to the original problem. We call the resulting separable problem the *binary decomposition.*

We first define formally what it means for two decision problems to be equivalent. Given a decision problem (A, u), let[7]

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

be the associated polyhedron containing all payoff vectors that are either achievable or weakly dominated by some mixed action. An example of $\mathcal{H}(A, u)$ is depicted in Figure 3.
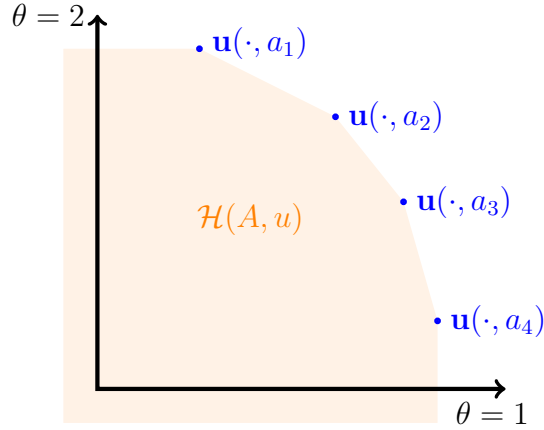

Figure 3: The shaded area represents $\mathcal{H}(A, u)$

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

$$V(P_1, \ldots, P_m; (A', u')) = V(P_1, \ldots, P_m; (A, u))$$

---

[7]Here and in what follows, whenever $+$ and $-$ are used in the operations of sets, they denote the Minkowski sum and difference.
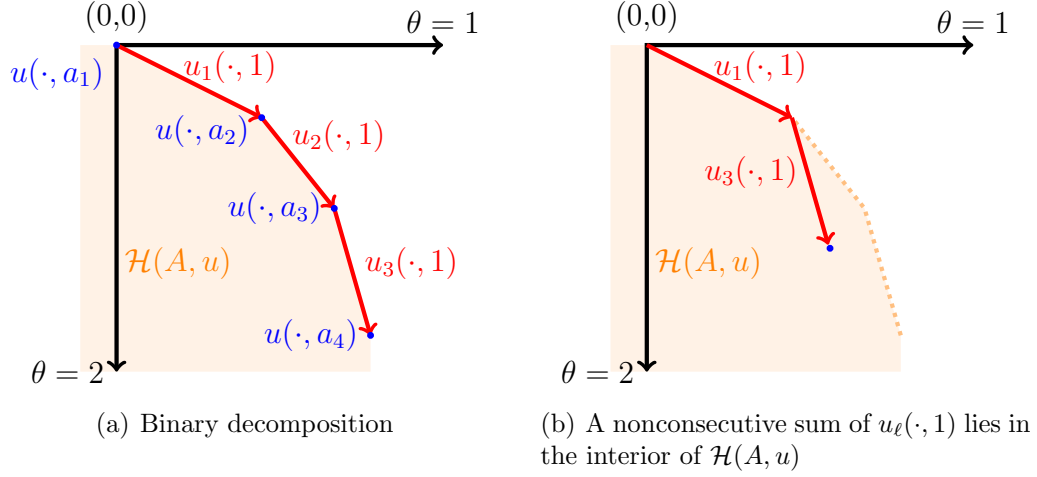
(a) Binary decomposition

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

Figure 4

for all Blackwell experiments $P_1, \ldots, P_m$, and so we call (A, u) and (A', u') *equivalent*.

Definition 6. *Two decision problems (A, u) and (A', u') are equivalent if $\mathcal{H}(A, u) = \mathcal{H}(A', u')$.*

Next we show by direct construction that, every binary-state decision problem is equivalent to a separable problem. We start with some normalization to simplify exposition. First we remove all weakly-dominated actions,[8] so that actions can be ordered such that

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

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

Definition 7. *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

---

[8]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.

actions. This can be visualized in Figure 4(a) for an example with four actions. 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 decomposed subproblem can be interpreted as choosing whether to "move forward" to the next action.

Notice that every feasible payoff vector in the original problem can be replicated in the binary decomposition. This is 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 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 $\mathcal{H}(A, u) = \mathcal{H}\big(\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)\big)$.

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

Proof. See Appendix A.3. □

Lemma 3 and Lemma 4 permit 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 robustly optimal strategy for (A_ℓ, u_ℓ). Then*

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

2. *There exists $\sigma^* : \mathbf{Y} \to \Delta A$ such that $u(\cdot, \sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{n-1} u_\ell(\cdot, \sigma_\ell(\mathbf{y}))$ for all $\mathbf{y}$. Moreover, any such σ^* is a robustly optimal strategy for (A, u).*

Proof. See Appendix A.4. □

Theorem 2 allows us to construct 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 optimal by Theorem 1).

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

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 more than three actions. This is because there could be multiple σ^* that satisfies $u(\cdot, \sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{n-1} u_\ell(\cdot, \sigma_\ell(\mathbf{y}))$ for all $\mathbf{y}$. For example, in the Covid example with two 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 the No Treatment with $T_1 + T_2$ does not decrease the guaranteed value, thus yielding another robustly optimal strategy. The reason that a robustly optimal strategy may play a dominated action is that, when such a strategy is played, the worst-case correlation structure will put probability 0 on the symptom realization $(Cough_-, Fever_-)$.

The theorem delivers two immediate corollaries about the (robust) value of different marginal information sources.

Corollary 2. *Suppose $\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)$ is the binary decomposition of (A, u). For any j,*

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

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 2 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 decomposed problems. In particular, an experiment that performs reasonably well across all decomposed problems can be completely ignored if, for each decomposed problem, there is some other, more specialized experiment that is the best. The next example shows that even when an experiment is the best single source, it can 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.*

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

	+	−
Covid	0.72	0.28
Flu	0.28	0.72

Table 5: Headache symptom

that Headache is the best information source when used in isolation, because it yields a value of $8+8 = 16$, which is greater than 15.5, the value of using either the Cough or Fever symptom alone. However, this symptom will not be used in a robustly optimal strategy: it 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.

Corollary 3. *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 $|J| \leq |A| - 1$, such that*

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

Corollary 3 implies that in an n-action decision problem, an agent needs to use at most $n - 1$ sources of information. 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.

5 General-State Decision Problems

Our previous analysis focuses on binary-state decision problems. The cornerstone of our approach is the decomposition of a complex decision problem into "elementary" binary-action problems. By aggregating the simple solution of these binary-action subproblems, we can derive a solution to the initial, more complex problem. A natural question is whether this approach can be extended into environments with more states. Unfortunately, it fails for multiple reasons: First, with more states, it is unclear how to decompose a general decision problem into more "elementary" ones. Second, the non-existence of the Blackwell supremum implies that in the Nature's minmax problem Eq. (1), there may no longer be a single experiment that uniformly minimize the agent's value across all decision problems, exacerbating the complexity

of the analysis (see Footnote 6). Lastly, an agent may want to use multiple information sources even in simple binary-action decision problems, as illustrated in Example 2 below.

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.*

<table>
<tr><td colspan="3" align="center">P_X</td><td></td><td colspan="3" align="center">P_Y</td></tr>
<tr><td>$P_X(x|\theta)$</td><td>x_1</td><td>x_2</td><td></td><td>$P_X(y|\theta)$</td><td>y_1</td><td>y_2</td></tr>
<tr><td>θ_1</td><td>1</td><td>0</td><td></td><td>θ_1</td><td>1</td><td>0</td></tr>
<tr><td>θ_2</td><td>1</td><td>0</td><td></td><td>θ_2</td><td>0</td><td>1</td></tr>
<tr><td>θ_3</td><td>0</td><td>1</td><td></td><td>θ_3</td><td>0</td><td>1</td></tr>
</table>

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:[9]

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

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 $1 + 0 + 1 = 2$.

Due to the difficulties highlighted above, an explicit construction of robustly optimal strategies remains an open question. However, our main point that robustly optimal strategies ignore many information sources remains valid even in general decision problems with larger state spaces. To this end, our main result of this section provides an upper bound on the number of information sources that are used by an agent under a robustly optimal strategy. Similar to Corollary 3, this upper bound depends only on the decision problem at hand, and does not depend on the fine details of the marginal information sources being observed by the agent.

---

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

To state our upper bound, we first examine the structure of the *interim value function* associated with the decision problem. Recall that a decision problem is a tuple $\Gamma \equiv (\Theta, \mu_0, A, \rho)$ with a finite state space Θ, a prior $\mu_0 \in \Delta\Theta$, a finite action space A, and a utility function $\rho : \Theta \times A \to \mathbb{R}$. For a given decision problem, the corresponding interim value function, $v : \Delta(\Theta) \to \mathbb{R}$, is defined as

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

Given a value function $v : \Delta(\Theta) \to \mathbb{R}$, its epigraph is defined as $\mathrm{epi}(v) = \{(\mu, w) : w \geq v(\mu), \mu \in \Delta(\Theta)\}$. It can be easily seen that the set of extreme points of the epigraph, denoted by $\mathrm{ext}(\mathrm{epi}(v))$, is finite and contains $\{(\delta_1, v(\delta_1)), ..., (\delta_n, v(\delta_n))\}$, where δ_i denotes the Dirac measure on θ_i. We call the elements of $\mathrm{ext}(\mathrm{epi}(v))$, excluding those degenerate points $(\delta_i, v(\delta_i))$, the *kinks* of v.[10] Thus, *the number of kinks of v is* $|\mathrm{ext}(\mathrm{epi}(v))| - |\Theta|$. See Figure 5 for an illustration 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 the kinks.
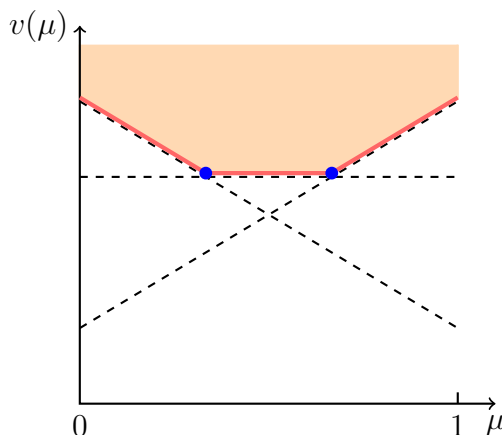

Figure 5: Interim value function and kinks

The following theorem provides a bound on the number of experiments that a agent would need, which is the number of kinks of the corresponding interim value function.

---

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

Theorem 3. *Consider any decision problem whose corresponding interim value function has k kinks. For any collection of experiments $\{P_j\}_{j=1}^m$, there exists a subset of marginal experiments $\{P_j\}_{j\in J\subseteq\{1,...,m\}}$ with $|J| \le k$, such that*

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

Proof. See Appendix A.5. □

Remark.   1. Theorem 3, along with (1), implies that there always exists a correlation among the information sources such that the marginal value of information sources outside J are all 0. This suggests that ignorance of information can also be rationalized by a Bayesian agent with a subjective correlation structure that is unobserved by an analyst.

2. It is easy to see that multiplying the utilities by a constant does not change the set of robustly optimal strategies. Consequently, Theorem 3 shows that a robustly optimal strategy attends to at most k information sources even when the stakes of the problem are arbitrarily large.

The full proof of Theorem 3 is deferred to the Appendix, but here we sketch the main steps. By the minmax theorem, it suffices to examine Nature's minmax problem:

$$V(P_1,\ldots P_m) = \min_{P\in\mathcal{J}(P_1,...,P_m)} V(P).$$

By Lemma 2, Nature's minmax problem can be relaxed into choosing an experiment among the set of all experiments that Blackwell dominate $P_1,...,P_m$. According to Blackwell's theorem (Theorem 1 in Blackwell (1953)), this is equivalent to choosing a posterior distribution that is a mean-preserving spread of the posterior distributions induced by $P_1,...,P_m$.

Next, note that the interim value function is convex and piecewise linear. Moreover, the "kinks" are the extreme points of those linear faces. Any non-extreme point in those linear faces can be expressed as a convex combination of extreme points. Thus, we can apply a mean-preserving spread to take any belief into extreme points while leaving the expected payoff unchanged. This allows us to further simplify the Nature's minmax value, by restricting attention to those experiments whose induced posterior distributions are supported on the extreme points. This set can be characterized by a k-dimensional polytope, where k is the number of kinks.

23

Now Nature's problem can be written as a k-dimensional linear program with k effective constraints. These k effective constraints must come from at most k marginal experiments. Consequently, the value of the problem is the same as the value of the problem with k experiments. Hence, the agent need not use more than k experiments.

The bound in Theorem 3 is based on the number of kinks of the interim value function, which may be hard to calculate when there are more than three states. The next corollary provides a bound that depends only on $|\Theta|$ and $|A|$. The key idea is that by the Upper Bound Theorem for polytopes, the maximum number of kinks can be derived as a function of the number of facets of the epigrapah, which in turns is connected to the number of actions.

Corollary 4. *Consider any decision problem. For any collection of experiments* $\{P_j\}_{j=1}^m$, *there exists a subset of marginal experiments* $\{P_j\}_{j \in J \subseteq \{1,...,m\}}$ *with*

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

such that

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

Proof. See Appendix A.6. $\qquad\square$

Remark. 1. When $|\Theta| = 2$, this bound reduces to $|A| - 1$, the bound given in Corollary 3;

2. The bounds in both Theorem 3 and Corollary 4 depend only on the decision problem. Therefore, as m grows large, the above theorem tells us that there exists a sequence of robustly optimal strategies for which the fraction of information sources that are ignored converges to 1.

Theorem 3 suggests one may ignore information sources due to robustness concerns. The following proposition further tells us which information sources may be ignored: if an information source P_m is not the best information source among $\{P_j\}_{j=1}^m$ for any decision problem, then it can be safely ignored in a robustly optimal strategy.

Proposition 1. *Suppose that* $V(P_m; (A, u)) \leq \max_{j=1,...,m-1} V(P_j; (A, u))$ *for all decision problems* (A, u). *Then, for any decision problem* (A', u'),

$$V(P_1, ..., P_m; (A', u')) = V(P_1, ..., P_{m-1}; (A', u')).$$

Proof. See Appendix B.1. □

The condition in Proposition 1 is weaker than P_m being Blackwell dominated by one of the other experiments $P_1, ..., P_{m-1}$, because the experiment that outperforms P_m may depend on the particular decision problem (A, u). As shown in Cheng and Borgers (2023), this condition is equivalent to P_m being dominated by a convex combination of $P_1, ..., P_{m-1}$. This characterization will be useful in our proof.[11]

Similar to Corollary 2, this proposition highlights a sense in which it is beneficial to gather information from multiple information sources that are specialized: the agent prefers to pay attention only to those information sources that perform the best in isolation in some decision problem.

6 Discussion

This section discusses some extensions of our model. Section 6.1 discusses the implications of additional knowledge about the correlation structure. Section 6.2 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 6.3 considers the case where the information sources available to the agent have already been processed by experts.

6.1 Knowledge of Correlation

6.1.1 Common Source

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

---

[11]In the proof, we established a slightly stronger result than Proposition 1: experiment P_m can be ignored if it is dominated by all correlation structures among $P_1, ..., P_{m-1}$.

Formally, we say a joint experiment $P \in \mathcal{J}(P_1, ..., P_m)$ is *rationalizable by a common source* 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 fundamental information source, and the experiments $P_1, ..., P_j$ are generated by independent garblings of signals from Q.

An immediate observation is that every $P \in \mathcal{J}(P_1, ..., P_m)$ is rationalizable by a common source. This can be seen by letting the common source Q be P itself, and the garblings γ_j be the deterministic functions that project each vector $y_1, ..., y_m$ into y_j. Therefore, this additional knowledge does not exclude any possible correlation.

6.1.2 Partial Knowledge of Correlations

In certain situations, an agent may understand the correlation among some information sources, even if they do not comprehend all of them. For example, in medical diagnosis, 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, $\Pi = \{S_1, \ldots, S_k\}$, of $\{1, 2, \ldots, m\}$ 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 separate marginal experiments, and then our analysis extends in a straightforward manner to this envi-

ronment.

However, our analysis does not immediately extend to other more complex situations. In particular, when the knowledge on the correlations span across non-disjoint subsets, the set of possible joint experiments can no longer 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 a causality diagram (see Pearl (2009) and Spiegler (2016)).

6.2   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:

$$V(\mathcal{P}_1, \ldots, \mathcal{P}_m) := \max_{\sigma: \mathbf{Y} \to \Delta(A)} \min_{P \in \mathcal{J}(\mathcal{P}_1, \ldots, \mathcal{P}_m)} \sum_{\theta \in \Theta} \sum_{y_1, \ldots, y_m \in \mathbf{Y}} P(y_1, \ldots, y_m | \theta) u(\theta, \sigma(y_1, \ldots, y_m)).$$

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

Proposition 2. *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.2. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

6.3 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 $P_j :$ $\Theta \to Y_j$, 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:

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

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, compressing the 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 3. *When* $|\Theta| = 2$, *for any* (A, u),

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

Proof. See Appendix B.3. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

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 Conclusion

We have shown how concerns about correlations could lead the agent to disregard seemingly relevant information. The most striking result emerges when there are only two states and two actions: the agent must rely on a single information source,

ignoring all others. Moreover, binary-state binary-action decision problems serve as building blocks for any general binary-state decision problem, enabling us to explicitly characterize the robustly optimal strategies. Such strong results come with the strong binary-state assumption, which certainly restricts their applicability. Nevertheless, there are interesting settings where the uncertainty can be naturally modeled with two states, such as in simple hypothesis testing, disease diagnoses, and presidential elections.

With more than two states, we do not have a closed-form characterization of the optimal strategy, but we establish a bound on the number of information sources that are used. Crucially, this bound depends only on the decision problem and does not depend on the number of information sources or their specific details. Thus, the bound specifies a limit to the addition of relevant information sources — at some point, any new relevant information source introduced will necessarily make at least one existing source redundant.

Aside from providing normative guidelines for constructing robust strategies, our findings offer an alternative explanation for the ignorance of information, which has distinct implications compared to existing explanations. For example, in models of rational inattention (see Maćkowiak, Matějka, and Wiederholt (2023) for a survey), more information is acquired and used when stakes are raised. In contrast, in our model, multiplying the utility function by any constant does not alter the set of information sources attended to. This distinction may help explain the ignorance of information even in high-stake decision problems.[12]

Throughout the paper, we have interpreted our model as applying to situations where the agent lacks information about the correlation among sources. An alternative interpretation is that this information is available, but the agent lacks the mental capacity to utilize this information. Indeed, there is empirical evidence that people often make mistakes when trying to use information about correlations—a phenomenon known as *correlation neglect* (see, e.g., Enke and Zimmermann (2019)). A sophisticated agent who is aware of their potential bias might try to protect themselves by robustly optimizing.

Our analysis leaves open some interesting questions. As discussed in Section 6.1 and Section 6.2, other sets of joint experiments can be conceived. Investigating such

---

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

alternatives could give insight into what kinds of information are more or less valuable depending on the kind of ambiguity present. In addition, our paper does not make any assumptions about the marginal experiments. In certain applications, parametric assumptions might be natural, such as assuming Gaussian distributions. This extra structure could allow reaching stronger conclusions. Lastly, in networks, agents' actions are influenced by observing their neighbors, creating a complex interdependence shaped by the network structure. It seems natural to model agents as robust optimizers in such environments where the correlation among sources is important, but difficult to determine.

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 Borgers. Dominance and optimality. *Working paper*, 2023.

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.

Benjamin Enke and Florian Zimmermann. Correlation neglect in belief formation. *The Review of Economic Studies*, 86(1):313–332, 2019.

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.

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

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.

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

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

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

Rakesh V Vohra. *Advanced mathematical economics*. Routledge, 2004.

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

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. $\square$

A.2 Proof of Lemma 3

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

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

Moreover, by Theorem 1 and Corollary 1,

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

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

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

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.

$\square$

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 otherwise 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 $\ell' > \ell$,

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

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 $\ell \geq \ell^*$.
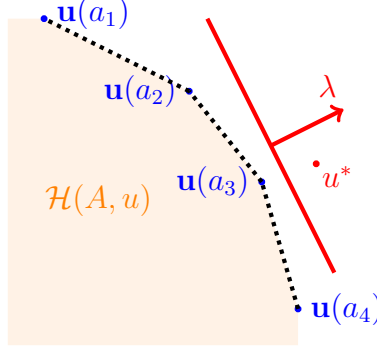
Figure 6

Thus

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

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).

$\square$

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.

For each $\mathbf{y}$, $\sum_{\ell=1}^{n-1} u_\ell(\cdot, \sigma_\ell(\mathbf{y})) \in \mathcal{H}\left(\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)\right) = \mathcal{H}(A, u)$. So there exists

36

$\sigma^*(\mathbf{y})$ such that $u(\cdot, \sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{n-1} u_\ell(\cdot, \sigma_\ell(\mathbf{y}))$. Now for any $P \in \mathcal{J}(P_1, ..., P_m)$,

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

where the second line follows from Lemma 3 and the third line follows from Lemma 4. So σ^* is a robustly optimal strategy. $\qquad\square$

A.5 Proof of Theorem 3

We shall start with some preliminary definitions and lemmas.

A.5.1 Definitions

For a Blackwell experiment $P : \Theta \to \Delta Y$, the induced posterior distribution $\tau^P \in \Delta(\Delta(\Theta))$ is defined as

$$
\tau^P(\mu) = \sum_{y \in Y_\mu} \sum_\theta \mu_0(\theta) P(y|\theta)
$$

where

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

Since we assume that Y is finite, τ^P will always have a finite support. For convenience, when we sum over μ a term that multiplies $\tau^P(\mu)$, it is to be understood that the sum is over the support of τ^P.

For each $a \in A$, let

$$
M_a = \left\{ \mu \in \Delta(\Theta) \,\middle|\, \sum_\theta \mu(\theta) \rho(\theta, a) \geq \sum_\theta \mu(\theta) \rho(\theta, a') \text{ for all } a' \in A \right\}
$$

denote the set of beliefs that action a best responds to. It is easy to check that M_a is convex, compact, and with finitely many extreme points. Let $E_a = \text{ext}(M_a)$.

Given an interim value function $v : \Delta(\Theta) \to \mathbb{R}$, let E denote the projection of $\text{ext}(\text{epi}(v))$ on $\Delta(\Theta)$.

A.5.2 Lemmas

Lemma 5. *For every a, $E_a \subseteq E$.*

Proof. Suppose, by contradiction, that there exists some a, $\mu \in E_a$, and $\mu \notin E$.

Since $\mu \notin E$, there exists $(\mu', w'), (\mu'', w'') \in \mathrm{epi}(v)$ and $\lambda \in (0, 1)$ such that $\mu' \neq \mu''$, and

$$(\mu, v(\mu)) = \lambda(\mu', w') + (1 - \lambda)(\mu'', w'').$$

Note that $(\mu, v(\mu))$ is a boundary point of $\mathrm{epi}(v)$, so by the supporting hyperplane theorem, there exists $h \in \mathbb{R}^{|\Theta|}$, and $c \in \mathbb{R}$ such that

$$h \cdot (\mu, v(\mu)) = c \quad \text{and} \quad h \cdot (\hat{\mu}, \hat{w}) \geq c, \quad \forall (\hat{\mu}, \hat{w}) \in \mathrm{epi}(v).$$

Moreover, the last element in h must be positive, because $\mathrm{epi}(v)$ is upper unbounded in its last dimension.

Now we claim that both (μ', w') and (μ'', w'') must be on this supporting hyperplane. That is, $h \cdot (\mu', w') = h \cdot (\mu'', w'') = c$. Otherwise, if either $h \cdot (\mu', w') > c$ or $h \cdot (\mu'', w'') > c$, we have $h \cdot (\mu, v(\mu)) = \lambda h \cdot (\mu', w') + (1 - \lambda) h \cdot (\mu'', w'') > c$, a contradiction.

Moreover, since the last element of h is positive, we must have $w' = v(\mu')$ and $w'' = v(\mu'')$, otherwise $h \cdot (\mu, v(\mu)) > \lambda h \cdot (\mu', v(\mu')) + (1 - \lambda) h \cdot (\mu'', v(\mu'')) \geq c$, again a contradiction.

Now we have $v(\mu) = \lambda v(\mu') + (1 - \lambda) v(\mu'')$. By the definition of v, $v(\mu') \geq \sum_\theta \mu'(\theta) \rho(\theta, a)$ and $v(\mu'') \geq \sum_\theta \mu''(\theta) \rho(\theta, a)$. So $v(\mu) \geq \lambda \sum_\theta \mu'(\theta) \rho(\theta, a) + (1 - \lambda) \sum_\theta \mu''(\theta) \rho(\theta, a)) = \sum_\theta \mu(\theta) \rho(\theta, a) = v(\mu)$. This means that $v(\mu') = \sum_\theta \mu'(\theta) \rho(\theta, a)$ and $v(\mu'') = \sum_\theta \mu''(\theta) \rho(\theta, a)$, which implies $\mu', \mu'' \in M_a$. This contradicts the assumption that $\mu \in E_a$ and concludes the proof. $\qquad\square$

Given a finite collection of Blackwell experiments $P_1, ..., P_m$, recall that $\mathcal{D}(P_1, ..., P_m)$ denotes the set of Blackwell experiments that dominate $P_1, ..., P_m$. Let $\mathcal{E}$ denote the set of all experiments with the induced posterior distribution supported in E and $\mathcal{E}(P_1, ..., P_m) = \mathcal{D}(P_1, ..., P_m) \cap \mathcal{E}$ denote the set of experiments that are more informative than $P_1, \ldots, P_m$ and have their support in E.

Lemma 6. *For any experiment P, there exists $Q \in \mathcal{E}(P)$ such that $V(P) = V(Q)$.*

Proof. Given a belief μ, let a be an action such that $\mu \in M_a$. Thus μ can be written as a convex combination of points in $E_a = \text{ext}(M_a)$, as

$$\sum_{\nu \in E_a} \gamma(\nu|\mu)\nu = \mu.$$

This defines a mean-preserving spread $\gamma : \Delta(\Theta) \to \Delta(\Delta(\Theta))$ satisfying the property that beliefs in M_a are taken to beliefs still in M_a. Now let Q be an experiment such that τ^Q is a mean-preserving spread of τ^P through γ, that is,

$$\tau^Q(\nu) = \sum_{\mu} \gamma(\nu|\mu)\tau^P(\mu).$$

By Theorem 1 in Blackwell (1953), $Q \in \mathcal{D}(P)$ and since Q has its support in E, we have that $Q \in \mathcal{E}(P)$. It remains to show that $V(P) = V(Q)$. Notice that, for each action $a \in A$, v is linear within M_a. Thus

$$\begin{aligned}
V(P) &= \sum_{\mu} \tau^P(\mu)v(\mu) = \sum_{\mu} \tau^P(\mu)\, v\left(\sum_{\nu \in E} \gamma(\nu|\mu)\nu\right) \\
&= \sum_{\mu} \tau^P(\mu) \sum_{\nu \in E} \gamma(\nu|\mu)v(\nu) = \sum_{\nu \in E} \sum_{\mu} \tau^P(\mu)\gamma(\nu|\mu)v(\nu) \\
&= \sum_{\nu \in E} \tau^Q(\nu)v(\nu) = V(Q).
\end{aligned}$$

$\square$

Lemma 7.
$$V(P_1, ..., P_m) = \min_{P \in \cap_{j=1}^m \mathcal{E}(P_j)} V(P)$$

Proof. First note that, by Lemma 2

$$V(P_1, ..., P_m) = \min_{P \in \mathcal{J}(P_1,...,P_m)} V(P) = \min_{P \in \mathcal{D}(P_1,...,P_m)} V(P) \leq \min_{Q \in \mathcal{E}(P_1,...,P_m)} V(Q)$$

since $\mathcal{E}(P_1, ..., P_m) \subseteq \mathcal{D}(P_1, ..., P_m)$. By Lemma 6, for any $P \in \mathcal{D}(P_1, ..., P_m)$, there exists a $Q \in \mathcal{E}(P) \subseteq \mathcal{E}(P_1, ..., P_m)$ with $V(P) = V(Q)$, so the inequality above must actually be an equality. The proof is concluded by noting that

$$\mathcal{E}(P_1, ..., P_m) = \mathcal{D}(P_1, ..., P_m) \cap \mathcal{E} = \cap_{j=1}^m \mathcal{D}(P_j) \cap \mathcal{E} = \cap_{j=1}^m (\mathcal{D}(P_j) \cap \mathcal{E}) = \cap_{j=1}^m \mathcal{E}(P_j).$$

$\square$

For any belief $\mu \in \Delta(\Theta)$, let

$$N(\mu) = \left\{ \eta \in \mathbb{R}^E \;\middle|\; \begin{array}{l} \sum_{\nu \in E} \eta(\nu)\nu = \mu \\ \eta \in \Delta(E) \end{array} \right\}$$

be the set of distributions that spread the belief μ into points in E. Since the constraints that define $N(\mu)$ are linear and $\Delta(E)$ is bounded, $N(\mu)$ is a polytope.

Given an experiment P, we define N^P as the τ^P-weighted Minkowski sum of $)N(\mu)$:

$$N^P = \sum_\mu \tau^P(\mu) N(\mu).$$

Note that N^P is also a polytope in $\mathbb{R}^E$.

Lemma 8. *Let $Q \in \mathcal{E}$. Then $Q \in \mathcal{E}(P)$ if and only if $\left(\tau^Q(\nu)\right)_{\nu \in E} \in N^P$.*

Proof. $Q \in \mathcal{E}(P)$ if and only τ_Q is a mean-preserving spread of τ_P, i.e., there exists an $\eta : supp(\tau^P) \to \Delta E$, such that, for any $\mu \in supp(\tau^P)$ and $\nu \in E$,

$$\mu = \sum_{\nu \in E} \eta(\nu|\mu)\nu \tag{7}$$

$$\tau^Q(\nu) = \sum_\mu \eta(\nu|\mu)\tau^P(\mu). \tag{8}$$

Condition (7) corresponds to the requirement that $\eta(\cdot|\mu) \in N(\mu)$. Thus, saying that τ_Q is a mean-preserving spread of τ_P is equivalent to saying that there exist $\eta(\cdot|\mu) \in N(\mu)$ such that (8) holds for all ν or, in other words, $\left(\tau^Q(\nu)\right)_{\nu \in E} \in N^P$. $\square$

The set E includes the extreme points of the simplex itself, $\text{ext}\,(\Delta(\Theta))$; the remaining elements belong to the set of "kinks" K. We now show that a measure in with support in E is uniquely determined by its values on K

Lemma 9. *Let $\eta, \eta' \in N(\mu)$ be such that $\eta(\nu) = \eta'(\nu)$ for all $\nu \in K$. Then $\eta = \eta'$.*

Proof. By definition, $E \backslash K = \text{ext}\,(\Delta(\Theta)) = \{\delta_\theta | \theta \in \Theta\}$, where δ_θ is the Dirac measure putting probability one in the state θ. Given $\theta \in \Theta$, we have

$$\mu(\theta) = \sum_{\nu \in E} \eta(\nu)\nu(\theta) = \sum_{\nu \in K} \eta(\nu)\nu(\theta) + \eta\,(\delta_\theta).$$

40

Thus, if $\eta(\nu) = \eta'(\nu)$ for all $\nu \in K$, we must have $\eta(\delta_\theta) = \eta'(\delta_\theta)$ for all $\theta \in \Theta$, so $\eta = \eta'$. $\qquad \square$

Now let $Z(\mu)$ be the projection of $N(\mu)$ on $\mathbb{R}^K$, so Z^P is also the projection of N^P on $\mathbb{R}^K$. As the projection of a polytope, Z^P is also a polytope and we can write

$$Z^P = \sum_\mu \tau^P(\mu) Z(\mu).$$

The following lemma shows that $\mathcal{E}(P)$ is characterized by Z^P.

Lemma 10. *Let $Q \in \mathcal{E}$. Then $Q \in \mathcal{E}(P)$ if and only if $\left(\tau^Q(\nu)\right)_{\nu \in K} \in Z^P$.*

Proof. By Lemma 8, $Q \in \mathcal{E}(P)$ if and only if $\left(\tau^Q(\nu)\right)_{\nu \in E} \in N^P$, that is, there exist $\eta(\cdot|\mu) \in N(\mu)$ for each μ in the support of τ^P such that

$$\tau^Q(\nu) = \sum_\mu \tau^P(\mu)\eta(\nu|\mu) \quad \forall \nu \in E.$$

Clearly, if this holds, then $\left(\tau^Q(\nu)\right)_{\nu \in K} \in Z^P$, by definition of Z^P. Now suppose that $\left(\tau^Q(\nu)\right)_{\nu \in K} \in Z^P$, that is, there exist $\eta'(\cdot|\mu) \in Z(\mu)$ such that

$$\tau^Q(\nu) = \sum_\mu \tau^P(\mu)\eta'(\nu|\mu) \quad \forall \nu \in K.$$

By Lemma 9, for each μ there is a unique $\eta(\cdot|\mu) \in N(\mu)$ whose projection in $\mathbb{R}^K$ is $\eta'(\cdot|\mu)$. Thus $\tau^Q(\nu)$ must equal $\sum_\mu \tau^P(\mu)\eta(\nu|\mu)$ for $\nu \in K$ and applying Lemma 9 to $\left(\tau^Q(\nu)\right)_{\nu \in E} \in N(\mu_0)$, we conclude that it must equal for all $\nu \in E$, so that $\left(\tau^Q(\nu)\right)_{\nu \in E} \in N^P$. $\qquad \square$

Proof of Theorem 3. Let $k = |K|$. Our goal is to show that there is a robustly optimal strategy using at most k experiments. By Lemma 7,

$$V(P_1, ..., P_m) = \min_{P \in \cap_{j=1}^m \mathcal{E}(P_j)} V(P)$$

Given an experiment Q, we define $\tau_K^Q := (\tau^Q(\nu))_{\nu \in K}$ and $\tau^Q(\delta_\theta) = \left[\mu_0(\theta) - \right.$

$\sum_{\ell=1}^{k} \tau^Q(\nu_\ell)\nu(\theta)\Big]$. By Lemma 10, the problem can be written as

$$V(P_1,...,P_m) = \min_{\tau_K^Q \in \bigcap\limits_{j=1}^{m} Z^{P_j}} \left(\sum_{\nu \in K} \tau^Q(\nu)v(\nu) + \sum_{\theta \in \Theta} \tau^Q(\delta_\theta)v(\delta_\theta) \right) \tag{9}$$

Since each Z^{P_j} is a polytope, so is their intersection, which is also non-empty because a fully informative information structure is always in $\mathcal{E}(P_j)$. Moreover, the objective function is affine in $\tau_K^Q \in \mathbb{R}^K$, so (9) can be written as a linear program:

$$V(P_1,...,P_m) = \max_{x \in \mathbb{R}^k} c \cdot x + constant \tag{10}$$
$$s.t. \quad Ax \le b,$$

for some $c \in \mathbb{R}^k$. Because the set of constraints is non-empty and bounded, the problem has a solution x^*.

Now, from a standard observation in linear programming (see Lemma 15 in Appendix B.4), we can keep only k effective constraints in the k dimensional linear program (10), without affecting the value of the problem. Let $I \subseteq \{1,...,N\}$ be an index set of these effective constraints, with $|I| = k$. Let $A[I]$ denote the $k \times k$ submatrix of A with the rows in I. Similarly, let $b[I]$ denote the $k \times 1$ subvector of b with the rows in I. Then we have

$$V(P_1,...,P_m) = \max c \cdot x + constant$$

$$s.t. \quad A[I]x \le b[I].$$

Each constraint comes from some Z^{P_j}. Since we have k constraints, there is a set $J \subseteq \{1,\ldots,m\}$ with $|J| \le k$ such that every constraint comes from a Z^{P_j} with $j \in J$. Not all constraints associated with each Z^{P_j} are present in this linear program, but adding them back does not alter the value. This means that the program the problem can be reformulated back to

$$\min_{\tau_K^Q \in \bigcap\limits_{j \in J} Z^{P_j}} \left(\sum_{\nu \in K} \tau^Q(\nu)v(\nu) + \sum_{\theta \in \Theta} \tau^Q(\delta_\theta)v(\delta_\theta) \right) = V(\{P_j\}_{j \in J}).$$

Therefore, $V(P_1,...,P_m) = V(\{P_j\}_{j \in J})$. $\qquad\square$

A.6   Proof of Corollary 4

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:

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

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 the number of kinks can be no more than

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

$\square$

B Online Appendix

B.1 Proof of Proposition 1

To prove the proposition, it is useful to introduce the "dominated by a convex combination" notion in Cheng and Borgers (2023). Let $\{P_1, ..., P_k\}$ be a collection of Blackwell experiments, with disjoint signal spaces $Y_1, ..., Y_k$. A convex combination of these Blackwell experiments, denoted by $\bigoplus_{j=1}^{k} \alpha_j P_j$, is a single Blackwell experiment with a signal space $Y_1 \cup \cdots \cup Y_k$:

$$\bigoplus_{j=1}^{k} \alpha_j P_j(z|\theta) = \alpha_j P_j(z|\theta) 1_{z \in Y_j}$$

where $\alpha_j \geq 0$ and $\sum_j \alpha_j = 1$.

The following lemma directly follows from the "if" direction of Proposition 1 in Cheng and Borgers (2023).

Lemma 11. *If for any decision problem* (A, u), $V(P_m; (A, u)) \leq \max_{j=1,...,m-1} V(P_j; (A, u))$, *then* P_m *is Blackwell dominated by a convex combination of* $\{P_1, ..., P_{m-1}\}$.

The next lemma shows that any convex combination of $\{P_1, ..., P_k\}$ is dominated by any joint experiments with marginals $P_1, ..., P_k$.

Lemma 12. *For any* $P \in \mathcal{J}(P_1, ..., P_k)$ *and any weights* $\{\alpha_j\}_{j=1}^{k}$, P *Blackwell dominates* $\bigoplus_{j=1}^{k} \alpha_j P_j$.

Proof. For any $P \in \mathcal{J}(P_1, ..., P_k)$, we construct the following garbling: $\gamma : Y_1 \times ... \times Y_k \to \Delta(Y_1 \cup \cdots \cup Y_k)$:

$$\gamma(y|y_1, ..., y_k) = \begin{cases} \alpha_j & \text{if } y = y_j, \\ 0 & \text{otherwise.} \end{cases}$$

Then for any j and $y \in Y_j$,

$$\sum_{y_1,...,y_k} \gamma(y|y_1,...,y_k)P(y_1,...,y_k|\theta) = \sum_{y_{-j}} \alpha_j P(...,y_{j-1},y,y_{j+1}...|\theta)$$

$$= \alpha_j P(y|\theta)$$

$$= \bigoplus_{j=1}^{k} \alpha_j P_j(y|\theta),$$

so P Blackwell dominates $\bigoplus_{j=1}^{k} \alpha_j P_j$. $\qquad \square$

Proof of Proposition 1. For any decision problem (A,u), let P^* be the joint experiment solving

$$\min_{P \in \mathcal{J}(P_1,...,P_{m-1})} V(P;(A,u)).$$

From Lemma 12 and the transitivity of the Blackwell order, P^* dominates P_m. So there exists $\gamma : Y_1 \times ... \times Y_{m-1} \to \Delta Y_m$ such that

$$P_m(y_m|\theta) = \sum_{y_1,...,y_{m-1}} \gamma(y_m|y_1,...,y_{m-1})\tilde{P}(y_1,...,y_{m-1}|\theta).$$

Now we construct the following $Q \in \mathcal{J}(P_1,...,P_m)$:

$$Q(y_1,...,y_m|\theta) = \gamma(y_m|y_1,...,y_{m-1})P^*(y_1,...,y_{m-1}|\theta)$$

which by construction is Blackwell equivalent to P^*. Therefore,

$$\begin{aligned}
V(P_1,...,P_m;(A,u)) &= \min_{P \in \mathcal{J}(P_1,...,P_m)} V(P;(A,u)) \\
&\leq V(Q;(A,u)) \\
&= V(P^*;(A,u)) \\
&= V(P_1,...,P_{m-1};(A,u)) \\
&\leq V(P_1,...,P_m;(A,u))
\end{aligned}$$

which proves the proposition. $\qquad \square$

B.2   Proof of Proposition 2

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

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

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 $P \in \mathcal{J}(P_1, \ldots, P_m)$:

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

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

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

$$= \max_{j=1,\ldots,m} V(\underline{P}_j)$$

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 $V(\underline{P}_{j^*}) = \max_j V(\underline{P}_j) \geq V(\mathcal{P}_1, \ldots, \mathcal{P}_m)$. $\quad\square$

B.3   Proof of Proposition 3

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

$$\mu'(\theta_1) u(\theta_1, a_j) + \mu'(\theta_2) u(\theta_2, a_j) < \mu'(\theta_1) u(\theta_1, a_k) + \mu'(\theta_2) u(\theta_2, a_k). \tag{11}$$

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

$$\mu'(\theta_1)u(\theta_1, a_j) + \mu'(\theta_2)u(\theta_2, a_j) \leq \mu'(\theta_1)u(\theta_1, a_i) + \mu'(\theta_2)u(\theta_2, a_i) \qquad (12)$$

The inequalities (11) and (12) 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. $\qquad \square$

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 $R_{P_j}^\ell$ for (A_ℓ, u_ℓ) and showing that $V(R_{P_j}; (A_\ell, u_\ell)) \geq V(R_{P_j}^\ell; (A_\ell, u_\ell)) = V(P_j; (A_\ell, u_\ell))$.

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,

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

From Lemma 13, if $a_i \in \text{argmax}_{a \in A} \sum_\theta P_j(y_j|\theta)u(\theta, a)$, 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 $i \leq \ell$ 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,

$$\gamma_\ell(\sigma^*(y_j)) \in \underset{a \in A_\ell}{\arg\max} \sum_{\theta, y_j} P_j(y_j|\theta) u_\ell(\theta, a),$$

so the experiment $R_{P_j}^\ell$, 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 $V(R_{P_j}^\ell; (A_\ell, u_\ell)) = V(P_j; (A_\ell, u_\ell))$. Moreover, by construction, R_{P_j} Blackwell dominates $R_{P_j}^\ell$, so $V(R_{P_j}; (A_\ell, u_\ell)) \geq V(R_{P_j}^\ell; (A_\ell, u_\ell))$.

<div align="right">□</div>

Proof of Proposition 3. Consider a binary decomposition of (A, u): $\{(A_1, u_1), \ldots, (A_k, u_k)\}$. From Theorem 2 and Lemma 14,

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

<div align="right">□</div>

B.4 Effective Constraints in Linear Programming

The following lemma, which we use in the proof of Theorem 3, states that a k-dimensional linear programming problem has at most k effective constraints.

Lemma 15. *Consider a feasible and bounded linear programming problem*

$$V = \max_{x \in \mathbb{R}^k} c \cdot x$$

$$s.t. \quad Ax \leq b$$

where $c \in \mathbb{R}^k$ and A is an $m \times k$ matrix with rank k, and b is a $m \times 1$ vector. There exists a full-rank $k \times k$ submatrix $\tilde{A}$ of A with the corresponding $k \times 1$ subvector $\tilde{b}$ such that

$$V = \max_{x \in \mathbb{R}^k} c \cdot x$$

49

$$s.t. \quad \tilde{A}x \leq \tilde{b}$$

Proof. The dual problem of the linear programming problem is

$$V = \min_{y \in \mathbb{R}^m} b \cdot y$$

$$s.t. \quad y^T A = c$$

$$y \geq 0$$

From Lemma 4.6 and Theorem 4.7 of Vohra (2004), a solution to this dual problem is a basic feasible solution, so there exists a full-rank $k \times k$ submatrix $\tilde{A}$ of A with the corresponding $k \times 1$ subvector $\tilde{b}$ such that

$$V = \min_{y \in \mathbb{R}^k} \tilde{b} \cdot y$$

$$s.t. \quad y^T \tilde{A} = c$$

$$y \geq 0$$

Taking the dual again, we have

$$V = \max_{x \in \mathbb{R}^k} c \cdot x$$

$$s.t. \quad \tilde{A}x \leq \tilde{b}.$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \square$

B.5 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 uniquely best marginal information source: $V(P_1; (A^{bi}, u^{bi})) > V(P_j; (A^{bi}, u^{bi}))$ for $j \neq 1$.

B.5.1 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 8. *A 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 $\mathbf{P}(\mathbf{y}) = P(\mathbf{y}|\cdot) \in \mathbb{R}^{|\Theta|}$ denote 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 16. *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

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

$\square$

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 17. *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

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

$\square$

B.5.2 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. 7.

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}).
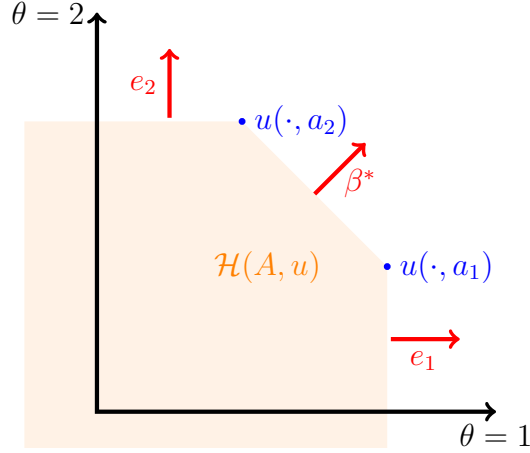

Figure 7: Payoff polyhedron for a binary-state binary-action problem

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

Definition 9. *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 18. *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. 8 for an illustration.
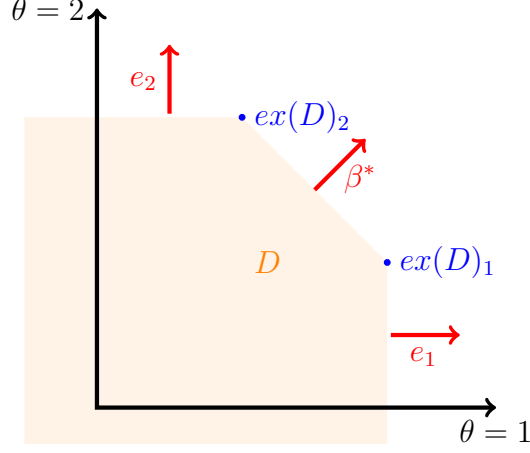
53

Figure 8: 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:

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

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,

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

55

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

$\square$

B.5.3 $\mathcal{B}$-cover

For any payoff set D, we define the smallest $\mathcal{B}$-shape polyhedron that covers D as its $\mathcal{B}$-cover. See Fig. 9 for an illustration.

Definition 10. *For any payoff set D, its $\boldsymbol{\mathcal{B}}$-**cover** is defined as*

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

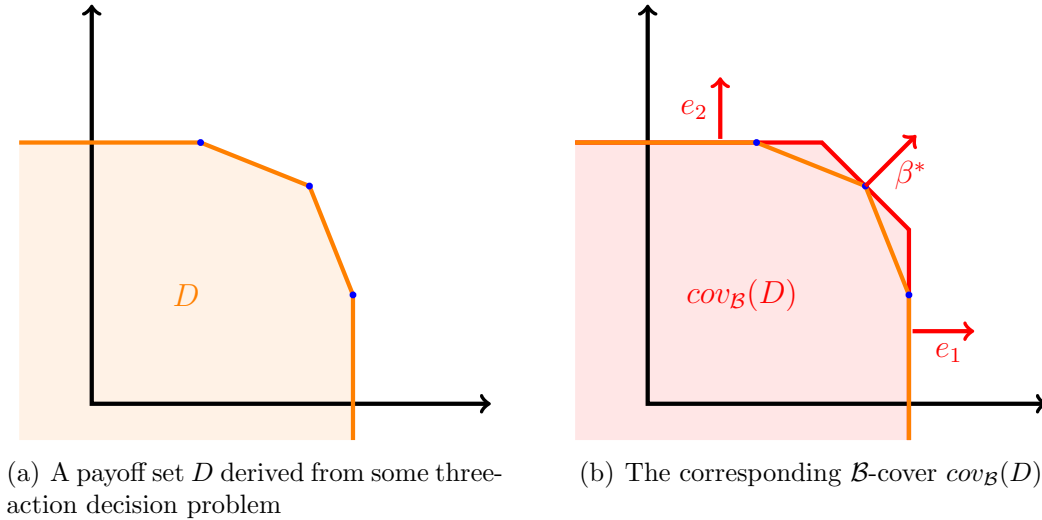*where $\rho_D(\beta) = \sup_{v \in D} \beta \cdot D$ is the support function of D.*

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

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

Figure 9

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

Lemma 19.   *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 $v' \in cov_\beta(D')$ such that $\tilde{v} = v + v'$. Since $v \in cov_\mathcal{B}(D)$ and $v' \in cov_\beta(D')$, 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.

   $\square$

B.5.4   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 20. *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

$$t : \mathbf{Y} \to D$$

$$y \mapsto \sum_{\ell=1}^k t_\ell(y).$$

Then

$$W(P_1, ..., P_m; D) \geq \min_{P \in \mathcal{J}} \sum_y \mathbf{P}(y) \cdot t(y)$$

$$= \min_{P \in \mathcal{J}} \sum_y \mathbf{P}(y) \cdot \sum_{\ell=1}^k t_\ell(y)$$

$$= \min_{P \in \mathcal{J}} \sum_{\ell=1}^k \sum_y \mathbf{P}(y) \cdot t_\ell(y)$$

$$\geq \sum_{\ell=1}^k \min_{P \in \mathcal{J}} \sum_y \mathbf{P}(y) \cdot t_\ell(y)$$

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

$\square$

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

Lemma 21. *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 19,

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

From property 3 in Lemma 19,

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

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 20, second inequality follows from Lemma 16, the third inequality follows from Lemma 18, 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 $j \neq 1$. $\qquad\square$

B.5.5 Common Support of the Blackwell Supremum

Lemma 22. *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, $\Lambda_{P^*} = co(\Lambda_{P_1} \cup \cdots \cup \Lambda_{P_m})$, which also should not contain such points, leading to a contradiction. $\qquad\square$

B.5.6 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:

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

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

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

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

From Lemma 21, $cov(D_2), ..., cov(D_m)$ must be trivial, and property 4 of Lemma 19 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

the Nature's MinMax Problem. From the minmax theorem, P^* is a solution to

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

Lemma 22 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

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

Since $\tilde{\phi}_j$ is constant for $j \ge 2$, the left-hand-side of the two equations above are the same, which implies $u(\cdot, \sigma^*(y_1, \bar{y}_{-1})) \le u(\cdot, \sigma^*(y_1, y_{-1}))$. 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, $u^{bi}(\cdot, \sigma^*(y_1, \bar{y}_{-1})) \le u^{bi}(\cdot, \sigma^*(y_1, y_{-1}))$ 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. $\qquad \square$