Robust Aggregation of Correlated Information*

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

June 20, 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.

*We are grateful to Nageeb Ali, Alex Bloedel, Alex Frankel, Marc Henry, Asen Kochov, Jiangtao Li, Elliot Lipnowski, George Mailath, Pietro Ortoleva, Collin Raymond, Shamim Sinnar, Rakesh Vohra, Leeat Yariv, and participants at various conferences and seminars for valuable comments.

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

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

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

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. However, collating and analyzing data from all sources can be taxing. To save time and effort, we may turn to a select few sources deemed reliable. In this paper, we show that limiting our sources of information has another, less obvious merit: *it leads to robust decisions when we lack knowledge about correlations between various information sources.*

Different information sources are often correlated: doctors may base their recommendation on the same study; financial analysts have a tendency to echo each other. Understanding the correlation structure between multiple sources is hard. In a scientific study, for example, determining the correlation between multiple variables requires an exponentially increasing sample size (the curse of dimensionality). Moreover, any misunderstanding of these correlations frequently leads to wrong inference and consequently inefficient decisions (see e.g. Enke and Zimmermann (2019)).

Our paper studies optimal decision making under ambiguity of correlations between information sources. Formally, a decision maker chooses among finitely many actions whose payoffs depend on a finite set of unknown states. Before deciding on an action, the decision maker observes the realizations of m signals from m different information sources, modeled as Blackwell experiments. To focus the analysis on ambiguity about correlations, in the baseline model, we assume that the decision maker knows every information source in isolation, but conceives of any possible joint information structures whose marginals are consistent with these information sources. To guard against this lack of knowledge, the decision maker chooses a strategy that performs well even under the worst possible correlation structure.

A simple strategy that protects against ambiguous correlation is a *best-source strategy*, which selects a *single* information source — the best one when considered individually — and best responds to it, while ignoring all other information sources. Since the resulting payoff from such a strategy is determined solely by the selected information source, this strategy guarantees a payoff that is independent of the correlation between information sources. Of course, this strategy completely forfeits the potential benefits from observing multiple information sources. Could the decision maker do better by using some more sophisticated strategy that makes use of multiple information sources? Surprisingly, Theorem 1 shows that, in any decision problem with two states and two actions, the answer is no: best-source strategies are always robustly optimal. Moreover, under additional mild assumptions, all robustly optimal strategies are indeed best-source strategies.[1]

[1] For only the converse direction of the theorem, we assume that the information sources each satisfy a full

With more than two actions, best-source strategies are no longer always optimal, and robustly optimal strategies will typically use multiple information sources. Theorem 2 constructs robustly optimal strategies under two states and multiple actions as follows. Without loss of generality, we can first eliminate all dominated actions since there always exists robustly optimal strategies that never use such actions. We then order the remaining n actions according to how much utility the actions generate in the first state. Given this ordering, a decision problem can be decomposed into two stages. In the first stage, we conduct $n-1$ "local" comparisons between each pair of consecutive actions. In line with Theorem 1, a best-source strategy is used in each local comparison to select a recommended action. Then in the second stage, the robustly optimal strategy maps each profile of recommendations in each of the local comparisons into a (possibly mixed) action in the original problem. Such a robustly optimal strategy uses an information source if and only if it is a best source in one of the local comparisons, and so it utilizes multiple information sources precisely when the best sources across local comparisons are not the same. In addition, in an n-action decision problem, the constructed robustly optimal strategy uses no more than $n-1$ information sources.

A full characterization of the robustly optimal strategy in decision problems with more than three states is more complex. However, as in Theorem 1 and Theorem 2, Theorem 3 establishes a bound, N, on the number of necessary information sources. In other words, there is a robustly optimal strategy that uses at most N information sources. Crucially, this bound again depends only on the decision problem, meaning that as the number of information sources grows large, the fraction of information sources used under robustly optimal strategies converges to 0.

Ignorance of readily available information is a well-established phenomenon, which can carry a significant cost. Handel and Schwartzstein (2018) describe the literature and divide the current explanations into two categories: frictions and mental gaps. Frictions are costs of acquiring or processing information. Mental gaps describe psychological distortions from rationality in information gathering or processing. This paper demonstrates robustness to correlations as an alternative explanation for this phenomenon. This explanation has distinct counterfactual implications from the other two, so it is important to determine which one is the most relevant before any intervention. For instance, a decision-maker 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 according to our model would not react to such an incentive.

Finally, our baseline model makes two assumptions. First, the decision maker has perfect knowledge of each of the information sources in isolation. tka Second, given the knowledge of

support assumption and that there is an information source that is in isolation, strictly optimal relative to all of the other information sources for the decision problem at hand.

the information sources in isolation, the decision maker has no knowledge about the correlation of the different information sources. We relax both of these assumptions in turn in Section 6. We first show that in binary state, binary action problems, Theorem 1 extends to a setting in which the decision maker additionally faces ambiguity about the marginal information sources. Secondly, our results extend in a straightforward fashion to some settings in which there is further knowledge about the correlations between the information sources.

Related Literature: Our paper studies robust decision making under uncertain correlations between information sources. The practice of finding robust strategies traces back at least to Wald (1950) and our modeling of information structures follows that of Blackwell (1953). 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 ambiguity aversion on correlation structures.

Learning from multiple information sources has gained considerable attention in recent literature. For instance, Börgers, Hernando-Veciana, and Krähmer (2013) study when two information structures are complements or substitutes and provide an explicit characterization. Ichihashi (2021) looks at how a firm purchases data from consumers with potentially correlated information source. Liang and Mu (2020) examine a social learning setting where agents' information is complementary. Liang, Mu, and Syrgkanis (2022) study an agent's optimal dynamic allocation of attention to multiple correlated information sources. In contrast to this work, our paper assumes the decision maker does not know the correlation structure and targets for a decision plan robust to all possible correlations.

There is a classic literature on "combining forecasts" going all the way back to the 1960's (for an early survey, see Clemen (1989)). Its theoretical portion (e.g. McConway (1981), Dawid, DeGroot, Mortera, Cooke, French, Genest, Schervish, Lindley, McConway, and Winkler (1995), and Levy and Razin (2020b)) assumes that one has access to experts' beliefs, but not to the raw information informing those beliefs. One must then try to combine those beliefs into a single one. In our framework, this amounts to an *interim approach*, where the aggregation of beliefs takes place after signals have realized. Using this interim approach, Levy and Razin (2020a) also consider ambiguity about the correlation structure. For each realized profile of forecasts, they consider all joint experiments satisfying a correlation bound and search for the worst case that can rationalize such a profile. In contrast, our paper uses an *ex-ante approach*, by considering an ex-ante strategy plan for all possible signal realizations.

A closely related paper is Arieli, Babichenko, and Smorodinsky (2018). They also consider the ex-ante strategy and allow for ambiguity about the information structure. However, they consider other sets of joint experiments, such as two experiments where one is Blackwell more informative than the other, but the agent does not know which. Moreover, they look at a specific decision problem with quadratic loss. Also closely related is Arieli, Babichenko, Talgam-Cohen, and Zabarnyi (2023), who study a minmax regret version of the problem. Complementing our Theorem 1, they show that when the marginal experiments are symmetric, the optimal aggregation rule follows a single random information source.

Robustness to correlations has also been studied in other contexts, such as mechanism design problems. 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 between different agents.

2 Model

An agent faces a decision problem $\Gamma \equiv (\Theta, \nu, A, \rho)$ with a finite state space Θ, a prior $\nu \in \Delta\Theta$, a finite action space A, and a utility function $\rho : \Theta \times A \to \mathbb{R}$. To later simplify notation, define $u(\theta, a) = \nu(\theta)\rho(\theta, a)$, which represents 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. 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**:

$$\mathcal{P}(P_1, ..., P_m) = \left\{ P : \Theta \to \Delta(Y_1 \times ... \times Y_m) : \sum_{y_{-j}} P(y_1, \dots, y_m|\theta) = P_j(y_j|\theta) \text{ for all } \theta, j, y_j \right\}.$$

To simplify notations, let $\mathbf{Y} = Y_1 \times \cdots \times Y_m$ denote the set of all possible profiles of signal realizations.

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:

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

We call a solution to the problem a **robustly optimal** strategy.

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

Note that for any decision problem, one simple strategy is to choose exactly one experiment $Q \in \{P_1, \ldots, P_m\}$ and play the optimal strategy that uses that information alone, ignoring the signal realizations of all other experiments. This strategy guarantees an expected value of $V(Q)$ regardless of the joint experiment $P \in \mathcal{P}(P_1, \ldots, P_m)$. By choosing Q optimally, the agent achieves an expected payoff of $\max_{j=1,\ldots,n} V(P_j)$. We call such a strategy a **best-source strategy**.

In some special cases, it is easy to see that a best-source strategy is robustly optimal. For example, if the marginal experiments are identical, then the worst case information structure would perfectly correlate the signals to make the signals of all but one information source redundant. Similarly, if P_1 Blackwell dominates $P_2, \ldots, P_m$, then nature can correlate the signals according to the corresponding Blackwell garblings to ensure that $P_2, \ldots, P_m$ contain no additional information beyond P_1. The interesting case is when the marginal experiments are not Blackwell ranked. In this case, any correlation structure $P \in \mathcal{P}(P_1, ..., P_m)$ would be strictly more informative than any individual marginal experiment.[2] As will be shown, in simple decision problems — those with binary states and binary actions — the agent can never do better than a best-source strategy. In more complicated problems, however, the decision maker may need to use more sophisticated strategies to robustly aggregate information from multiple sources.

3 Binary State Environment

For this section, we consider the special case in which $|\Theta| = 2$. We characterize both the robustly optimal strategies and values in this environment.

3.1 Binary-State Binary-Action Problems

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

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

In addition, if the marginal experiments have 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 unique, then all robustly optimal strategies are best-source strategies.*

[2] In fact, any correlation structure has to dominates the "Blackwell supremum" of $\{P_1, ..., P_m\}$, which will be discussed in more details in Section 4.2.

Theorem 1 presents a simple solution to any binary-state, binary-action decision problem: identify the best marginal information source and best respond to it accordingly. Moreover, except in the case where there are multiple best information sources, as long as the information sources satisfy full support, a strategy that uses more than one source is strictly suboptimal. In other words, against such a strategy, nature can choose some correlation structure that will yield a strictly lower expected utility than that guaranteed by the best-source strategy.

We present the proof of Theorem 1 in detail in Section 4. Clearly, $\max_{j=1,\dots,m} V(P_i;(A;u))$ is a lower bound on the robustly optimal value. In order to show the reverse inequality, we construct a joint information structure, $\overline{P}(P_1,\dots,P_m)$, in which an optimal strategy of the agent is to best respond to the signal of the best marginal information source alone. In the proof, we show additionally that $\overline{P}(P_1,\dots,P_m)$ can be chosen *uniformly* across all binary state, binary action decision problem. This is a feature which plays an important role in the analysis of general decision problems in the binary state environment.

While only using **one** information source is sufficient in binary action, binary state decision problems, the following example demonstrates that an agent may benefit from using multiple sources of information in more complex decision problems.

Example 1. *An investor can invest in two assets whose outputs depend on an unknown binary state $\theta \in \{1,2\}$. Outputs from each asset are given by:*

Asset 1

	Invest	*Not Invest*
$\theta = 1$	2	0
$\theta = 2$	-1	0

Asset 2

	Invest	*Not Invest*
$\theta = 1$	-1	0
$\theta = 2$	2	0

The investor's payoff is the sum of outputs from both assets. This can be written as a decision problem with $A = \{I, NI\} \times \{I, NI\}$ and $u(\theta,a) = u_1(\theta,a_1) + u_2(\theta,a_2)$ where $a_1, a_2 \in \{I, NI\}$ and u_1, u_2 are the outputs function given in the table above.[3]

Suppose the investor has access to two experiments P_1, P_2:

P_1

	$y_1 = 1$	$y_1 = 0$
$\theta = 1$	0.9	0.1
$\theta = 2$	0.5	0.5

P_2

	$y_2 = 1$	$y_2 = 0$
$\theta = 1$	0.5	0.5
$\theta = 2$	0.9	0.1

By paying attention to one experiment, for example P_1, the optimal strategy is to invest in both assets if $y_1 = 1$ and only asset 2 if $y_1 = 0$. The expected payoff from this strategy is thus

[3] Recall that $u(\theta,a) = \nu(\theta)\rho(\theta,a)$, so the payoffs here have been weighted by the prior.

$0.9 \cdot 1 + 0.1 \cdot (-1) + 0.5 \cdot 1 + 0.5 \cdot 2 = 2.3$.[4]

Now suppose the investor makes the investment decision of asset 1 based on experiment P_1, and asset 2 based on experiment P_2. Then for asset $i = 1, 2$, the optimal strategy is to invest iff $y_i = 1$. "Adding up" these two strategies yield:

	$y_2 = 1$	$y_2 = 0$
$y_1 = 1$	*Invest in both*	*Invest in asset 1*
$y_1 = 0$	*Invest in asset 2*	*No investment*

This strategy guarantees an expected output of $0.9 \cdot 2 + 0.1 \cdot 0 + 0.5 \cdot (-1) + 0.5 \cdot 0 = 1.3$ *from each asset regardless of the correlations, which gives a total output of* $2.6 > 2.3$. *So the agent strictly benefits from utilizing information from both information sources. In fact, as we will show in the next section, this strategy is a robustly optimal strategy.*

It is clear why paying attention to only one experiment is clearly suboptimal in the above decision problem: the most informative experiment (P_i) for the investment decision pertaining to asset $i \in \{1, 2\}$ are distinct. Thus, the conclusion from Theorem 1 of using only a **single** information source is very specific to binary action-binary state decision problems.

Nevertheless, we do see that Theorem 1 does indeed serve as the foundation for the robustly optimal strategy: decide whether or not to invest in asset i on the basis of P_i alone. We now generalize this idea.

3.2 Separable Problems

Motivated by the previous example, we consider a class of decision problems featuring two special properties: (1) the action space is a product of binary action spaces and (2) the payoff function can be expressed in an additively separable form of binary-action problems.

Definition 1. *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*

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

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

[4] Symmetrically, by paying attention to only P_2, the optimal strategy is to invest in both assets if $y_2 = 1$ and only asset 2 if $y_2 = 0$. The expected payoff is also 2.3.

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 1. *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* $\sigma : \mathbf{Y} \to \Delta(A_1 \times ... \times A_k)$ *defined by*

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

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

Proof. See Section A.1. □

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

$$V\left(P_1, \ldots, P_m; \bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)\right) \geq \sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j; (A_\ell, u_\ell)). \tag{2}$$

The equality in Proposition 1 follows as a result of the special property highlighted in the discussion after Theorem 1—that in binary state environments, there exists a single $\mathcal{P}(P_1, \ldots, P_m)$ that uniformly minimizes the agent's value across all binary action problems.[5]

3.3 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 demonstrate 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 use these subproblems to construct the corresponding separable problem that is equivalent to the original problem. We call the resulting separable problem the *binary decomposition.*

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

We first define formally what it means for two decision problems to be equivalent. Given a decision problem (A, u), let[6]

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

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 1.

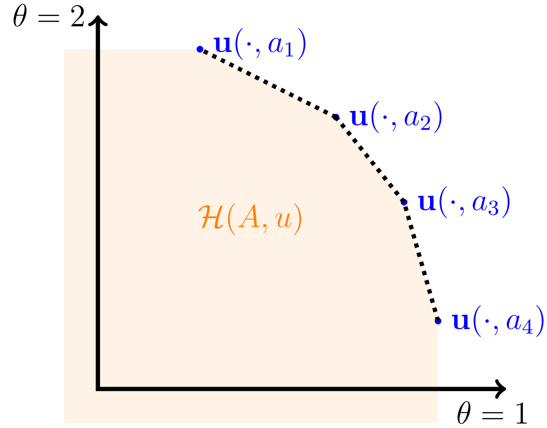


Figure 1: The shaded area represents $\mathcal{H}(A, u)$

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

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

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

Definition 2. *Two decision problems (A, u) and (A', u') are equivalent if $\mathcal{H}(A, u) = \mathcal{H}(A', u')$.*

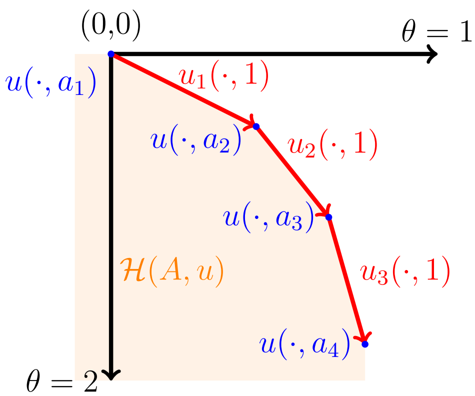
Next we show by direct construction that, every binary-state decision problem is equivalent to a separable problem. We start with some normalization to simplify exposition. First we remove all weakly*-dominated actions,[7] so that actions can be ordered such that

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

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

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

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



(a) Binary decomposition

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

Figure 2

Definition 3. *Given a decision problem (A, u), the* ***binary decomposition*** *of (A, u) is a separable problem $\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)$ where*

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

The key idea underlying the binary decomposition is to decompose the original problem into binary-action decision problems that compare each pair of consecutive actions. This can be visualized in Figure 2(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 action 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 $\mathcal{H}(A, u) \subset \mathcal{H}\big(\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)\big)$. By contrast, the binary decomposition $\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)$ could introduce additional payoff vectors. To illustrate, take the example in Figure 2(b). Here, by taking $\boldsymbol{\delta} = (1, 0, 1)$, the separable problem induces an additional payoff vector that does not belong to the original problem. However, this additional action lies in the interior of $\mathcal{H}(A, u)$, and thus is dominated by one of the original (possibly mixed) actions. This observation is not a coincidence. As shown in the next lemma, any additional payoff vectors induced in the binary decomposition will always lie within $\mathcal{H}(A, u)$, so $\mathcal{H}(A, u) = \mathcal{H}\big(\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)\big)$.

Lemma 2. *The binary decomposition of (A, u) is equivalent to (A, u).*

Proof. See Section A.4. □

Lemma 2 and Lemma 1 permit us to derive a robustly optimal strategy for any decision problem (A, u) through its binary decomposition.

Theorem 2. *Let $(A_1, u_1), \ldots, (A_{n-1}, u_{n-1})$ 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 Section A.2. □

Theorem 2 allows us to construct a robustly optimal strategy for any decision problem (A, u) in two steps: 1) For each subproblem, (A_ℓ, u_ℓ), only one (the best) marginal experiment needs to be considered, and a robustly optimal strategy σ_ℓ^* can be chosen to be measurable with respect to this experiment alone; 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, the marginal experiments, $Y_1, \ldots, Y_m$, influence the robustly optimal strategy only through its effect on the choice of $\sigma_\ell^*(\mathbf{y})$ in each of the subproblems.

The theorem delivers two immediate corollaries.

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

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

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

Corollary 1 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.

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

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

Corollary 2 implies that in an n-action decision problem, an agent needs to use at most $n-1$ sources of information.

4 Proof of Theorem 1

In this section, we return to the proof of Theorem 1 and show that a best-source strategy is robustly optimal. Establishing the uniqueness of this robustly optimal strategy requires a different technique, and we defer the proof to Appendix A.8.

We first begin with some preliminary remarks regarding the Blackwell order when $|\Theta| = 2$.

4.1 The Blackwell Order

It will be useful to rank experiments according to how much information they convey. We will review the Blackwell order in this subsection for the sake of completeness. Readers familiar with the Blackwell order may choose to skip this subsection.

Definition 4. *$P : \Theta \to \Delta(Y)$ 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.*

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

Definition 5. *$Q : \Theta \to \Delta(Z)$ is a garbling of $P : \Theta \to \Delta(Y)$ if there exists a function $g : Y \to \Delta(Z)$ (the "garbling") such that $Q(z|\theta) = \sum_y g(z|y)P(y|\theta)$.*

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

Definition 6. *Given a set of actions A and an experiment $P : \Theta \to \Delta(Y)$, the feasible set of P is*

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

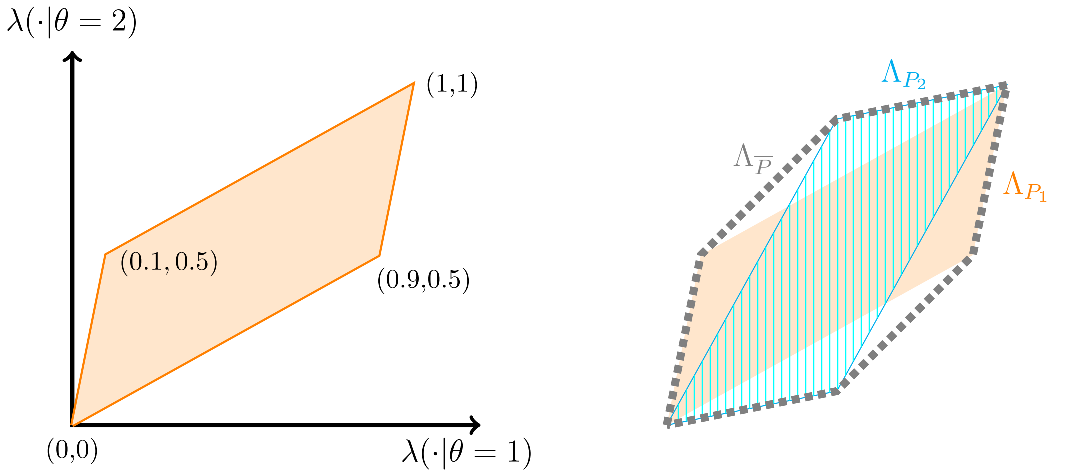
The feasible set of an experiment specifies what conditional action distributions can be obtained by some choice of strategy σ. One might then say that more information allows for a larger set.

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;*



(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 3

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 a set A with $|A| = 2$, $\Lambda_Q(A) \subseteq \Lambda_P(A)$.

This last equivalent condition gives us a simple graphical representation of Blackwell experiments when $|\Theta| = 2$. See Figure 3(a) for an illustration. Since $|A| = 2$, to characterize $\Lambda_P(A)$, 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 y-axis the probability in state 2. Clearly $(0, 0), (1, 1) \in \Lambda_P(A)$ for all P, because these two points represent taking the same actions regardless of the signal realizations. With the information obtained from the Blackwell experiment, additional points can be obtained. For example, the point $(0.1, 0.5)$ in Figure 3(a) can be achieved if the decision maker has access to a signal that realizes with probabiliy 0.1 in state 1 and probability 0.5 in state 0, and takes action $a = 1$ when observing such a signal realization. Symmetrically, she can also take action $a = 0$ when observing the same signal realization, which yields the point $(0.5, 0.9)$. Moreover, randomization convexifies the set and thus $\Lambda_P(A)$ is a convex and rotational symmetric polytope in $[0, 1]^2$. Conversely, any convex and rotational symmetric polytope in $[0, 1]^2$ correspond to $\Lambda_P(A)$ for some P.

4.2 The Blackwell Supremum

Our analysis will use some lattice properties of the Blackwell order. In particular, the concept of a Blackwell supremum will be useful.

Definition 7. *Let P_1 and P_2 be two arbitrary experiments. We say that $\overline{P}$ is the Blackwell supremum of P_1 and P_2 if*

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

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

The definition extends to any number of experiments. By definition, if there are two Blackwell suprema, they must Blackwell dominate each other. This means that by looking at the equivalence class of experiments with the same level of information, we can say that the Blackwell supremum is unique.

Under binary state, the Blackwell supremum always exists and can be characterized using the feasible set, as illustrated in Figure 3(b). If $\overline{P}$ is the Blackwell supremum of P_1 and P_2, we know from Blackwell's Theorem that $\Lambda_{\overline{P}}$ must contain both Λ_{P_1} and Λ_{P_2}.[8] Moreover, any P' that is more informative than P_1 and P_2 must be more informative than $\overline{P}$ as well, so $\Lambda_{P'}$ must also contain $\Lambda_{\overline{P}}$. Hence the feasible set of the Blackwell supremum should be the smallest feasible set containing $\Lambda_{P_1} \cup \Lambda_{P_2}$. The feasible set is always convex, so the $\overline{P}$ that corresponds to $\Lambda_{\overline{P}} = co(\Lambda_{P_1} \cup \Lambda_{P_2})$ is the Blackwell supremum. This observation yields the following lemma:[9]

Lemma 3. *When $|\Theta| = 2$, the Blackwell supremum always exists. An experiment $\overline{P}$ is the Blackwell supremum of P_1 and P_2 if and only if $\Lambda_{\overline{P}} = co(\Lambda_{P_1} \cup \Lambda_{P_2})$.*

When $|\Theta| \geq 3$, a Blackwell supremum may not exist, as illustrated in example 18 of Bertschinger and Rauh (2014). The proof of existence fails becasue in a higher dimensional space, the convex hull of $\Lambda_{P_1} \cup \Lambda_{P_2}$ might not correspond to any Blackwell experiment.

4.3 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 problem, of choosing the worst possible correlation structure.

First note that since the objective function is linear in both σ and P, and the choice sets of σ and P are both convex and compact, the minimax theorem implies that

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

[8] For ease of notation, we omit the dependence of $\Lambda_P(A)$ on the set A when $|A| = 2$.

[9] For a formal proof, see e.g., Kertz and Rösler (1992) or Bertschinger and Rauh (2014).

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 $\mathcal{P}(P_1, \ldots, P_m)$ to minimize a Bayesian agent's value in the decision problem.

Observe that every experiment in $\mathcal{P}(P_1, \ldots, P_m)$ must be more informative than every P_j, since the projection into the jth coordinate defines a garbling. So if we let $\mathcal{D}(P_1, \ldots, P_m)$ denote the set of Blackwell experiments that dominates $P_1, ..., P_j$, then $\mathcal{P}(P_1, \ldots, P_m) \subseteq \mathcal{D}(P_1, \ldots, P_m)$. The set $\mathcal{D}(P_1, \ldots, P_m)$ is in general a larger set, because not every experiment that dominate $P_1, ..., P_m$ can be represented as a joint experiments with marginals $P_1, ..., P_m$.[10] However, the next lemma shows that relaxing 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 4.

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

The idea underlying Lemma 4 is that in the relaxed problem, Nature would only choose the experiments that are Blackwell minimal—those that do not dominate any other experiment in $\mathcal{D}(P_1, \ldots, P_m)$. In additional, any Blackwell minimal element in the set can be represented as a joint experiment, as shown in Appendix A.3.

Lemma 4 is particularly useful when the state is binary. Under binary states, the Blackwell supremum $\overline{P}$ of $P_1, ..., P_m$ exists, and it is the minimum element in $\mathcal{D}(P_1, \ldots, P_m)$. Therefore, $\overline{P}$ solves Nature's problem regardless of the decision problem, which yields the following corollary.

Corollary 3. *When* $|\Theta| = 2$,

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

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

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

We can now prove Theorem 1.

Proof of Theorem 1. By Corollary 3, it suffices to show that $V(\overline{P}(P_1, ..., P_m)) = \max_{j=1,\ldots,m} V(P_j)$. By Lemma 3, an experiment $\overline{P}$ is the Blackwell supremum of $P_1, \ldots, P_m$ if and only if

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

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

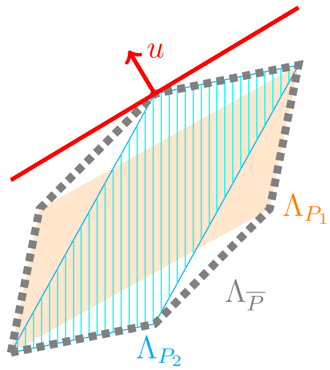


Figure 4: The maximum is achieved at an extreme point that belongs to Λ_{P_2}

Now, the maximum utility achievable given Blackwell experiment $\overline{P}(P_1, \ldots, P_m)$ 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 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 $\bar{P}$ is more informative than every P_j, we also have $V(\bar{P}) \geq \max_{j'=1,\ldots,m} V(P_{j'})$, concluding the proof.

□

The idea of Theorem 1 can be visualized in Figure 4 for two marginal experiments. Each marginal Blackwell experiment P_1, P_2 can be represented by $\Lambda_{P_1}, \Lambda_{P_2}$, the set of feasible state-action distribution 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.

5 General-State Decision Problems

Our previous analyses focus 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 in a few ways. First, with more states, it is unclear how to decompose a general decision problem into the more "elementary" ones. Second, the nonexistence of the Blackwell supremum implies that in the Nature's minmax problem, there may no longer be a single experiment that uniformly minimize the agent's value across all decision problems, which significantly exacerbates the complexity of the analysis (see Footnote 5). Lastly, an agent may want to use multiple information sources even in a binary-action decision problem, 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, y_1, y_2, and given by Table 1.*

P_X

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

P_Y

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

Table 1

Intuitively, experiment P_X tells the agent whether the state is θ_3 or not and experiment P_Y tells the agent 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 obtains the perfect information payoff.

Let $A = \{1, 0\}$ and suppose that the utilities are as follows:[11]

$$
\begin{aligned}
&u(\theta, a=1) = \mathbf{1}\left(\theta \in \{\theta_1, \theta_3\}\right) - 0.9 * \mathbf{1}\left(\theta = \theta_2\right), \\
&u(\theta, a=0) = 0.
\end{aligned}
$$

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

In this section, we develop a different technique, using the piecewise linearity of the interim value function to simplify the set of Blackwell experiments Nature would use. This allows us to provide a general bound on the number of experiments an agent needs to use.

Recall that a decision problem is a tuple $\Gamma \equiv (\Theta, \nu, A, \rho)$ with a finite state space Θ, a prior $\nu \in \Delta\Theta$, a finite action space A, and a utility function $\rho : \Theta \times A \to \mathbb{R}$. For a given decision problem Γ, the corresponding *interim value function*, $v^\Gamma : \Delta(\Theta) \to \mathbb{R}$, is defined as

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

[11]Recall that $u(\theta, a) = \nu(\theta)\rho(\theta, a)$, so the payoffs here have been weighted by the prior.

Given a value function $v : \Delta(\Theta) \to \mathbb{R}$, its epigraph is defined as 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 ext(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. The *kinks* of v form the set of extreme points of its epigraph, excluding those point-mass beliefs $(\delta_i, v(\delta_i))$. Thus, *the number of kinks of* v is $|\text{ext(epi }(v))| - |\Theta|$. 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. I'm not sure if "kink" is a word that is used formally. Maybe find a different way to say? Also, is there an intuitive way of describing the kinks?

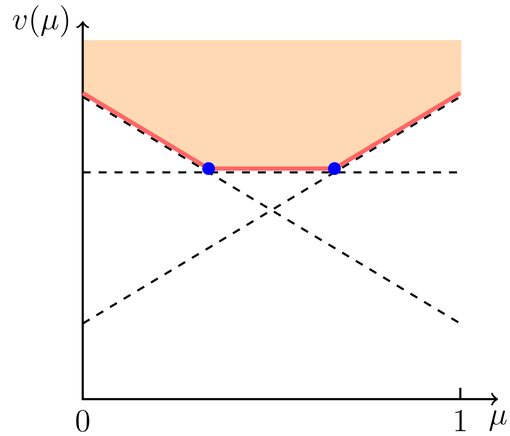


Figure 5: Interim value function and kinks

The following theorem provides a bound on the number of experiments that a decision maker would need, which is the number of kinks of the corresponding interim value function. The important feature of this upper bound is that it does not depend on the set of experiments in any way; it only depends on the decision problem.

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 $\{P_j\}_{j \in J \subset \{1,...,m\}}$ with $|J| \leq k$, such that*

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

Remark. Notice that the bound, k, in Theorem 3 depends only on the decision problem, and is independent of both the set of information sources as well as the number of information sources, m. 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.

The full proof of Theorem 3 is deferred to Appendix A.5, but here we sketch the main steps. By the minmax theorem, it suffices to examine Nature's minmax problem:

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

By Lemma 4, 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, this is equivalent to choosing a posterior distribution that is a mean-preserving spread to 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.

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 number of 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.

Theorem 3 suggests one may ignore information sources due to robustness concerns. The following proposition further tells us which information sources will always be ignored: if an information source P_m is never the best information source among $\{P_j\}_{j=1}^m$, then it can always be ignored in a robustly optimal strategy.

Proposition 1. *If for any decision problem* (A, u)*,* $V(P_m; (A, u)) \leq \max_{j=1,\dots,m-1} V(P_j; (A, u))$*, 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 A.6. □

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 $P_1, ..., P_{m-1}$. Such characterization will be useful in our proof.[12]

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. In other words, there may be information sources that perform reasonably well across all decision problems, but which the agent chooses to ignore because for each decision problem, there is at least one other experiment that performs better.

6 Discussions

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 decision maker 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 decision maker 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, inevitably leading to correlations between their recommendations. If we know that a common information source is the *only* possible channel generating the correlation between information sources, does this additional knowledge help the decision maker to restrict the presumed set of correlations? In other words, what types of correlation structures can be rationalized by sharing a common source.

Formally, we say a joint experiment $P \in \mathcal{P}(P_1, ..., P_m)$ 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.

[12] 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 between $P_1, ..., P_{m-1}$.

An immediate observation is that every $P \in \mathcal{P}(P_1, ..., P_m)$ 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 correlations.

6.1.2 Partial Knowledge of Correlations

In certain situations, a decision maker may understand the correlation between some information sources, even if they do not comprehend all of them. For example, diagnostic imaging such as X-rays and MRI are frequently used together and exhibit well-established correlations. On the other hand, genomic sequencing technologies, which have been more recently adopted, 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 $\mathcal{P}(P_1, ..., P_m)$. For example, suppose the decision maker knows that the correlation between P_1 and P_2 is given by $P_{12} : \Theta \to \Delta(Y_1 \times Y_2)$, with marginals consistent with P_1 and P_2. The conceived joint experiment with this additional knowledge is thus

$$\Big\{ P : \Theta \to \Delta(Y_1 \times ... \times Y_m) : \sum_{y_{-j}} P(y_1, \ldots, y_m|\theta) = P_j(y_j|\theta) \text{ for all } \theta, j, y_j, \\ \sum_{y_3, \ldots, y_m} P(y_1, y_2, \ldots, y_m|\theta) = P_{12}(y_1, y_2|\theta) \text{ for all } \theta, y_1, y_2 \Big\}.$$

It is easy to see that imposing this additional constraint is equivalent to treating $\{P_1, P_2\}$ as a single information source P_{12}, and that all our results apply following such adaptation.

The same argument applies to cases where the decision maker knows the correlations within a few different subsets of information sources, provided that these subsets are disjoint. However, if the subsets overlap, the problem becomes significantly more complicated. For example, suppose there are three information sources, given by $\{P_1, P_2, P_3\}$. If the decision maker 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 is

$$\Big\{ P : \Theta \to \Delta(Y_1 \times Y_2 \times Y_3) : \sum_{y_3} P(y_1, y_2, y_3|\theta) = P_{12}(y_1, y_2|\theta) \text{ for all } \theta, y_1, y_2 \\ \sum_{y_1} P(y_1, y_2, y_3|\theta) = P_{23}(y_2, y_3) \text{ for all } \theta, y_2, y_3 \Big\}.$$

The constraints on $\mathcal{P}(P_1, ..., P_m)$ can no longer be treated by replacing a subset of experiments with a single experiment, and our existing results no longer apply. A detail investigation of this problem is beyond the scope of this paper, and we view it as an interesting direction for future research.

6.2 Ambiguity about Marginals

Our model so far assumes that the decision maker 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:

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

The agent's decision problem is similarly defined:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

□

6.3 Aggregating Experts' opinions

In certain instances, a decision maker 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 additional, we call the induced *recommendation information structure*, denoted by $R_{P_j;(A,u)} : \Theta \to \Delta A$, as the information structure derived by a garbling σ_j^*:[13]

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

When the decision maker has access to only a single source of information, it is clear that compressing the information through reporting beliefs or action recommendations does not hurt the decision maker; that is, $V(P_j; (A, u)) = V(B_{P_j}; (A, u)) = V(R_{P_j;(A,u)}; (A, u))$ for any j. This is because beliefs also action recommendations already contain all the information needed for the decision making.

However, with multiple available information sources, compressing information could potentially hurt because some information, which may not be useful on its own, could become valuable when combined with other sources. This begs the question of whether the decision maker could still achieve the same value as if she had access to the raw information sources. In other words, does

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

hold when $m > 1$?

First, we observe that $V(P_1, ..., P_m; (A, u)) = V(B_{P_1}, ..., B_{P_m}; (A, u))$: note that P_j is Blackwell equivalent to B_{P_j} for all j, and so Lemma 4 implies the values $V(P_1, ..., P_m; (A, u))$ and $V(B_{P_1}, ..., B_{P_m}; (A, u))$ must be equal. The relationship between $V(R_{P_1;(A,u)}, ..., R_{P_m;(A,u)}; (A, u))$ and $V(P_1, ..., P_m; (A, u))$ is more interesting: we will show that when $|\Theta| = 2$, these two values coincide, but in general, we could have $V(R_{P_1;(A,u)}, ..., R_{P_m;(A,u)}); (A, u)) < V(P_1, ..., P_m; (A, u))$.

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

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

Proof. See Appendix A.7. □

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 degenerated 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

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

$V(P_X, P_Y) = 1 + 0 + 1 = 2 > V(R_{P_X}, R_{P_Y})$.

7 Conclusion

Our results show that ambiguity about the correlation between information sources can lead a decision maker to ignore seemingly relevant information. To highlight the role of this concern, we assumed that the agent knew perfectly each marginal experiment, while knowing nothing about the correlation between them. This tractable structure allowed us to characterize robustly optimal strategies under two states and prove a general bound on the number of sources used for more than two states. However, for some applications, it may be desirable to relax those assumptions.

In Section 6.2, we showed that our approach also applies to the case where the agent considers a set of possible distributions for each marginal. In that case, the set of joint experiments is still tractable and the conclusion of Theorem 1 still holds. Intuitively, the added ambiguity can only increase the incentive to be conservative and ignore information.

It would be interesting to analyze intermediate assumptions about the correlation structure. Our approach still applies to some simple cases. For example, if the decision maker knows the joint distribution of P_1 and P_2 (they could be independent), but does not know anything else about the correlation of those experiments and P_3, P_4, etc, we can treat the vector (P_1, P_2) as a single experiment and our results will apply. More generally, the same logic applies if the experiments can be partitioned into a collection of disjoint sets, with the experiments within each set being mutually independent, but allowing any dependence between experiments of different sets. Beyond these simple cases, our analysis does not immediately apply as it relies on this particular product structure of the sets. Understanding the implications of richer specifications for the set of allowed correlations is an interesting avenue for future research.

In some applications, it might be interesting to directly use the robustly optimal strategy. For example, in AI it is common to use different methods to arrive at a solution to a problem and then somehow aggregate the multitude of answers (give citations). In such cases, a full characterization of the robustly optimal strategies could be quite useful. Theorem 2 shows such complete characterization, but we do not have a full characterization for the case with a larger state space. In general, the logic of Theorem 2 can still be used, in the sense that we can find an equivalent decomposition of the original problem into subproblems, where in each subproblem it is optimal to use only one information source. The difficulty lies in finding the correct decomposition—there is no canonical decomposition, and the optimal decomposition can depend on the information sources. A full characterization of robustly optimal strategies for the general case remains an open question.

Finally, whilex our paper offers an alternative rationale for the ignorance of information, our main results complement the literature on costly information acquisition. Our main theorems show that there always exist some correlations of marginal information sources such that a standard Bayesian agent who knew the information structures to be correlated in this way would find it (weakly) optimal to ignore many of the available information sources. Moreover, under perfect knowledge of this particular correlation, the agent would find it strictly optimal to ignore many of the available information sources as long as the cost of observing/processing each information source is strictly positive.

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. Universally robust information aggregation for binary decisions. *arXiv preprint arXiv:2302.03667*, 2023.

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.

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.

Robert T Clemen. Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 5(4):559–583, 1989.

Alexander P Dawid, Morris H DeGroot, Julia Mortera, Roger Cooke, S French, C Genest, MJ Schervish, DV Lindley, KJ McConway, and RL Winkler. Coherent combination of experts' opinions. *Test*, 4:263–313, 1995.

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.

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.

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

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

Gilat Levy and Ronny Razin. The drowning out of moderate voices: a maximum likelihood approach to combining forecasts. *Theoretical Economics*, 2020b.

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.

Kevin J McConway. Marginalization and linear opinion pools. *Journal of the American Statistical Association*, 76(374):410–414, 1981.

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

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.

A Appendix

A.1 Proof of Lemma 1

Proof. By definition of σ,

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

Moreover, by Theorem 1 and Corollary 3,

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

Together, these inequalities prove our claim that

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

and that σ is a robustly optimal strategy.

□

A.2 Proof of Theorem 2

Proof. From Lemma 2, (A, u) is equivalent to $\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)$, so

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

where the second equality follows from Lemma 1. This establishes the first statement of the theorem.

For each $\mathbf{y}$, $\sum_{\ell=1}^{n-1} u_\ell(\cdot, \sigma_\ell(\mathbf{y})) \in \mathcal{H}\Big(\bigoplus_{\ell=1}^{n-1}(A_\ell, u_\ell)\Big) = \mathcal{H}(A, u)$. So there exists $\sigma^*(\mathbf{y})$ such that $u(\cdot, \sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{n-1} u_\ell(\cdot, \sigma_\ell(\mathbf{y}))$. Now for any $P \in \mathcal{P}(P_1, ..., P_m)$,

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

where the second line follows from Lemma 1 and the third line follows from Lemma 2. So σ^* is a robustly optimal strategy. □

A.3 Proof of Lemma 4

Proof. The first equality follows from the minmax theorem. To prove the second equality, it suffices to show that for any $Q \in \mathcal{D}(P_1, ..., P_m)$, there exists $P \in \mathcal{P}(P_1, ..., P_m)$ such that $\tilde{Q}$ 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 $P : \Theta \to \Delta(Y_1 \times ... \times Y_m)$:

$$P(y_1, ..., y_m|\theta) = \sum_x \prod_{j=1}^{m} \gamma_j(y_j|x)Q(x|\theta). \tag{4}$$

Clearly, $P \in \mathcal{P}(P_1, ..., P_m)$ 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 $\tilde{Q}$ is Blackwell Dominated by Q. □

A.4 Proof of Lemma 2

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 a convex and closed, by Corollary 11.4.2 of Rockafellar (1970), there exists $\lambda \in \mathbb{R}^2 \setminus \{0\}$ such that

$$\lambda \cdot u^* > \sup_{v \in \mathcal{H}(A,u)} \lambda \cdot v. \tag{5}$$

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 strictly decreasing in ℓ. Therefore, for any $\ell' > \ell$,

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

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 $\ell \geq \ell^*$.

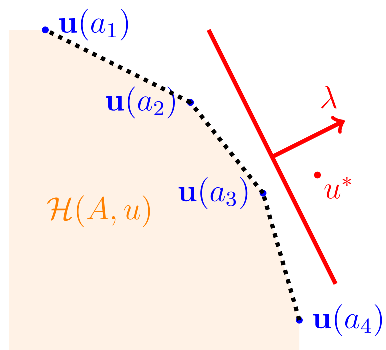


Figure 6

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 (5).

□

A.5 Proof of Theorem 3

We shall start with some preliminary definitions and lemmas.

A.5.1 Definitions

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

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

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

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 $\hat{\mathcal{D}}(P_1, ..., P_m) = \mathcal{D}(P_1, ..., P_m) \cap \{P : supp(\tau_P) \in E\}$ denote the subset of $\mathcal{D}$ such that the induced posterior distribution is supported in E.

Recall that we assumed every action $a_i \in A$ is a best response to some belief. Let $\Xi_i = \{\mu \in \Delta(\Theta) | \sum_\theta \mu(\theta)\rho(\theta, a_i) \geq \sum_\theta \mu(\theta)\rho(\theta, a') \text{ for all } a' \in A\}$ denote the set of beliefs that action a_i is a best response to. It is easy to verify that Ξ_i is nonempty, compact, and convex.

A.5.2 Lemmas

Lemma 5. *For every i, ext$(\Xi_i) \subset E$.*

Proof. Suppose by contradiction that there exists $x \in \text{ext}(\Xi_i)$ and $x \notin E$.

Since $x \notin E$, $(x, v(x))$ is not an extreme point of epi (v), so there exists $(x', r'), (x'', r'') \in$ *epi*(v) and $\lambda \in (0, 1)$ such that $(x', r') \neq (x'', r'')$ and

$$(x, v(x)) = \lambda(x', r') + (1 - \lambda)(x'', r'').$$

Observes that $x' \neq x''$, otherwise either $r' < v(x)$ or $r'' < v(x)$, which contradicts to $(x', r'), (x'', r'') \in epi(v)$.

Since $(x, v(x))$ is a boundary point of *epi*(v), by the supporting hyperplane theorem, there exists $h \in \mathbb{R}^n$ and $c \in \mathbb{R}$ such that

$$h \cdot (x, v(x)) = c \quad \text{and} \quad h \cdot y \geq c \quad \text{for all } y \in epi(v).$$

Notice that both (x', r') and (x'', r'') must be on this hyperplane, otherwise

$$h \cdot (x, v(x)) = \lambda h \cdot (x', r') + (1 - \lambda) h \cdot (x'', r'') > c$$

which leads to a contradiction. Moreover, $r' = v(x')$ and $r'' = v(x'')$, otherwise

$$\begin{aligned} h \cdot (x, v(x)) &= \lambda h \cdot (x', r') + (1-\lambda) h \cdot (x'', r'') \\ &> \lambda h \cdot (x', v(x')) + (1-\lambda) h \cdot (x'', v(x'')) \\ &= h \cdot [\lambda(x', v(x')) + (1-\lambda)(x'', v(x'')] \\ &\geq c \end{aligned}$$

where the last inequality follows from $\lambda(x', v(x')) + (1-\lambda)(x'', v(x'')) \in \text{epi } (v)$.

So

$$\begin{aligned} v(x) &= \lambda \sum_\theta v(x') + (1-\lambda) v(x'') \\ &\geq \lambda \sum_\theta x'(\theta)\rho(\theta, a_i) + (1-\lambda) \sum_\theta x''(\theta)\rho(\theta, a_i) \\ &= \sum_\theta x(\theta)\rho(\theta, a_i) \\ &= v(x) \end{aligned} \tag{6}$$

Moreover, by the definition of the interim value function, we have $\sum_\theta x'(\theta)\rho(\theta, a_i) \leq v(x')$ and $\sum_\theta x''(\theta)\rho(\theta, a_i) \leq v(x'')$. Therefore, for equation (6) to hold, we must have $\sum_\theta x'(\theta)\rho(\theta, a_i) = v(x')$ and $\sum_\theta x''(\theta)\rho(\theta, a_i) = v(x'')$, which implies $x', x'' \in \Xi_i$. This contradicts to $x \in \text{ext}(\Xi_i)$.

□

Lemma 6. *For any P, there exists $\tilde{P} \in \hat{\mathcal{D}}(P)$ such that $V(P) = V(\tilde{P})$.*

Proof. For any belief μ, there exists i such that $\mu \in \Xi_i$, and we let $i(\mu)$ be any such i. Observe that v is linear on Ξ_i for each i.

By the definition of $\text{ext}(\Xi_i)$, for each μ, there exists $\gamma(\cdot|\mu) \in \Delta(\text{ext}(\Xi_{i(\mu)}))$ such that

$$\sum_{\mu' \in \text{ext}(\Xi_{i(\mu)})} \gamma(\mu'|\mu)\mu' = \mu.$$

We construct the following posterior distribution:

$$\tilde{\tau}(\mu') = \sum_\mu \tau(\mu)\gamma(\mu'|\mu).$$

From Lemma 5, $\text{ext}(\Xi_i) \subset E$, so $\tilde{\tau}$ is supported on E. Moreover, by construction, $\tilde{\tau}$ is a mean-preserving spread of τ. From Blackwell (1953), there exists $\tilde{P}$ inducing $\tilde{\tau}$ and $\tilde{P}$ Blackwell dominates P. Therefore, $\tilde{P} \in \hat{\mathcal{D}}(P)$, and we will show that $V(P) = V(\tilde{P})$, and the lemma follows.

Now

$$
\begin{aligned}
V(P) &= \sum_{\mu \in supp(\tau_P)} \tau(\mu) v(\mu) \\
&= \sum_{\mu \in supp(\tau_P)} \tau(\mu)\, v\left(\sum_{\mu' \in E} \gamma(\mu'|\mu)\mu'\right) \\
&= \sum_{\mu \in supp(\tau_P)} \tau(\mu) \sum_{\mu' \in E} \gamma(\mu'|\mu) v(\mu') \\
&= \sum_{\mu' \in E} \sum_{\mu \in supp(\tau_P)} \tau(\mu)\gamma(\mu'|\mu) v(\mu') \\
&= \sum_{\mu' \in E} \tilde{\tau}(\mu') v(\mu') \\
&= V(\tilde{P})
\end{aligned}
$$

where the third equality holds because for each μ, $\gamma(\cdot|\mu)$ is supported on $\Xi_{i(\mu)}$ and v is linear on $\Xi_{i(\mu)}$.

□

Lemma 7.

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

Proof. First note that

$$
\begin{aligned}
V(P_1, ..., P_m) &= \min_{P \in \mathcal{P}(P_1,...,P_m)} V(P) \\
&= \min_{P \in \mathcal{D}(P_1,...,P_m)} V(P) \\
&\leq \min_{P \in \hat{\mathcal{D}}(P_1,...,P_m)} V(P)
\end{aligned}
$$

where the second equality holds from Lemma 4, the inequality holds because $\hat{\mathcal{D}}(P_1, ..., P_m) \subset \mathcal{D}(P_1, ..., P_m)$.

Now we show that $V(P_1, ..., P_m) \geq \min_{P \in \hat{\mathcal{D}}(P_1,...,P_m)} V(P)$. Let $P^* \in \text{argmin}_{P \in \mathcal{D}(P_1,...,P_m)} V(P)$. From Lemma 6, there exists $\tilde{P} \in \hat{D}(P^*) \subset \hat{\mathcal{D}}(P_1, ..., P_m)$ such that $V(\tilde{P}) = V(P^*)$. Therefore, $V(P_1, ..., P_m) = V(P^*) = V(\tilde{P}) \geq \min_{P \in \hat{\mathcal{D}}(P_1,...,P_m)} V(P)$, where the inequality holds because $\tilde{P} \in \hat{\mathcal{D}}(P_1, ..., P_m)$. Therefore $V(P_1, ..., P_m) = \min_{P \in \hat{\mathcal{D}}(P_1,...,P_m)} V(P)$.

Finally, let $\mathcal{T}$ denote the set of experiments with induced posteriors with support in E. Then $\hat{\mathcal{D}}(P_1, ..., P_m) = \mathcal{D}(P_1, ..., P_m) \cap \mathcal{T} = \cap_{j=1}^m \mathcal{D}(P_j) \cap \mathcal{T} = \cap_{j=1}^m (\mathcal{D}(P_j) \cap \mathcal{T}) = \cap_{j=1}^m \hat{\mathcal{D}}(P_j)$, which concludes the proof. □

The next lemma shows that set $\hat{\mathcal{D}}(P)$ can be characterized by a k-dimensional polytope, where k is the number of kinks. To simplify the statement of the result, we introduce a few definitions.

Let $T = E \backslash \{\delta_1, ..., \delta_n\}$ denote the set of kinks, and let $k \doteq |T|$. We can list the elements in T by $\{t_1, ..., t_k\}$.

For any belief $\mu \in \Delta(\Theta)$, define the set $X(\mu)$ to be the set of $x \in \Delta(\Theta) \subset \mathbb{R}^k$ such that

$$x_1 t_1 + x_2 t_2 + \cdots + x_k t_k \leq \mu$$

$$x_1 + ... + x_k \leq 1$$

$$x_\ell \geq 0 \text{ for } \ell = 1, .., k$$

which is a k-dimensional polytope.

Lemma 8. *An experiment $Q \in \hat{\mathcal{D}}(P)$ if and only if $\big(\tau_Q(t_1), ..., \tau_Q(t_k)\big) \in \bigoplus_{\mu \in supp(\tau_P)} \tau(\mu) X(\mu)$.*

Proof. "$\Rightarrow$": Suppose an experiment $Q \in \hat{\mathcal{D}}(P)$, then τ_Q is a mean-preserving spread of τ_P. By definition, there exists a stochastic mapping $\eta : supp(\tau_P) \to \Delta E$, such that for any $\mu \in supp(\tau_P)$ and $\nu \in supp(\tau_Q)$,

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

$$\tau_Q(\nu) = \sum_{\mu} \eta(\nu|\mu)\tau_P(\mu).$$

So for each $\mu \in supp(\tau_P)$,

$$\begin{aligned} \mu &= \sum_{\nu \in E} \eta(\nu|\mu)\nu \\ &= \sum_{\ell=1}^{k} \eta(t_\ell|\mu)t_\ell + \sum_{i=1}^{n} \eta(\delta_i|\mu)\delta_i \end{aligned}$$

which implies $\sum_{\ell=1}^{k} \eta(t_\ell|\mu)t_\ell \leq \mu$, so $\big(\eta(t_1|\mu), ..., \eta(t_k|\mu)\big) \in X(\mu)$. Since $\tau_Q(\nu) = \sum_\mu \eta(\nu|\mu)\tau_P(\mu)$, for any $\ell = 1, ..., k$,

$$\tau_Q(t_\ell) = \sum_{\mu} \tau_P(\mu)\eta(t_\ell|\mu)$$

which implies $\big(\tau_Q(t_1), ..., \tau_Q(t_k)\big) \in \bigoplus_{\mu \in supp(\tau_P)} \tau_P(\mu)X(\mu) \subseteq [0, 1]^k$.

"$\Leftarrow$": Suppose an experiment Q generates a posterior distribution $\tau_Q \in \Delta(\Delta(E))$ with $\big(\tau_Q(t_1), ..., \tau_Q(t_k)\big) \in \bigoplus_{\mu \in supp(\tau_P)} \tau_P(\mu)X(\mu)$, we show that τ_Q is a mean-preserving spread of τ_P.

Since $\big(\tau_Q(t_1), ..., \tau_Q(t_k)\big) \in \bigoplus_{\mu \in supp(\tau_P)} \tau_P(\mu) X(\mu)$, there exists $x(\mu) \in X(\mu) \subseteq [0,1]^k$ such that

$$\big(\tau_Q(t_1), ..., \tau_Q(t_k)\big) = \sum_{\mu \in supp(\tau_P)} \tau_P(\mu) x(\mu)$$

Let $x_\ell(\mu)$ denote the ℓ-th element of $x(\mu)$, then by the definition of $X(\mu)$,

$$x_1(\mu)t_1 + x_2(\mu)t_2 + \cdots + x_k(\mu)t_k \leq \mu$$

$$x_1(\mu) + ... + x_k(\mu) \leq 1.$$

Define $\eta : supp(\tau_P) \to \Delta(E)$ as follows:

$$\eta(t_\ell|\mu) = x_\ell(\mu) \quad \text{for } \ell = 1, ..., k$$

$$\eta(\delta_i|\mu) = [\mu - \big(x_1(\mu)t_1 + x_2(\mu)t_2 + \cdots + x(\mu)_k t_k\big)]_i$$

where $[\mu - \big(x_1(\mu)t_1 + x_2(\mu)t_2 + \cdots + x(\mu)_k t_k\big)]_i$ denote the i-th element of the vector.

Notice that

$$\begin{aligned} \sum_{i=1}^{n} \eta(\delta_i|\mu) &= \sum_{\theta} [\mu(\theta) - \big(x_1(\mu)t_1(\theta) + x_2(\mu)t_2(\theta) + \cdots + x(\mu)_k t_k(\theta)\big)] \\ &= 1 - \sum_{\ell=1}^{k} \eta(t_\ell|\mu) \end{aligned}$$

so $\sum_{\ell=1}^{k} \eta(t_\ell|\mu) + \sum_{i=1}^{n} \eta(\delta_i|\mu) = 1$, which shows η is indeed a stochastic mapping. Moreover, it is easy to verify that η preserves the mean, i.e., $\sum_{\ell=1}^{k} \eta(t_\ell|\mu)t_\ell + \sum_{i=1}^{n} \eta(\delta_i|\mu)\delta_i = \mu$.

The last thing we need to show is that $\tau_Q(\delta_i) = \sum_\mu \eta(\delta_i|\mu)\tau_P(\mu)$, for all $i = 1, ..., n$. Notice that

$$
\begin{aligned}
\tau_Q(\delta_i) &= \big[\mu_0 - \sum_{\ell=1}^{k} \tau_Q(t_\ell) t_\ell \big]_i \\
&= \big[\mu_0 - \sum_{\ell=1}^{k} \sum_{\mu} \tau_P(\mu) x_\ell(\mu) t_\ell \big]_i \\
&= \big[\mu_0 - \sum_{\mu} \tau_P(\mu) \sum_{\ell=1}^{k} x_\ell(\mu) t_\ell \big]_i \\
&= \big[\mu_0 - \sum_{\mu} \tau_P(\mu) \left(\mu - \sum_{i=1}^{n} \eta(\delta_i|\mu)\delta_i \right) \big]_i \\
&= \big[\sum_{\mu} \tau_P(\mu) \sum_{i=1}^{n} \eta(\delta_i|\mu)\delta_i \big]_i \\
&= \sum_{\mu} \tau_P(\mu) \eta(\delta_i|\mu).
\end{aligned}
$$

Now we have shown that τ_Q is a mean-preserving spread of τ_P with support in E, so $Q \in \hat{\mathcal{D}}(P)$.

□

The following lemma is a standard result in linear programming, stating that a k-dimensional linear programming problem has at most k effective constraints.

Lemma 9. *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 a $m \times k$ 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

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

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

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 eixsts a 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

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

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

□

Proof of Theorem 3. Recall that

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

Given Lemma 8, the problem can be written as

$$V(P_1, ..., P_m) = \min_{(\tau_Q(t_1), ..., \tau_Q(t_k)) \in \cap_{j=1}^m \mathcal{E}(P_j)} \sum_{\ell=1}^{k} \tau_Q(t_\ell) v(t_l) + \sum_{i=1}^{n} \tau_Q(\delta_i) v(\delta_i)$$

where $\mathcal{E}(P_j) = \bigoplus_{\mu \in supp(\tau_{P_j})} \tau_{P_j}(\mu) X(\mu)$, and $\tau_Q(\delta_i) = \left[\mu_0 - \sum_{\ell=1}^k \tau_Q(t_\ell) t_\ell\right]_i$.

Since the objective function is affine in $(\tau_Q(t_1), ..., \tau_Q(t_k))$, and the constraint set is a polytople, the problem can be reformulated as a linear programing problem:

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

$$s.t. \quad A_1 x \leq b_1$$

$$A_2 x \leq b_2$$

$$\cdots$$

$$A_m x \leq b_m$$

for some $c \in \mathbb{R}^k$, and A_j, b_j are the constraints from $\mathcal{E}(P_j)$. Let $A = [A_1; ...; A_m]$ and $b = [b_1; ...; b_m]$, the constraint can be written as $Ax \leq b$. We index the raws by $i = 1, ..., N$.

The constraint set is non-empty because fully informative information structure is always in $\hat{\mathcal{D}}(P_j)$, so the problem is feasible. Moreover, the constraint set is bounded so the problem has a solution. Let x^* be the solution to the problem.

For every index set $I \subseteq \{1, ..., N\}$, let $A[I]$ denote the $|I| \times k$ submatrix of A with the rows in I. Similarly let $b[I]$ denote the $|I| \times 1$ subvector of b with the raws in I.

From Lemma 9, the there exists $I \subseteq \{1, ..., N\}$ such that

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

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

Since the k number of constraints (rows) at most come from k different A_j, $j = 1, ...m$, so there exists J such that $|J| \leq k$ and

$$V(P_1, ..., P_m) = V(\{P_j\}_{j \in J}) = \min_{(\tau_Q(t_1), ..., \tau_Q(t_k)) \in \cap_{j \in J} \mathcal{E}(P_j)} \sum_{\ell=1}^{k} \tau_Q(t_\ell) v(t_l) + \sum_{i=1}^{n} \tau_Q(\delta_i) v(\delta_i)$$

which concludes the proof. □

A.6 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 signal spaces $Y_1, ..., Y_k$ where $Y_j \cap Y_{j'} = \emptyset$ for all j, j'. 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 10. *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 11. *For any $P \in \mathcal{P}(P_1, ..., P_k)$ and any weights $\{\alpha_j\}_{j=1}^k$, P Blackwell dominates $\bigoplus_{j=1}^k \alpha_j P_j$.*

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

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

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

$$\begin{aligned} \sum_{y_1,...,y_k} \gamma(y|y_1, ..., y_k)P(y_1, ..., y_k|\theta) &= \sum_{y_{-j}} \alpha_j P(..., y_{j-1}, y, y_{j+1}...|\theta) \\ &= \alpha_j P(y|\theta) \\ &= \bigoplus_{j=1}^{k} \alpha_j P_j(y|\theta), \end{aligned}$$

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

Proof of Proposition 1. For any decision problem (A, u), let $P^*_{A,u}$ solves

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

From Lemma 11 and the transitivity of the Blackwell order, $P^*_{A,u}$ 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 $Q \in \mathcal{P}(P_1, ..., P_m)$:

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

which by construction is Blackwell equivalent to $P^*_{A,u}$. Therefore,

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

which proves the proposition.

□

A.7 Proof of Proposition 3

Lemma 12 (Single-Peaked Property)**.** *Suppose in a 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 $\mu \in \Delta(\Theta)$,

$$a_i \in \underset{a \in A}{\text{argmax}} \sum_{\theta} \mu(\theta) u(\theta, a)$$

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

$$\mu'(\theta_1) u(\theta_1, a_j) + \mu'(\theta_2) u(\theta_2, a_j) < \mu'(\theta_1) u(\theta_1, a_k) + \mu'(\theta_2) u(\theta_2, a_k). \tag{7}$$

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

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

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

$$\mu'(\theta_1)u(\theta_1,a_j)+\mu'(\theta_2)u(\theta_2,a_j)\leq \mu'(\theta_1)u(\theta_1,a_i)+\mu'(\theta_2)u(\theta_2,a_i) \tag{8}$$

The inequalities (7) and (8) together imply that a_j is never a unique best response to any belief, contradicting our assumption.

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

Lemma 13. *For any subproblem (A_ℓ,u_ℓ) in a binary decomposition of (A,u),*

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

Proof. Recall that P_j Blackwell dominates $R_{P_j;(A,u)}$, so $V(P_j;(A_\ell,u_\ell))\geq V(R_{P_j;(A,u)};(A_\ell,u_\ell))$. Let $R_{P_j;(A_\ell,u_\ell)}$ be the recommendation information structure with respect to the decision problem (A_ℓ,u_ℓ). By definition, we have $V(P_j;(A_\ell,u_\ell))=V(R_{P_j;(A_\ell,u_\ell)};(A_\ell,u_\ell))$. Next, we will show that $V(R_{P_j;(A,u)};(A_\ell,u_\ell))\geq V(R_{P_j;(A_\ell,u_\ell)};(A_\ell,u_\ell))$, which then completes the proof.

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

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

From Lemma 12, if $a_i\in\operatorname{argmax}_{a\in A}\sum_\theta P_j(y_j|\theta)u(\theta,a)$, for all $i\leq \ell\leq n-1$, $\sum_\theta P_j(y_j|\theta)u(\theta,a_\ell)\geq \sum_\theta P_j(y_j|\theta)u(\theta,a_{\ell+1})$, and for all $2\leq \ell\leq i$, $\sum_\theta P_j(y_j|\theta)u(\theta,a_\ell)\geq\sum_\theta P_j(y_j|\theta)u(\theta,a_{\ell-1})$.

Now we construct another garbling $\gamma:A\to\{0,1\}$:

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

By construction, for each y_i in the support,

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

which then implies the information structure induced by the garbling $\gamma\circ\sigma^*$ is a recommendation information structure for the decision problem (A_ℓ,u_ℓ). Moreover, by construction, $R_{P_j;(A,u)}$ Blackwell dominates $R_{P_j;(A_\ell,u_\ell)}$, so $V(R_{P_j;(A,u)};(A_\ell,u_\ell))\geq V(R_{P_j;(A_\ell,u_\ell)};(A_\ell,u_\ell))$.

□

Proof of Proposition 3. Consider a binary decomposition of (A,u): $\{(A_1,u_1),\dots,(A_k,u_k)\}$.

From Theorem 2 and Lemma 13,

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

□

A.8 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 $j \neq 1$.

A.8.1 Payoff Sets

Recall that as in Section 3.3, 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 her to freely dispose utils. Such polyhedron is upper bounded, convex, closed, and has a finite number of extreme points. For the sake of our proof, it will be convenient to directly work on these payoff vector sets.

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:

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

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

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 14. *For any decision problem D,*

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

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

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

□

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

Lemma 15. *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

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

□

A.8.2 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 subspaces:

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

where $\Lambda_{(A^{bi}, u^{bi})} = \{e_1, e_2, \lambda^*\}$ with $e_1 = (1, 0)$, $e_2 = (0, 1)$, and $\lambda^* \in \mathbb{R}^2_{++}$ 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 $k_{\lambda^*} \in \mathbb{R}$ are the constant terms. This can be visualized in Fig. 7.

The set of normal vectors, $\Lambda_{(A^{bi}, u^{bi})}$, depends on the binary action decision problem we are considering. Since the decision problem (A^{bi}, u^{bi}) is fixed, for notational simplicity, we will henceforth omit the dependence of Λ on (A^{bi}, u^{bi}).

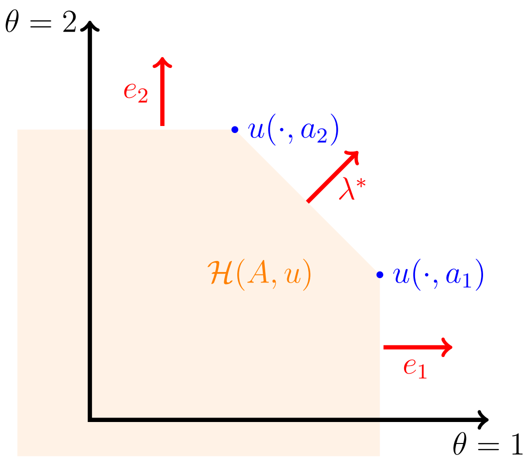


Figure 7: Payoff polyhedron for a binary-state binary-action problem

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 Λ-shape polyhedron if*

$$D = \bigcap_{\lambda \in \Lambda} \{v \in \mathbb{R}^2 : \lambda \cdot v \leq k_\lambda\}$$

for some constants $\{k_\lambda\}_{\lambda \in \Lambda} \in \mathbb{R}$.

Note that the constraint $\lambda^* \cdot v \leq k_{\lambda^*}$ may be redundant in a Λ-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 Λ-shape polyhedron *trivial*.

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

Lemma 16. *If D is a non-trivial $\Lambda-$shape polyhedron, then $W(P_1; D) > \max_{j \neq 1} W(P_j; D)$.*

Proof. Any non-trivial Λ-shape polyhedron D has two extreme points, denoted by $ex(D)_1$ and $ex(D)_2$. See Fig. 8 for an illustration.

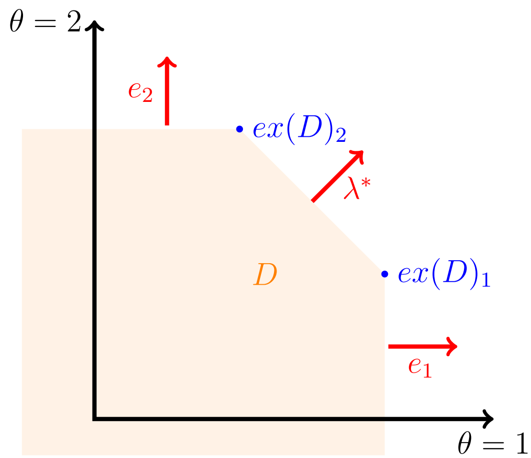


Figure 8: Extreme points of a non-trivial Λ-polyhedron

The two extreme points are defined by two linear equations:

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

with the closed-form solutions $ex(D)_1 = \begin{pmatrix} k_{e_1} \\ \frac{k_{\lambda^*} - \lambda_1^* k_{e_1}}{\lambda_2^*} \end{pmatrix}$ and $ex(D)_2 = \begin{pmatrix} \frac{k_{\lambda^*} - \lambda_2^* k_{e_2}}{\lambda_1^*} \\ k_{e_2} \end{pmatrix}$. A useful observation is that $(ex(D)_2 - ex(D)_1) = (k_{\lambda^*} - k_{e_1}\lambda_1^* - k_{e_2}\lambda_2^*) \begin{pmatrix} -\frac{1}{\lambda_1^*} \\ \frac{1}{\lambda_2^*} \end{pmatrix}$. That is, λ^* determines the direction of the vector $(ex(D)_2 - ex(D)_1)$, and the constant terms k_λ only affect the scalar multiplier. Moreover, the multiplier $(k_{\lambda^*} - k_{e_1}\lambda_1^* - k_{e_2}\lambda_2^*) > 0$, because $(k_{e_1}, k_{e_2}) \in int(D)$ and $k_{\lambda^*} = \max_{v \in D} \lambda^* \cdot v$.

For any Λ-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

$$t_j^*(y_j) = \begin{cases} ex(D)_1 & \text{if } \mathbf{P}_j(y_j) \cdot \begin{pmatrix} -\frac{1}{\lambda_1^*} \\ \frac{1}{\lambda_2^*} \end{pmatrix} \leq 0 \\ ex(D)_2 & \text{if } \mathbf{P}_j(y_j) \cdot \begin{pmatrix} -\frac{1}{\lambda_1^*} \\ \frac{1}{\lambda_2^*} \end{pmatrix} > 0. \end{cases}$$

For each P_j, let $\tilde{Y}_j = \{y \in Y_j : \mathbf{P}_j(y_j) \cdot \begin{pmatrix} -\frac{1}{\lambda_1^*} \\ \frac{1}{\lambda_2^*} \end{pmatrix} \leq 0\}$, 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

$$\begin{aligned} W(P_1; D) - W(P_j; D) &= (\mathbf{x}_{P_j} - \mathbf{x}_{P_1}) \cdot (ex(D)_2 - ex(D)_1) \\ &= (k_{\lambda^*} - k_{e_1}\lambda_1^* - k_{e_2}\lambda_2^*)(\mathbf{x}_{P_j} - \mathbf{x}_{P_1}) \cdot \begin{pmatrix} -\frac{1}{\lambda_1^*} \\ \frac{1}{\lambda_2^*} \end{pmatrix} \end{aligned}$$

Note that for different non-trivial Λ-shape polyhedra D (i.e., different parameters $k_{e_1}, k_{e_2}, k_{\lambda^*}$), 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 Λ-shape polyhedron, the value is also strictly positive for any non-trivial Λ-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 Λ-shape polyhedron. Therefore,

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

for any non-trivial Λ-shape polyhedron.

□

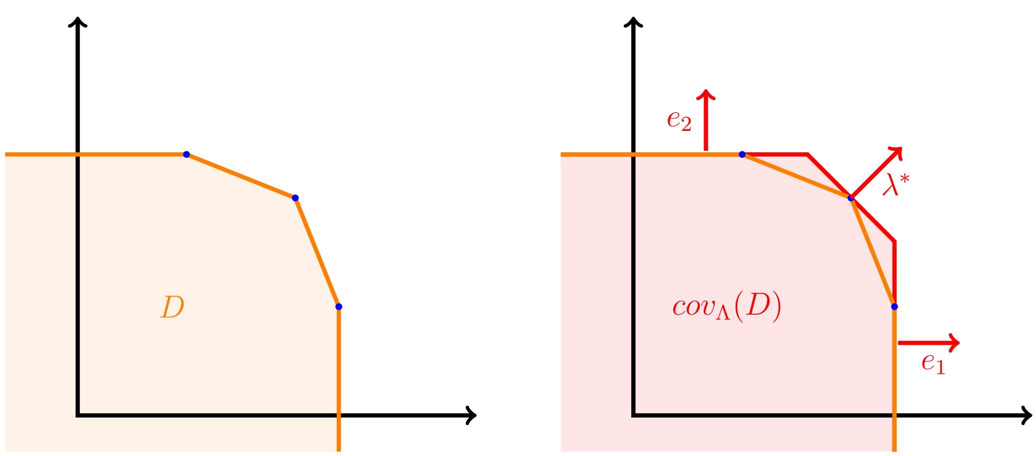
A.8.3 Λ-cover

For any payoff set D, we define the smallest Λ-shape polyhedron that covers it as its Λ-*cover*. See Fig. 9 for an illustration.

Definition 10. *For any payoff set D, its* Λ-***cover*** *is defined as*

$$cov_\Lambda(D) \doteq \bigcap_{\lambda \in \Lambda} \{v : \lambda \cdot v \leq \rho_D(\lambda)\},$$

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



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

(b) The corresponding Λ-cover $cov_\Lambda(D)$

Figure 9

We state a few properties of Λ-cover that will be useful in our analysis.

Lemma 17. *1. (Monotonicity) If $D \subseteq D'$, $cov_\Lambda(D) \subseteq cov_\Lambda(D')$.*

2. *(Reflexive) If D is a Λ-shape polyhedron, $cov_\Lambda(D) = D$.*

3. *(Superadditivity) $cov_\Lambda(D + D') \supseteq cov_\Lambda(D) + cov_\Lambda(D')$*

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

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

$$\bigcap_{\lambda \in \Lambda} \{v : \lambda \cdot v \leq \rho_D(\lambda)\} \subseteq \bigcap_{\lambda \in \Lambda} \{v : \lambda \cdot v \leq \rho_{D'}(\lambda)\}.$$

2. Clearly $D \subseteq cov_\Lambda(D)$, because for every $v \in D$ and every $\lambda \in \Lambda$, $\lambda \cdot v \leq \rho_D(\lambda)$.

   Now consider any Λ-shape polyhedron, represented by

$$D = \bigcap_{\lambda \in \Lambda} \{v \in \mathbb{R}^2 : \lambda \cdot v \leq k_\lambda\}$$

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

$$cov_\Lambda(D) = \bigcap_{\lambda \in \Lambda} \{v : \lambda \cdot v \leq \rho_D(\lambda)\} \subseteq \bigcap_{\lambda \in \Lambda} \{v : \lambda \cdot v \leq k_\lambda\} = D,$$

   which implies $cov_\Lambda(D) = D$.

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

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

   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 $\max_{v \in D} \lambda^* \cdot v < \lambda^* \cdot \bar{v}$. 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 $\bar{v} \notin \{v : \lambda^* \cdot v \leq \rho_D(\lambda^*)\}$, contradicting to the constraint $\lambda^* \cdot v \leq \rho_D(\lambda^*)$ being redundant. Thus, $\bar{v} \in D$ and for all $v \in D$, $v \leq \bar{v}$, which concludes the proof.

□

A.8.4 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 18. *If $\{D_\ell\}_{\ell=1}^k$ is dominated by D,*

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

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

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

Then

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

□

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

Lemma 19. *Suppose a collection of decision problems $D_1, ..., D_m$ is dominated by a Λ-shape* *polyhedron D, and satisfies*

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

Then $cov_\Lambda(D_j)$ must be trivial for all $j \neq 1$.

Proof. Since $D_1 + \cdots + D_m \subseteq D$, from properties 1 and 2 in Lemma 17,

$$cov(D_1 + \cdots + D_m) \subseteq cov(D) = D.$$

From property 3 in Lemma 17,

$$cov(D_1) + \cdots + cov(D_m) \subseteq cov(D_1 + \cdots + D_m),$$

so $cov(D_1)$, $\cdots$, $cov(D_m)$ is also dominated by D.

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

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

where the first inequality follows from Lemma 18, second inequality follows from Lemma 14, the third inequality follows from Lemma 16, 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 $j \neq 1$. □

A.8.5 Commom Support of the Blackwell Supremum

Lemma 20. *Suppose $P_j(y_j|\theta) > 0$ for all j, y_j, θ, and $P^* \in \mathcal{P}(P_1, ..., P_m)$ 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 3, $\Lambda_{P^*} = co(\Lambda_{P_1} \cup \cdots \cup \Lambda_{P_m})$, which also should not contain such points, leading to a contradiction. □

A.8.6 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

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

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

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

From Lemma 19, $cov(D_2), ..., cov(D_m)$ must be trivial, and property 4 of Lemma 17 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 4 and Corollary 3, a Blackwell supremum $P^* \in \mathcal{P}(P_1, ..., P_m)$ solves the

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

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

Lemma 20 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.

□