Robust Aggregation of Correlated Information*

Henrique de Oliveira† Yuhta Ishii‡ Xiao Lin§

October 3, 2023

Abstract

An agent makes decisions with multiple sources of information. In isolation, each source is well understood, but jointly 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, Pietro Ortoleva, Collin Raymond, Shamim Sinnar, Rakesh Vohra, 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 many information sources available. Treatment decisions can be made by consulting multiple doctors. Retirement plans can follow the advice of numerous financial experts. Indeed, for nearly any decision we face, there are news, websites, books, articles and other relevant sources at our disposal. But collating and analyzing data from different sources can be both practically and mentally 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 their correlations.*

Different information sources are usually correlated: doctors may base their recommendation on the same study; financial analysts can have an incentive to echo each other, as shared errors are more forgivable. Understanding the correlation 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. Even when information about the correlation is readily available, behaviorally biased agents may fail to properly account for it.[1] Thus, agents may look to take decisions that do not leave them vulnerable to a misspecification in correlation.

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, 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 the best *single* information source and best responds to it, ignoring all other information sources. Since the resulting payoff from such a strategy is determined solely by the information source selected, this strategy guarantees a payoff that is independent of the correlation between information sources. Of course, such a strategy completely forfeits the potential benefits from observing multiple information sources. Could we 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

[1] This phenomenon, known as correlation neglect, is well documented in the behavioral literature (see e.g. Enke and Zimmermann (2019)).

answer is no: best-source strategies are always robustly optimal.

With more than two actions or more than two states, best-source strategies are no longer always optimal, and robustly optimal strategies may combine different sources of information. Theorem 2 characterizes robustly optimal strategies under two states and multiple actions, as follows: First, we decompose a decision problem with n actions into $n-1$ binary actions subproblems, in which the actions are ordered and each binary action subproblem constitutes a "local" comparison of consecutive actions. In line with Theorem 1, a best-source strategy is used within each binary-action subproblem to select a recommended action. Then the robustly optimal strategy maps each profile of recommended actions for the subproblems into an action in the original problem. Theorem 2 shows that the robustly optimal strategy uses signals from multiple information sources precisely when the best sources across subproblems are not the same. However, the strategy given by Theorem 2 for an n-action decision problem uses at most $n-1$ information sources. Moreover, the robustly optimal strategy uses an information source if and only if that information source is a best source in one of the local subproblems.

A full characterization of the robustly optimal strategy in decision problems with more than three states is more complex and beyond the scope of this paper. Indeed, we show an example of a binary-action problem with three states in which the robustly optimal strategy necessarily uses multiple information sources. However, as in Theorem 1 and Theorem 2, Theorem 3 establishes a bound on the number of necessary information sources. Crucially, this bound again depends only on the decision problem, and it could be relatively small when many information sources are available.

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.

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 these works, 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, but before the true state is revealed.

Using this interim approach, Levy and Razin (2020a) also consider ambiguity about the correlation structure. They allow for joint experiments that satisfy a bound of correlation, and look for the worst case that can rationalize the realized prediction. In contrast, our paper uses an *ex-ante approach*, by considering the strategy for all signal realizations and its ensuing expected payoff. This results in a "combined forecast" that depends on the payoff structure as well.

Perhaps the most closely related paper to ours 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) where, similarly to our Theorem 1, the optimal aggregation rule turns out to pick a single information source. However, their agent's objective is to minimize regret, so different techniques are necessary.

Robustness to correlations have also been studied in other contexts, such as mechanism design. In particular, 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. With similar robustness concerns regarding the correlations of values between different bidders, He and Li (2020) study an auctioneer's robust design problem when selling a single indivisible good to a group of bidders.

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.

A marginal experiment $P_j : \Theta \to \Delta Y_j$ maps each state to a distribution over some finite signal set Y_j. The agent can observe the realizations of multiple marginal experiments $\{P_j\}_{j=1}^m$, but does not have detailed knowledge of the joint. To simplify notation, let $\mathbf{Y} = Y_1 \times \cdots \times Y_m$ denote the set of possible observations the agent can see. Thus, the agent conceives of the following set of joint experiments:

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

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

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

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

Clearly if only one experiment $P : \Theta \to \Delta(Y)$ is considered ($m = 1$), $V(P)$ 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.

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. As a starting point for our analysis, 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. By choosing Q optimally, the agent achieves an ex-ante expected payoff of $\max_{j=1,\ldots,n} V(P_j)$, regardless of the actual joint experiment $P \in \mathcal{P}(P_1, \ldots, P_m)$. We call such a strategy a *best-source strategy*.

In some special cases, it is easy to see that the best-source strategy is robustly optimal. For example, if the marginal experiments are identical, Nature can perfectly correlate the signals to make each additional information source completely uninformative beyond the first. Similarly, if P_1 Blackwell dominates $P_2, \ldots, P_m$, then nature can correlate the signals according to the corresponding Blackwell garblings to make the additional informational content of $P_2, \ldots, P_m$ zero, and so the agent can never guarantee a robust value more than $\max_{j=1,\ldots,n} V(P_j)$. But in general, so long as the marginal experiments are not Blackwell ranked, any correlation structure $P \in \mathcal{P}(P_1, ..., P_m)$ would be strictly more informative than every marginal experiments, in the Blackwell sense.[2]

Somewhat surprisingly, we show in Theorem 1 that in simple decision problems—those with binary states and binary actions — the best-source strategy is the best that the agent can do even beyond these special cases.

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

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

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. We present the proof of Theorem 1 in detail in Section 4. Clearly, $\max_{j=1,\ldots,m} V(P_i; (A; u))$ is a lower bound on the robustly optimal value. In order to show the reverse inequality, we construct a joint information structure, $\overline{P}(P_1, \ldots, P_m)$, 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 $\overline{P}(P_1, \ldots, P_m)$ 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:*

[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.

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* $0.9 \cdot 1 + 0.1 \cdot (-1) + 0.5 \cdot 1 + 0.5 \cdot 2 = 2.3$.[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]Recall that $u(\theta, a) = \nu(\theta)\rho(\theta, a)$, so the payoffs here have been weighted by the prior.

[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.

3.1 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 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.

Proposition 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

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

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]

3.2 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*.

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.

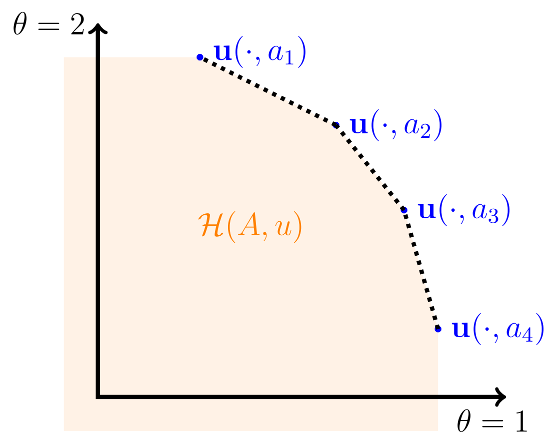


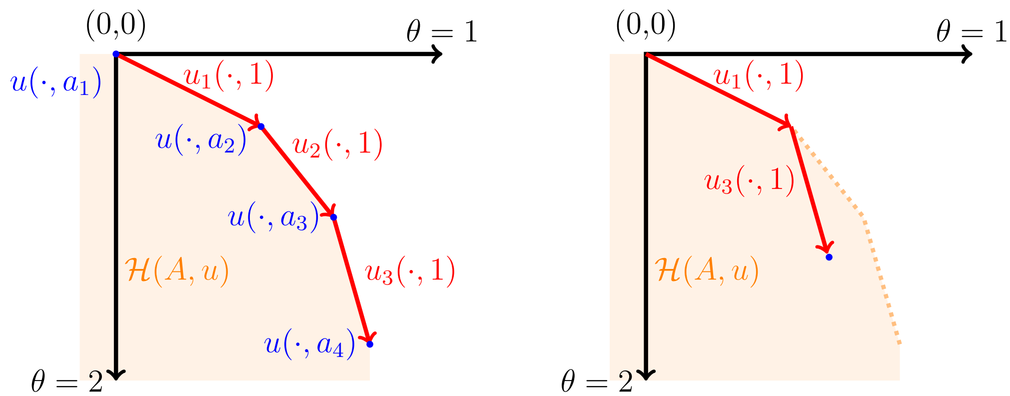
Figure 1: The shaded area represents $\mathcal{H}(A,u)$

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

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

[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.

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



(a) Binary decomposition

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

Figure 2

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 $u(\cdot, a_1) = (0, 0)$.

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.

[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.

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

Proof. See Section A.4. □

Lemma 1 and Proposition 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 $\{P_j\}_{j\in S\subset\{1,\dots,m\}}$ 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

We now return to the proof of Theorem 1. 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;*

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

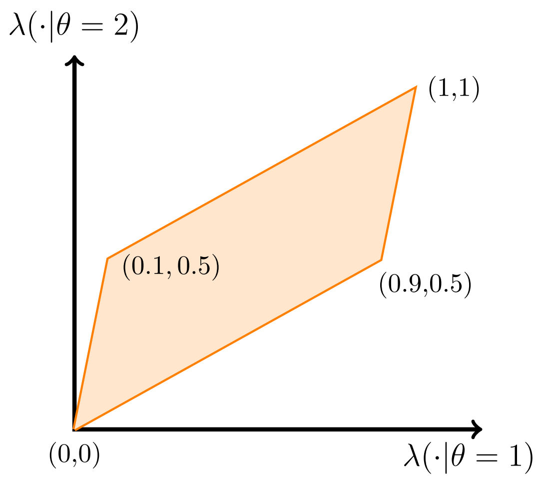
This last equivalent condition gives us a simple graphical representation of Blackwell experiments when $|\Theta| = 2$. See Figure 3(a) for an illustration. Since $|A| = 2$, to characterize $\Lambda_P(A)$, it suffies to specify the probability of taking one of the two actions. The x-axis denotes the 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;*



(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

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 2. *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.

[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).

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

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

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 3.

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

The idea underlying Lemma 3 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 3 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* $|\Theta| = 2$,

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

[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.

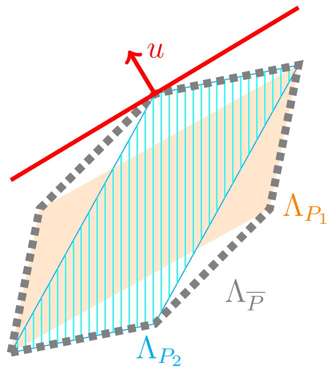


Figure 4: The maximum is achieved at an extreme point that belongs to Λ_{P_2}

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 $V(\overline{P}(P_1, ..., P_m)) = \max_{j=1,...,m} V(P_j)$. By Lemma 2, an experiment $\overline{P}$ is the Blackwell supremum of $P_1, \ldots, P_m$ if and only if

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

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. $\square$

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 non-existence 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. Of course, 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:

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

Then the agent's maxmin value from marginals P_X, P_Y is her perfect information payoff: $0 + 1 + 1 = 2$.

By using only one information source (either P_X or P_Y), $a = 0$ is always a best response to any signal realization, so the agent's expected payoff is 0.

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

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

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 is 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 blue dots are its kinks.

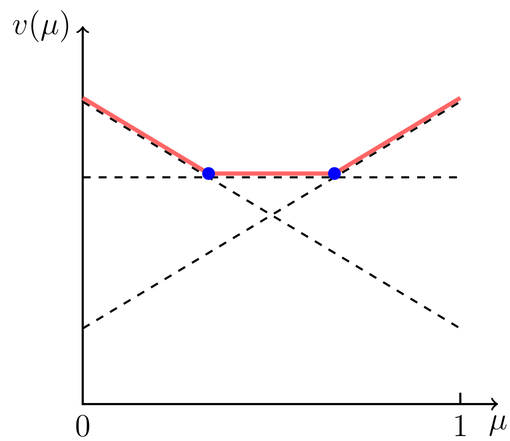


Figure 5: Interim value function and kinks

The following theorem provides a bound on the number of experiments that a decision maker would need, which is precisely 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*

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

The full proof of Theorem 3 is deferred to Appendix A.5, but here we will sketch the main steps. We will prove the theorem by examining Nature's minmax problem. From Lemma 3, Nature's minmax problem can be relaxed into choosing an experiment among the set of all experiments that Blackwell dominate $P_1, ..., P_m$.

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 2. *If for any decision problem* (A, u), $V(P_m; (A, u)) \leq \max_{j=1,...,m-1} V(P_j; (A, u))$, *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 2 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.[11]

[11] In the proof, we established a slightly stronger result than Proposition 2: experiment P_m can be ignored if it is dominated by all correlation structures between $P_1, ..., P_{m-1}$.

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.

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.

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

A Appendix

A.1 Proof of Proposition 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 1, (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 Proposition 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 Proposition 1 and the third line follows from Lemma 1. So σ^* is a robustly optimal strategy. $\square$

A.3 Proof of Lemma 3

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 $\tilde{Q} \in \mathcal{P}(P_1, ..., P_m)$ 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 $\tilde{Q} : \Theta \to \Delta(Y_1 \times ... \times Y_m)$:

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

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

A.4 Proof of Lemma 1

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 $\ell \geq \ell^*$.

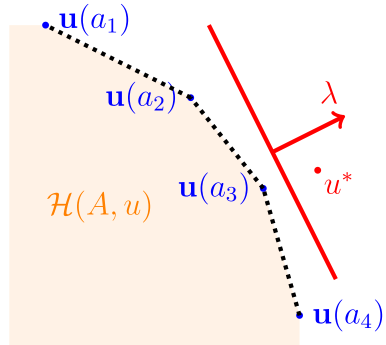


Figure 6

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 4. *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 5. *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 4, $\text{ext}(\Xi_i) \subset E$, 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 6.

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

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 3, 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 5, 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 will need a few more 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 7. *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 $Q \in \hat{\mathcal{D}}(P)$. $\square$

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 8. *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 7, 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 8, 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 2

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 9. *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 $\{P_1, ..., P_{m-1}\}$; that is,*

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 10. *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: $\gamma : Y_1 \times ... \times Y_k \to \Delta(Y_1 \cup \cdots \cup Y_k)$:

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

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

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

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

Proof of Proposition 2. 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 10 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,

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

which proves the proposition.

□