Robust Merging of Information[*]

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

May 31, 2021

Abstract

When multiple sources of information are available, any decision must take into account their correlation. If information about this correlation is lacking, an agent may find it desirable to make a decision that is robust to possible correlations. Our main results characterize the strategies that are robust to possible hidden correlations. In particular, with two states and two actions, the robustly optimal strategy pays attention to a single information source, ignoring all others. More generally, the robustly optimal strategy may need to combine multiple information sources, but can be constructed quite simply by using a decomposition of the original problem into separate decision problems, each requiring attention to only one information source. An implication is that an information source generates value to the agent if and only if it is best for at least one of these decomposed problems.

1 Introduction

During the COVID-19 pandemic, testing has been essential in effectively monitoring the transmission of the virus. Two prevalent diagnostic tests are the molecular and antigen tests, which differ in checking the virus's genetic materials or specific proteins.[1] It might then be appealing to use both tests.[2] However, in order to correctly interpret the joint pair of results from the two tests, knowledge of their correlation is crucial. For example, conditional on the molecular test producing a false negative, what is the probability that the antigen test also yields a false negative? Although the likelihoods of false positives and false negatives for each test are well-understood, data regarding the correlations between these tests is scarce.[3] With such limited information about these correlations, how is a health authority supposed to make use of the results of both tests?

Such unclear correlation between information sources is a common difficulty in practical decision problems. For example, someone might have access to the opinions of multiple experts (such as doctors), but these experts might use similar specialized sources (such as a flawed study). In this paper, we assume that the agent fully understands each information source in isolation, but has no knowledge about the correlations between different information sources. We then look for strategies that are robust to such correlation, by considering the worst possible correlation that could occur.

Our main results characterize robustly optimal strategies. The simplest characterization occurs when we have two states and two actions. In that case, to guard against hidden correlation, one must resort to a rather extreme measure: the optimal robust strategy involves paying attention to a single information source, ignoring all others. In the example of the health authority, since the relevant state is whether the patient is infected with covid or not, if the decision to be made is whether to put the patient in quarantine or not, our result implies that only one test should be considered. Even if both tests have already been administered or are completely costless, the health authority should still ignore one of them.

In more general settings, this extreme measure is no longer necessary and it can be beneficial to use multiple information sources. However, we show a method of finding robust strategies that consists of decomposing a decision problem into subproblems, each requiring the use of a single information source. This shows the precise way in which information sources should be merged. In general, this decomposition can depend on the information sources, but we also show that, with two states, there is a canonical decomposition of a decision problem into binary action problems that is independent of information sources.

Finally, these characterizations of the robustly optimal strategy provide normative guidelines for constructing strategies that are robust to potentially misspecified correlations and reduce the computational burden of finding such strategies. They also provide an alternative explanation for some behavioral patterns documented empirically, such as when decision makers ignore free information in making their decisions.

---

[1] For more information regarding these tests, see for example https://www.fda.gov/health-professionals/closer-look-covid-19-diagnostic-testing.

[2] Taking both tests is indeed recommended by FDA: "(for antigen test) positive results are usually highly accurate, ... negative results may need to be confirmed with a molecular test." Some medical providers always require one to take both tests.

[3] See for example Dinnes et al. [2020].

2 Related Literature

Our paper provides practical robust strategies to deal with possible hidden correlation. The practice of finding robust strategies dates back at least to Wald [1950] and our modeling of information structures follows that of Blackwell [1953].

Our way of modeling robustness, by considering the worst case scenario, also is in line with the literature on ambiguity aversion, going back to Gilboa and Schmeidler [1989]. More recently, Epstein and Halevy [2019] run an experiment that documents ambiguity aversion on correlation structures.

More closely related, some papers consider strategies that are robust to unknown correlations in different contexts. 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.

A recent thread of related literature similarly studies how a decision maker combines forecasts from multiple sources. Levy and Razin [2020a] consider a model where the decision maker can consult multiple forecasts (posterior beliefs), but is uncertain about the information structures that generate these forecasts. Levy and Razin [2020b] study a maximum likelihood approach of combining forecasts, and derive a novel result that only extreme forecasts will be used. A key distinction is that the aforementioned papers consider robust optimality from an interim approach, while we study the decision maker's robustly optimal ex-ante decision plan. Finally, Arieli, Babichenko, and Smorodinsky [2018] also study features of the robustly optimal ex-ante decision plans. An important difference is that they study robust aggregation in a specific decision problem while we characterize the robustly optimal ex-ante decision plan in general decision problems.[4] Moreover, Arieli, Babichenko, and Smorodinsky [2018] study robust aggregation when the decision maker has limited knowledge of the distribution of posteriors/signals generated by each expert. In contrast, in order to focus our analysis on robustness concerns about correlations between information sources, we assume in our model that the decision maker possesses a perfect understanding of the marginal distributions of signals of each expert/information source in isolation.

3 Model

An agent faces a decision problem $\Gamma \equiv (\Theta, \nu, A, \varrho)$ with **binary** state space $\Theta = \{1, 2\}$, prior $\nu \in \Delta\Theta$, finite action space A, and utility function $\varrho : \Theta \times A \to \mathbb{R}$. 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 $\mathbf{Y} = Y_1 \times \cdots \times Y_m$

---

[4]More specifically, they study the robustly optimal decision plan in the sense of worst-case regret when the decision maker must make a forecast $x \in [0, 1]$ in the unit interval, and incurs a quadratic loss of $-(\omega - x)^2$ conditional on the true state, $\omega \in \{0, 1\}$.

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 his/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.

3.1 The Blackwell order

It will be useful to rank experiments according to how much information they convey. For that, we will use the Blackwell order, which we review in this subsection for completeness. The reader familiar with the Blackwell order may skip this part.

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

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

Definition 2. *$Q : \Theta \to \Delta(Z)$ is a garbling of $P : \Theta \to \Delta(Y)$ if there exists a function $g : Y \to \Delta(Z)$ such that*

$$Q(z|\theta) = \sum_y g(z|y) P(y|\theta).$$

The function g is then called "the garbling".

Thus Q is a garbling of P when one can replicate Q by "adding noise" to the signal generated from P. Another notion of informativeness can be obtained by considering the strategies that are feasible given the experiment.

Definition 3. *Given a set of actions A and an experiment $P : \Theta \to \Delta(Y)$, the feasible set of actions given P is*

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

The feasible set of an experiment specifies what conditional action distributions can be obtained by some choice of strategy σ (see Figure 1). One might then try to rank the informativeness of experiments according to the size of the feasible set.

4

Blackwell's Informativeness Theorem states that these three rankings of informativeness are equivalent (for a proof, see 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, $\Lambda_Q \subseteq \Lambda_P$.*

It should be clear from the definitions that the Blackwell order is not complete—it is possible for two experiments to be unranked. It is also possible for two different experiments to be *equivalent*, in the sense that each Blackwell dominates the other. For example, we can change the labels in the signal set while keeping the probabilities the same. This lack of uniqueness is easily remedied by considering equivalence classes of experiments when necessary.

3.2   The Blackwell supremum

In the next section, we will use some lattice properties of the Blackwell order. In particular, the concept of a Blackwell supremum will be useful.

Definition 4. *Let P and Q be two arbitrary experiments. We say that R is the Blackwell supremum of P and Q if*

1. *R is more informative than P and Q*

2. *If S is more informative than P and Q then S is also more informative than R*

The definition generalizes immediately to any number of experiments. It is immediate from the definition that, if there are two Blackwell suprema, they must Blackwell dominate each other. Hence, by considering the equivalence class of equally informative experiments, we can say that the Blackwell supremum is unique. However, when $|\Theta| > 2$ a Blackwell supremum may not exist (see Bertschinger and Rauh [2014], example 18). Fortunately for us, the Blackwell order does form a lattice when $|\Theta| = 2$. In particular, the existence of a Blackwell supremum will be useful.

Lemma 1. *When $|\Theta| = 2$, the Blackwell supremum always exists.*

For a proof, see Bertschinger and Rauh [2014], proposition 16.

The Blackwell supremum can also be characterized using the feasible set. If R is the Blackwell supremum of P and Q, we know from Blackwell's Theorem that Λ_R must contain both Λ_P and Λ_Q. Moreover, if S is more informative than P and Q, it must be more informative than R as well, so Λ_S must also contain Λ_R. Hence the feasible set of the Blackwell supremum should be the smallest feasible set containing $\Lambda_P \cup \Lambda_Q$. Since the feasible set is always convex, a candidate feasible set is $co(\Lambda_P \cup \Lambda_Q)$. If such an R exists satisfying $\Lambda_R = co(\Lambda_P \cup \Lambda_Q)$, it must be the Blackwell supremum (see Figure 1). The difficulty lies in showing that such an R exists, and that's where the proof of existence fails when $|\Theta| > 2$.

Another useful property of the Blackwell order when $|\Theta| = 2$ is that it is characterized by the feasible set with only two actions—P is more informative than Q if for any set A with $|A| = 2$, we have $\Lambda_Q \subseteq \Lambda_P$ (see Blackwell [1953], Theorem 10). We can use this property to obtain a characterization of the Blackwell supremum.
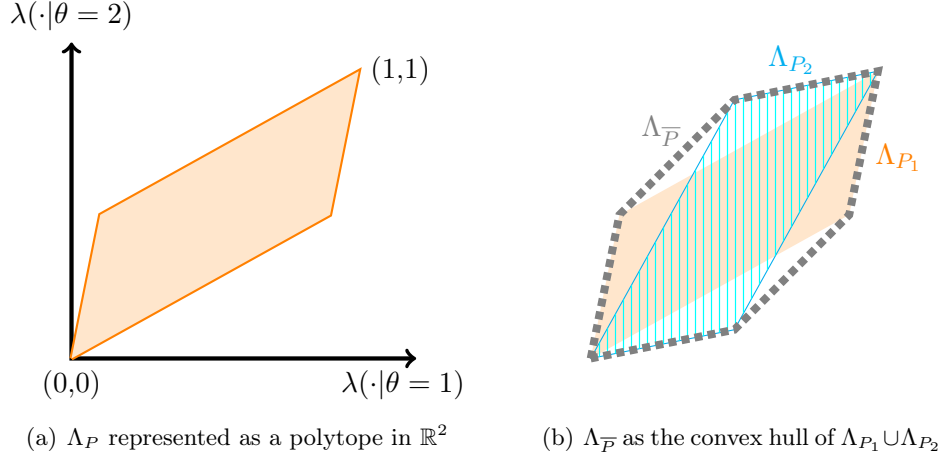
(a) Λ_P represented as a polytope in $\mathbb{R}^2$ (b) $\Lambda_{\overline{P}}$ as the convex hull of $\Lambda_{P_1} \cup \Lambda_{P_2}$

Figure 1

Lemma 2. *Suppose $|\Theta| = 2$ and $|A| = 2$. Then R is the Blackwell supremum of P and Q if and only if $\Lambda_R = co(\Lambda_P \cup \Lambda_Q)$*

In fact, this result can be used to show the existence of the Blackwell supremum when $|\Theta| = 2$.

4   Nature's 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

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

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.

An immediate observation is that if there exists a Blackwell least informative element in the set $\mathcal{P}(P_1, \ldots, P_m)$, it would solve Nature's problem—any other information structure would yield a higher value for the agent. Notice 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. By Lemma 1, there is a *Blackwell supremum*—the least informative experiment that Blackwell dominates every P_j. The only question that remains is whether this Blackwell supremum can be expressed as a joint distribution with marginals $P_1, \ldots, P_m$. This is proved in the following lemma.

Lemma 3. *For any collection of experiments $\{P_j\}_{j=1}^m$, there exists a Blackwell supremum $\overline{P}(P_1, \ldots, P_m) \in \mathcal{P}(P_1, \ldots, P_m)$ so that for any $P \in \mathcal{P}(P_1, \ldots, P_m)$, $V(\overline{P}(P_1, \ldots, P_m)) \leq V(P)$.*

Proof. See Appendix A.1. □

Immediately from the lemma, we have the following proposition.

Proposition 1.
$$V(P_1, ..., P_m) = V(\overline{P}(P_1, ..., P_m))$$

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

Thus, 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.

5 Binary Action Decision Problems

While Proposition 1 provides a useful characterization of the agent's value, it still does not answer our main question: what are the robust strategies? This is because a strategy may be a best response to the Blackwell supremum $\bar{P}(P_1, \ldots, P_m)$, without being a robustly optimal strategy. In particular, the Blackwell supremum typically specifies a probability of zero for many signal realizations, so that any action is a best response to those signal realizations. But if we fix a strategy that chooses a particularly bad action after such a signal realization, it might be a best response for Nature to make it occur with positive probability. So we now turn to the question of finding the optimal robust strategies.

For any decision problem, one simple strategy that can always be used 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 particular actual joint experiment $P \in \mathcal{P}(P_1, \ldots, P_m)$. Theorem 1 shows that if the decision problem has binary action, this is indeed a robustly optimal strategy.

Theorem 1. *If $|A| = 2$, then*

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

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

$$\Lambda_{\overline{P}} = co\left(\Lambda_{P_1} \cup \cdots \cup \Lambda_{P_m}\right) \tag{1}$$

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 (1), an extreme point of $\Lambda_{\overline{P}}$ must belong to some Λ_{P_j}. Hence, we have

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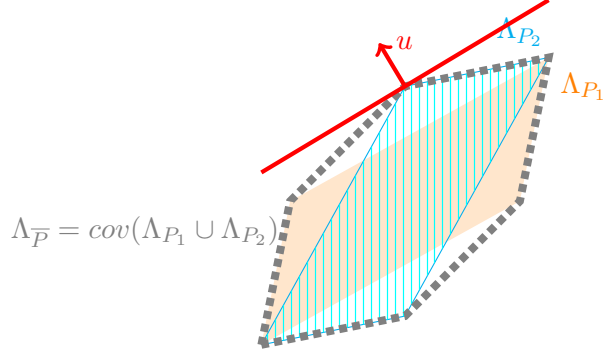
□

Figure 2: *The maximum is achieved at an extreme point*

The idea of Theorem 1 can be visualized in Figure 2 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.

6   General Decision Problems

Theorem 1 allows us to solve any binary action decision problem in a fairly simple way: finding the best marginal information source and best responding to it. For decision problem with more actions, a robustly optimal strategy may need to use multiple information sources.

To understand how optimal robust strategies work in general, we start our discussion in Section 6.1 and Section 6.2 with a simple class of decision problems: those that can be written as a composition of multiple binary action problems. For these problems, we show that the optimal strategy can be obtained by simply "adding up" the optimal strategies for the isolated binary action problems. Finally, in Section 6.3 and Section 6.4, we show that this simple class of decision problems is exhaustive—**any** finite action decision problem can be decomposed into binary action decision problems.

6.1   An Example

We start with an example which showcases how an agent can benefit from using information from multiple sources when she faces a more complex problem.

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:*

<div>

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

</div>

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.[5]

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

<table>
<tr><td colspan="3" align="center">P_1</td></tr>
<tr><td></td><td>$y_1 = 1$</td><td>$y_1 = 0$</td></tr>
<tr><td>$\theta = 1$</td><td>0.9</td><td>0.1</td></tr>
<tr><td>$\theta = 2$</td><td>0.5</td><td>0.5</td></tr>
</table>

<table>
<tr><td colspan="3" align="center">P_2</td></tr>
<tr><td></td><td>$y_2 = 1$</td><td>$y_2 = 0$</td></tr>
<tr><td>$\theta = 1$</td><td>0.5</td><td>0.5</td></tr>
<tr><td>$\theta = 2$</td><td>0.9</td><td>0.1</td></tr>
</table>

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

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.

The strategy constructed in Example 1 is in fact a robustly optimal strategy. There are two special structures of this example: 1. The action space is a product space of binary action spaces; 2. The payoff function can be written in an additively separable form of binary action problems. These two features enable us to find a robustly optimal strategy in a fairly simple way: find the robustly optimal strategy for each binary action problem via Theorem 1, and then "add them up". We will generalize and formalize this idea in the next section.

6.2   Composition of Decision Problems

Recall that we define a decision problem as $\Gamma \equiv (\Theta, \nu, A, \rho)$, which can simply be summarized by (A, u) where $u(\theta, a) = \nu(\theta)\rho(\theta, a)$. Since in this section we are going to alter the decision problems along the analyses, we use $V(P_1, ..., P_m; (A, u))$ to denote the agent's value in decision problem (A, u).

Definition 5. *Given a finite collection of decision problems $(A_1, u_1), ..., (A_n, u_n)$, their composition, denoted by $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$, is a decision problem with action space $A = (A_1 \times ... \times A_k)$ and $u(\theta, \mathbf{a}) = \sum_{\ell=1}^{k} u_\ell(\theta, a_\ell)$.*

Thus, the composition of decision problems is a single decision problem that has a specific additively separable structure. Notice that the decision problem in Example 1 is a composition

---

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

of two decision problems $A_1 = \{I_1, N_1\}$, $A_2 = \{I_2, N_2\}$ and $u_1(\cdot, I_1) = (2, -1)$, $u_1(\cdot, N_1) = (0, 0)$, $u_2(\cdot, I_2) = (-1, 2)$, $u_2(\cdot, N_2) = (0, 0)$.

Consider a finite collection of **binary** action problems, $(A_1, u_1), \ldots, (A_k, u_k)$, and consider the composition of these problems $(\bar{A}, \bar{U}) := \bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$. In this decision problem, a simple, robust strategy that an agent can always use is to choose exactly one experiment $Q_\ell \in \{P_1, \ldots, P_m\}$ for every binary problem ℓ and play the optimal strategy that uses that information alone, ignoring the signal realizations of all other experiments. Furthermore, by choosing this Q_ℓ optimally for each ℓ, regardless of the actual joint experiment $P \in \mathcal{P}(P_1, \ldots, P_m)$, the agent can achieve a total ex-ante utility of $\sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j, (A_\ell, u_\ell))$, which is typically strictly greater than $\max_{j=1,\ldots,m} V(P_j, (\bar{A}, \bar{U}))$. The following lemma shows that this is indeed the best that the agent can do in $(\bar{A}, \bar{U})$.

Lemma 4. *Let $(A_1, u_1), \ldots, (A_k, u_k)$ be a finite collection of **binary** action problems. Then*

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

Moreover, let $\sigma_\ell : \mathbf{Y} \to \Delta A_\ell$ be a robustly optimal strategy for decision problem (A_ℓ, u_ℓ). Then $\sigma : \mathbf{Y} \to \Delta(A_1 \times \ldots \times A_k)$ defined by

$$\sigma(y_1, \ldots, y_m) = \left(\sigma_\ell(y_1, \ldots, y_m)\right)_{\ell=1}^{k} \qquad \text{for all } y_1, \ldots, y_m \tag{2}$$

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

Proof. Using Proposition 1, $V\left(P_1, \ldots, P_m; \bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)\right) = V\left(\overline{P}(P_1, \ldots, P_m); \bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)\right)$. By Theorem 1, we then have:

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

To see the second statement, for any $P \in \mathcal{P}(P_1, \ldots, P_m)$, the agent's payoff from strategy σ is

$$\sum_{\theta \in \Theta} \sum_{y_1, \ldots, y_m} P(y_1, \ldots, y_m | \theta) \sum_{\ell=1}^{k} u_\ell(\theta, \sigma_\ell(y_1, \ldots, y_m)) = \sum_{\ell=1}^{k} \sum_{\theta \in \Theta} \sum_{y_1, \ldots, y_m} P(y_1, \ldots, y_m | \theta) u_\ell(\theta, \sigma_\ell(y_1, \ldots, y_m))$$

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

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

Since σ guarantees the maxmin value regardless of P, it is a robustly optimal strategy. $\square$

Lemma 4 provides a simple solution to any problem that can be expressed as a composition of binary action problem: For each binary action problem, one can derive a robustly optimal strategy by paying attention to the best marginal experiment and best responding to it. Then assembling these strategies as in (2) yields a robustly optimal strategy for the composite problem.

6.3 Decomposition of decision problems

In the previous section, we saw how a problem that is a composition of binary action problems can be solved by combining the solutions of each binary action problem. It is natural to ask then what problems can be decomposed into binary action problems. As we will see in the next section, it turns out that such a decomposition is possible for any decision problem. Before we get to that, we must define precisely what it means to decompose a decision problem and for that we need a notion of equivalence between decision problems.

For each decision problem (A, u), define the associated polyhedron containing all payoff vectors that are either achievable or weakly dominated by some mixed action:[6]

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

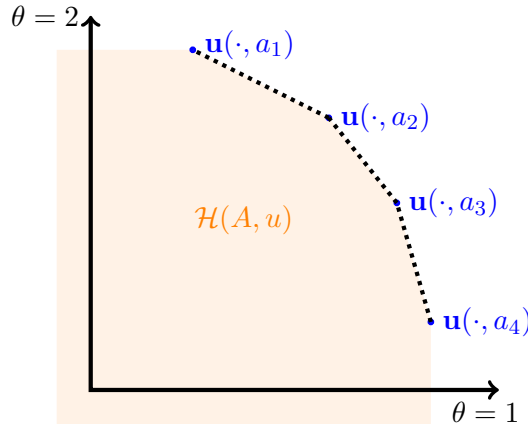An example of $\mathcal{H}(A, u)$ is depicted in Figure 3.

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

Whenever $\mathcal{H}(A', u') = \mathcal{H}(A, u)$, the two decision problems $\mathcal{H}(A', u')$ and $\mathcal{H}(A, u)$ give the same value under any information environment, so we call them *equivalent*.

Definition 6. *A decision problem (A, u) is equivalent to another decision problem (A', u') if*

$$\mathcal{H}(A', u') = \mathcal{H}(A, u).$$

Now we can define decomposition, the inverse operation of composition:

Definition 7. *A decision problem (A, u) admits a decomposition $\{(A_\ell, u_\ell)\}_{\ell=1}^k$ if (A, U) is equivalent to $\bigoplus_{\ell=1}^k (A_\ell, u_\ell)$.*

Example 2. *Consider two decision problems $A_1 = \{I_1, N_1\}$, $u_1(I_1) = (2, -1)$, $u_1(N_1) = (0, 0)$ and $A_2 = \{I_2, N_2\}$, $u_2(I_2) = (-1, 2)$, $u_2(N_2) = (0, 0)$. The associated polyhedra are the blue/red shaded areas in Figure 4(a). Their composition $(A_1, u_1) \bigoplus (A_2, u_2)$ consists of four actions, which are depicted in Figure 4(b).*

---

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

Now we consider a three-action decision problem $A = \{a_1, a_2, a_3\}$ with $u(a_1) = (-1, 2)$, $u(a_2) = (1, 1)$, and $u(a_3) = (2, -1)$. Notice that $\mathcal{H}(A, u) = \mathcal{H}((A_1, u_1) \bigoplus (A_2, u_2))$ as the shaded area in *Figure 4*(b), so (A, u) is equivalent to $(A_1, u_1) \bigoplus (A_2, u_2)$. Therefore, $(A_1, u_1), (A_2, u_2)$ is a decomposition of (A, u).

(a) Polyhedra induced by (A_1, u_1) and (A_2, u_2) (b) Polyhedron induced by $(A_1, u_1) \bigoplus (A_2, u_2)$
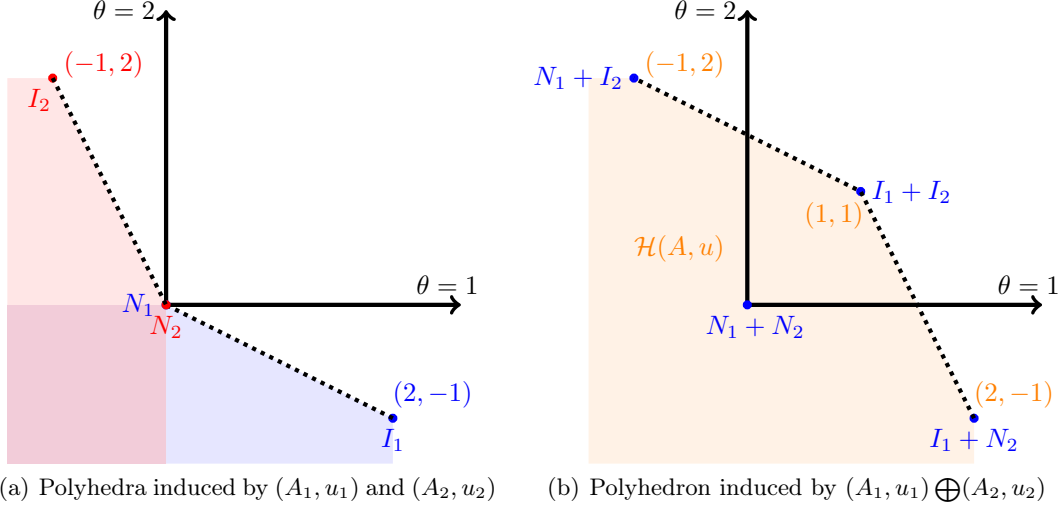
Figure 4

The analyses in the previous section give some hint on how to find robustly optimal strategies for general decision problems. If a given decision problem (A, u) admits a decomposition $(A_1, u_1), \ldots, (A_k, u_k)$ where each (A_ℓ, u_ℓ) is a **binary** action problem, then it is immediately clear by Lemma 4 and $\mathcal{H}(A, u) = \mathcal{H}\left(\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell) \right)$ that

$$V(P_1, \ldots P_m; (A, u)) = V\left(P_1, \ldots, P_m; \bigoplus_{\ell=1}^{k}(A_\ell, u_\ell) \right) = \sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j; (A_\ell, u_\ell)). \quad (3)$$

Moreover, the robustly optimal strategy for $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$, defined in (2), allows us to characterize robustly optimal strategies for (A, u) as by the following lemma.

Lemma 5. *Suppose (A, u) is equivalent to $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$, and $\sigma : \mathbf{Y} \to \Delta(A_1 \times \ldots \times A_k)$ is a robustly optimal strategy for $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$, then there exists $\sigma^* : \mathbf{Y} \to \Delta A$ such that*

$$u(\sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{k} u_\ell(\sigma_\ell(\mathbf{y})), \quad \text{for all } \mathbf{y} \in \mathbf{Y}.$$

Moreover, any such σ^* is a robustly optimal strategy for (A, u).

Proof. For each $\mathbf{y}$, $\sum_{\ell=1}^{k} u_\ell(\sigma_\ell(\mathbf{y})) \in \mathcal{H}\left(\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell) \right) = \mathcal{H}(A, u)$. So there exists $\sigma^*(\mathbf{y})$ such that $u(\sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{k} u_\ell(\sigma_\ell(\mathbf{y}))$. Moreover, since σ^* guarantees a higher value in (A, u) than σ in $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$, and $V(P_1, \ldots P_m; (A, u)) = V\left(P_1, \ldots, P_m; \bigoplus_{\ell=1}^{k}(A_\ell, u_\ell) \right)$, σ^* is a robustly optimal strategy for (A, u). $\square$

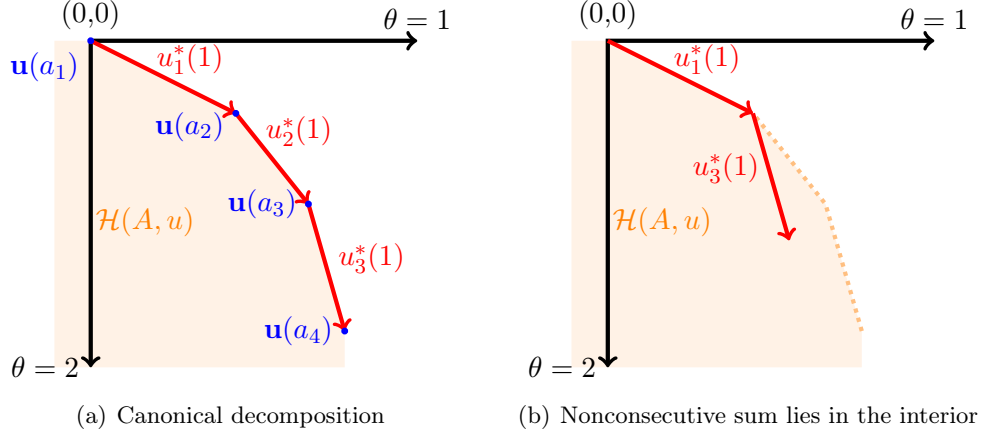(a) Canonical decomposition   (b) Nonconsecutive sum lies in the interior

Figure 5

If a decision problem (A, u) admits a decomposition into binary action problems, Lemma 4 and Lemma 5 characterize a set of robustly optimal strategies. However, it is not immediately clear what kind of decision problem admits a decomposition into binary action problems. Interestingly, we show by direct construction that, **any** decision problem admits a decomposition into binary action problems.

6.4   Canonical Decomposition

We are now ready to show that any decision problem can be decomposed into binary-action problems. Given an arbitrary decision problem (A, u), we start with some normalization to simplify exposition. First we remove all weakly*-dominated actions,[7] so that actions can be ordered as

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

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

Definition 8. *Given a decision problem (A, u), the **canonical decomposition** of (A, u) is the following collection of $n - 1$ binary action problems $(A_1^*, u_1^*), \ldots, (A_{n-1}^*, u_{n-1}^*)$:*

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

The canonical decomposition can be visualized in Figure 5 for an example with four actions. To see that a canonical decomposition is a decomposition, first notice that for any $i = 1, ..., n$, $u(a_i) = \sum_{\ell=1}^{i-1} u_\ell^*(1) + \sum_{\ell=i}^{n-1} u_\ell^*(0)$, so $\mathcal{H}(A, u) \subset \mathcal{H}\big(\bigoplus_{\ell=1}^{n-1}(A_\ell^*, u_\ell^*)\big)$. For the other direction, we need to show that for any $\boldsymbol{\delta} \in \{0, 1\}^{n-1}$, $\sum_{\ell=1}^{n-1} \boldsymbol{\delta}_\ell u_\ell^*(1) \in \mathcal{H}(A, u)$. The idea is that any nonconsecutive sum of $u_\ell^*(1)$ always lies in the interior of $\mathcal{H}(A, u)$, as illustrated in the example

---

[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.

in Figure 5(b).

Lemma 6. *The canonical decomposition is a decomposition.*

Proof. See Appendix A.2. □

Finally Lemma 4, Lemma 5, and Lemma 6 immediately imply Theorem 2.

Theorem 2. *Let $(A_1^*, u_1^*), \ldots, (A_{n-1}^*, u_{n-1}^*)$ be the canonical 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*

$$u(\sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{n-1} u_\ell^*(\sigma_\ell^*(\mathbf{y})), \quad \textit{for all } \mathbf{y}.$$

Moreover, any such σ^* is a robustly optimal strategy for (A, u).

Theorem 2 allows us to construct a robustly optimal strategy for any decision problem (A, u) in two steps: 1. For each (A_ℓ^*, u_ℓ^*), only one (the best) marginal experiment needs to be considered, and an robustly optimal strategy σ_ℓ^* only need to be measurable with respect to this experiment; 2. For each realization $\mathbf{y}$, pick a (mixed) action $\sigma(\mathbf{y}) \in \Delta(A)$ such that $u(\sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{n-1} u_\ell^*(\sigma_\ell^*(\mathbf{y}))$.

The theorem features two interesting corollaries.

Corollary 1. *For any decision problem (A, u) with the canonical decomposition (A_1^*, u_1^*), ..., (A_{n-1}^*, u_{n-1}^*) and any collection of marginal experiments $\{P_j\}_{j=1}^m$, 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 describes when an additional marginal experiment robustly improves the agent's value, which happens if and only if it outperforms all other marginal experiment in at least one of the canonically decomposed problem.

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 any n-action decision problem, it is not beneficial to use more than $n - 1$ experiments. Theorem 1 can be viewed as a special case where $n = 2$.

7 Extensions and Discussions

7.1 General state space

Our previous analyses focus on binary state decision problems. A natural question is whether those results can be extended into environments with more states. Unfortunately, when $|\Theta| > 3$,

Theorem 1 no longer holds as we will show in Example 3 below. Since Theorem 1 is the building blocks of other previous results, the same methodology doesn't work for general state space. Nevertheless, using the duality approach, we show that the central idea of "decomposition" extends, in the sense that the original decision problem can still be decomposed into a collection of "subproblems", and in each of these "subproblem", only one information source needs to be used. We will explain in the end of the section why this decomposition result is weaker than what we have for binary state environment.

Example 3. *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.*

<div>

P_X

	x_1	x_2
θ_1	1	0
θ_2	1	0
θ_3	0	1

P_Y

	y_1	y_2
θ_1	1	0
θ_2	0	1
θ_3	0	1

</div>

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.

The example illustrates that even in a binary action decision problem, the agent would like to use more than one information sources, which draws a contrast with Theorem 1.

In what follows, we allow for a more general state space Θ with $|\Theta| < \infty$. The notations we defined for binary state environment naturally generalize to general finite state environment, so we use the same notations without redefining them.

We first introduce the following definition of weak decomposition.

Definition 9. *A decision problem (A, u) admits a weak decomposition $\{(A_\ell, u_\ell)_{\ell=1}^k\}$ if*

$$\mathcal{H}(\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)) \subseteq \mathcal{H}(A, u)$$

That is, the polyhedron induced by a weak decomposition needs to be contained in the payoff polyhedron induced by (A, u). The following theorem characterizes the idea of decomposition

15

in general state environment.

Theorem 3. *Fix a decision problem (A, u) and Blackwell experiments $P_1, \ldots, P_m$. There exists a **weak** decomposition $((A_1^*, u_1^*), \ldots, (A_k^*, u_k^*))$ of (A, u) for which*

$$V(P_1, \ldots, P_m; (A, u)) = \sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j; (A_\ell^*, u_\ell^*)). \tag{4}$$

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

Theorem 3 can be seen as a generalization of Theorem 2 from the binary state environment. In particular, when the state space is binary, we showed in the previous section that by representing a decision problem equivalently as $\bigoplus_{\ell=1}^{n-1}(A_\ell^*, u_\ell^*)$ corresponding to the canonical decomposition, indeed the constructed robustly optimal strategy in the latter decision problem guarantees the payoff $\sum_{\ell=1}^{n-1} \max_{j=1,\ldots,n} V(P_j, (A_\ell^*, u_\ell^*))$. Moreover, similar to Theorem 2, one can derive a robustly optimal strategy based on a weak decomposition in two steps: 1. For each (A_ℓ^*, u_ℓ^*), the agent chooses a strategy that uses only the best marginal experiment and best responses to it, denoted by $\sigma_\ell^*(\mathbf{y})$; 2. For each realization $\mathbf{y}$, the agent picks a (mixed) action $\sigma(\mathbf{y}) \in \Delta(A)$ such that $u(\sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{k} u_\ell^*(\sigma_\ell^*(\mathbf{y}))$.

The decomposition in Theorem 3 is weaker than Theorem 2 in the following aspects: 1. the decomposition here may depend on the marginal experiments, while the canonical decomposition in Theorem 2 only depends on the decision problem; 2. the decomposition is weak so the payoff polyhedrons of the original problem and decomposed problem might not exactly coincide, though they lead to the same value; 3. in each subproblem under the weak decomposition, it might contain more than two actions, while in the canonical decomposition in Theorem 2, every subproblem is a binary action problem.

7.2 Large Datasets

Our main results in the binary state learning environment of Section 5 and Section 6 show that the robustly optimal strategy only uses a select-few information sources ignoring the signal realizations of all other information sources. Theorem 1 shows that when the decision problem involves just a binary choice, the robustly optimal strategy takes a particularly stark form where the decision maker pays attention to only a *single* information source. On the other hand, in general decision problems involving more actions, Theorem 2 demonstrates that robustly optimal strategy makes use of multiple information sources.

In Theorem 4 we show that the robustly optimal strategy again takes the form of paying attention to only a single information source in **any** decision problem when the state space is binary and the information sources are each individually sufficiently informative.

Let us again assume throughout this section that the state space is binary, i.e. $\theta \in \Theta := \{1, 2\}$. Consider any Blackwell experiment, P, with a corresponding signal space Y. Let P^t denote the Blackwell experiment with signal space Y^t, where $P^t(\cdot \mid \theta)$ consists of t i.i.d draws from $P(\cdot \mid \theta)$: for all $(y_1, \ldots, y_t) \in Y^t$ and each $\theta \in \{1, 2\}$,

$$P^t(y_1, \ldots, y_t \mid \theta) = \prod_{\tau=1}^{t} P(y_\tau \mid \theta).$$

16

Because we want to study the setting in which each information source is individually, very informative, we study the robustly optimal strategy of a decision maker when the decision maker has access to information sources $P_1^t, \ldots, P_m^t$ when t is large.

To state our result, let us define the quantity $w(P)$ for a given Blackwell experiment P with signal space Y:

$$w(P) = \max_{\theta \in \Theta} \min_{\nu \in \Delta(Y)} KL(\nu, P_\theta),$$

where $KL(\nu, \nu') = \sum_{y \in Y} \nu(y) \log \frac{\nu(y)}{\nu'(y)}$ is the Kullback-Leibler divergence between probability measures ν and ν'. Notice that $w(P)$ assigns a positive real number to each Blackwell experiment P and also coincides with the well-known Chernoff distance between the probability measures $P_0 \in \Delta(Y)$ and $P_1 \in \Delta(Y)$.

Theorem 4. *Suppose that $\Theta = \{0, 1\}$ and let $P_1, \ldots, P_m$ be a finite collection of Blackwell experiments. Suppose that*

$$w(P_1) > \max\{w(P_i) : i \neq 1\}.$$

Then given any decision problem (A, u), there exists some t^* such that for all $t \geq t^*$,

$$V(P_1^t, \ldots, P_m^t; (A, u)) = V(P_1^t; (A, u)).$$

Proof. Let $(A_1^*, u_1^*), \ldots, (A_{n-1}^*, u_{n-1}^*)$ be the canonical decomposition of (A, u). Then Moscarini and Smith [2002] show that for each (A_ℓ^*, u_ℓ^*) there exists some t_ℓ^* such that for all $t \geq t_\ell^*$,

$$V(P_1^t; (A_\ell^*, u_\ell^*)) > \max_{j \neq 1} V(P_j^t; (A_\ell^*, u_\ell^*)).$$

Now consider any $t \geq t^* := \max_\ell t_\ell^*$. By Theorem 2,

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

$\square$

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.

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.

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

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

Jacqueline Dinnes, Jonathan J Deeks, Ada Adriano, Sarah Berhane, Clare Davenport, Sabine Dittrich, Devy Emperador, Yemisi Takwoingi, Jane Cunningham, Sophie Beese, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of sars-cov-2 infection. *Cochrane Database of Systematic Reviews*, (8), 2020.

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.

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

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.

Giuseppe Moscarini and Lones Smith. The law of large demand for information. *Econometrica*, 70(6):2351–2366, 2002.

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

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

A   Appendix

A.1   Proof of Lemma 3

Proof. Consider a collection of experiments $\{P_j\}_{j=1}^m$ and their Blackwell supremum $\overline{P} : \Theta \to \Delta Z$. Since $\overline{P}$ Blackwell dominates P_j for all j, there exists garblings $g_j : Z \to \Delta(Y_m)$, $j = 1, ..., m$, such that for all $y_j \in Y_j$,

$$P_j(y_j|\theta) = \sum_{z \in Z} g_j(y_j|z)\overline{P}(z|\theta).$$

Construct the following experiment $\tilde{P} : \Theta \to \Delta(Y_1 \times \ldots \times Y_m)$:

$$\tilde{P}(y_1, \ldots, y_m|\theta) = \sum_{z \in Z} \prod_{j=1}^m g_j(y_j|z)\overline{P}(z|\theta). \tag{5}$$

Notice that $\sum_{-j} \tilde{P}(y_1, \ldots, y_m|\theta) = \sum_{z \in Z} g_j(y_j|z)\overline{P}(z|\theta) = P_j(y_j|\theta)$, so $\tilde{P} \in \mathcal{P}(P_1, \ldots, P_m)$. Moreover, (5) implies $\tilde{P}$ is a garbling of $\overline{P}$ so $\overline{P}$ Blackwell dominates $\tilde{P}$. From the definition of Blackwell supremum, $\tilde{P}$ Blackwell dominates $\overline{P}$, so $\overline{P} = \tilde{P} \in \mathcal{P}(P_1, \ldots, P_m)$. □

A.2   Proof of Lemma 6

Proof. We first show that $(A_1^*, u_1^*), \ldots, (A_{n-1}^*, u_{n-1}^*)$ is a weak decomposition. Suppose otherwise so that there exists some $(a_1^*, \ldots, a_n^*)$ for which $u^* := u(a_1^*) + \cdots + u(a_n^*) \notin \mathcal{H}(A, u)$. By Corollary 11.4.2 of Rockafellar [1970], there exists $\lambda \in \mathbb{R}^2 \setminus \{0\}$ such that

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

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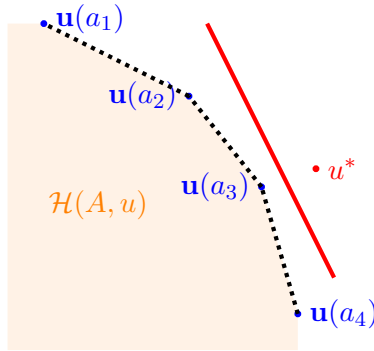

Figure 6

Given the canonical decomposition, for any $\ell' > \ell$,

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

Let $\ell^* = \min\{\ell : \lambda \cdot u_\ell(1, \cdot) \le 0\}$, where we use the convention that $\min \emptyset = n$. Then

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

$$= \sum_{\ell=1}^{\ell^*-1} \lambda \cdot (u_\ell^*(1) - u_\ell^*(a_\ell^*)) + \sum_{\ell=\ell^*}^{n-1} \lambda \cdot (u_\ell^*(0) - u_\ell^*(a_\ell^*)) \ge 0.$$

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

It remains to show that $(A_1^*, u_1^*), \ldots, (A_{n-1}^*, u_{n-1}^*)$ is an exact decomposition, but this is straightforward since it suffices to show that $\{u(a, \cdot) : a \in A\} \subseteq \mathcal{H}\left(\bigoplus_{\ell=1}^{n-1}(A_\ell^*, u_\ell^*)\right)$. Clearly this is the case since for every action $a_k \in A$, $u(a_k) = \sum_{\ell=1}^{k-1} u_\ell^*(1)$. $\qquad\square$

A.3   Proof of Theorem 3

We start from proving the following lemma that will be useful in the proof.

Lemma 7. *If $(A_\ell, u_\ell)_{\ell=1}^k$ is a weak decomposition of (A, u),*

$$V(P_1, \ldots, P_m; (A, u)) \ge \sum_{\ell=1}^k V(P_1, ..., P_m; (A_\ell, u_\ell)).$$

Proof. Suppose P^* is a solution to

$$\min_{P \in \mathcal{P}(P_1,...,P_m)} \max_{\sigma:\mathbf{Y}\to\Delta A} \sum_{\theta\in\Theta} \sum_{\mathbf{y}\in\mathbf{Y}} P(\mathbf{y} \mid \theta) u(\theta, \sigma(\mathbf{y})).$$

From minmax theorem,

$$V(P_1, \ldots, P_m; (A, u)) = \max_{\sigma:\mathbf{Y}\to\Delta A} \sum_{\theta\in\Theta} \sum_{\mathbf{y}\in\mathbf{Y}} P^*(\mathbf{y} \mid \theta) u(\theta, \sigma(\mathbf{y})).$$

For $\ell = 1, ..., k$, define σ_ℓ^* be a solution to

$$\max_{\sigma_\ell:\mathbf{Y}\to\Delta A_\ell} \sum_{\theta\in\Theta} \sum_{\mathbf{y}\in\mathbf{Y}} P^*(\mathbf{y} \mid \theta) u_\ell(\theta, \sigma_\ell(y)).$$

Now we have

$$\sum_{\ell=1}^k V(P_1, ..., P_m; (A_\ell, u_\ell)) = \sum_{\ell=1}^k \max_{\sigma_\ell:\mathbf{Y}\to\Delta A_\ell} \min_{P \in \mathcal{P}(P_1,...,P_m)} \sum_{\theta\in\Theta} \sum_{\mathbf{y}\in\mathbf{Y}} P(\mathbf{y} \mid \theta) u_\ell(\theta, \sigma_\ell(\mathbf{y}))$$

$$\le \sum_{\ell=1}^k \max_{\sigma_\ell:\mathbf{Y}\to\Delta A_\ell} \sum_{\theta\in\Theta} \sum_{\mathbf{y}\in\mathbf{Y}} P^*(\mathbf{y} \mid \theta) u_\ell(\theta, \sigma_\ell(\mathbf{y})) \qquad (7)$$

$$= \sum_{\ell=1}^k \sum_{\theta\in\Theta} \sum_{\mathbf{y}\in\mathbf{Y}} P^*(\mathbf{y} \mid \theta) u_\ell(\theta, \sigma_\ell^*(\mathbf{y}))$$

From the definition of weak decomposition, there exists $\hat{\sigma} : \mathbf{Y} \to A$ such that $u(\theta, \hat{\sigma}(\mathbf{y})) \ge$

20

$\sum_{\ell=1}^{k} u_\ell(\theta, \sigma_\ell^*(\mathbf{y}))$ for all $\mathbf{y}$ and θ.

Therefore,

$$
\begin{aligned}
\sum_{\ell=1}^{k} \sum_{\theta \in \Theta} \sum_{\mathbf{y} \in \mathbf{Y}} P^*(\mathbf{y} \mid \theta) u_\ell(\theta, \sigma_\ell^*(\mathbf{y})) &= \sum_{\theta \in \Theta} \sum_{\mathbf{y} \in \mathbf{Y}} P^*(\mathbf{y} \mid \theta) \sum_{\ell=1}^{k} u_\ell(\theta, \sigma_\ell^*(\mathbf{y})) \\
&\leq \sum_{\theta \in \Theta} \sum_{\mathbf{y} \in \mathbf{Y}} P^*(\mathbf{y} \mid \theta) u(\theta, \hat{\sigma}(\mathbf{y})) \\
&\leq \max_{\sigma: \mathbf{Y} \to \Delta A} \sum_{\theta \in \Theta} \sum_{\mathbf{y} \in \mathbf{Y}} P^*(\mathbf{y} \mid \theta) u(\theta, \sigma(\mathbf{y})) \\
&= V(P_1, ..., P_m; (A, u))
\end{aligned}
\tag{8}
$$

The statement of the lemma follows immediately from (7) and (8) $\qquad\square$

Proof of Theorem 3. Consider a robustly optimal strategy $\sigma^* : \mathbf{Y} \to \Delta(A)$ so

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

By considering the dual of the above linear program (Kantorovich dual), we obtain:

$$
V(P_1, \ldots, P_m, (A, u)) = \max_{\phi_1: \Theta \times Y_1 \to \mathbb{R}, \ldots, \phi_m: \Theta \times Y_m \to \mathbb{R}} \sum_{j=1}^{m} \sum_{\theta \in \Theta} \sum_{y_j \in Y_j} P_j(y_j \mid \theta) \phi_j(\theta, y_j)
$$

$$
\text{s.t.} \quad \sum_{j=1}^{m} \phi_j(\theta, y_j) \leq u(\theta, \sigma^*(\mathbf{y})) \text{ for all } (\theta, \mathbf{y}) \in \Theta \times \mathbf{Y}
\tag{9}
$$

Let $\phi_1^*, ..., \phi_m^*$ be the solution to the dual program. Define the collection of decision problems $\{(A_j, u_j)\}_{j=1}^{m}$ such that $A_j = Y_j$ and $u_j = \phi_j^*$. From the constraint (9), $((A_1, u_1), \ldots, (A_m, u_m))$ forms a weak decomposition of (A, u).

Moreover, in every "sub-problem" (A_j, u_j), by playing the strategy

$$
\sigma^j(y_1, ..., y_j, ..., y_m) = y_j \quad \text{for all } (y_1, ..., y_m),
$$

the agent achieves exactly a payoff of

$$
\begin{aligned}
\sum_{\theta \in \Theta} \sum_{(y_1, ..., y_m) \in \mathbf{Y}} P(y_1, \ldots, y_m \mid \theta) u_j(\theta, \sigma^j(y_1, \ldots, y_m)) &= \sum_{\theta \in \Theta} \sum_{y_j \in Y_j} P_j(y_j \mid \theta) u_j(\theta, \sigma^j(y_1, \ldots, y_m)) \\
&= \sum_{\theta \in \Theta} \sum_{y_j \in Y_j} P_j(y_j \mid \theta) \phi_j^*(\theta, y_j),
\end{aligned}
$$

which implies $V(P_1, ..., P_m; (A_j, u_j)) \geq \max_{\ell=1, ..., m} V(P_\ell; (A_j, u_j)) \geq \sum_{\theta \in \Theta} \sum_{y_j \in Y_j} P_j(y_j | \theta) \phi_j^*(\theta, y_j)$.

Summing over all $j = 1, ..., m$ we have

$$
\sum_{j=1}^{m} V(P_1, ..., P_m; (A_j, u_j)) \geq \sum_{j=1}^{m} \max_{\ell=1, ..., m} V(P_\ell; (A_j, u_j)) \geq \sum_{j=1}^{m} \sum_{\theta \in \Theta} \sum_{y_j \in Y_j} P_j(y_j \mid \theta) \phi_j^*(\theta, y_j).
\tag{10}
$$

Now from Lemma 7, equation (10), and $\{\phi_j^*\}_{j=1}^{m}$ being the solution to the dual program, we

have

$$V(P_1, \ldots, P_m; (A, u)) \geq \sum_{j=1}^{m} V(P_1, ..., P_m; (A_j, u_j))$$

$$\geq \sum_{j=1}^{m} \max_{\ell=1,\ldots,m} V(P_\ell; (A_j, u_j))$$

$$\geq \sum_{j=1}^{m} \sum_{\theta \in \Theta} \sum_{y_j \in Y_j} P_j(y_j \mid \theta) \phi_j^*(\theta, y_j)$$

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

which implies $V(P_1, \ldots, P_m; (A, u)) = \sum_{j=1}^{m} \max_{\ell=1,\ldots,m} V(P_\ell; (A_j, u_j))$.

$\square$