Robust Merging of Information[*]

Henrique de Oliveira Yuhta Ishii Xiao Lin

February 13, 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.

[*]We are grateful to Nageeb Ali, Marc Henry, Pietro Ortoleva, and Shamim Sinnar for valuable comments.

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]

In order to correctly interpret the joint pair of results from the two tests, knowledge of the 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 two tests is much more scarce.[3] In the presence of such limited information about these correlations, how does a decision maker make use of the results of both tests?

To answer this question, consider a health authority who must use the results of these two tests to make a decision of whether to recommend quarantine or not to a potentially infected patient. Due to limited information about the correlations of these two tests, the authority evaluates policies according to their worse case guarantee among all possible correlations. Our main results imply that it is *never* beneficial to base this quarantine decision on the results of both tests. In this regard, conducting both tests is never helpful in decision problems involving a simple choice between two actions. In contrast, if the decision problem involves more than three actions, such as designing a full treatment plan, then using the results from both tests may robustly improve the agent's payoff.

Aside from the COVID example, there are many other settings in which an agent makes decisions based on multiple information sources where data regarding the correlations may be limited. For example, a consumer can acquire information from different review platforms before buying a product, a graduate student often seeks advice from multiple faculty members when pursuing a new project, and an investor often solicits the recommendations of different financial consultants. We study a general model where an agent confronts a decision problem after observing signals generated from distinct information sources. To study the impact of robustness concerns regarding correlations, 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 study the robustly optimal strategy of such an agent who chooses a decision plan that maximizes expected payoff with respect to the worst possible correlation.

Our main results characterize the set of all 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 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

[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].

consists of decomposing a decision problem into subproblems, each requiring the use of a single information source. This shows the precise (and restricted) 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.

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.

Another thread of related literature 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 mentioned papers consider robust optimality using an interim approach, while we study the decision maker's robustly optimal ex-ante decision plan.

3 Model

An agent faces a decision problem $\Gamma \equiv (\Theta, \nu, A, \varrho)$ with finite state space Θ, prior $\nu \in \Delta\Theta$, finite action space A, and utility function $\varrho : \Theta \times A \to \mathbb{R}$. To later simplify notation, define $u : A \to \mathbb{R}^{|\Theta|}$ such that $u(a) = \big(\nu(\theta)\varrho(\theta, a)\big)_{\theta\in\Theta}$. Since (A, u) is the only relevant part, we simply call (A, u) a decision problem.[4]

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

[4] While it is redundant to include A, it helps in simplifying notations later.

following set of joint experiments:

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

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):[5]

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

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 Payoff Polyhedron

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]

$$\mathcal{H}(A, u) = co\{u(a) : a \in A\} - \mathbb{R}^{|\Theta|} \subset \mathbb{R}^{|\Theta|}.$$

The following figure depicts an example of $\mathcal{H}(A, u)$ when $|\Theta| = 2$.

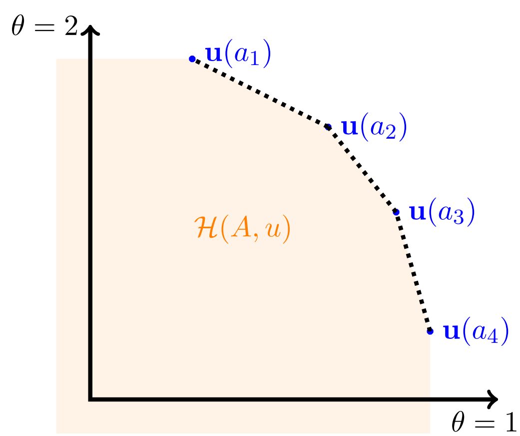


Figure 1: *The shaded area represents* $\mathcal{H}(A, u)$

We define the following order on decision problems.

[5]Whenever there is no confusion about the decision problem, we omit (A, u) from the argument of V.

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

Definition 1. *A decision problem (A, u) **contains** another decision problem (A', u') if*

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

*Two decision problems (A, u) and (A', u') are **equivalent** if they contain each other.*

3.2 Composition of Decision Problems

One of the key ideas that we will use in our analysis is how a decision problem can be broken down in a way that allows us to find robust strategies. In order to do this, we now define a central operation of this paper: the composition of decision problems.

Definition 2. *Given a finite collection of decision problem $(A_1, u_1), ..., (A_n, u_n)$, their composition, denoted by $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$, is a decision problem with action space $A = (A_1 \times \ldots \times A_k)$ and $u(\mathbf{a}) = \sum_{\ell=1}^{k} u_\ell(a_\ell)$.*

Thus, the composition of decision problems is a single decision problem that has a specific additively separable structure. We will see later that this structure facilitates the search for robust strategies.

With composition of decision problems defined, we can define its inverse operation:

Definition 3. *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 1. *Consider two decision problems $A_1 = \{I_1, N_1\}$, $u_1(I_1) = (2, 0)$, $u_1(N_1) = (0, 1)$ and $A_2 = \{I_2, N_2\}$, $u_2(I_2) = (0, 2)$, $u_2(N_2) = (1, 0)$. The associated polyhedra are the blue/red shaded areas in Figure 2(a). Their composition $(A_1, u_1) \bigoplus (A_2, u_2)$ consists of four actions, which are depicted in Figure 2(b).*

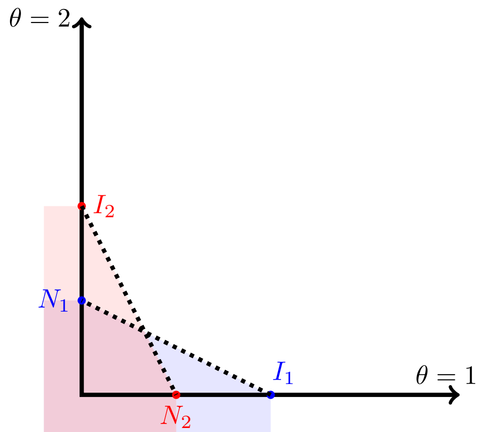
Now we consider a three-action decision problem $A = \{a_1, a_2, a_3\}$ with $u(a_1) = (3, 0), u(a_2) = (2, 2)$, and $u(a_3) = (0, 3)$. Notice that $\mathcal{H}(A, u) = \mathcal{H}((A_1, u_1) \bigoplus (A_2, u_2))$ as the shaded area in Figure 2(b), so (A, u) is equivalent to $(A_1, u_1) \bigoplus (A_2, u_2)$. Therefore, $(A_1, u_1), (A_2, u_2)$ is a decomposition of (A, u).

4 Binary State Decision Problems

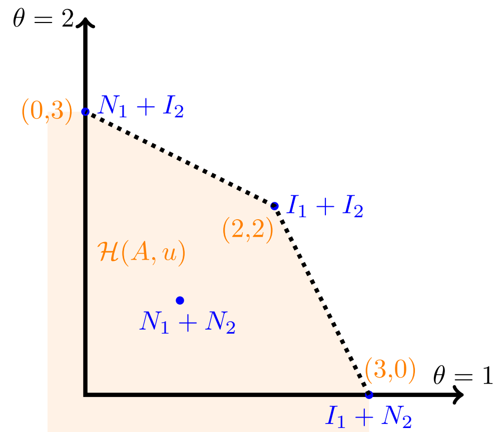
In this section, we study the agent's decision problem while restricting attention to binary state decision problems (i.e. $|\Theta| = 2$). Binary state environments yield nice properties such as a natural monotone ordering on the set of undominated actions, and the existence of the Blackwell minimum element in the set $\mathcal{P}(P_1, ..., P_m)$. These properties allow us to provide a clean and simple characterization to the agent's robustly optimal strategy. The insights from the binary state setup will be extended to general state decision problems in Section 5.

4.1 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.



(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 2

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

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

That is, the value of the agent's maxmin problem equals the value of a problem where Nature chooses an experiment in set $\mathcal{P}(P_1, \dots, P_m)$ 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 $\mathcal{P}(P_1, \dots, P_m)$, it would solve Nature's problem—any other information structure would yield a higher value for the agent. Notice that every experiment in $\mathcal{P}(P_1, \dots, P_m)$ must be more informative than every P_j. By Proposition 16 in Bertschinger and Rauh [2014], the Blackwell order defines a lattice on the set of experiments under binary state. In particular, 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, denoted by $\overline{P}(P_1, \dots, P_m)$, can be expressed as a joint distribution with marginals $P_1, \dots, P_m$. This is proved in the following lemma.

Lemma 1. *For any collection of experiments* $\{P_j\}_{j=1}^m$, $\overline{P}(P_1, \dots, P_m) \in \mathcal{P}(P_1, \dots, P_m)$.

Proof. See Appendix A.1. ☐

Immediately from the lemma, we have the following proposition.

Proposition 1. *For any decision problem* (A, u),

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

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

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

4.2 Binary Action 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 an 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.

Consider any decision problem (A, u). 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 $\max_{j=1,\ldots,n} V(P_j; (A, u))$, regardless of the particular actual joint experiment $P \in \mathcal{P}(P_1, \ldots, P_m)$. Theorem 1 shows that if (A, u) is a binary action problem, this is indeed an optimal robust strategy.

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

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

Proof. By Proposition 1, it suffices to show that $V(\overline{P}(P_1, ..., P_m)) = \max_{j=1,\ldots,m} V(P_j)$.

For any experiment $P : \Theta \to \Delta Y$, let

$$\Lambda_P = \left\{ \lambda : \Theta \to \Delta A | \lambda(a|\theta) = \sum_y \sigma(a|y) P(y|\theta) \right\} \subset \mathbb{R}^2.$$

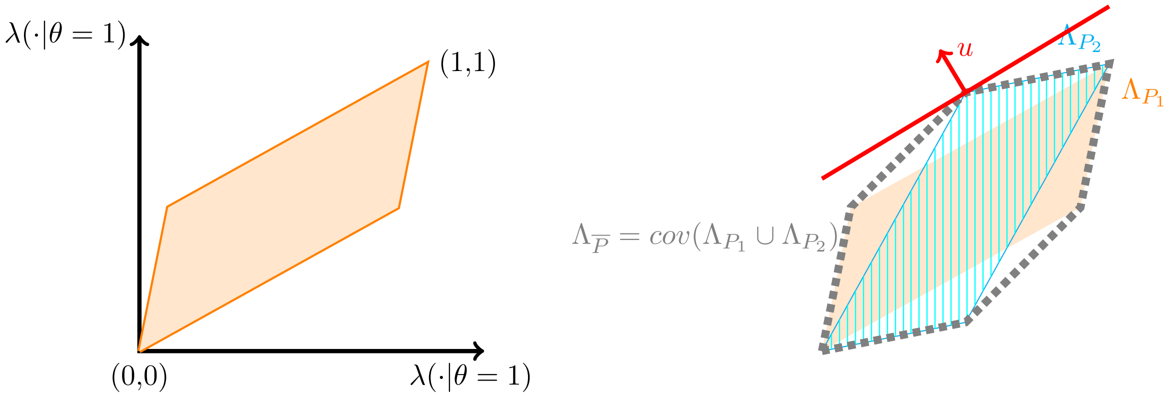
The set belongs to $\mathbb{R}^2$ because $|A| = 2$ so $\lambda(\cdot|\theta)$ can be represented by a number in $[0, 1]$. One can interpret Λ_P as the feasible state-action distribution generated by experiment P. Geometrically, Λ_P is a Zonotope, as depicted in Figure 3(a).

By Proposition 16 in Bertschinger and Rauh [2014], an experiment $\overline{P}$ is the Blackwell supremum of $P_1, \ldots, P_m$ if and only if

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

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

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



(a) Λ_P represented as a Zonotope

(b) Graphical representation of Λ_{P_1}, Λ_{P_2}, and $\Lambda_{\overline{P}}$

Figure 3

□

The idea of Theorem 1 can be visualized in Figure 3(b) 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.

4.3 General Decision Problems

In light of Theorem 1, a natural question arises of whether $\max_{j=1,\ldots,m} V(P_j;(A,u))$ is always the agent's optimal value for all decision problems (A,u). The example below shows that in general this does not hold.

Example 2. *Suppose there are two assets whose outputs depend on an unknown state $\theta \in \{0,1\}$. The output vectors are given by $X_1 = (2,-1)$ and $X_2 = (-1,2)$ where the first element denotes the output from state 1. An investor can choose whether or not to invest one unit in each of the assets, and her payoff is the sum of outputs from each asset she invested. Notice that this decision problem is exactly the composition $(A_1,u_1)\bigoplus(A_2,u_2)$, where (A_ℓ,u_ℓ) corresponds to the decision problem of investing asset ℓ.*

The investor holds equal prior on states and has access to two experiments P_1, P_2:

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

P_1

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

P_2

By paying attention to one experiment, for example P_1, the optimal strategy is to invest in both asset if $y_1 = 1$ and only asset 2 if $y_1 = 0$. The expected payoff from this strategy is thus $\frac{1}{2}[0.9 \cdot 1 + 0.1 \cdot (-1)] + \frac{1}{2}[0.5 \cdot 1 + 0.5 \cdot 2] = 1.15$.

Instead suppose the investor uses the following strategy that makes use of the information of both experiments. She considers investment decision of the two assets separately: in deciding whether to invest in asset 1, she only looks at experiment 1 and invests iff $y_1 = 1$. *Similarly when deciding whether to invest in asset 2, she looks at experiment 2 and invests iff* $y_2 = 0$. *This strategy can be written as:*

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

This strategy guarantees an expected output of $\frac{1}{2}[0.9 \cdot 2 + 0.1 \cdot 0] + \frac{1}{2}[0.5 \cdot (-1) + 0.5 \cdot 0] = 0.65$ *from each asset regardless of the correlations, which gives a total output of* $1.3 > 1.15$.

The strategy constructed in Example 2 is in fact a robustly optimal strategy. A special structure here is that the decision problem is a composition of two binary action decision problems. This allows the agent to optimally use only one for each asset, which guarantees robustness. But different experiments are used for different assets, which provides a payoff greater than $\max_{j=1,\dots,m} V(P_j, (A, u))$. This idea extends to composition of any finite collection of binary action problems.

4.3.1 Composition of binary action problems

Slightly more generally, consider a finite collection of **binary** action problems, $(A_1, u_1), \dots, (A_k, u_k)$, 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 $Q_\ell \in \{P_1, \dots, P_m\}$ for every task 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 task ℓ, regardless of the actual joint experiment $P \in \mathcal{P}(P_1, \dots, P_m)$, the agent can achieve a total ex-ante utility of $\sum_{\ell=1}^{k} \max_{j=1,\dots,m} V(P_j, (A_\ell, u_\ell))$, which is typically strictly greater than $\max_{j=1,\dots,m} V(P_j, (\bar{A}, \bar{U}))$. Moreover, the following lemma shows that indeed this is the best that the agent can do in $(\bar{A}, \bar{U})$.

Lemma 2. *Let* $(A_1, u_1), \dots, (A_k, u_k)$ *be a finite collection of* ***binary*** *action problems. Then*

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

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

$$\sigma(y_1, \dots, y_m) = \Big(\sigma_\ell(y_1, \dots, y_m)\Big)_{\ell=1}^{k} \quad \text{for all } y_1, \dots, y_m \tag{2}$$

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

Proof. Using Proposition 1, $V\left(P_1, \dots, P_m; \bigoplus_{\ell=1}^{k} (A_\ell, u_\ell)\right) = V\left(\overline{P}(P_1, \dots, P_m); \bigoplus_{\ell=1}^{k} (A_\ell, u_\ell)\right)$.

By Theorem 1, we then have:

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

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

$$\begin{aligned}\sum_{\theta\in\Theta}\sum_{y_1,\dots,y_m}P(y_1,...,y_m|\theta)\sum_{\ell=1}^{k}u_\ell(\sigma_\ell(y_1,...,y_m);\theta)&=\sum_{\ell=1}^{k}\sum_{\theta\in\Theta}\sum_{y_1,\dots,y_m}P(y_1,...,y_m|\theta)u_\ell(\sigma_\ell(y_1,...,y_m);\theta)\\&\geq\sum_{\ell=1}^{k}V(P_1,...,P_m;(A_\ell,u_\ell))\\&=V\left(P_1,\dots,P_m;\bigoplus_{\ell=1}^{k}(A_\ell,u_\ell)\right)\end{aligned}$$

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

Lemma 2 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.

4.3.2 Decomposition into binary action problems

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 $(A_1,u_1),\dots,(A_k,u_k)$ where each (A_ℓ,u_ℓ) is a **binary** action problem, then it is immediately clear by Lemma 2 and $\mathcal{H}(A,u)=\mathcal{H}\left(\bigoplus_{\ell=1}^{k}(A_\ell,u_\ell)\right)$ that

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

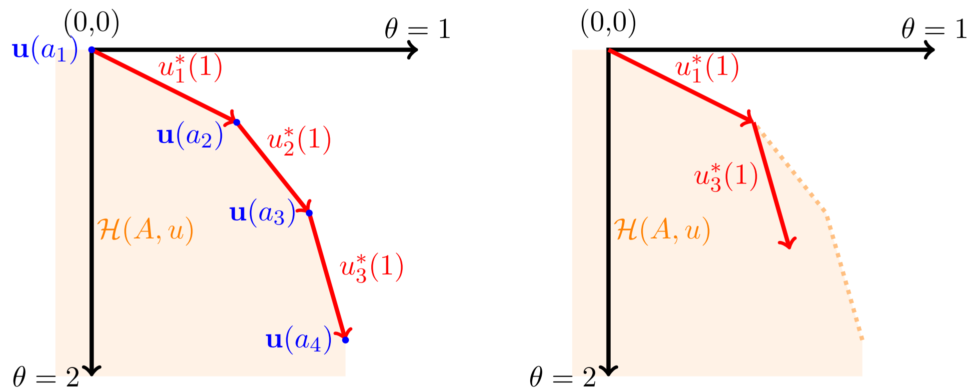
Moreover, the robustly optimal strategy for $\bigoplus_{\ell=1}^{k}(A_\ell,u_\ell)$, defined in (2), allows us to characterize robustly optimal strategies for (A,u) as by the following lemma.

Lemma 3. *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*

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

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



(a) Canonical decomposition

(b) Nonconsecutive sum lies in the interior

Figure 4

σ in $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$, and $V(P_1, \dots P_m; (A,u)) = V\left(P_1, \dots, P_m; \bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)\right)$, σ^* is a robustly optimal strategy for (A,u). $\square$

If a decision problem (A,u) admits a decomposition into binary action problems, Lemma 2 and Lemma 3 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.

Given any 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

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

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

Definition 4. *Given a decision problem (A,u), the **canonical decomposition** of (A,u) is the following collection of $n-1$ binary action problems $(A_1^*, u_1^*), \dots, (A_{n-1}^*, u_{n-1}^*)$:*

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

The canonical decomposition can be visualized in Figure 4 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 in Figure 4(b).

Lemma 4. *The canonical decomposition is a decomposition.*

[7] An action $a \in A$ is weakly*-dominated if there exists $\alpha \in \Delta A$ such that $u(a) \leq u(\alpha)$.

Proof. See Appendix A.2. □

Finally Lemma 2, Lemma 3, and Lemma 4 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*

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

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^*), 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 $n = 2$.

5 General State Decision Problems

We now turn our attention to general decision environments beyond the simple binary state decision setup. The building blocks of the binary state results are Theorem 1 and the idea of decomposition. The former shows that with binary action, the optimal strategy is to simply use one marginal experiment, and the latter allows us to tackle any binary state decision problem by decomposing it into binary action ones. Unfortunately, when $|\Theta| > 3$, Theorem 1 no longer holds, as can be seen in Example 3. Nevertheless, the idea of decomposition remains. We show that a robustly optimal strategy takes the following form: there exists some **weak** decomposition of the decision problem into k distinct decision problems $A_1, \ldots, A_k$, each of which we call a task. In each of these tasks, A_ℓ, the agent chooses an optimal strategy, σ_ℓ, that uses only the information from one single experiment, say Y_{b_ℓ}, among those experiments available to the agent, $Y_1, \ldots, Y_m$. The agent then aggregates these "decomposed strategies" into a strategy for the original decision problem in a natural way. To state our formal results, we first need a modified definition of a decopmosition.

Definition 5. *A decision problem (A, u) admits a **weak** decomposition $\{(A_\ell, u_\ell)\}_{\ell=1}^k$ if (A, U) contains $\bigoplus_{\ell=1}^k (A_\ell, u_\ell)$.*

Our main theorem of Section 5 is the following.

Theorem 3. *Fix a decision problem (A, u) and Blackwell experiments $P_1, \ldots, P_m$. Then for all **weak** decompositions $((A_1, u_1), \ldots, (A_k, u_k))$ of (A, u),*

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

*Moreover, there exists a **weak** decomposition $((A_1^*, u_1^*), \ldots, (A_k^*, u_k^*))$ of (A, u) for which*

$$V(P_1, \ldots, P_m; (A, u)) = \sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j; (A_\ell^*, u_\ell^*)). \tag{5}$$

Notice that (5) of 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, the first part of the above theorem clarifies that this is indeed the *optimal* decomposition in the sense that for all other weak decompositions $((A_1, u_1), \ldots, (A_k, u_k))$ of (A, u),

$$\sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j, (A_\ell, u_\ell)) \leq \sum_{\ell=1}^{k} \max_{j=1,\ldots,m} V(P_j, (A_\ell^*, u_\ell^*)).$$

However, a key difference with the binary state setting is that the optimal **weak** decomposition that underlies (5) need not be a decomposition, i.e. $\mathcal{H}\left(\bigoplus_{\ell=1}^k (A_\ell, u_\ell)\right) \subsetneq \mathcal{H}(A, u)$.

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.*

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.

P_X

	x_1	x_2
$\theta = 1$	1	0
$\theta = 2$	1	0
$\theta = 3$	0	1

P_Y

	y_1	y_2
$\theta = 1$	1	0
$\theta = 2$	0	1
$\theta = 3$	0	1

Table 1

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

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

Suppose that the agent begins with a prior that is uniform. Notice that $V(P_X, P_Y, (A, u)) = \frac{2}{3}$, which is the perfect information payoff.

What is the ***optimal weak decomposition*** *in this decision problem? Define*

$$A_1 := \{0, 1\}, u_1(0, \cdot) = (0, 0, 0), u_1(1, \cdot) = (0, -1, 1);$$
$$A_2 := \{0, 1\}, u_2(0, \cdot) = (0, 0, 0), u_2(1, \cdot) = (1, -1, 0).$$

It is easy to see that (A_1, u_1) and (A_2, u_2) form a weak decomposition of (A, u) that is ***not*** *a decomposition. To see that this is indeed an optimal weak decomposition, note that P_X is better than P_Y in decision problem (A_1, u_1) while P_Y is better than P_X in decision problem (A_2, u_2). Moreover, $V(P_X; (A_1, u_1)) = V(P_Y; (A_2, u_2)) = \frac{1}{3}$ and thus,*

$$V(P_X, P_Y; (A, u)) = V(P_X; (A_1, u_1)) + V(P_Y; (A_2, u_2)) = \frac{2}{3}.$$

As perhaps Theorem 3 already suggests, there is a way to characterize robustly optimal strategies in terms of weak decompositions. We have the following characterization of all robustly optimal strategies.

Corollary 3. *Fix (A, u) a decision problem and $P_1, \ldots, P_m$ experiments. Let $\sigma : \mathbf{Y} \to \Delta(A)$ be a strategy. Then the following are equivalent:*

1. *σ is robustly optimal;*
2. *There exists some weak decomposition $((A_1, u_1), \ldots, (A_k, u_k))$ of (A, u) and some $\sigma_\ell \in B^*(P_1, \ldots, P_m, (A_\ell, u_\ell))$ for each $\ell = 1, 2, \ldots, m$ such that*

$$u(\sigma(\mathbf{y})) \geq \sum_{\ell=1}^{k} u_\ell(\sigma_\ell(\mathbf{y}))$$

for all $\mathbf{y} \in \mathbf{Y}$.

5.1 Proof of Theorem 3

Again the same ideas that are central to the binary state analysis are at the heart of Theorem 3. Let $(A_1, u_1), \dots, (A_k, u_k)$ be a finite collection of decision problems and consider again the richer decision problem formed by composing these decision problems, $\bigoplus_{\ell=1}^{k}(A_k, u_k)$.

Again this decision problem admits a simple and natural strategy that is robust in the sense that it attains the same constant ex-ante payoff regardless of the actual correlations between the available experiments $P_1, \dots, P_m$. For every task A_ℓ, choose an optimal pure strategy, σ_ℓ that is optimal among all strategies that only uses the signal realizations of exactly one experiment. Let $\sigma = (\sigma_1, \dots, \sigma_k)$ be the strategy in $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$ that plays σ_ℓ in each task (A_ℓ, u_ℓ).

Because each agent perfectly understands the signal distributions of each marginal experiment, such a strategy attains a payoff of $\max_{j=1,\dots,m} V(P_j, (A_\ell, u_\ell))$ in each task (A_ℓ, u_ℓ). Moreover, because in $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$, payoffs are additive separable across tasks, this implies that σ achieves a payoff of $\sum_{\ell=1}^{k} \max_{j=1,\dots,m} V(P_j, (A_\ell, u_\ell))$ for all $P \in \mathcal{P}(P_1, \dots, P_m)$. This is summarized in the following claim.

Claim 1. *The strategy σ achieves the ex-ante payoff of $\sum_{\ell=1}^{k} \max_{j=1,\dots,m} V(P_j, A_\ell)$ in the decision problem $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$ for all $P \in \mathcal{P}(P_1, \dots, P_n)$.*

How do we make use of this observation for the construction of a robust strategy in the original decision problem (A, u)? If $(A_1, u_1), \dots, (A_k, u_k)$ is a **weak** decomposition of (A, u) so that the decision problem $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$ is contained in the decision problem (A, u), then the robust strategy σ constructed above in Claim 1 for the decision problem $\bigoplus_{\ell=1}^{k}(A_\ell, u_\ell)$ is weakly dominated by some strategy σ^* in the decision problem (A, u). Constructing such a strategy is simple since for every signal realization $\mathbf{y}$, we choose some $\sigma^*(\mathbf{y})$ such that

$$u(\sigma^*(\mathbf{y})) \geq \sum_{\ell=1}^{k} u_\ell(\sigma(\mathbf{y})).$$

Because σ^* weakly dominates σ, σ^* guarantees at least $\sum_{\ell=1}^{k} \max_{j=1,\dots,m} V(P_j, (A_\ell, u_\ell))$ for all possible joint experiments. This proves the following claim and the first part of Theorem 3.

Claim 2. *The strategy σ^* guarantees at least a payoff of $\sum_{\ell=1}^{k} \max_{j=1,\dots,m} V(P_j, (A_\ell, u_\ell))$ in the decision problem (A, u) for every $P \in \mathcal{P}(P_1, \dots, P_n)$. Consequently,*

$$V(P_1, \dots, P_m, (A, u)) \geq \sum_{\ell=1}^{k} \max_{j=1,\dots,m} V(P_j, (A_\ell, u_\ell)).$$

5.1.1 Completing the Proof of Theorem 3

To complete the proof of Theorem 3, it remains to show that there exists a weak decomposition that makes Equation 4 an equality. We now prove that this is indeed the case.

A standard duality argument from linear programming provides the key argument. Because of the first part of the theorem, it is sufficient to show that there exists some decomposition $((A_1, u_1), \dots, (A_k, u_k))$ of (A, u) such that

$$\sum_{\ell=1,\dots,k} \max_{j=1,\dots,m} V(P_j, (A_\ell, u_\ell)) \geq V(P_1, \dots, P_m, (A, u)).$$

To see this, consider a robustly optimal strategy $\sigma^* : \mathbf{Y} \to \Delta(A)$ such that

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

Then by considering the dual of the above linear program, we obtain:

$$\begin{aligned} V(P_1, \ldots, P_m, (A, u)) = \max_{\phi_1 : Y_1 \to \mathbb{R}^{|\Theta|}, \ldots, \phi_m : Y_m \to \mathbb{R}^{|\Theta|}} & \sum_{j=1}^{m} \sum_{\theta \in \Theta} \sum_{y_j \in Y_j} P_j(y_j \mid \theta) \phi_j(y_j, \theta) \\ \text{s.t.} \ & \sum_{j=1}^{m} \phi_j(y_j, \theta) \leq u(\sigma^*(\mathbf{y}), \theta) \text{ for all } (\theta, \mathbf{y}) \in \Theta \times \mathbf{Y}. \end{aligned} \tag{6}$$

Let $\phi_1^*, \ldots, \phi_m^*$ be solutions to the above optimization problem. Define (A_ℓ, u_ℓ) to be the decision problem consisting of actions indexed by $y_\ell \in Y_\ell$ where action a_{y_ℓ} yields the utility $u_\ell(a_{y_\ell}, \theta) = \phi_\ell^*(y_\ell, \theta)$. Because $\phi_1^*, \ldots, \phi_m^*$ satisfy the constraints (6), it is immediate that $((A_1, u_1), \ldots, (A_k, u_k))$ form a weak decomposition of (A, u).

Moreover, in every task (A_ℓ, u_ℓ), because the strategy that plays action a_{y_ℓ} whenever the realized signal in experiment Y_ℓ is y_ℓ exactly achieves a payoff of $\sum_{\theta \in \Theta} \sum_{y_\ell \in Y_\ell} P_\ell(y_\ell \mid \theta) \phi_\ell^*(y_\ell, \theta)$,

$$\max_{j=1,\ldots,m} V(P_j, (A_\ell, u_\ell)) \geq V(P_\ell, (A_\ell, u_\ell)) \geq \sum_{\theta \in \Theta} \sum_{y_\ell \in Y_\ell} P_\ell(y_\ell \mid \theta) \phi_\ell^*(y_\ell, \theta).$$

Summing across all $\ell = 1, 2, \ldots, n$,

$$\sum_{\ell=1}^{m} \max_{j=1,\ldots,m} V(P_j, (A_\ell, u_\ell)) \geq \sum_{\ell=1}^{m} \sum_{\theta \in \Theta} \sum_{y_\ell \in Y_\ell} P_\ell(y_\ell \mid \theta) \phi_\ell^*(y_\ell, \theta) = V(P_1, \ldots, P_m, (A, u)).$$

This completes the proof. □

References

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.

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.

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 1

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

$$\tilde{P}(y_1, \ldots, y_m|\theta) = \sum_{z \in Z} \prod_{j=1}^{m} g_j(y_j|z)\overline{P}(z|\theta). \tag{7}$$

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, (7) 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 4

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

$$\lambda \cdot u^* > \sup_{v \in \mathcal{H}(A,u)} \lambda \cdot v. \tag{8}$$

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

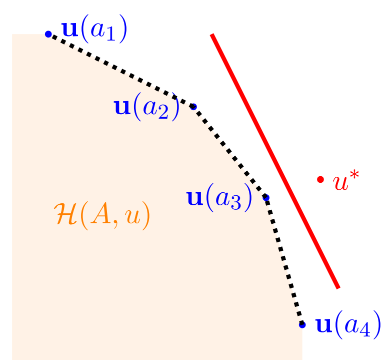


Figure 5

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

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

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

$$\begin{aligned} \lambda \cdot u(a_{\ell^*}, \cdot) - \lambda \cdot u^* &= \sum_{\ell=1}^{\ell^*-1} \lambda \cdot u_\ell^*(1) - \sum_{\ell=1}^{n-1} \lambda \cdot u_\ell^*(a_\ell^*) \\ &= \sum_{\ell=1}^{\ell^*-1} \lambda \cdot (u_\ell^*(1) - u_\ell^*(a_\ell^*)) + \sum_{\ell=\ell^*}^{n-1} \lambda \cdot (u_\ell^*(0) - u_\ell^*(a_\ell^*)) \geq 0. \end{aligned}$$

But $u(a_{\ell^*}, \cdot) \in \mathcal{H}(A, u)$, which contradicts Inequality (8).

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 $a_k \in A$, $u(a_k) = \sum_{\ell=1}^{k-1} u_\ell^*(1)$. $\square$