Essays on Information Economics

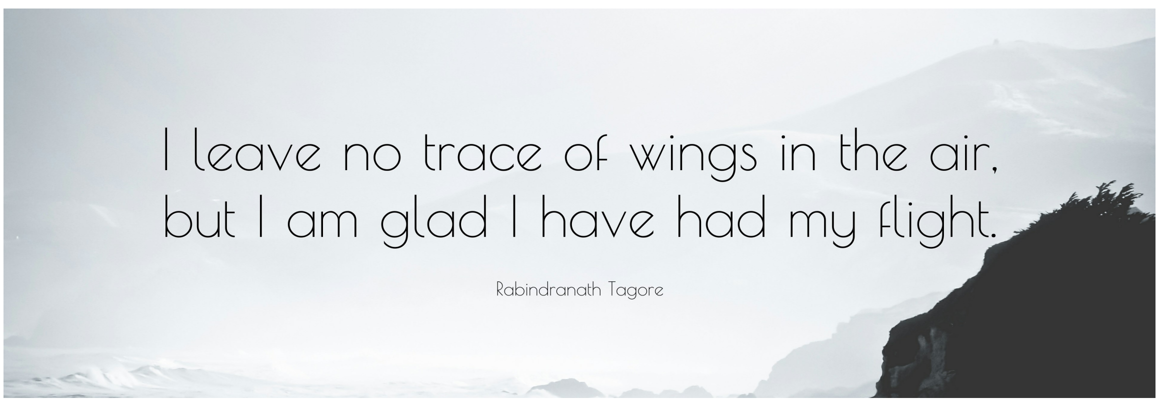
by

Xienan Cheng

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2022

Doctoral Committee:

Professor Tilman Börgers, Chair
Assistant Professor Heng Liu
Associate Professor David Miller
Professor Scott Page


I leave no trace of wings in the air,
but I am glad I have had my flight.
Rabindranath Tagore

Xienan Cheng
xienanc@umich.edu
ORCID iD: 0000-0001-7900-7245

ACKNOWLEDGMENTS

No man is an island entire of itself. Without the help of many people, this dissertation would not have been possible.

First and foremost, I am extremely indebted to Tilman Börgers for his constant support. His invaluable feedback and continued guidance helped me to overcome the difficulties in preparing this dissertation. His contagious passion for truth inspired me along the way to become a better researcher. It has been my good fortune to work with such a brilliant economist.

I am also deeply grateful to David Miller, Heng Liu, Scott Page, Mitchell Dudley, Francine Lafontaine, Pinar Yildirim, and Brian Wu for their insights and advice. I owe a special debt of gratitude to Shaowei Ke for his help during the difficult time.

I thank Zhiyuan Yao for picking me up from the airport at midnight and accommodating me for a couple of weeks before my apartment was ready when I arrived in Ann Arbor for the first time. I appreciate my roommate Shiting Li's kindness in everyday life.

Last but not least, I would like to thank my dearest mother and maternal grandmother for their unconditional love.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

ABSTRACT

This dissertation contains three essays on information economics that aim to understand the value and cost of information in the context of individual decision making and strategic interaction.

Chapter 1 investigates how a sender, by providing free information to a perfectly rational receiver, can manipulate the receiver's learning. Formally, I employ a Bayesian persuasion model with the additional assumption that the receiver has costly opportunities to acquire further information after receiving the information from the sender. It seems that the sender should make the receiver end up knowing more than she would otherwise (I call this "encouragement"). However, as is shown in the paper, the sender may also make the receiver end up knowing less (I call this "deterrence") or knowing different things than she would otherwise (I call this "diversion"). I identify the necessary condition for the feasibility of these manipulations, where two properties of the receiver's information acquisition cost function, that the recent literature on information acquisition calls Sequential Learning Proofness (SLP) and the more restrictive Indifference to Sequential Learning (ISL), play a vital role.

Chapter 2, which is joint work with Tilman Börgers, proposes a general theory of dominance among choices that encompasses strict and weak dominance among strategies in games, Blackwell dominance among experiments, and first or second order stochastic dominance among monetary lotteries. One choice dominates another if in a variety of situations the former choice yields higher expected utility than the latter. We then investigate whether, in a finite set of possible choices, all undominated choices are optimal in some situation. We present a formal framework in which the answer to this question is positive, and we show that within this framework the set of undominated choices is the smallest set to which the decision maker can restrict attention ex ante without running the risk of not having an optimal choice in the particular situation in which she finds herself. For this result it is crucial that the dominating alternatives are allowed to be convex combinations (in games: mixed strategies). A detailed analysis of dominance in game theory, Blackwell dominance, and first or second order stochastic dominance in one common framework also allows us to compare the properties of these concepts, and to obtain insights into why certain versions of our result apply only to some, but not all of these concepts.

Chapter 3, which is joint work with Tilman Börgers, investigates which joint distributions of two signals are not Blackwell dominated among all the joint distributions with fixed marginal distributions. For a special case with just two states and two signal realizations per signal, we provide a complete characterization of joint distributions that are not Blackwell dominated by any single joint distribution and the counterpart for any convex combination of joint distributions. For the general case, we present a necessary condition for a joint distribution not being Blackwell dominated by a convex combination of joint distributions. In all cases, the conditionally independent joint distribution is Blackwell dominated.

CHAPTER 1

Deterrence, Diversion, and Encouragement: How Freely Provided Information May Distort Learning

1.1 Introduction

When one sends information to another, the receiver can sometimes carry out further learning at some cost. Here are two examples. First, firms often publicly disclose some information through websites, annual reports, press releases, and fact books. For instance, oil companies reveal the test flow rates for new exploration wells. Their competitors can acquire further information using some tactics. For instance, they can hire competitive intelligence companies like *Aqute*, whose website claims, "Our original research helps you...learn about your competitors' wins, losses, ambitions, and concerns. You can understand the current reality inside your key competitors, giving you the time and the knowledge to successfully outflank your rivals." Second, in some juridical systems, after the prosecutors present the evidence, judges can conduct further investigation. For instance, Article 283 of the code of criminal procedure of France (as of 2006) says, "The president[1] may order any investigatory step he deems useful if the investigations appear to him to be incomplete, or if further matters have come to light since it was concluded. Such steps are taken either by the president, by one of his assessors, or by an investigating judge he delegates for this purpose." Article 196 of the criminal procedure law of China says, "During a court hearing, if the collegial panel has doubts about the evidence, it may announce an adjournment, in order to carry out an investigation to verify the evidence."

These examples motivate my research question: how can a sender, by providing free information to a perfectly rational receiver, manipulate the receiver's learning (change the receiver's learning outcome)? Formally, I employ a Bayesian persuasion model with the additional assumption that the receiver has costly opportunities to acquire further information

[1]"President" refers to the president of the court.

after receiving the information from the sender. When the sender reveals no information, the receiver learns by herself and ends up with a learning outcome. When the sender reveals some information, the receiver learns further by herself and ends up with another learning outcome. I investigate the treatment effect of information by comparing these two learning outcomes. It seems that the latter should be strictly more Blackwell informative (Blackwell (1951), Blackwell (1953))[2] than the former (the receiver ends up knowing more than she would otherwise), and I call this "encouragement". However, as is shown in the paper, the latter can be strictly less Blackwell informative than the former (the receiver ends up knowing less than she would otherwise), and I call this "deterrence". The two learning outcomes may also be Blackwell incomparable (the receiver ends up knowing different things than she would otherwise), and I call this "diversion". I also call these cases collectively "distortion." So surprisingly, more information does not necessarily make one more informed. Moreover, even when the sender and receiver's preferences are diametrically opposed, the sender may benefit from revealing free information to the receiver.

Then a relevant question is: under what condition are these manipulations feasible? I find that certain features of the receiver's information cost function are crucial for the sender's ability to manipulate the receiver's learning. Two properties that the recent literature on information acquisition calls Sequential Learning Proofness (SLP) and the more restrictive Indifference to Sequential Learning (ISL) play a vital role. I identify the necessary condition for the feasibility of deterrence, diversion, and encouragement. When the cost function satisfies ISL, deterrence is infeasible. With Uniformly Posterior Separable cost functions, if there are two actions and two states, then diversion is infeasible. When the cost function satisfies SLP, and the sender reveals all the information that the receiver would learn if the sender revealed no information, encouragement is infeasible.

Related Literature.—The Bayesian persuasion framework has been extended to include that the sender endogenously acquires costly information (Gentzkow and Kamenica, 2014) and that the receiver has exogenous information (Kolotilin et al., 2017). My model enables a receiver to adjust the information sent by the sender endogenously. The literature with this framework includes Bloedel and Segal (2021), Lipnowski et al. (2020), Lipnowski et al. (2022), Wei (2021), Matysková and Montes (2021), and Bizzotto et al. (2020). In Bloedel and Segal (2021), Lipnowski et al. (2020), Lipnowski et al. (2022), and Wei (2021), the receiver cannot digest all the information sent by the sender. In other words, the sender sets an upper bound on the receiver's information. In Matysková and Montes (2021), Bizzotto et al. (2020), and my work, the receiver can digest all the information sent by the sender

[2]A signal is strictly more Blackwell informative than another signal if its resulting payoff is higher or equal for any decision problem and higher for some decision problem.

and can learn even more. In other words, the sender sets a lower bound on the receiver's information. But the research question of Matysková and Montes (2021) and Bizzotto et al. (2020) is whether the receiver always benefits from a better learning technology, which is different from the present study. In addition, they only consider specific cost functions while I consider general cost functions.

The present study is also related to the literature on rational inattention (Sims, 2003). Single-agent rational inattention decision problems with a Uniformly Posterior Separable (UPS) cost function have been studied by Matějka and McKay (2015) and Caplin and Dean (2013). This paper studies the counterpart with a partition cost function (non-UPS). There is also literature considering rational inattention in a strategic context: Yang (2012) studies a coordination game in which both players can costly acquire information; Martin (2017) and Ravid (2019) study seller-buyer games in which the buyer can costly acquire information but the seller does not directly send signals.

Morris and Strack (2019), Hébert and Woodford (2021), and Bloedel and Zhong (2021) focus on the sequential sampling foundations for cost of information. This paper investigates the strategic implication of an information cost with some features related to such a foundation.

The rest of the paper is organized as follows. Section 1.2 introduces the model. Section 1.3 formally defines deterrence, diversion, and encouragement. Section 1.4 discusses the features of information acquisition cost functions (SLP and ISL). Section 1.5 investigates the feasibility of distortion in general, and then Section 1.6, Section 1.7, and Section 1.8 investigate the feasibility of deterrence, diversion, and encouragement respectively. Finally, Section 1.9 discusses some additional points and concludes.

1.2 Model

Assume there are two players, a sender (he) and a receiver (she). The set of states of the world is $\Omega = \{\omega_1, ..., \omega_n\}$. Both players share the same prior $\mu \in \Delta(\Omega)$, where $\Delta(\Omega)$ indicates the set of all probability distributions on Ω. The set of available actions is $A = \{a_1, ..., a_m\}$. The receiver's gross utility is $u(a, \omega)$, while that of the sender is $v(a, \omega)$. Assume the values of the gross utilities are all finite. In addition to her gross utility, the receiver incurs a cost of information acquisition, which I will later specify in detail.

The timing of the game is as follows. First, the sender commits to a signal with finite realizations and sends the realizations to the receiver, inducing her to form interim beliefs. Unlike in the classical Bayesian persuasion model, the receiver can then choose whether to costly acquire another signal. Her choice of the additional signal can be contingent on the realizations of the signal from the sender. After observing the realizations of the new signal and forming posterior beliefs, the receiver takes action.

Now I reformulate the sender's strategies and the receiver's information acquisition strategies to simplify the problem. The distribution of the receiver's interim beliefs induced by the signal from the sender determines the receiver's subsequent learning choices; thus, it determines the final joint distribution of the receiver's actions and the states, which matters for the sender's payoff. So from the sender's perspective, any signals that induce the same distribution of the receiver's interim beliefs must imply the same payoffs. To utilize this idea, I introduce some notation. Denote the set of distribution over beliefs by $\Delta(\Delta(\Omega))$ and a generic element of it by I. Let $\succsim_B$ denote "weakly more Blackwell informative than[3]," and $\mathcal{B}(I)$ denote the set of distributions over beliefs that are weakly more Blackwell informative than I. Note that all the elements in $\mathcal{B}(I)$ have the same expected value with I. Use δ_γ to denote the degenerated distribution over beliefs at $\gamma \in \Delta(\Omega)$. Specifically, let $I_\varnothing \equiv \delta_\mu$. Therefore, the sender's choice of the signal is equivalent to the choice of $I_S \in \mathcal{B}(I_\varnothing)$.

For the receiver, given the sender chooses I_S, observing the realization of the signal sent by the sender implies the receiver updates her belief to $\xi \in \text{supp } I_S$, where supp I_S indicates the support of I_S. Then, similar to the analysis for the sender, for each ξ, the receiver's learning is equivalent to choosing a distribution over beliefs in $\mathcal{B}(\delta_\xi)$. So the receiver's learning strategy given I_S can be characterized by a family of conditional distribution over beliefs, that is, $L : \text{supp } I_S \to \Delta(\Delta(\Omega))$. Note I_S and L imply an unconditional distribution over the posterior beliefs $I_{SR} \in \mathcal{B}(I_S)$.

In sum, the timing of the game is represented by Figure 1.1. For simplicity, this graph only illustrates the two-state case. A point on the line segment represents a belief. First, the sender chooses $I_S \in \mathcal{B}(I_\varnothing)$. Then the receiver chooses L, which together with I_S implies $I_{SR} \in \mathcal{B}(I_S)$. Last, the receiver takes action for each $\gamma \in \text{supp } I_{SR}$.

[3] A signal is weakly more Blackwell informative than another signal if its resulting payoff is higher or equal for any decision problem.

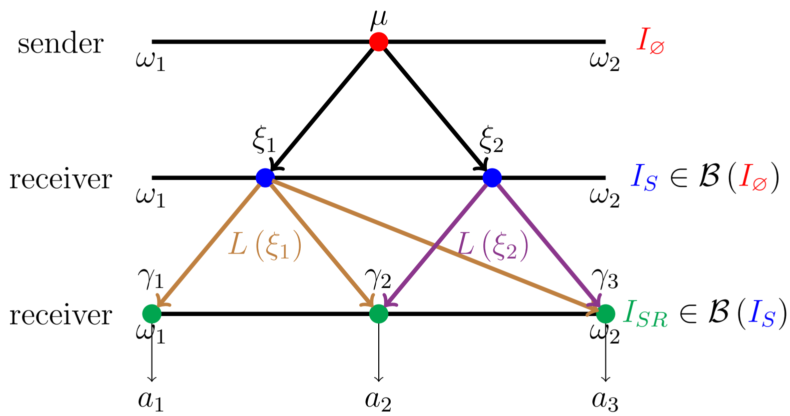


Figure 1.1: The timing of the game

Next, I analyze the receiver's optimization problem given that the sender reveals I_S. We know that for a profit maximization problem where the firm chooses the factors of production, it can be solved in two steps: first, for a specific quantity, choose the factors of production to minimize the cost; second, choose an optimal quantity. I can utilize a similar method here.

First, consider a specific unconditional distribution over beliefs $I_{SR} \in \mathcal{B}(I_S)$. On the one hand, for every $\gamma \in \text{supp } I_{SR}$, the receiver chooses the action to solve the following problem:

$$\max_{a \in A} \int_\Omega u(a, \omega)\, d\gamma(\omega)$$

Denote the resulting indirect gross utility function of γ as $\hat{u}(\gamma)$. Then, the corresponding indirect gross utility function of I_{SR}, denoted by U, is

$$U(I_{SR}) \equiv \int_{\Delta(\Omega)} \hat{u}(\gamma)\, dI_{SR}(\gamma)$$

On the other hand, there may be more than one L that leads to I_{SR}. The receiver chooses the one that minimizes the expected learning cost. The resulting minimal expected learning cost for the receiver from I_S to I_{SR} is denoted by $C(I_S, I_{SR})$. I allow its value to be infinity, which means it is infeasible to learn to the degree of I_{SR} starting from I_S[4].

Second, the receiver chooses I_{SR} to solve the following problem:

$$\max_{I_{SR} \in \mathcal{B}(I_S)} \quad U(I_{SR}) - C(I_S, I_{SR})$$

I denote the set of solutions to the above problem as $BR(I_S)$.

[4]If I_{SR} is strictly less Blackwell informative than I_S or I_{SR} is Blackwell incomparable with I_S, then $C(I_S, I_{SR})$ must be infinity. If I_{SR} is weakly more Blackwell informative than I_S, then $C(I_S, I_{SR})$ may be infinity.

Given the receiver's best response for all I_S, the sender chooses I_S^* to maximize his payoff. In sum, the equilibrium concept is perfect Bayesian equilibrium (henceforth referred to as equilibrium for short): that is, beliefs are updated using Bayes' rule, and each player chooses strategies to maximize his or her expected payoff given the other player's strategy and his or her beliefs at the corresponding stage of the game.

As indicated above, the equilibrium involves the receiver's best response and the sender's optimal choice. These two ingredients are relevant for the following two questions, respectively. (i) The receiver's best response addresses the question of whether it is feasible for the sender to distort the learning outcome of the receiver by revealing information. (ii) The sender's optimal choice addresses a further question of whether it is optimal for the sender to exert a distortion. In this paper, I mainly focus on the first question and address the second question whenever relevant.

1.3 Three Types of Manipulations

Now I investigate how the receiver's learning outcome varies with the freely provided information by comparing the case where the sender reveals no information and the case where the sender reveals some information. The receiver's best response correspondence BR is the building block of the following analysis. Let

$$\mathcal{P} \equiv \left\{ (I_\varnothing, I_R, I_S, I_{SR}) \in (\mathcal{B}(I_\varnothing))^4 \,|I_R \in BR(I_\varnothing) \text{ and } I_{SR} \in BR(I_S) \right\}$$

where $(\mathcal{B}(I_\varnothing))^4$ represents the quaternary Cartesian power of $\mathcal{B}(I_\varnothing)$.

Definition 1.1 (Distortion)**.** $(I_\varnothing, I_R, I_S, I_{SR}) \in \mathcal{P}$ *is a* distortion *if* $I_{SR} \notin BR(I_\varnothing)$.

I_R indicates the distribution over beliefs the receiver ends up with if the sender reveals no information, while I_{SR} indicates the counterpart when the sender reveals information. So distortion means the sender guides the receiver to a learning outcome that she would not reach without this additional information. According to the relationship between I_{SR} and I_R, I can further divide distortions into three types.

Definition 1.2 (Deterrence)**.** $(I_\varnothing, I_R, I_S, I_{SR})$ *is a* deterrence *if it is a distortion and* $I_{SR} \prec_B I_R$.

where $\prec_B$ means "strictly less Blackwell informative than."

Definition 1.3 (Diversion)**.** $(I_\varnothing, I_R, I_S, I_{SR})$ *is a* diversion *if it is a distortion and* $I_{SR} \not\gtrless_B I_R$.

where $\not\gtrless_B$ means "Blackwell incomparable with."

Definition 1.4 (Encouragement)**.** $(I_\varnothing, I_R, I_S, I_{SR})$ *is an* encouragement *if it is a distortion and* $I_{SR} \succ_B I_R$.

where $\succ_B$ means "strictly more Blackwell informative than." The deterrence exists if there is a signal such that, in response to it, the receiver ends up knowing less than what she would learn if she did not get any signal. The counterparts for the diversion and encouragement are knowing differently and more, respectively. Note that for a given best response correspondence of the receiver, different types of distortion can exist simultaneously. That is, $(I_\varnothing, I_R, I_S, I_{SR})$ is one type of distortion, while $(I_\varnothing, I_R, I'_S, I'_{SR})$ is another type of distortion. The definitions of deterrence and diversion immediately imply the following results.

Lemma 1.1. *If* $(I_\varnothing, I_R, I_S, I_{SR})$ *is a deterrence, then* $I_S \prec_B I_R$.

Proof. By Definition 1.2, $I_{SR} \prec_B I_R$. $I_{SR} \in BR(I_S)$ implies $I_S \precsim_B I_{SR}$, where $\precsim_B$ denotes "weakly less Blackwell informative than." So $I_S \prec_B I_R$. □

Lemma 1.1 means in a deterrence, the information revealed by the sender must be strictly less Blackwell informative than what the receiver would acquire if she did not get any signal.

Lemma 1.2. *If* $(I_\varnothing, I_R, I_S, I_{SR})$ *is a diversion, then* $I_S \prec_B I_R$ *or* $I_S \not\gtrless_B I_R$.

Proof. Suppose $I_S \succsim_B I_R$. $I_{SR} \in BR(I_S)$ implies $I_{SR} \succsim_B I_S$. Then $I_{SR} \succsim_B I_R$, which contradicts Definition 1.3. So $I_S \prec_B I_R$ or $I_S \not\gtrless_B I_R$. □

Lemma 1.2 means in a diversion, the information revealed by the sender must be strictly less Blackwell informative than or Blackwell incomparable with what the receiver would acquire if she did not get any signal. There is not a similar constraint for encouragement. By definition, the existence of the distortions only depends on the best response of the receiver, and I am interested in how it is affected by the features of the cost function, which is a primitive of the receiver's optimization problem. In sum, the existence of a deterrence, diversion, or encouragement implies it is feasible for the sender to distort the learning outcome of the receiver by revealing information.

Accordingly, when $(I_\varnothing, I_R, I^*_S, I_{SR})$ is a deterrence/diversion/encouragement, it implies a deterrence/diversion/encouragement equilibrium.

1.4 The Features of the Information Acquisition Cost Functions

The features of the cost function C, which stems from the receiver's optimization problem, play an important role in distortion. Consider the following features of cost functions proposed by Bloedel and Zhong (2021).

Definition 1.5 (Sequential Learning Proofness, SLP)**.** *A cost function C satisfies* Sequential Learning Proofness *if*

$$C\left(I, I''\right) \leqslant C\left(I, I'\right) + C\left(I', I''\right)$$

for all $I, I', I'' \in \mathcal{B}\left(I_{\varnothing}\right)$ such that $I'' \succsim_B I' \succsim_B I$.

One may be tempted to interpret this feature as "to learn in one shot is weakly less costly than to learn sequentially," but this is not completely right. In this interpretation, $C\left(I', I''\right)$ is treated as the subsequent cost to I'' after the receiver learns by herself to reach I'. But in my definition of the cost function, $C\left(I', I''\right)$ refers to the cost of reaching I'' for the receiver when the sender guides her to I'. Example 1.3 below illustrates the difference between these two costs. Under the additional condition that the subsequent cost of learning at a distribution over beliefs is independent of how the receiver arrives at it, this interpretation works. But I would also like to include the case where this condition does not hold in the paper, so this interpretation is only partially correct. If this condition holds and thus $C\left(I', I''\right)$ can also be interpreted as the subsequent cost to I'' after the receiver learns by herself to reach I, then as suggested in Bloedel and Zhong (2021), every cost function should satisfy this feature. Otherwise, the receiver can find a sequential path from I to I'' with a lower cost, which contradicts the notion of $C\left(I, I''\right)$ being the minimum cost of reaching I'' from I.

Some cost functions may even have the following stronger feature.

Definition 1.6 (Indifference to Sequential Learning, ISL)**.** *A cost function C satisfies* Indifference to Sequential Learning *if*

$$C\left(I, I''\right) = C\left(I, I'\right) + C\left(I', I''\right)$$

for all $I, I', I'' \in \mathcal{B}\left(I_{\varnothing}\right)$ such that $I'' \succsim_B I' \succsim_B I$.

That is, the inequality in SLP is always binding. Again, if the above condition holds, this feature can be interpreted as "to learn in one shot is as costly as to learn sequentialy". And thus all different ways of learning sequentially are equally costly. In other words, the cost of information acquisition is path-independent.

Based on these two features, all the cost functions can be divided into three types: ISL cost functions, non-ISL but SLP cost functions, and non-SLP cost functions. Below I present one example for each type of cost function.

Example 1.1 (A cost function that satisfies ISL)**.**

Following the convention in the literature, I start with another form of cost function $\tilde{C}(\xi, L)$ with $\xi \in \Delta(\Omega)$ and $L(\xi) \in \mathcal{B}(\delta_\xi)$, which represents the cost from ξ to $L(\xi)$. $\tilde{C}$ is uniformly posterior separable (UPS, Caplin et al. (2019b)) if

$$\tilde{C}(\xi, L(\xi)) = \mathbb{E}_{L(\xi)}(G(\gamma)) - G(\xi)$$

where $\gamma \in \text{supp } L(\xi)$ and $G(\cdot)$ is strictly convex. A notable example is the entropy cost function, with a special G, that is, $G_E(\gamma) = \sum_{j=1}^{n} \gamma_j \ln \gamma_j$ and extended to boundary points using the limit condition $\lim_{\gamma_j \to 0} \gamma_j \ln \gamma_j = 0$, where γ_j refers to the jth entry of the belief (the probability of ω_j). This is the most commonly used cost function in the literature of rational inattention. Then I derive the corresponding cost function in the format of this paper. Let each $\xi \in \text{supp } I$ correspond to a conditional distribution $L(\xi)$ over beliefs. I and L imply an unconditional distribution I' over the beliefs. This implies

$$\begin{aligned}
C(I, I') &= \mathbb{E}_I\left(\tilde{C}(\xi, L(\xi))\right) \\
&= \mathbb{E}_I\left(\mathbb{E}_{L(\xi)}(G(\gamma)) - G(\xi)\right) \\
&= \mathbb{E}_I\left(\mathbb{E}_{L(\xi)}(G(\gamma))\right) - \mathbb{E}_I(G(\xi)) \\
&= \mathbb{E}_{I'}(G(\gamma)) - \mathbb{E}_I(G(\gamma))
\end{aligned}$$

Now I check whether a UPS cost function satisfies ISL.

$$\begin{aligned}
C(I, I'') &= \mathbb{E}_{I''}(G(\gamma)) - \mathbb{E}_I(G(\gamma)) \\
&= \mathbb{E}_{I'}(G(\gamma)) - \mathbb{E}_I(G(\gamma)) + \mathbb{E}_{I''}(G(\gamma)) - \mathbb{E}_{I'}(G(\gamma)) \\
&= C(I, I') + C(I', I'')
\end{aligned}$$

So a UPS cost function satisfies ISL. ■

Example 1.2 (A cost function that does not satisfy ISL but satisfies SLP).

Assume there are three states and $p_1 > p_2 > p_3 > 0$ where p_i is the probability of ω_i. The receiver can choose any subset of her current information set and pay a fixed cost of c to know whether the true state is in this subset or not. One can think of this as asking a "yes or no" question. She can continue paying the cost for further refinement for as many times as she wants. In other words, she can conduct sequential binary partitions. Next, I illustrate that the cost function derived from this learning technology satisfies SLP but does not satisfy ISL. Let $I_\varnothing = \{\{\omega_1, \omega_2, \omega_3\}\}$ and $I''' = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}\}$. If $\tilde{I}$ is not a partition of $\{\omega_1, \omega_2, \omega_3\}$, then $C\left(I_\varnothing, \tilde{I}\right) = +\infty$. Let $I = \{\{\omega_1\}, \{\omega_2, \omega_3\}\}$, $I' = \{\{\omega_2\}, \{\omega_1, \omega_3\}\}$,

$I'' = \{\{\omega_3\}, \{\omega_1, \omega_2\}\}$. There are three feasible ways of reaching I''' from $I_\varnothing$, as illustrated in Figure 1.2.

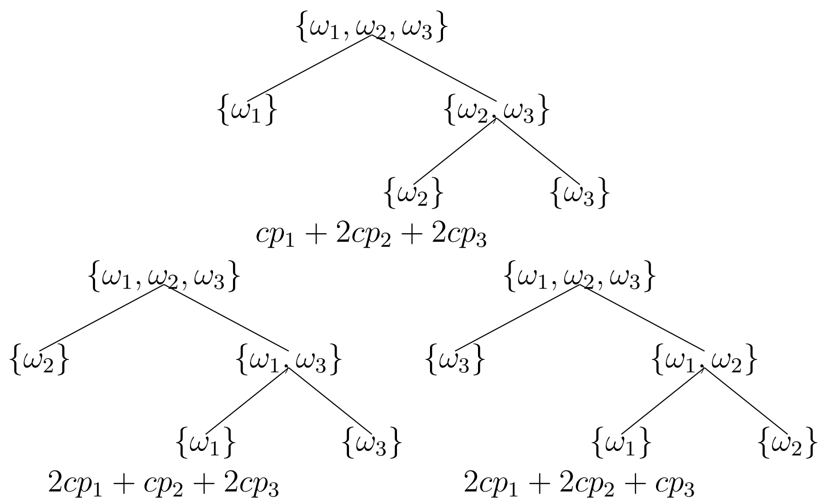


Figure 1.2: Three ways from $\{\{\omega_1, \omega_2, \omega_3\}\}$ to $\{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}\}$

For the first alternative, the receiver first pays the cost to learn whether the true state is ω_1 or not. If it is ω_1, she stops learning; if it is not, she pays the cost again to distinguish ω_2 and ω_3. It can be treated as going from $I_\varnothing$ via I to I'''. Since the subsequent cost of learning at I is independent of how the receiver arrives at it, the total expected cost is equal to $C(I_\varnothing, I) + C(I, I''')$, where $C(I_\varnothing, I) = c$ and $C(I, I''') = c(p_2 + p_3)$. The analyses of the other two alternatives are similar. So

$$C(I_\varnothing, I''') = \min_{\tilde{I} \in \{I, I', I''\}} C\left(I_\varnothing, \tilde{I}\right) + C\left(\tilde{I}, I'''\right)$$

It can be verified that I is optimal. So

$$C(I_\varnothing, I''') = C(I_\varnothing, I) + C(I, I''') \tag{1.1}$$

Note $C(I_\varnothing, I') = C(I_\varnothing, I'') = c$, $C(I', I''') = c(p_1 + p_3)$ and $C(I'', I''') = c(p_1 + p_2)$. Then

$$C(I_\varnothing, I''') < C(I_\varnothing, I') + C(I', I''') \tag{1.2}$$

and

$$C(I_\varnothing, I''') < C(I_\varnothing, I'') + C(I'', I''') \tag{1.3}$$

(1.1), (1.2), and (1.3) indicate this partition cost function satisfies SLP but does not satisfy ISL. ■

Example 1.3 (A cost function that does not satisfy SLP)**.**

Following Bizzotto et al. (2020), assume there are two states, and the receiver can perform a binary test at a fixed cost of c'. The precision of this test is fixed, that is, $\mathbb{P}(s = \omega|\omega) = e > \frac{1}{2}$, where $\mathbb{P}$ is the probability and s is the result of the test. Suppose the receiver can only access the test once. Now I illustrate that this cost function does not satisfy SLP.

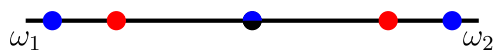


Figure 1.3: Non-SLP cost function

In Figure 1.3, a point on the line segment represents a belief. Let the prior be 50-50, which is represented by the black point. (It is half black half blue, which means a black point overlaps with a blue point.) If the receiver accesses the test at the prior, she ends up with a distribution over beliefs represented by the two red points. Denote it by I'. If the sender reveals I' and the receiver accesses the test at each realization of I', then the receiver ends up with a distribution over beliefs represented by the three blue points. Denote it by I''. This implies $C(I_\varnothing, I') = c'$, $C(I', I'') = c'$, and $C(I_\varnothing, I'') = +\infty$. The last equation is because two tests are required to reach I'' from $I_\varnothing$, but the receiver can only access it once. Therefore,

$$C(I_\varnothing, I'') > C(I_\varnothing, I') + C(I', I'')$$

which violates SLP. Finally, I clarify a possible confusion. When the sender reveals I' to the receiver, the receiver can still access the test. But when the receiver learns to I' by herself, she can no longer access the test. As a result, the cost from I' to I'' is $+\infty$. So if one takes the latter as $C(I', I'')$, which is inconsistent with my definition, she may argue SLP holds in this case by mistake. ■

In the following sections, I investigate how the features of the cost functions affect the feasibility of distortion.

1.5 Distortion

First I establish some results regarding distortion.

Note SLP is characterized by a triangle inequality. Now I consider a special case of SLP where the inequality is binding locally. That is, $\exists I_\varnothing, I_S, I_R \in \mathcal{B}(I_\varnothing)$ such that $C(I_\varnothing, I_S) + C(I_S, I_R) = C(I_\varnothing, I_R)$ with $I_R \succsim_B I_S \succsim_B I_\varnothing$.

Lemma 1.3. *Suppose the cost function C satisfies Sequential Learning Proofness. If $(I_\varnothing, I_R, I_S, I_{SR})$ satisfies $I_S \precsim_B I_R$ and $C(I_\varnothing, I_S) + C(I_S, I_R) = C(I_\varnothing, I_R)$, then it is not a distortion.*

Roughly speaking, this result means, with an SLP cost function, when the sender reveals *a specific part* (the part making the SLP constraint binding) of the information that the receiver would learn if the sender revealed no information, the receiver accepts the message and learns exactly the same thing as before. Intuitively, the message sent by the sender saves weakly more cost for the original optimal learning outcome than other learning outcome. Therefore, the original optimal learning outcome remains optimal. So distortion is infeasible.

Lemma 1.3 implies that a binding SLP constraint makes the distortion infeasible. Since the SLP constraint is binding globally for an ISL cost function, the following result applies.

Lemma 1.4. *Suppose the cost function C satisfies Indifference to Sequential Learning. If $(I_\varnothing, I_R, I_S, I_{SR})$ satisfies $I_S \precsim_B I_R$, then it is not a distortion.*

Informally, this result means, with an ISL cost function, when the sender reveals *any* part of the information that the receiver would learn if the sender revealed no information, the receiver accepts the message and learns exactly the same thing as before.

1.6 Deterrence

In this section, first I present a condition under which deterrence is impossible. Then I demonstrate an example where the condition does not hold and thus deterrence is possible.

Observe that Lemma 1.1 and Lemma 1.4 imply the following result.

Theorem 1.1. *If the cost function C satisfies Indifference to Sequential Learning, then deterrence is infeasible.*

The reason why it is true is as follows. Lemma 1.1 indicates, to make the receiver learn less, the sender must reveal less information than what the receiver would learn if the sender revealed no information. But Lemma 1.4 suggests, with an ISL cost function, this kind of information cannot induce any distortion, including deterrence. As a result, deterrence is impossible when the cost function satisfies ISL.

On the other hand, if the cost function does not satisfy ISL, then deterrence is possible. Here is an example.

Example 1.4 (Deterrence with a non-ISL but SLP cost function)**.**

Suppose there are three states, $p_1 = p = 0.8$, $p_2 = p_3 = q = 0.1$. So there is one likely state and two unlikely states, and the two unlikely states are of the same probability. Assume there are three available actions, and $u(a_i, \omega_j) = 1$ when $i = j$, $u(a_i, \omega_j) = 0$ when $i \neq j$. The cost is the partition cost illustrated in Example 1.2 and $c = 0.15$.

When the sender reveals no information, one can verify that the receiver's unique optimal strategy is as follows. The receiver first pays the cost to learn whether the true state is the likely state or not, because this minimizes the probability of paying the cost again. If the partition indicates it is not the likely state, given that the remaining two states are of the same probability, she has a strong enough incentive to further distinguish between the two. This strategy is shown in Figure 1.4, where the big black dot indicates the likely state, while the small grey dots indicate the unlikely states. As a result, the receiver learns the states perfectly, that is, $I_R = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}\}$.

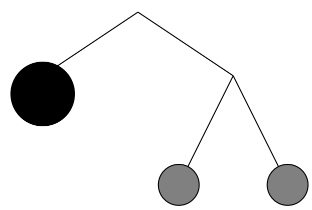


Figure 1.4: The receiver's optimal learning strategy when the sender reveals no information

Next, I investigate the receiver's optimal strategy when $I_S = \{\{\omega_2\}, \{\omega_1, \omega_3\}\}$. Given that $c > \frac{q}{p+q}$, this implies $I_{SR} = I_S$. That is, as the likely state is much more probable than the unlikely state, the receiver does not learn further and directly guesses the likely state.

As $I_R \succ_B I_{SR}$, $(I_\varnothing, I_R, I_S, I_{SR})$ is a deterrence. The intuition is that the order of learning affects the extent of learning (this is because the partition cost does not satisfy ISL). So the sender can make the receiver end up knowing less by disrupting the receiver's order of learning.

It is shown above that deterrence is feasible. Furthermore, assume the sender and receiver's utility functions are "zero-sum", that is, $v(a_i, \omega_1) = -u(a_i, \omega_j)$, $\forall a_i, \omega_j$. One can verify that the sender has a higher payoff when revealing I_S than revealing nothing. So surprisingly, this example implies that even when the interests of the sender and the receiver are diametrically opposite, revealing information may be better than withholding it for the sender. ■

1.7 Diversion

In this section, first I present a condition under which diversion is impossible. Then I demonstrate examples where the condition does not hold and thus diversion is possible.

I focus on UPS cost. The analysis of the receiver's optimal strategy is as follows, which is discussed in Caplin and Dean (2013). Define the net utility function for action a_i as

$$N^{a_i}(\gamma) = \sum_{j=1}^{n} \gamma_j u(a_i, \omega_j) - G(\gamma)$$

Given an interim belief ξ, it can be shown that the receiver solves the following optimization problem:

$$\max_{L(\xi)\in\mathcal{B}\left(\delta_\xi\right),\sigma:\Delta(\Omega)\to A} \quad \mathbb{E}_{L(\xi)}\left(N^{\sigma(\gamma)}(\gamma)\right)$$

where $L(\xi)$ specifies the learning strategy, that is, a distribution over posteriors; σ specifies the action strategy, that is, which action to take at each posterior. It can be shown that in the optimal strategy, σ must be an injection. In other words, any action can be taken at no more than one posterior in the optimal strategy. The geometric approach to finding the optimal solution is to figure out the posteriors and actions whose associated net utility functions support the highest chord above the interim belief. This concavification method is familiar from Kamenica and Gentzkow (2011). Here is an example.

Example 1.5 (Receiver's optimal strategies with a UPS cost function)**.**

This example is illustrated in Figure 1.5. There are two states and two actions. The two red points characterize an interval. When the interim belief is in the interior of this interval (e.g., the left black point), the unique optimal strategy is to learn, and the two red points are the resulting posteriors. At the left red point, a_2 is taken. At the right red point, a_1 is taken. When the interim belief is outside the interval or on the boundary (e.g., the right black point), the unique optimal strategy is not to learn and to take a_1.

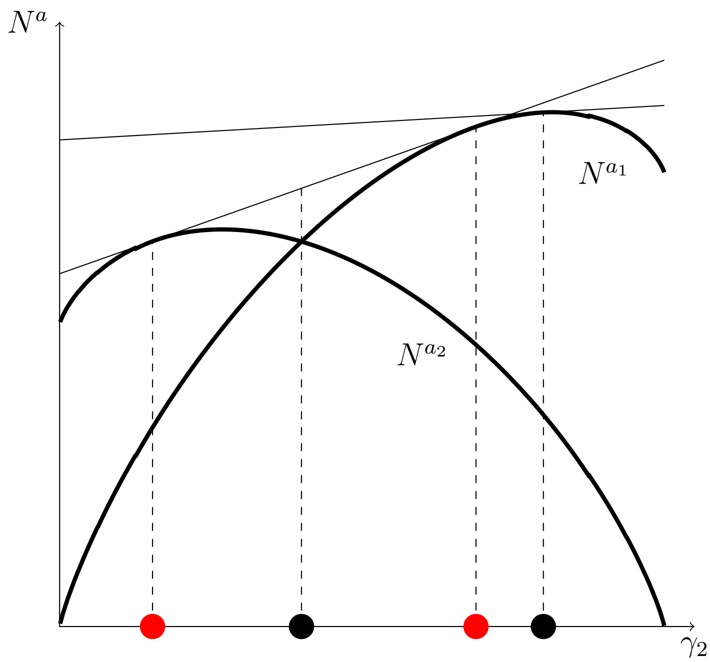


Figure 1.5: Receiver's optimal strategies with a UPS cost function when there are two states and two actions.

■

Based on the formulation of the receiver's optimal strategy, here is a result about the impossibility of diversion.

Theorem 1.2. *Suppose the cost function C is Uniformly Posterior Separable. If there are two states and two actions, then diversion is infeasible.*

The reasoning of this result can be seen through Figure 1.5. When the prior belief is outside the interval or on the boundary, $I_R = I_\varnothing$ and thus $I_{SR} \succsim_B I_S \succsim_B I_R$, which violates the definition of diversion. When the prior belief is within the interval, the support of I_R are the two red points. No matter what I_S is, all the elements in the support of I_{SR} must be outside the interval or on the boundary. As a result, $I_{SR} \succsim_B I_R$, which violates the definition of diversion. In sum, diversion is infeasible.

But by relaxing the above conditions, diversion exists with the UPS cost. First, I introduce a lemma that can facilitate the judgment of diversion.

Lemma 1.5. $\forall I, I' \in \mathcal{B}(I_\varnothing)$*, if supp* $I \nsubseteq$ *conv supp* I' *and supp* $I' \nsubseteq$ *conv supp* I*, then* $I \not\succsim_B I'$, where conv means the convex hull. This result is from Wu (2018). It indicates if the convex hull of I and that of I' are not nested, then I and I' are Blackwell incomparable.

On the one hand, when there are two states and three actions, diversion is feasible with the UPS cost. Here is an example.

Example 1.6 (Diversion with a UPS cost function when there are two states and three actions)**.**

This example is illustrated in Figure 1.6. The line touches three net utility functions simultaneously. At the prior belief denoted by the black point ($I_\varnothing$), one of the optimal strategies for the receiver is to choose the two red points (I_R). When the sender reveals the blue points (I_S), one of the optimal strategies of the receiver is not to learn (so I_{SR} is the same as I_S). According to Lemma 1.5, I_R and I_{SR} are Blackwell incomparable. So this is a diversion. The intuition is that with two states and two actions, when the belief is between the two red points, the receiver learns. But now I add a safe action (a_3) whose payoff does not vary severely with the states. As a result, this action is optimal when the probabilities of the two states are close. So it is feasible for the receiver not to learn even when the belief is between the two red points, which supports the diversion.

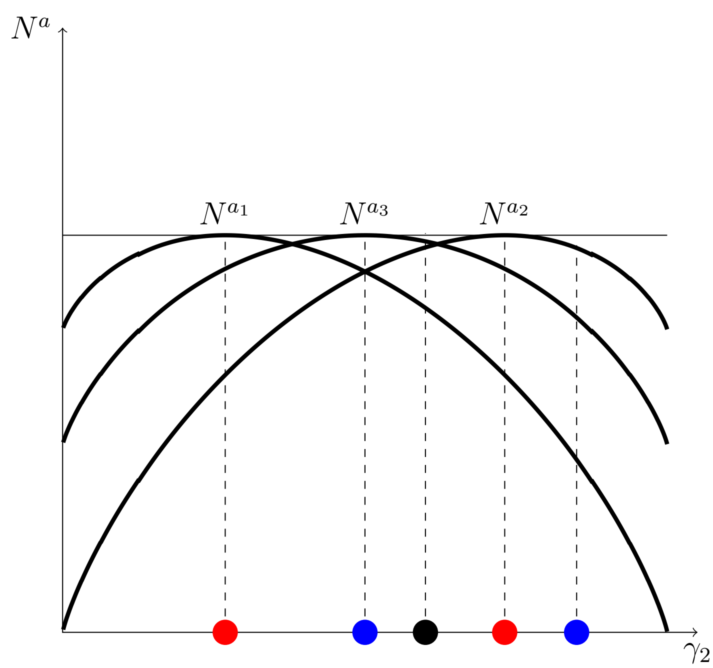


Figure 1.6: Diversion with a UPS cost function when there are two states and three actions

■

On the other hand, when there are three states and two actions, diversion is also feasible with the UPS cost. Here is an example.

Example 1.7 (Diversion with a UPS cost function when there are three states and two actions)**.**

Suppose $u(a_1, \omega_1) = 2$, $u(a_2, \omega_1) = 0$, $u(a_1, \omega_2) = 0$, $u(a_2, \omega_2) = 1$, $u(a_1, \omega_3) = 1$, and $u(a_2, \omega_3) = 0$; the cost is entropy cost; a diversion can occur as illustrated in Figure 1.7. The prior belief is the black point, and I_R is denoted by the two red points. I_S is the same as I_{SR}, which is denoted by the two blue points. According to Lemma 1.5, I_R and I_{SR} are Blackwell incomparable. So this is a diversion. The intuition is that, similar to the case of two states and two actions, the receiver learns when the interim belief is in the middle, but the boundary is a line rather than a point. This additional dimension enables the diversion.

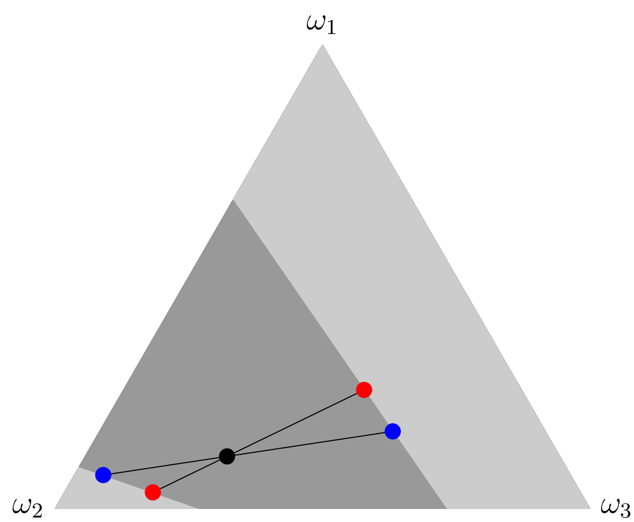


Figure 1.7: Diversion with a UPS cost function when there are three states and two actions

■

Note UPS cost functions satisfy ISL. So unlike deterrence, the ISL cost function is compatible with diversion.

There also exists a diversion equilibrium. Here is an example.

Example 1.8 (Diversion equilibrium)**.**

Suppose a headhunter, as a sender, tries to convince a recruiting firm, who is the receiver, that a candidate is competent. There are three states: The candidate being good at technology, being good at sales, and being incompetent. There are three actions for the firm: to hire the candidate as a technician, to hire the candidate as a salesperson, and to not to hire. If its action matches the state, it gets a payoff of 1. Otherwise, it gets 0. As for the headhunter, when the firm hires the candidate as either a technician or a salesperson, he gets a payoff of 1 (commission). When the firm does not hire, he gets 0. The headhunter and the firm share a prior belief, which is a uniform distribution over three states. The firm's information acquisition cost is entropy cost.

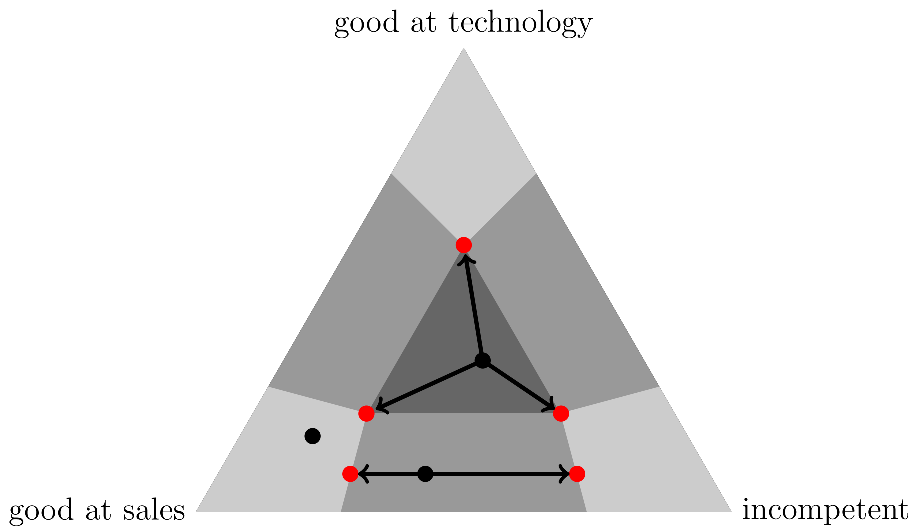


Figure 1.8: Firm's optimal strategies

The firm's optimal strategy is characterized in Figure 1.8 following Caplin et al. (2019a). When its interim belief (the black point) is in the dark gray region, the firm reaches the posterior beliefs represented by the three red points. This means that when it is not quite sure about the true state, it learns about all the states partially. When its interim belief is in the medium gray region, the firm reaches the posterior beliefs represented by the two red points. This means that when it is fairly certain that the true state is not one of the states, it learns more about the other two states. When its interim belief is in the light gray region, which means it believes the true state is very likely to be one specific state, the firm does not learn.

When the headhunter reveals no information, the firm acquires a signal with three realizations. In each of the realizations, the firm believes all three states are possible, but one of them is more probable than the other two, so it takes the action matching that state. As a result, the firm may make three types of mistakes: not to hire when the candidate is competent, to assign the wrong position (for example, hire a candidate who is good at technology as a salesperson), to hire when the candidate is incompetent.

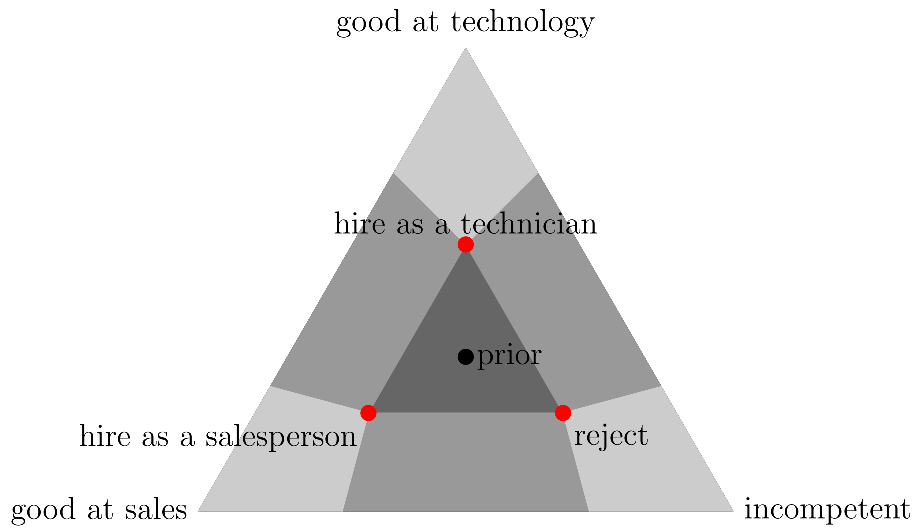


Figure 1.9: The firm's best response when the headhunter reveals no information

On the other hand, the optimal signal for the headhunter is as follows: he tailors the evidence such that when the candidate is incompetent, the report suggests the candidate is incompetent, good at technology, or good at sales; when the candidate is good at some aspect, the report suggests the candidate is good at that aspect. So this signal also has three realizations. The firm optimally chooses not to learn further after receiving this signal. In one of the realizations, the firm knows for sure that the candidate is incompetent and it rejects the candidate. So it never misses a competent candidate. In another realization, the firm believes the candidate is either good at technology or incompetent, and the former is more probable. So it hires the candidate as a technician. Note the firm knows for sure that the candidate is not good at sales. So it never assigns the wrong position. However, it hires an incompetent candidate more often. In sum, by providing free information, the headhunter makes the firm make two mistakes less often but one mistake more often. That's why this is a diversion. The intuition is that, when the firm hires a candidate, the headhunter helps the firm to make the mistake of assigning the wrong position less often, in exchange for the firm tolerating more mistake of hiring an incompetent candidate. This is like a "cross subsidy".

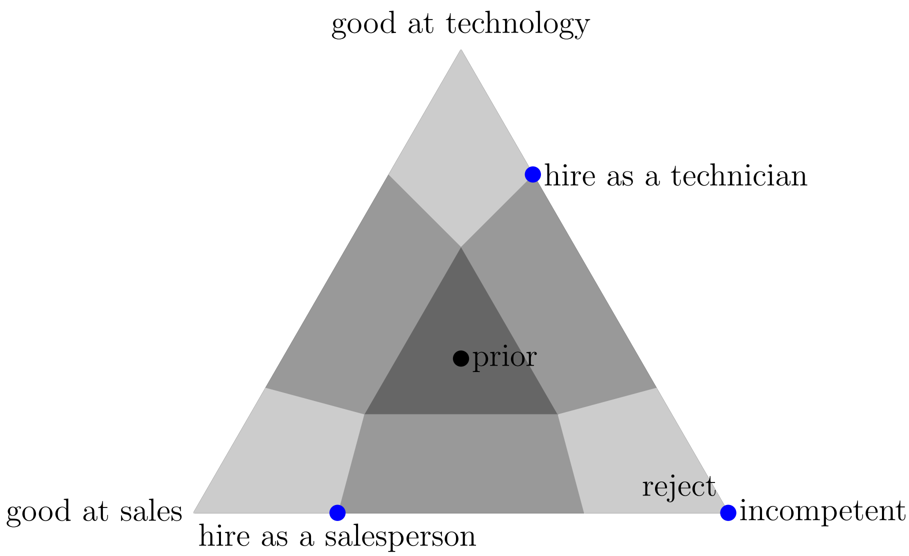


Figure 1.10: The headhunter's optimal signal

■

1.8 Encouragement

In this section, first I present a condition under which encouragement is impossible. Then I demonstrate examples where the condition does not hold and thus encouragement is possible.

First, let's revisit Lemma 1.3. The condition for it is $I_S \precsim_B I_R$ and $C(I_\varnothing, I_S) + C(I_S, I_R) = C(I_\varnothing, I_R)$. Lemma 1.4 addresses the case they are satisfied by $I_S \precsim_B I_R$ and ISL. $I_S = I_R$ can also make them hold. This implies

Lemma 1.6. *Suppose the cost function C satisfies Sequential Learning Proofness. If $(I_\varnothing, I_R, I_S, I_{SR})$ satisfies $I_S = I_R$, then it is not a distortion.*

Roughly speaking, this result means, with an SLP cost function, when the sender reveals all the information that the receiver would learn if the sender revealed no information, the receiver accepts the message and does not learn any further. This is the idea of the "non-learning equilibrium" in Matysková and Montes (2021). They get this result using UPS cost functions, which satisfy ISL. But my result suggests it is still true even when the cost functions only satisfy SLP.

Note Lemma 1.1 and Lemma 1.2 indicate that $I_S = I_R$ has already ruled out deterrence and diversion. So the role played by SLP is to rule out encouragement.

Theorem 1.3. *Suppose the cost function C satisfies Sequential Learning Proofness. If $(I_\varnothing, I_R, I_S, I_{SR})$ satisfies $I_S = I_R$, then it is not an encouragement.*

But by relaxing the above conditions, encouragement is possible. Here are examples.

Example 1.9 (Encouragement with an SLP cost function and $I_S \prec_B I_R$)**.**

Suppose there are four states, $p_1 = p = 0.5$, $p_2 = p_3 = w = 0.2$, and $p_4 = q = 0.1$. So there is one likely state, two unlikely states, and one rare state. Assume there are four available actions, and $u(a_i, \omega_j) = 1$ when $i = j$, $u(a_i, \omega_j) = 0$ when $i \neq j$. The cost is the partition cost and $c = 0.26$. In this case, $I_\varnothing = \{\{\omega_1, \omega_2, \omega_3, \omega_4\}\}$, and $I_R = \{\{\omega_1, \omega_4\}, \{\omega_2\}, \{\omega_3\}\}$. When $I_S = \{\{\omega_1, \omega_2, \omega_4\}, \{\omega_3\}\}$, $I_{SR} = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}\}$. It can be shown that $I_{SR} \notin BR(I_\varnothing)$. So $(I_\varnothing, I_R, I_S, I_{SR})$ is an encouragement. ■

Example 1.10 (Encouragement with an SLP cost function and $I_S \not\gtrless_B I_R$)**.**

This example is illustrated in Figure 1.11. The cost function is the entropy cost. Suppose the prior is μ. If the sender reveals no information, the receiver ends up with I_R whose support are γ^1 and γ^2. If the sender reveals I_S, whose supports are γ^3 and γ^4, the receiver keeps learning and ends up with I_{SR}, whose supports are γ^5, γ^6, γ^7, and γ^8. According to Lemma 1.5, it is clear that I_S and I_R are Blackwell incomparable. But the distribution over beliefs, whose supports are γ^5 and γ^7, is a mean preserving spread of γ^1, and the distribution over beliefs, whose supports are γ^6 and γ^8, is a mean preserving spread of γ^2. So $I_{SR} \succ_B I_R$, and thus it is an encouragement.

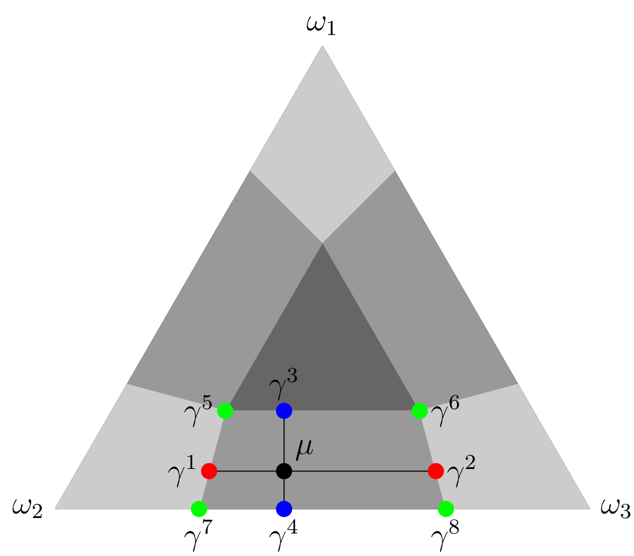


Figure 1.11: Encouragement with an SLP cost function and I_S being Blackwell incomparable with I_R

■

Example 1.11 (Encouragement with a non-SLP cost function and $I_S = I_R$).

Figure 1.12: Encouragement with a non-SLP cost function and $I_S = I_R$

Consider the cost function illustrated in Example 1.3. Suppose $e = 0.7$, $c' = 0.05$ and $\mu = 0.5$. Let the utility function of receiver be $u(a_i, \omega_j) = 1$ when $i = j$, $u(a_i, \omega_j) = 0$ when $i \neq j$. It turns out that if the belief is within the interval represented by the two bars, the receiver pays the cost to access the test. If the belief is outside the interval, the receiver does not learn. Note that the prior, which is represented by the black point, is within the interval. So starting with it, the receiver accesses the test and ends up with the distribution over beliefs that is represented by the two red points and denoted by I_R. Now consider $I_S = I_R$. At the realization represented by the left red point, it is outside the interval, so the receiver does not learn further. At the other realization, the receiver accesses the test. In sum, the receiver ends up with a distribution over beliefs represented by the three blue points, denoted by I_{SR}. Clearly $I_{SR} \succ_B I_R$. So $(I_\varnothing, I_R, I_S, I_{SR})$ is an encouragement. ■

1.9 Discussion and Conclusion

An interesting question remaining is whether the receiver is better off or not in a distortion. On the one hand, the receiver may be worse off, as is shown in the following example.

Example 1.12 (receiver may be worse off in a distortion).

Suppose $U(I_1) = 6$, $U(I_2) = 3$, $C(I, I') = +\infty$ for all $I \neq I'$ except $C(I_\varnothing, I_1) = 2$ and $C(I_\varnothing, I_2) = 1$.

$$U(I_1) - C(I_\varnothing, I_1) > U(I_2) - C(I_\varnothing, I_2)$$

implies $BR(I_\varnothing) = \{I_1\}$. $C(I_2, I_1) = +\infty$ implies $BR(I_2) = \{I_2\}$. As a result, $(I_\varnothing, I_1, I_2, I_2)$ is a distortion. The receiver is worse off because

$$U(I_2) < U(I_1) - C(I_\varnothing, I_1)$$

■

On the other hand, here are some sufficient conditions for the receiver being weakly better off in a distortion.

Proposition 1.1. *If* $(I_\varnothing, I_R, I_S, I_{SR})$ *is a distortion with* $I_S \prec_B I_R$ *and* $C(I_S, I_R) \leqslant C(I_\varnothing, I_R)$*, then the receiver is weakly better off.*

Proposition 1.2. *If* $(I_\varnothing, I_R, I_S, I_{SR})$ *is an encouragement with* $I_S \succsim_B I_R$*, then the receiver is weakly better off.*

A natural avenue for future research is to characterize the condition under which the receiver is better or worse off.

In sum, this paper investigates how a sender, by providing free information to a perfectly rational receiver, can manipulate the receiver's learning. The sender may make the receiver end up knowing more than she would otherwise but, as is shown in the paper, the sender may also make the receiver end up knowing less or different things than she would otherwise. On the one hand, the above finding indicates the motivations for voluntary disclosure may be to disrupt learning: a criminal suspect may take the initiative to reveal some clues to the detective; a cheating cartel participant may turn itself in to the cartel facilitator[5]; a country may voluntarily disclose intelligence to an enemy country. On the other hand, this paper implies that mandatory disclosure does not necessarily make the stakeholders more informed as the regulators expect.

I also find that the feasibility of distortions relies on two properties of the receiver's information acquisition cost function: Sequential Learning Proofness (SLP) and the more restrictive Indifference to Sequential Learning (ISL). When the cost function satisfies ISL, deterrence is infeasible. With Uniformly Posterior Separable cost functions, if there are two actions and two states, then diversion is infeasible. When the cost function satisfies SLP, and the sender reveals all the information that the receiver would learn if the sender revealed no information, encouragement is infeasible.

[5]An entity that fosters collusion, for example, by providing a place to meet for the cartelists, collecting and sharing market data, and working to secure compromises between members of the cartel.

CHAPTER 2

Dominance and Optimality

joint work with Tilman Börgers

2.1 Introduction

Economic theory has developed many notions of "dominance" of some choice over another. For example, in game theory the notions that one strategy "strictly dominates" another one, or that it "weakly dominates" another one, are fundamental concepts (see, for example, Pearce (1984)). In the theory of choice among monetary lotteries the concepts of "first order stochastic dominance" (Quirk and Saposnik, 1962) and "second order stochastic dominance" (Rothschild and Stiglitz, 1970) are frequently used. In the theory of information "Blackwell dominance" among experiments (Blackwell (1951), Blackwell (1953)) is an important concept.

All these concepts of dominance provide partial orders of sets of alternatives among which a decision maker chooses. One alternative dominates another alternative if it is the better choice regardless of certain aspects of the decision maker's decision problem. For example, one strategy dominates another one if it is a better choice regardless of the player's belief about the other players' strategy choices. One monetary lottery first order stochastically dominates another lottery if it yields higher expected utility regardless of the decision maker's utility function, provided that this utility function is increasing. One monetary lottery second order stochastically dominates another lottery if it yields higher expected utility regardless of the decision maker's utility function, provided that this utility function is increasing and concave. One experiment Blackwell dominates another experiment if it allows the decision maker to achieve higher expected utility regardless of which decision problem the decision maker faces.[1]

[1] In this paragraph we have been deliberately vague about whether we refer to "strict" or "weak dominance" and whether we mean by "higher expected utility" that the utility is strictly or weakly larger. We shall, of course, be more precise about these issues later in the paper.

Many famous results of economic theory provide equivalent characterizations of dominance relations that are easier to check than the original definition. For example, to check whether one lottery first order stochastically dominates another lottery one may equivalently compare the two lotteries' cumulative distribution functions (this is the main result in Quirk and Saposnik (1962)). To determine whether one experiment Blackwell dominates another experiment one may equivalently check whether the latter experiment can be obtained by "garbling" the former experiment (Theorem 5 in Blackwell (1953)). Also, to check whether one strategy dominates another one one can equivalently check whether for every pure strategy combination that the other players might choose the former strategy yields higher utility than the latter.[2]

When we eliminate dominated choices in a decision problem we narrow down the options that a rational decision maker might choose without completely specifying all characteristics of this decision maker. In this paper we will ask in a variety of contexts whether ruling out dominated choices is the best we can do without specifying the decision maker's characteristics further, or whether there are other, not dominated choices that a rational decision maker nonetheless will never choose. We prove an abstract theorem that shows that nothing more than dominated choices can be ruled out. Our main result is of the following form: for every not dominated alternative there exists some specification of the decision maker's problem in which a rational decision maker will choose this alternative.

The precise details matter for this result, however. In particular, the result is not true unless we consider the possibility that an alternative is dominated by *a convex combination* of the other alternatives. The potential relevance of convex combinations of alternatives is familiar from game theory: A strategy is a best response to some belief if and only if it is not strictly dominated by any of the other strategies nor by any convex combination of the other strategies (i.e. nor by any mixed strategy). It is well-known that this result would not be true if we had not included the possibility that the dominating strategy is a mixed strategy. In other words: there are well-known examples of games in which a strategy is not strictly dominated by any other pure strategy, yet it is not a best response to any belief of the player. Our main theorem is built on this insight, but applies to a much more general setting than just games.

One such setting is the choice among experiments where one might wish to use Blackwell dominance to rule out some choices. Suppose a decision maker can choose one experiment from a finite set of available experiments. Assume that all experiments are available at no cost. Consider an experiment that is not Blackwell dominated by any other experiment.

[2]Unlike the other results mentioned in this paragraph, this last result is a trivial observation.

Does there exist a decision problem in which it is optimal to choose this experiment from the set of available experiments? It turns out that the answer to this question is "yes" only if we allow for the possibility that the dominating experiment is an appropriately defined convex combination of the other experiments. We give in Section 2.5 a counterexample that demonstrates how the result otherwise fails. On the other hand, once convex combinations of experiments are considered, the result is an immediate implication of our main theorem.

Other details matter. Our most general result will characterize those alternatives that are, for some specification of the details of the decision problem, *the only* optimal choice of the decision maker. We call such alternatives "uniquely optimal." Thus, our main result establishes an equivalence between an alternative not being dominated by convex combinations of other alternatives, and an alternative being a uniquely optimal choice. We also establish equivalences between an alternative not being dominated and the alternative being one optimal choice, or the alternative being an optimal choice in some "non knife-edge" circumstances. These versions of our result are the ones that are familiar from game theory, but they are only applicable under a set of assumptions that are satisfied in finite strategic games, but that are not satisfied when, say, comparing experiments.

One might question our focus on uniquely optimal choices. Of course, sometimes decision makers will face situations in which they are indifferent between several optimal choices. But we shall show that, when there are multiple optimal choices in a decision problem, then at least one of those choices is a uniquely optimal choice in some other decision problem. Therefore, the set of uniquely optimal choices is the minimal set of alternatives to which a decision maker may restrict attention such that this set includes an optimal choice regardless of the specifics of the decision maker's decision problem.

We expand in this paper on the argument explained in the previous paragraph. In a very general decision problem we introduce "minimally sufficient" sets of alternatives, that is, sets of alternatives that are *sufficient* in the sense that whatever the particulars of the decision problem the sets always contain at least one optimal choice, and that are *minimal* in the sense that they have no subset that is also sufficient. We prove that under some conditions the set of uniquely optimal choices is the only minimally sufficient set of alternatives. We motivate minimally sufficient sets as the sets of alternatives that a decision maker would restrict attention to if attention is costly for the decision maker, but the decision maker is not willing to give up any material payoff in the decision problem in return for lower attention cost.

The general result on which we build our analysis is presented in Section 2.2. We motivate a focus on uniquely optimal actions in Section 2.3. In the subsequent sections we apply our general analysis first to dominance relations in games, then to dominance relations among experiments, and finally to dominance relations among monetary lotteries. We mention related literature in each of the applications sections. The general framework in Section 2.2 is conceptually and formally related to some ideas that appeared in Fishburn (1975). In particular, we use one of Fishburn's separation theorems. Fishburn compared optimality of sets of alternatives and dominance among the elements of sets of alternatives. In particular he focused on choice among lotteries. His result specializes to ours in Section 2.6 if one assumes in Fishburn's setting that one of the two sets of alternatives consists of only one element.

2.2 General Results

Let X and Y be two non-empty sets, and let $u : X \times Y \to \mathbb{R}$. Here, X is the set of actions x that a decision maker can choose from, while Y is the set of "situations" y that this decision maker might find herself in. $u(x, y)$ is the decision maker's utility if choosing action x in situation y. The decision maker first observes the situation $y \in Y$ and then chooses an action $x \in X$. Throughout this section we shall make the following assumption:

Assumption 2.1. *X is finite. Y is a convex subset of a topological vector space. u is linear and continuous in y.*

The assumption that X is finite greatly simplifies the analysis below. In the applications that we shall consider, Y is, depending on the context, the set of beliefs the decision maker might hold, or the set of value functions corresponding to the decision problems the decision maker might face, or the set of utility functions the decision maker might have. This is why it is convenient to let Y be a subset of a topological vector space. Y will be convex in all our applications. The linearity of u will reflect that decision makers in our applications are expected utility maximizers. The convexity of Y together with the linearity of u will allow us to use separating hyperplane theorems in our proofs.

Definition 2.1. *An action $x \in X$ is* optimal *if there exists a $y \in Y$ such that:*

$$u(x, y) \geq u(x', y) \text{ for all } x' \in X.$$

We denote by X_O the set of optimal actions.

Definition 2.2. *An action $x \in X$ is* interior optimal *if there exists a $y \in ri(Y)$ (where $ri(Y)$ denotes the relative interior of Y) such that:*

$$u(x, y) \geq u(x', y) \text{ for all } x' \in X.$$

We denote by X_{IO} the set of interior optimal actions. It may not seem intuitively obvious why it is relevant whether the situation y in which an action x is a best response is interior or not. However, when Y represents a set of beliefs, then the relative interior of Y will represent the set of full support beliefs. Having beliefs with full support has been interpreted in the game theoretic literature as sign of caution by the decision maker. This is why interior situations y will receive special attention in this section.

Definition 2.3. *An action $x \in X$ is* uniquely optimal *if there exists a $y \in Y$ such that:*

$$u(x, y) > u(x', y) \text{ for all } x' \in X \text{ such that } x \neq x'.$$

We denote by X_{UO} the set of uniquely optimal actions. It is not immediately obvious why the set of uniquely optimal actions should receive special attention. We address this issue therefore in detail in the next section.

Our objective in this section is to characterize the sets of optimal, interior optimal, and uniquely optimal actions in terms of dominance notions. We therefore next introduce the dominance notions that we are considering.

Definition 2.4. *An action $x \in X$ is* strictly dominated *if there are a set $\{x_1, x_2, \ldots, x_n\} \subseteq X \setminus \{x\}$ and a vector $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n_+$ with $\sum_{i=1}^n \lambda_i = 1$ such that:*

$$\sum_{i=1}^{n} (\lambda_i u(x_i, y)) > u(x, y) \text{ for all } y \in Y.$$

We denote by X_{NSD} the set of all actions that are not strictly dominated. It is essential that we consider the possibility here that an action is dominated not by a single action but by a convex combination of actions. We shall illustrate this point in the applications that we consider later in the paper. One may think of the convex combination of actions as a "mixed action" in analogy to mixed strategies in game theory.

Definition 2.5. *An action* $x \in X$ *is* weakly dominated *if there are a set* $\{x_1, x_2, \ldots, x_n\} \subseteq X \setminus \{x\}$ *and a vector* $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n_+$ *with* $\sum_{i=1}^n \lambda_i = 1$ *such that:*

$$\sum_{i=1}^{n} (\lambda_i u(x_i, y)) \geq u(x, y) \text{ for all } y \in Y$$

with strict inequality for at least one $y \in Y$.

We denote by X_{NWD} the set of all actions that are not weakly dominated.

Strict and weak dominance are standard notions that are familiar from game theory. We will use a third concept that is less familiar, but that will prove crucial for some of our results.

Definition 2.6. *An action* $x \in X$ *is* redundant *if there are a set* $\{x_1, x_2, \ldots, x_n\} \subseteq X \setminus \{x\}$ *and a vector* $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n_+$ *with* $\sum_{i=1}^n \lambda_i = 1$ *such that:*

$$\sum_{i=1}^{n} (\lambda_i u(x_i, y)) = u(x, y) \text{ for all } y \in Y.$$

Thus, an action is redundant if it is equivalent in expected utility to a convex combination of the other actions. Denote by X_{NR} the set of actions that are not redundant.

We are now ready to state our main result in this section. This result generalizes a number of results familiar from the literature on game theory.

Theorem 2.1. *(i) If an action is optimal, then it is not strictly dominated:* $X_O \subseteq X_{NSD}$.

(ii) If Y *is compact, then an action that is not strictly dominated is optimal:* $X_{NSD} \subseteq X_O$.

(iii) If an action is interior optimal then it is not weakly dominated: $X_{IO} \subseteq X_{NWD}$.

(iv) If Y *is finite dimensional, then an action that is not weakly dominated is interior optimal:* $X_{NWD} \subseteq X_{IO}$.

(v) An action is uniquely optimal if and only if it is not weakly dominated and not redundant: $X_{UO} = X_{NWD} \cap X_{NR}$.

Note that results (ii) and (iv) are based on assumptions regarding Y that go beyond those made in Assumption 1. Our proofs use these additional assumptions. Of course, the theorem does not claim that these additional assumptions are necessary.

Our proof of Theorem 1 below is based on elementary separating hyperplane theorems in finite dimensional Euclidean space. Thus, it is a simple and geometric proof. The ideas on which some parts of the proof are based are related to ideas in Fishburn (1975).

Proof. STEP 1: We first prove that $X_O \subseteq X_{NSD}$. The proof is indirect. Suppose x were optimal for some $\bar{y} \in Y$, but that x were strictly dominated. Let $\{x_1, x_2, \ldots, x_n\} \subseteq X \backslash \{x\}$ be the actions in the support of the strictly dominating convex combination, and let $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ be the corresponding weights. We have: $\sum_{i=1}^{n} (\lambda_i u(x_i, y)) > u(x, y)$ for all $y \in Y$. But this implies that for some i we have: $u(x_i, \bar{y}) > u(x, \bar{y})$, contradicting that x is optimal in situation $\bar{y}$.

We next prove that $X_{IO} \subseteq X_{NWD}$. The proof is indirect. Suppose x were optimal for some $\bar{y} \in ri(Y)$, but that x were weakly dominated. Let $\{x_1, x_2, \ldots, x_n\} \subseteq X \setminus \{x\}$ be the actions in the support of the weakly dominating convex combination, and let $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ be the corresponding weights. We then have: $\sum_{i=1}^{n} (\lambda_i u(x_i, y)) \geq u(x, y)$ for all $y \in Y$ and $\sum_{i=1}^{n} (\lambda_i u(x_i, y^*)) > u(x, y^*)$ for some $y^* \in Y$. Because x is optimal at $\bar{y}$, we must have: $\sum_{i=1}^{n} (\lambda_i u(x_i, \bar{y})) - u(x, \bar{y}) = 0$. Define: $\hat{y} \equiv (1 + \varepsilon)\bar{y} - \varepsilon y^*$. Because $\bar{y}$ is in the relative interior of Y, we have $\hat{y} \in Y$ for sufficiently small $\varepsilon > 0$.. Now note that:

$$
\begin{aligned}
& \sum_{i=1}^{n} (\lambda_i u(x_i, \hat{y})) - u(x, \hat{y}) \\
= \;& (1 + \varepsilon) \left(\sum_{i=1}^{n} (\lambda_i u(x_i, \bar{y})) - u(x, \bar{y}) \right) - \varepsilon \left(\sum_{i=1}^{n} (\lambda_i u(x_i, y^*)) - u(x, y^*) \right) \\
= \;& -\varepsilon \left(\sum_{i=1}^{n} (\lambda_i u(x_i, y^*)) - u(x, y^*) \right) < 0,
\end{aligned}
$$

This contradicts the assumption that the convex combination weakly dominates x.

We finally prove that $X_{UO} \subseteq X_{NWD} \cap X_{NR}$. The proof is indirect. Suppose x were uniquely optimal for some particular $\bar{y} \in Y$. but that x were weakly dominated. Let $\{x_1, x_2, \ldots, x_n\} \subseteq X \setminus \{x\}$ be the actions in the support of the weakly dominating convex combination, and let $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ be the corresponding weights. We have: $\sum_{i=1}^{n} (\lambda_i u(x_i, y)) \geq u(x, y)$ for all $y \in Y$. But this implies that for some i we have: $u(x_i, \bar{y}) \geq u(x, \bar{y})$, contradicting that x is uniquely optimal in situation $\bar{y}$. The same argument proves that a uniquely optimal action cannot be redundant.

STEP 2: We now prove the converses of the statements in Step 1. We first prove that $X_{NSD} \subseteq X_O$ if Y is compact. The proof is indirect. Suppose x were not optimal for any $y \in Y$. We prove that then a convex combination of actions strictly dominates x. Consider the following set.

$$C \equiv \{(u(x_1, y) - u(x, y), u(x_2, y) - u(x, y), ..., u(x_n, y) - u(x, y)) \,|y \in Y\},$$

where we take $x_1, x_2, \ldots, x_n$ to be an enumeration of the set $X \setminus \{x\}$. If x is not optimal in any situation $y \in Y$, then:

$$C \cap \mathbb{R}^n_- = \emptyset.^3$$

Observe that $\mathbb{R}^n_-$ is a closed and convex set, and that C is convex and compact (convex because of the linearity of u and the convexity of Y and compact because of the continuity of u and the compactness of Y). We can then apply the following hyperplane theorem:

Separating Hyperplane Theorem 1: Suppose $C \subseteq \mathbb{R}^n$ is convex and compact. If $C \cap \mathbb{R}^n_- = \emptyset$ then there exists $\lambda \in \mathbb{R}^n_+$ with $\lambda \neq 0$ such that $\lambda \cdot x > 0$ for all $x \in C$.

For completeness, we briefly derive this result from a standard separating hyperplane theorem:

Proof. The strict separating hyperplane theorem (for example Theorem 3.7 in Vohra (2005)) implies that there are $\lambda \in \mathbb{R}^n$ with $\lambda \neq 0$ and $\varepsilon \in \mathbb{R}$ such that $\lambda \cdot x < \varepsilon$ for all $x \in \mathbb{R}^n_-$ and $\lambda \cdot x > \varepsilon$ for all $x \in C$. It easily follows that $\lambda \in \mathbb{R}^n_+$ and that $\varepsilon > 0$. Therefore, $\lambda \cdot x \geq \varepsilon$ for all $x \in C$ implies $\lambda \cdot x > 0$ for all $x \in C$. □

Obviously, we may normalize the vector λ to which the theorem refers so that its components add up to 1. Applying the theorem to our setting, we therefore find that there is a convex combination of actions such that:

$$\sum_{i=1}^{n} [\lambda_i (u(x_i, y) - u(x, y))] > 0 \text{ for all } y \in Y.$$

This means that the convex combination of the actions $x_1, x_2, \ldots, x_n$ with weights $\lambda_1, \lambda_2, \ldots, \lambda_n$ strictly dominates x.

Next we prove that $X_{NWD} \subseteq X_{IO}$ if Y is finite dimensional. If Y is finite dimensional then it is without loss of generality to assume that it is a subset of a finite dimensional

[3] We denote by $\mathbb{R}^n_-$ the set of all vectors $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ such that $x_i \leq 0$ for all i.

Euclidean space. Our proof of the result is again indirect. Suppose x were not optimal for any $y \in ri(Y)$. Define the set C as before. Because u is linear in y, the relative interior of C is the image of the relative interior of Y. Therefore we have:

$$ri(C) \cap \mathbb{R}^n_- = \emptyset.$$

We now apply the following separating hyperplane theorem:

Separating Hyperplane Theorem 2: Suppose $C \subseteq \mathbb{R}^n$ is convex. If $ri(C) \cap \mathbb{R}^n_{\leq 0} = \emptyset$ then there exists $\lambda \in \mathbb{R}^n_+$ with $\lambda \neq 0$ such that $\lambda \cdot x \geq 0$ for all $x \in C$ and $\lambda \cdot x > 0$ for at least one $x \in C$.

Proof. By Theorem 6.2 of Rockafellar (1970), $ri(C)$ is non-empty and convex. By Lemma 5 in Fishburn (1975), there exists a $\lambda \in \mathbb{R}^n_+$ with $\lambda \neq 0$ such that $\lambda \cdot x \geq 0$ for all $x \in ri(C)$ and $\lambda \cdot x > 0$ for at least one $x \in ri(C)$. By continuity, $\lambda \cdot x \geq 0$ for all x in the topological closure of $ri(C)$, and by Theorem 6.3 in Rockafellar (1970), the topological closure of $ri(C)$ is a superset of C. Therefore, $\lambda \cdot x \geq 0$ for all $x \in C$. Finally, because $ri(C) \subseteq C$, we have $\lambda \cdot x > 0$ for at least one $x \in C$. □

Normalizing again the vector λ to which the theorem refers so that its components add up to 1, we find that there is a convex combination of actions such that:

$$\sum_{i=1}^{n} \left[\lambda_i \left(u(x_i, y) - u(x, y)\right)\right] \geq 0 \text{ for all } y \in Y,$$

with strict inequality for at least one $y \in Y$. This means that the convex combination of the actions $x_1, x_2, \ldots, x_n$ with weights $\lambda_1, \lambda_2, \ldots, \lambda_n$ weakly dominates x.

We finally prove that $X_{NWD} \cap X_{NR} \subseteq X_{UO}$. The proof is indirect. Suppose x were not uniquely optimal for any $y \in Y$. Defining the set C as before, this means that:

$$C \cap \mathbb{R}^n_{<0} = \emptyset.^4$$

We now apply the following separating hyperplane theorem:

Separating Hyperplane Theorem 3: Suppose $C \subseteq \mathbb{R}_n$ is convex. If $C \cap \mathbb{R}^+_{<0} = \emptyset$ then there exists $\lambda \in \mathbb{R}^n_+$ with $\lambda \neq 0$ such that $\lambda \cdot x \geq 0$ for all $x \in C$.

[4] We denote by $\mathbb{R}^n_{<0}$ the set of all vectors $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ such that $x_i < 0$ for all i.

Proof. Minkowski's separating hyperplane theorem (for example Theorem 3.6 in Vohra (2005) implies that there are $\lambda \in \mathbb{R}^n$ with $\lambda \neq 0$ and $\varepsilon \in \mathbb{R}$ such that $\lambda \cdot x \leq \varepsilon$ for all $x \in \mathbb{R}^n_{<0}$ and $\lambda \cdot x \geq \varepsilon$ for all $x \in C$. It easily follows that $\lambda \in \mathbb{R}^n_+$ and $\varepsilon \geq 0$. Therefore, $\lambda \cdot x \geq \varepsilon$ for all $x \in C$ implies $\lambda \cdot x \geq 0$ for all $x \in C$. □

Normalizing again the vector λ to which the theorem refers so that its components add up to 1, we find that there is a convex combination of actions such that:

$$\sum_{i=1}^{n} \lambda_i u(x_i, y) \geq u(x, y) \text{ for all } y \in Y.$$

Thus, the convex combination of $\{x_1, \ldots, x_n)$ with weights $\lambda_1, \ldots, \lambda_n$ either weakly dominates x or is equivalent to x, which contradicts the assumption with which we began this indirect proof. □

2.3 Limited Attention Without Loss of Optimality

We now provide a rationale for focusing on uniquely optimal actions as defined in the previous section. We provide conditions under which the set of uniquely optimal actions is the smallest set that the decision maker can limit attention to if she wants to choose an optimal action in every situation. We thus envisage the following scenario: the decision maker first restricts attention to some subset $\widehat{X}$ of the set of all actions. Then she observes the situation y, and then she picks an action x from the subset of actions to which she has restricted attention. We assume that attention is costly: the decision maker wants to restrict attention to a set that is small (in terms of set-inclusion). Finally, attention cost are of second order importance: the decision maker's first priority is to take an optimal action in every situation.

To formalize this, we introduce some additional notation. For every $y \in Y$ the set of optimal actions is:

$$O(y) = \{x \in X | u(x, y) \geq u(x', y) \text{ for all } x' \in X\}.$$

In the following definition, the key term that we wish to define is "minimal sufficiency,"

Definition 2.7. *A set* $\widehat{X} \subseteq X$ *is* sufficient *if for every* $y \in Y$*:*

$$O(y) \cap \hat{X} \neq \emptyset.$$

A set $\widehat{X} \subseteq X$ *is* minimally sufficient *if it is sufficient and there is no sufficient set* $\bar{X} \subseteq X$ *such that:*

$$\bar{X} \subsetneq \hat{X}.$$

A minimally sufficient set of actions is thus a smallest set of actions to which the decision maker may restrict attention if she wants to choose optimally in every situation y.

The following result provides sufficient conditions for the set of uniquely optimal actions to be the unique minimally sufficient subset of the set of all actions. We emphasize that this result does not rely on Assumption 2.1.

Theorem 2.2. *If* X *is finite, if for any* $x, x' \in X$ *with* $x \neq x'$ *there is a* $y \in Y$ *such that* $u(x, y) \neq u(x', y)$*, and if:*

$$X_{UO} = X_{NWD} \cap X_{NR},$$

then X_{UO} *is the unique minimally sufficient subset* $\widehat{X}$ *of* X*.*

Among the three assumptions of this theorem, the condition $X_{UO} = X_{NWD} \cap X_{NR}$ is not formulated in terms of the primitives of our model. However, Theorem 2.1 shows assumptions for the primitives of our model that imply that $X_{UO} = X_{NWD} \cap X_{NR}$ holds.

Proof. It is obvious that every sufficient set must include X_{UO}. What remains to be shown is that X_{UO} contains for every situation $y \in Y$ an optimal action. Because the proposition assumes that $X_{UO} = X_{NWD} \cap X_{NR}$, this is equivalent to the statement that $X_{NWD} \cap X_{NR}$ contains for every situation $y \in Y$ an optimal action.

To prove this, we first observe that $X_{NWD} \cap X_{NR}$ can be constructed by the following algorithm. Set $X^0 = X$. For $k = 1, 2, \ldots, n$, if no action $x \in X^{k-1}$ is either weakly dominated by, or equivalent to, a convex combinations of the actions in $X^{k-1} \setminus \{x\}$, then set $X^{k-1} = X^k$. Otherwise, pick arbitrarily some such $x \in X^{k-1}$, and set $X^k = X^{k-1} \setminus \{x\}$. Note that this algorithm ends after at most n steps. We claim that the final set is $X^n = X_{NWD} \cap X_{NR}$. It is clear that X^n includes all actions in $X_{NWD} \cap X_{NR}$. It thus remains to show that any action $x \notin X_{NWD} \cap X_{NR}$ is eliminated in some step of this algorithm.

To prove the claim we show that if x is weakly dominated by or equivalent to a convex combination of the actions in $X^k \setminus \{x\}$, and if $x \in X^{k+1}$, then x it is also weakly dominated or equivalent to a convex combination of the actions in $X^{k+1} \setminus \{x\}$. Suppose that the convex combination of actions in $X^k \setminus \{x\}$ that weakly dominate x include x^k. Without loss of generality let the elements of $X^k \setminus \{x\}$ be $\{x_1, x_2, \ldots, x_m\}$, and let the weights of the convex combination be $\lambda_1, \lambda_2, \ldots, \lambda_m$. Without loss of generality assume that x_1 is eliminated in step k. This means that x_1 is either weakly dominated or equivalent to, a convex combination of the elements of the set $\{x_2, \ldots, x_m\} \cup \{x\}$. Let the weights of the convex combination be: $\hat{\lambda}_2, \ldots, \hat{\lambda}_m, \hat{\lambda}_x$. It is then obvious that x is also weakly dominated, or equivalent to, a convex combination of $\{x_2, \ldots, x_m\} \cup \{x\}$ with weights: $\lambda_2 + \lambda_1\hat{\lambda}_2, \ldots, \lambda_m + \lambda_1\hat{\lambda}_m, \lambda_1\hat{\lambda}_x$.

It remains to show $\lambda_1\hat{\lambda}_x < 1$. Suppose the opposite: $\lambda_1\hat{\lambda}_x = 1$, hence $\lambda_1 = \hat{\lambda}_x = 1$. This means that x_1 is either weakly dominated or equivalent to x, and that x is either weakly dominated or equivalent to x_1. But this means that x_1 and x are duplicates, which is a case that we ruled out in the assumptions of Theorem 2.2.

We can now infer that, if x is weakly dominated by, or equivalent to, a convex combination of $\{x_2, \ldots, x_m\} \cup \{x\}$ with weights:

$$\lambda_2 + \lambda_1\hat{\lambda}_2, \ldots, \lambda_m + \lambda_1\hat{\lambda}_m, \lambda_1\hat{\lambda}_x$$

it is also weakly dominated by , or equivalent to, a convex combination of $\{x_2, \ldots, x_m\}$ with weights:

$$\frac{\lambda_2 + \lambda_1\hat{\lambda}_2}{1 - \lambda_1\hat{\lambda}_x}, \ldots, \frac{\lambda_m + \lambda_1\hat{\lambda}_m}{1 - \lambda_1\hat{\lambda}_x}.$$

Now consider any situation $y \in Y$ and suppose $x \in X$ is optimal in situation y. Obviously, $x \in X^0$. Also, either $x \in X^1$, or x is weakly dominated by, or equivalent to, a convex combination of the actions in X^1. Then one of the actions in X^1 must also be a best response to y. Iterating this argument leads to the conclusion that one of the actions in X^n is optimal in situation y. □

2.4 Dominance and Optimality of Strategies in Games

We now explain how to apply the results of Section 2.2 to games. We focus on a player i in a strategic game who has to choose one strategy from a finite set of strategies S_i. There are finitely many other players $j \neq i$ and the Cartesian product of their strategy sets is S_{-i}. Player i's utility function is $u_i : S_i \times S_{-i} \to \mathbb{R}$. Player i's belief about the other players' choices is a probability measure μ_i on S_{-i}. Denote the set of all such probability measures by $\Delta(S_{-i})$.[5] Player i's expected utility when she has belief μ_i and chooses strategy s_i is:

$$\sum_{s_{-i} \in S_{-i}} u_i(s_i, s_{-i})\mu_i(s_{-i}).$$

In the notation of Section 2.2 in this application the set X of actions is the set S_i of strategies, and the set Y of situations is the set $\Delta(S_{-i})$ of beliefs. Note that the set of beliefs is a compact and convex subset of $\mathbb{R}^{|S_{-i}|}$, and that the utility function u_i is linear (and therefore continuous) in a player's beliefs.

The notions of strict and weak dominance introduced in Section 2.2 correspond to the thus named notions in game theory. Note that when checking dominance relations among two strategies s_i and s_j in games we typically only compare expected utility for any given pure strategy combination of the other players, that is, we only consider beliefs that are Dirac measures on S_{-i}. This is sufficient because $\Delta(S_{-i})$ the convex hull of the set of Dirac measures on S_{-i} and because of the linearity of expected utility. This is a general point: The properties of actions $x \in X$ defined in Definitions 2.4, 2.5, and 2.6 would not change if one replaced the expression "for all $y \in Y$" by the expression "for all $y \in Y^*$" where Y^* is a subset of Y such that the convex hull of Y^* equals Y.

All parts of Theorem 2.1 hold in this setting. In particular, note that part (ii) applies because the set $\Delta(S_{-i})$ is compact. Parts (i) and (ii) of Theorem 2.1, if applied to finite strategic games, are the same as Lemma 3 in Pearce (1984) whose proof was different from ours, however. Pearce's proof was built on the existence of Nash equilibria in zero sum games. As regards parts (iii) and (iv) of Theorem 2.1 note that an element of the relative interior of $\Delta(S_{-i})$ is a full support belief, and therefore parts (iii) and (iv) of Theorem 2.1 correspond to Lemma 4 in Pearce (1984). Again, the proof in Pearce (1984) is different from ours. Finally, part (v) of Theorem 2.1, if applied to strategic games, is a special case of Proposition 3 in Weinstein (2020), who, like Pearce, presents a proof that is built on the theory of zero-sum games.

It is well-known that the results listed above for strategic games would not be true if one considered dominance by pure strategies only, not by mixed strategies. Mixed strategies are

[5]From now on, for any finite or compact set A, we denote by $\Delta(A)$ the set of all (Borel-) probability distributions on A.

the equivalent of the "convex combinations" in Section 2.2. As an example, consider the two player game in Table 2.1, where only player 1's utility is shown. Player 1's strategy B is strictly dominated, but not by any pure strategy. It is strictly dominated by, for example, the mixed strategy that places probability 0.5 on T and M.

	L	R
T	3	0
M	0	3
B	1	1

Table 2.1: A strategy that is not strictly dominated by any pure strategy may be strictly dominated by a mixed strategy

2.5 Dominance and Optimality of Experiments

In this section we apply the results of Section 2 to experiments. Let Ω be a finite set of states of the world, and let $\mu \in \Delta(\Omega)$ be a decision maker's prior belief about the state. The decision maker can observe a signal about the state of the world before making a decision. Here, we mean by a signal a mapping: $s : \Omega \to \Delta(M_s)$ a signal, where M_s is a finite set of signal realizations and $s(\omega) \in \Delta(M_s)$ is the distribution of signal realizations conditional on ω. There is a finite set $\mathcal{S}$ of such signals from which the decision maker must choose one. Signals are costless.

The decision maker faces a decision problem $(A, \mathfrak{u})$. Here, A is a finite set of actions and $\mathfrak{u} : A \times \Omega \to \mathbb{R}$ is a von Neumann Morgenstern utility function. We denote by $\mathcal{A}$ the set of all such decision problems.

For every signal $s \in \mathcal{S}$ we denote by $\mu_s \in \Delta(\Delta(\Omega))$ the corresponding distribution of posterior beliefs. For every decision problem $(A, \mathfrak{u})$ we denote by $v_{A,\mathfrak{u}} : \Delta(\Omega) \to \mathbb{R}$ the value function:

$$v_{A,\mathfrak{u}}(\nu) = \max_{a \in A} \sum_{\omega \in \Omega} \left(\mathfrak{u}(a, \omega)\nu(\omega)\right).$$

Here, ν stands for an arbitrary posterior belief of the decision maker. If the decision maker faces decision problem $(A, \mathfrak{u})$, has access to signal s before choosing an action, and chooses an action that maximizes her expected utility, she obtains ex ante expected utility:

$$\int_{\Delta(\Omega)} v_{A,\mathfrak{u}}(\nu) d\mu_s.$$

Blackwell introduced a partial order over signals. Blackwell (1951) and Blackwell (1953) showed various conditions all to be equivalent to the original definition of the Blackwell order. In Definition 2.8, we don't present Blackwell's original definition of the order, but we use one of the conditions that Blackwell showed to be equivalent to the original definition to define the Blackwell order. This is more in line with the way in which dominance orders in other areas of economics are conventionally defined.

Definition 2.8. *Signal s Blackwell dominates signal $\hat{s}$ if for every decision problem $(A, \mathfrak{u})$ in $\mathcal{A}$:*

$$\int_{\Delta(\Omega)} v_{A,\mathfrak{u}}(\nu) d\mu_s \geq \int_{\Delta(\Omega)} v_{A,\mathfrak{u}}(\nu) d\mu_{\hat{s}}.$$

We now explain how to fit Blackwell dominance into our framework. The set X is the set of signals among which the decision maker can choose. The set Y is the set of all value functions that correspond to a decision problem in $\mathcal{A}$. We endow this set with the standard vector space structure and with the topology of uniform convergence. Observe that the set Y is convex. To see this note that the convex combination of two value functions corresponding to decision problem $(A, \mathfrak{u})$ and decision problem $(A', \mathfrak{u}')$ with weights λ and $1-\lambda$ is the value function for the decision problem in which the decision maker chooses from $A \times A'$ and with probability λ the first choice matters, and utility is given by $\mathfrak{u}$, and with probability $1-\lambda$ only the second choice matters, and utility is given by $\mathfrak{u}'$.

The utility function $u(x, y)$ from Section 2 is in the setting of this section the expected utility $\int_{\Delta(\Omega)} v_{A,\mathfrak{u}}(\nu) d\mu_s$. Note that this utility function is linear and continuous in the value function.

We explain next how in this setting convex combinations of signals can be interpreted as signals in themselves. For the purposes of this discussion we assume that no two signals in $\mathcal{S}$ have overlapping message sets: $M_s \cap M_{\hat{s}} = \emptyset$ for every $s, \hat{s} \in \mathcal{S}$ with $s \neq \hat{s}$. This is not a substantial assumption. Rather, this assumption allows us to simplify the notation in the following definition.

Definition 2.9. *Suppose* $\{s_1, s_2, \ldots, s_n\} \subseteq \mathcal{S}^*$ *and assume that the vector* $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}_+$ *satisfies* $\sum_{i=1}^n \lambda_i = 1$. *The* convex combination of the signals $\{s_1, s_2, \ldots, s_n\}$ with weights $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ *is the signal*

$$s : \Omega \to \Delta\left(\bigcup_{i=1}^n M_{s_i}\right)$$

such that for every $\omega \in \Omega$, *every* $i \in \{1, 2, \ldots, n\}$, *and every* $m_{s_i} \in M_{s_i}$, *we have:*

$$s(m_{s_i}|\omega) = \lambda_i s_i(m_{s_i}|\omega).$$

Intuitively, a convex combination of the signals in set $\{s_1, s_2, \ldots, s_n\}$ with weights $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is the following signal: The decision maker observes the message of signal s_i with probability α_i. This signal yields for the decision maker the expected utility that we attribute to the convex combination of actions in Section 2. Therefore, in our context, we can interpret a convex combination of signals as another signal.

Observe that, in the terminology of Section 2.2 a signal is Blackwell dominated if and only if it is either weakly dominated by a convex combination of other signals or is redundant.

Items (i), (iii) and (v) of Theorem 2.1 apply directly to our setting. By contrast, items (ii) and (iv) do not apply. This is because the set of all value functions generated by finite action problems is neither compact nor finite dimensional. Observe that claims (i) and (ii) are, however, vacuously true for the Blackwell order. This is because no signal ever strictly dominates another signal. This is because we have not ruled out from consideration those decision problems $(A, \mathfrak{u}) \in \mathcal{A}$ in which the utility function u does not depend on the state ω. All signals are useless in such decision problems, and therefore, no signal is ever *strictly better* than another signal in *all* decision problems. Parts (iii) and (iv) are in our setting in this section of no interest because in the setting of this section it is easily checked that the relative interior of the set Y with respect to the topology of uniform convergence is empty. We therefore focus now on part (v) of Theorem 2.1.

For maximum clarity we translate the definition of unique optimality, and the assertion of part (v) of Theorem 2.1 to our setting.

Definition 2.10. *A signal* $s \in \mathcal{S}$ *is a uniquely optimal choice in decision problem* $(A, \mathfrak{u})$ *if:*

$$\int_{\Delta(\Omega)} v_{A,\mathfrak{u}}(\nu) d\mu_s > \int_{\Delta(\Omega)} v_{A,\mathfrak{u}}(\nu) d\mu_{\hat{s}} \text{ for all } \hat{s} \in \mathcal{S} \text{ with } \hat{s} \neq s.$$

Part (v) of Theorem 2.1 says in our setting:

Proposition 2.1. *Signal $s \in \mathcal{S}$ is a uniquely optimal choice in some decision problem $(A, \mathfrak{u})$ if and only if it is not Blackwell dominated by any convex combination of signals in $\mathcal{S} \setminus \{s\}$.*

We now show by means of an example that the result would not be true if we replace did not allow the Blackwell dominating signal to be a convex combination of the other signals, but required the Blackwell dominating signal to be one of the other signals. Consider the following example: $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\mu(\omega) = \frac{1}{3}$ for all $\omega \in \Omega$, $\mathcal{S} = \{s_1, s_2, s_3, s_4\}$. $M_i = \{m_i^1, m_i^2\}$ for $i = 1, 2, 3$, and $M_4 = \{m_4^1, m_4^2, \ldots, m_4^6\}$. For each of the signals s_1, s_2, and s_3, and for each state of the world, the corresponding row in Table 2.2 below indicates the conditional probability of observing each signal realization.

s_1	m_1^1	m_1^2
ω_1	1	0
ω_2	0	1
ω_3	0	1

s_2	m_2^1	m_2^2
ω_1	0	1
ω_2	1	0
ω_3	0	1

s_3	m_3^1	m_3^2
ω_1	0	1
ω_2	0	1
ω_3	1	0

Table 2.2: Conditional distributions of s_1, s_2, and s_3

Table 2.3 provides the same information for signal 4.

s_4	m_4^1	m_4^2	m_4^3	m_4^4	m_4^5	m_4^6
ω_1	$\frac{1}{4}$	0	0	$\frac{3}{8}$	$\frac{3}{8}$	0
ω_2	0	$\frac{1}{4}$	0	$\frac{3}{8}$	0	$\frac{3}{8}$
ω_3	0	0	$\frac{1}{4}$	0	$\frac{3}{8}$	$\frac{3}{8}$

Table 2.3: Conditional distributions of s_4

We claim that s_4 is not Blackwell dominated by any of s_1, s_2, s_3, but that it is Blackwell dominated by the convex combination of these three signals that places weight 1/3 on each of these signals. To see that s_4 is not Blackwell dominated by s_1 note that signal s_4 has a realization (m_4^2) which reveals that the true state is ω_2, whereas s_1 has no such realization. This implies that s_4 cannot be Blackwell dominated by s_1. Analogous arguments show that s_4 is not Blackwell dominated by s_2 or s_3.

To see that s_4 is Blackwell dominated by the convex combination of signals s_1, s_2, s_3 that places probability 1/3 on each of these signals we consider the distribution of posterior beliefs generated by s_4 and compare it to the distribution of posterior beliefs generated by the convex combination. For each state in Ω signal 4 generates with probability 1/12 a posterior belief that is a Dirac measure on this state. For each state in Ω the convex combination of signals s_1, s_2, s_3 generates with probability 1/9 ($> 1/12$) a posterior belief that is a Dirac measure on this state. Also, for each pair of states in Ω, signal 4 generates with probability 1/4 a posterior belief that places probability 1/2 on each of the two states in this pair. For each pair of states in Ω the convex combination of signals s_1, s_2, s_3 generates with probability 2/9 ($< 1/8$) a posterior belief that places probability 1/2 on two of the three states. One can now easily show that the distribution of posterior beliefs under the convex combination of signals is a mean-preserving spread of the distribution of posterior beliefs that is generated by signal s_4. By standard results, this implies that s_4 is Blackwell dominated by the convex combination of signals s_1, s_2, s_3 that places probability 1/3 on each of these signals.

Because s_4 is Blackwell dominated by a convex combination of s_1, s_2 and s_3 in every decision problem one of signals s_1, s_2, or s_3 yields at least as high expected utility as s_4. Yet s_4 is not Blackwell dominated by any single of the signals s_1, s_2 and s_3. The example in Tables 2.2 and 2.3 is therefore the analogue for signals of the example in Table 2.1.

We conclude this section by briefly considering briefly some alternatives to the Blackwell order and the applicability of the results in Theorem 2.1 to these orders. Suppose that the set of states Ω is a lattice. We can then restrict attention to decision problems such that set of actions A is a finite lattice and the utility function u is supermodular in the product lattice on $\Omega \times A$. Let us call such decision problems "monotone," and let us denote the set of all such decision problems by $\mathcal{M}$.

Definition 2.11. *Signal s* monotonically dominates *signal $\hat{s}$ if for every monotone decision problem $(A, \mathfrak{u}) \in \mathcal{M}$:*

$$\int_{\Delta(\Omega)} v_{A,\mathfrak{u}}(\nu) d\mu_s \geq \int_{\Delta(\Omega)} v_{A,\mathfrak{u}}(\nu) d\mu_{\hat{s}}.$$

We can apply part (v) of Theorem 2.1 to conclude:.

Proposition 2.2. *Signal $s \in \mathcal{S}$ is a uniquely optimal choice in some monotone decision problem $(A, \mathfrak{u})$ if and only if it is not monotonically dominated by any convex combination of signals in $\mathcal{S} \setminus \{s\}$.*

The order that we have introduced in Definition 2.11 is closely related to the orders introduced in Lehmann (1988) and Athey and Levin (2018) but it does not coincide with either of these. Lehmann, and also Athey and Levin, assume the action set to be a subset of the set of real numbers, and thus they assume the action set to be completely ordered. With this assumption the argument that we used above to show that the set of value functions is convex no longer applies. This is because that argument involved creating a new decision problem from two given decision problems in which the decision maker's action set was the Cartesian product of the original action sets. It was important that this new decision problem was included in the set of admissible decision problems. But if action sets have to be one-dimensional, this argument no longer holds. Kim (2022) allows multi-dimensional action sets but imposes joint conditions on signals and decision problems. A detailed consideration of his order is outside of the scope of this paper.

One might also modify the Blackwell order by considering not only the value functions that are generated by finite decision problems, but instead *all* value functions that are convex in the posterior, and in addition normalize value functions, so that every value function must assume 0 as the minimum expected utility and 1 as the maximum expected utility. Note that this rules out constant value functions. With this construction, the concept of strict dominance among signals is no longer vacuous. For example, a perfectly informative signal will strictly dominate a completely uninformative signal. A statistical characterization of this dominance relation among signals is left for future work.

2.6 Dominance and Optimality of Monetary Lotteries

Consider an expected utility maximizer who chooses one lottery from a finite set of lotteries. Here, a lottery is a probability distribution over $\mathbb{R}$. One lottery first order stochastically dominates another lottery if the former lottery yields at least as high expected utility as the latter provided that the decision maker's utility is non-decreasing in money. One lottery second order stochastically dominates another lottery if the former lottery yields at least as high expected utility as the latter provided that the decision maker's utility is non-decreasing and concave in money.[6]

To apply the results of Section 2.2 we let the set X of actions be the set of monetary lotteries that the decision maker can choose from, and we let the set Y be the set of non-decreasing (in the case of first order stochastic dominance), or the set of non-decreasing and concave (in the case of second order stochastic dominance) utility functions with domain $\mathbb{R}$. The utility function u from Section 2 is then the expected utility, that is, if x is the lottery with cumulative distribution function F, and y is the utility function $\mathfrak{u} : \mathbb{R} \to \mathbb{R}$, then

$$u(x, y) = \int \mathfrak{u}(z) dF.$$

In this specification, the set Y is a convex set, although it is not compact, and has an empty relative interior. The utility function u is linear in $\mathfrak{u}$ and continuous in the topology of uniform convergence.

As was the case with signals, convex combinations of lotteries have again an intuitive interpretation. The convex combination that attaches weight λ to the lottery with cumulative distribution function F_1 and weight $1-\lambda$ to the lottery with cumulative distribution function F_2 is the lottery with cumulative distribution function $\lambda F_1 + (1 - \lambda) F_2$. This new lottery yields exactly the expected utility specified in Section 2 for convex combinations.

A lottery is first or second-order stochastically dominated by a convex combination of other lotteries if it is either weakly dominated by the convex combination, or if it is made redundant by the convex combination of the other lotteries.

We can apply parts (i), (iii) and (v) of Theorem 2.1. As in the previous section, strict dominance among lotteries is a vacuous notion because we have allowed the constant utility function. Interior optimality is vacuous because the relative interior of Y is empty. Therefore, we focus on part (v) of the Theorem.

We illustrate in Table 2.4 why part (v) of the Theorem would not hold if we had not considered convex combinations. We display three lotteries, ℓ_1, ℓ_2 and ℓ_3. Each row corresponds to a Dollar amount, each column corresponds to a lottery, and the entry in the table indicates the probability of the given dollar amount.

[6]Our definitions of the stochastic dominance orders are taken from Mas-Colell et al. (1995), where they are Definitions 6.D.1 and 6.D.2. The definitions differ slightly from the original definition of first order stochastic dominance in Quirk and Saposnik (1962) and second order stochastic dominance in Rothschild and Stiglitz (1970).

	ℓ_1	ℓ_2	ℓ_3
0\$	$\frac{1}{2}$	0	$\frac{4}{9}$
1\$	0	1	$\frac{1}{3}$
2\$	$\frac{1}{2}$	0	$\frac{2}{9}$

Table 2.4: Payoffs of three lotteries

Note that neither lottery 1 nor lottery 2 first order stochastically dominates lottery 3. However, the convex combination of lotteries 1 and 2 that attaches weight 2/3 to lottery 1 and weight 1/3 to lottery 2 attaches probability 1/3 each to 0 Dollars, 1 Dollar, and 2 Dollars. This convex combination first order stochastically dominates lottery 3. Indeed, a decision maker with a strictly increasing utility function will choose either lottery 1 or lottery 2, depending on the particular utility function, but will never choose lottery 3.

Our discussion suggests that it would be interesting to modify the framework by considering only lotteries with some given compact support, and then to allow only strictly increasing utility functions that assume 0 as the minimum value and 1 as the maximum value. The notion of strict dominance would then not be vacuous, and in addition one might try to adapt the framework so that part (ii) of Theorem 2.1 applies. We leave this to future work.

2.7 Conclusion

This paper has uncovered both shared properties of dominance orders in economics as well as differences among these orders. The differences that we have found might motivate the introduction of new orders of experiments and of monetary lotteries.

Our paper has emphasized the important role of convex combinations of actions in dominance relations. In the theory of the optimal choice of investment portfolios it might be of interest to investigate "efficient" portfolios as those that are not dominated by convex combinations of other portfolios. In the theory of information acquisition, it might be of interest to investigate signals that are not Blackwell dominated by convex combinations of other signals. In a companion paper we tackle this latter problem in a setting in which a signal is two-dimensional, and the marginal distributions of signals are given and fixed. We characterize joint distributions are not Blackwell dominated by convex combinations of other signals. Results of this type yield insight into optimal choices of decision makers without relying on specific assumptions about their environment or their preferences.

CHAPTER 3

Blackwell Undominated Joint Distributions of Signals

joint work with Tilman Börgers

3.1 Introduction

Decision makers often have access to multiple information sources. Investors in financial markets can obtain information from corporate disclosures, news media, and social networks. After the first diagnosis, patients may approach another doctor for a second opinion. Before buying a car, a consumer may gather information from dealers, friends, and reviews on the internet. Health authorities may require the residents to have both molecular and antigen tests for COVID-19.

The opinion that independent information sources are favorable is prevalent across disciplines. In philosophy: "The general point is that an additional outside opinion should move one only to the extent that one counts it as independent from opinions one has already taken into account."(Elga, 2010) In journalism: "...good journalism rests on sources who are *independent*, *multiple*, verified, authoritative, informed and named."(O'Connor, 2014) On the other hand, economic models often assume that different information sources are stochastically independent conditional on the true state.

However, dependence may be helpful. Blom (1975) indicates that random sampling without replacement is better than random sampling with replacement in the sense that the estimator resulting from the former has a lower variance than the estimator resulting from the latter. Note that the samples are correlated in random sampling without replacement while the samples are independent in random sampling with replacement. As we will illustrate later, another way of presenting that random sampling without replacement is better than random sampling with replacement is that the signal resulting from random sampling without replacement Blackwell dominates (Blackwell (1951), Blackwell (1953)) the signal resulting from random sampling with replacement.

These observations inspire us to investigate the relationship between correlation among signals and informativeness of pooled signals. Formally, the research question is: among all the joint distributions of two signals with fixed marginal distributions, which ones are not Blackwell dominated? For a special case with just two states and two signal realizations per signal, we show that under a specific condition, more negative correlations in both states increase Blackwell informativeness. In addition, we provide a complete characterization of joint distributions that are not Blackwell dominated by any single joint distribution or any convex combination of joint distributions. That is, the joint distribution which is as negatively correlated as possible in both states, or is as positively correlated as possible in one state and as negatively correlated as possible in the other state is not Blackwell dominated. On the other hand, for the general case, we prove that every joint distribution that has full support conditional on each state is Blackwell dominated by a convex combination of some joint distributions. In all cases, the conditionally independent joint distribution is Blackwell dominated.

The only closely related paper that we are aware of is Clemen and Winkler (1985). It uses a normal distribution model and assumes that the degrees of correlation among signals are the same in all states. It finds that a more negatively correlated joint distribution Blackwell dominates a less negatively correlated joint distribution. In our two-state-two-realization model (a Bernoulli distribution model), when the degrees of correlation among signals are the same in both states, there is a similar result. But we find that when the degrees of correlation among signals are different across states, more negative correlations in both states may increase or decrease Blackwell informativeness.

This paper is organized as follows. Section 3.2 presents the general model and Section 3.3 presents the two-state-two-realization model. Section 3.4 illustrates an example of sampling. Section 3.5 provides a complete characterization of joint distributions of signals that are not Blackwell dominated by a single joint distribution for the two-state-two-realization model. Section 3.6 presents a necessary condition for a joint distribution not being Blackwell dominated by a convex combination of joint distributions in the general model. Section 3.7 provides a complete characterization of joint distributions that are not Blackwell dominated by a convex combination of joint distributions for the two-state-two-realization model. Finally, Section 3.8 concludes.

3.2 General Model

Let Ω be a finite set of states of the world and a generic state is denoted by ω. A signal is a mapping $s : \Omega \to \Delta(M)$, where M is a finite set of signal realizations, $\Delta(M)$ is the set of all probability distributions over M, and $s_\omega \in \Delta(M)$ is the distribution of signal realizations conditional on ω.

Given finite signals $s^1, s^2, ..., s^n$ with realizations in M^1, M^2, ..., M^n, let $\beta_i \in \mathbb{R}_+$ and $\sum_{i=1}^n \beta_i = 1$. A convex combination of these signals with weights $(\beta_1, \beta_2, ..., \beta_n)$ is the signal

$$s^\beta : \Omega \to \Delta\left(M^1 \cup M^2 \cup ... \cup M^n\right)$$

such that for every $\omega \in \Omega$, every $i = 1, 2, \ldots, n$, and every $m^i \in M^i$, we have:

$$s^\beta_\omega(m^i) = \beta_i s^i_\omega(m^i).$$

Consider two (marginal) signals s^1 and s^2 with corresponding sets of signal realizations M^1 and M^2. We now consider a joint signal $\tilde{s}$ with realizations in $M^1 \times M^2$. Denote for every $m^1 \in M^1$, $m^2 \in M^2$ and every $\omega \in \Omega$ the marginal distributions of $\tilde{s}$ by:

$$\tilde{s}^1_\omega(m^1) = \sum_{\tilde{m}^2 \in M^2} \tilde{s}_\omega(m^1, \tilde{m}^2) \quad \text{and} \quad \tilde{s}^2_\omega(m^2) = \sum_{\tilde{m}^1 \in M^1} \tilde{s}_\omega(\tilde{m}^1, m^2).$$

Denote by $J\left(s^1, s^2\right)$ the set of joint signals that satisfy for every $\omega \in \Omega$ and for all $m^1 \in M^1$ and $m^2 \in M^2$:

$$\tilde{s}^1_\omega(m^1) = s^1_\omega(m^1) \quad \text{and} \quad \tilde{s}^2_\omega(m^2) = s^2_\omega(m^2).$$

Denote by $\Delta(\Omega)$ the set of all probability distributions over Ω. The decision maker has a prior belief $\mu \in \Delta(\Omega)$. A decision problem D consists of a finite set of actions A and a Bernoulli utility function $u : A \times \Omega \to \mathbb{R}$. If the decision maker faces decision problem D, has access to signal s before choosing an action, and chooses actions to maximize her expected utility, she obtains ex ante expected utility::

$$V(D, s) = \sum_{m \in M} \max_{a \in A} \sum_{\omega \in \Omega} \mu\left(\omega\right) s_\omega\left(m\right) u\left(a, \omega\right) \tag{3.1}$$

Definition 3.1. *Signal s* weakly Blackwell dominates *signal s' if for all decision problems D:*

$$V(D, s) \geq V(D, s').$$

Definition 3.2. *Signal s* strictly Blackwell dominates *signal s' if s weakly Blackwell dominates*

s' and there exists a decision problems D such that

$$V(D, s) > V(D, s').$$

The notions of "strictly" and "weakly" are different from those in game theory. The reason is as follows. Essentially there are three cases when comparing signals (resp. strategies). The first is that one signal (resp. strategy) leads to a higher payoff compared with the other signal (resp. strategy) for all decision problems (resp. strategy profiles of other players). The second is that one signal (resp. strategy) leads to a higher or equal payoff compared with the other signal (resp. strategy) for all decision problems (resp. strategy profiles of other players), and a higher payoff for at least one decision problem (resp. strategy profile of other players). The third is that both signals (resp. strategies) lead to the same payoff for all decision problems (resp. strategy profiles of other players). In game theory, "dominance" refers to only the first and the second case, and thus the first case is called "strict dominance" while the second "weak dominance". However, for signals, the first case is impossible because, for a decision problem where the optimal action does not vary with the states of the world, all signals lead to the same payoff. So only the second and third cases are considered, and thus the second case is called "strict Blackwell dominance" while the third is called "weak Blackwell dominance".

A well-known result is that s strictly Blackwell dominates s' if and only if the distribution over posteriors induced by s is a mean preserving spread of the distribution over posteriors induced by s'.

3.3 Two-State-Two-Realization Model

In this section, we describe a notable special case of the general model. There are two states, that is, $\Omega = \{\omega_1, \omega_2\}$. Consider two identical signals whose sets of signal realizations are both $M = \{m, \hat{m}\}$. Conditional on ω_1 (resp. ω_2) each signal has realization m (resp. $\hat{m}$) with probability α and realization $\hat{m}$ (resp. m) with probability $1 - \alpha$. Without loss of generality, we assume $\alpha \in \left(\frac{1}{2}, 1\right)$. So observing m (resp. $\hat{m}$) alone raises the decision maker's belief that the state is ω_1 (resp. ω_2). α represents the informativeness of both signals which is assumed to be the same in each state. The higher α is, the more informative the signals are. When α approaches $\frac{1}{2}$, the signals convey almost no information. When α approaches 1, the signals are almost perfectly informative. Denote each signal by s^α. Each $\tilde{s} \in J\left(s^\alpha, s^\alpha\right)$ can be characterized by (x, y) in Table 3.1 where $x \in [0, 1-\alpha]$ and $y \in [0, 1-\alpha]$. As x (resp. y) increases, the signals are more positively correlated conditional on ω_1 (resp. ω_2).

	m	$\hat{m}$
m	$x+2\alpha-1$	$1-\alpha-x$
$\hat{m}$	$1-\alpha-x$	x

ω_1

	m	$\hat{m}$
m	y	$1-\alpha-y$
$\hat{m}$	$1-\alpha-y$	$y+2\alpha-1$

ω_2

Table 3.1: A generic joint signal distribution

We represent $J(s^\alpha, s^\alpha)$ as the points inside the square in Figure 3.1.

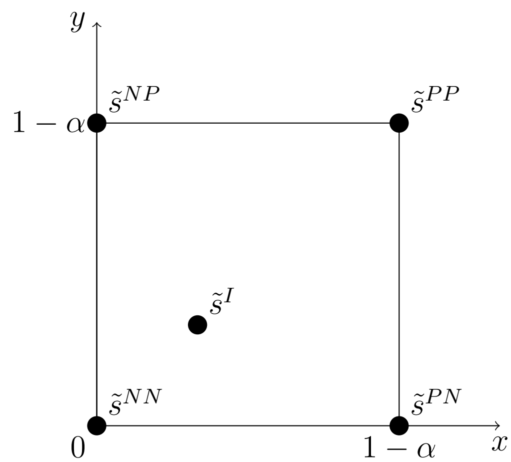


Figure 3.1: The representation of $J(s^\alpha, s^\alpha)$

Denote the joint signal in which the marginal signals are as positively correlated as possible in both states by $\tilde{s}^{PP}$ (the upper right corner), the joint signal in which the marginal signals are as positively correlated as possible in ω_1 while as negatively correlated as possible in ω_2 by $\tilde{s}^{PN}$ (the lower right corner), the joint signal in which the marginal signals are as negatively correlated as possible in both states by $\tilde{s}^{NN}$ (the lower left corner), the joint signal in which the marginal signals are as negatively correlated as possible in ω_1 while as positively correlated as possible in ω_2 by $\tilde{s}^{NP}$ (the upper left corner), and the joint signal in which the marginal signals are independent in both sates by $\tilde{s}^{I}$ (on the diagonal connecting the lower left and upper right corner, corresponding to $x = (1-\alpha)^2$ and $y = (1-\alpha)^2$).

3.4 Example

In this section, we state some conventional wisdom in statistics and then formalize it using the model we propose.

In a *probability sample*, each unit in the population has a known probability of selection, and a random number table or other randomization mechanism is used to choose the specific units to be included in the sample. Consider the following forms of probability sampling.

A *simple random sample with replacement* of size n from a population of N units can be thought of as drawing n independent samples of size 1. One unit is randomly selected from the population to be the first sampled unit, with probability $\frac{1}{N}$. Then the sampled unit is replaced in the population, and a second unit is randomly selected with probability $\frac{1}{N}$. This procedure is repeated until the sample has n units, which may include duplicates from the population.

If after each draw, the sampled unit is not replaced in the population, it is a *simple random sample without replacement*.

Certain population units may be associated with each other or belong to a particular group, called *clusters*. In a *two-stage cluster sampling*, some clusters are sampled in the first stage, and then within each sampled cluster, units are sampled.

There is received wisdom about the comparison of these samplings (Lohr, 2021). On the one hand, in finite population sampling, sampling the same unit twice provides no additional information. So sampling without replacement is more informative than sampling with replacement. On the other hand, units in the same cluster are not as likely to mirror the diversity of the population as well as units chosen randomly from the whole population. Thus, cluster sampling results in less information per observation than sampling from the population. In this paper, we propose a new perspective for this conventional wisdom by comparing these samplings in terms of Blackwell-informativeness and point out that their difference lies in the correlation between samples.

Specifically, we apply the model in Section 3.3. A researcher is investigating the political leanings of the population (6 persons). In ω_1, there are 3 female Democrats, 1 female Republican, 1 male Democrat, and 1 male Republican. In ω_2, there are 1 female Democrat, 3 female Republicans, 1 male Democrat, and 1 male Republican. The researcher's prior belief is that these two states are equally likely. An agent implements sampling and presents the selected samples to the researcher for investigation. The researcher can get two samples from the agent. Out of consideration of privacy, the agent refuses to provide any information about gender and inhibits the researcher from inferring it. (For example, the researcher must use an online questionnaire rather than phone calls.)

First consider simple random sampling with replacement, whose conditional distributions are shown in Table 3.2[1] where D represents that the sample is a Democrat while R represents that the sample is a Republican.

[1]In this case, $\alpha = \frac{2}{3}$, $x = y = \frac{1}{9}$.

		D	R
ω_1	D	$\frac{4}{9}$	$\frac{2}{9}$
	R	$\frac{2}{9}$	$\frac{1}{9}$

		D	R
ω_2	D	$\frac{1}{9}$	$\frac{2}{9}$
	R	$\frac{2}{9}$	$\frac{4}{9}$

Table 3.2: The conditional distributions of simple random sampling with replacement

In this case, two draws are conditionally independent. It implies the following posterior distribution: the probability of the posterior probability of ω_1 being $\frac{4}{5}$ (observing DD) equals $\frac{5}{18}$, the probability of the posterior probability of ω_1 being $\frac{1}{2}$ (observing DR or RD) equals $\frac{8}{18}$, and the probability of the posterior probability of ω_1 being $\frac{1}{5}$ (observing RR) equals $\frac{5}{18}$.

Then consider simple random sampling without replacement, whose conditional distributions are shown in Table 3.3[2].

		D	R
ω_1	D	$\frac{6}{15}$	$\frac{4}{15}$
	R	$\frac{4}{15}$	$\frac{1}{15}$

		D	R
ω_2	D	$\frac{1}{15}$	$\frac{4}{15}$
	R	$\frac{4}{15}$	$\frac{6}{15}$

Table 3.3: The conditional distributions of simple random sampling without replacement

In this case, two draws are conditionally correlated. For example, given that the first draw is a Democrat, the probability of the second draw being a Democrat is lower than the original proportion of Democrats in the population because the first draw is ruled out. The same logic applies to the first draw being a Republican. So the two draws are negatively correlated in both states. It implies the following posterior distribution: the probability of the posterior probability of ω_1 being $\frac{6}{7}$ (observing DD) equals $\frac{7}{30}$, the probability of the posterior probability of ω_1 being $\frac{1}{2}$ (observing DR or RD) equals $\frac{16}{30}$, and the probability of the posterior probability of ω_1 being $\frac{1}{7}$ (observing RR) equals $\frac{7}{30}$. This distribution over posteriors is a mean preserving spread of the counterpart for simple random sampling with replacement. So the joint distribution from sampling without replacement strictly Blackwell dominates the joint distribution from simple random sampling with replacement.

Finally, consider two-stage cluster sampling *with probability proportional to size*. In the first stage, the female cluster or male cluster is chosen where the probability of the female

[2]In this case, $\alpha = \frac{2}{3}$, $x = y = \frac{1}{15}$.

cluster being chosen is the proportion of females. In the second stage, two samples are drawn with replacement in the chosen cluster. This two-stage sampling can be implemented as follows. First, randomly select a sample. Second, pick the samples who are of the same gender as the selected sample in the first stage and replace the first selected sample in this sub-population. Then randomly select a sample from the sub-population. Note that the agent can observe the genders so that this sampling is feasible. But the agent does not reveal the genders of the samples to the researcher even though the agent ensures that the two samples are of the same gender. The conditional distributions are shown in Table 3.4[3].

	D	R
D	$\frac{11}{24}$	$\frac{5}{24}$
R	$\frac{5}{24}$	$\frac{3}{24}$

ω_1

	D	R
D	$\frac{3}{24}$	$\frac{5}{24}$
R	$\frac{5}{24}$	$\frac{11}{24}$

ω_2

Table 3.4: The conditional distributions of two-stage cluster sampling with probability proportional to size

In this case, two draws are conditionally correlated. Take ω_1 and the first draw being a Democrat as an example. Given that the first draw is a Democrat, the posterior probability of the first draw being a female is higher than the proportion of females in the population because the proportion of Democrats is higher in females than in males. As a result, the probability of the second draw being a Democrat is higher than the proportion of Democrats in the population. The same logic applies to ω_2 or the first draw being a Republican. So the two draws are positively correlated in both states. It implies the following posterior distribution: the probability of the posterior probability of ω_1 being $\frac{11}{14}$ (observing DD) equals $\frac{7}{24}$, the probability of the posterior probability of ω_1 being $\frac{1}{2}$ (observing DR or RD) equals $\frac{10}{24}$, and the probability of the posterior probability of ω_1 being $\frac{3}{14}$ (observing RR) equals $\frac{7}{24}$. The distribution over posteriors for simple random sampling with replacement is a mean preserving spread of the counterpart for two-stage cluster sampling with probability proportional to size. So the joint distribution from simple random sampling with replacement strictly Blackwell dominates the joint distribution from two-stage cluster sampling with probability proportional to size.

We can draw the joint distribution resulting from the above samplings within the square, as is shown in Figure 3.2, where $\tilde{s}^I$ represents the simple random sampling with replacement (the signals are independent conditional on each state), $\tilde{s}^N$ represents the simple random

[3]In this case, $\alpha = \frac{2}{3}$, $x = y = \frac{1}{8}$.

sampling without replacement (the signals are negatively correlated conditional on each state), $\tilde{s}^P$ represents the two-stage cluster sampling with probability proportional to size (the signals are positively correlated conditional on each state), and $\alpha = \frac{2}{3}$. It turns out that in this case all of them are on the diagonal connecting the lower left and upper right corner of the square, which implies their posterior beliefs of ω_1 after observing DR or RD are all $\frac{1}{2}$.

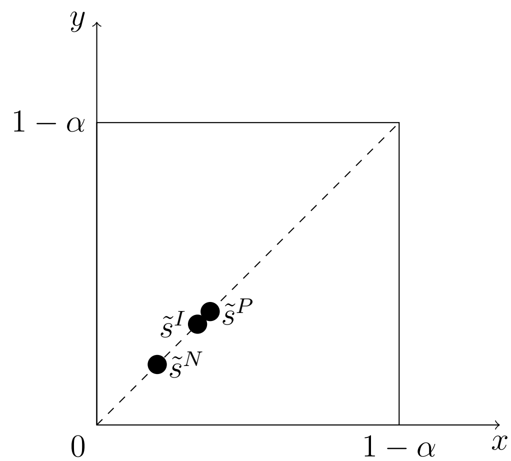


Figure 3.2: The representation of samplings

3.5 Blackwell Dominance by a Single Joint Signal: the Two-State-Two-Realization Model

In this section, we present a characterization of the joint signals in $J(s^\alpha, s^\alpha)$ that are not strictly Blackwell dominated by any joint signal in $J(s^\alpha, s^\alpha)$.

First, we present a result that generalizes the observation from the example of sampling in Section 3.4. Let

$$L = \left\{(x, y) \in [0, 1-\alpha]^2 \,|x = 0 \text{ or } y = 0\right\}$$

For all $(x, y) \in L$, let

$$C(x, y) = \left\{(x', y') \in [0, 1-\alpha]^2 \,|\, (x, y) = \beta(1-\alpha, 1-\alpha) + (1-\beta)(x, y), \beta \in [0, 1]\right\}$$

Denote by $S^{C(x,y)}$ the set of joint signals corresponding to elements in $C(x, y)$.

Lemma 3.1. *For all $(x, y) \in L$, for any two joint signals in $S^{C(x,y)}$ corresponding to β and β' with $\beta' < \beta$, the one corresponding to β' strictly Blackwell dominates the one corresponding to β.*

We prove it by showing that the distribution of posteriors implied by β' is a mean preserving spread of the distribution of posteriors implied by β. The details are in the appendix.

This result indicates that on any line segment with one end at the upper right corner and the other end on the left or lower side of the square, the joint signal that is farther away from the upper left corner strictly Blackwell dominates the joint signal that is closer to the upper left corner. For example, in Figure 3.3, $\tilde{s}'$ strictly Blackwell dominates $\tilde{s}$. It implies that Blackwell order is complete within $S^{C(x,y)} \subsetneq J(s^\alpha, s^\alpha)$.

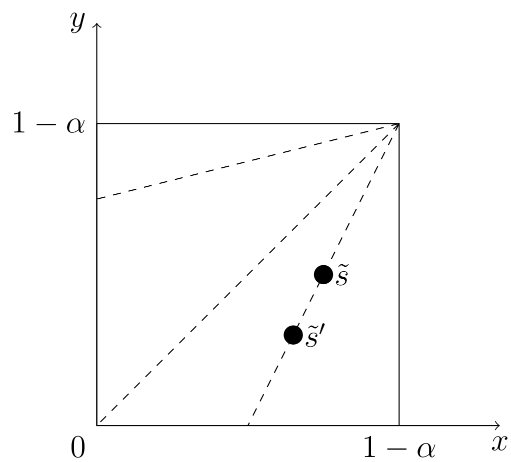


Figure 3.3: The representation of Lemma 3.1

It follows from Lemma 3.1 that

Corollary 3.1. *The conditionally independent joint signal is strictly Blackwell dominated by some joint signal in* $J(s^\alpha, s^\alpha)$.

Note that a smaller β implies a smaller or equal x and a smaller or equal y. That is, the marginal signals of the joint signal with a smaller β are more negatively correlated in both states. The intuition of why more negative correlations in both states are good is as follows. $\tilde{s}^{PP}$ is inferior because it is just a repetition of the marginal signals and thus provides no additional information. It is like "buy (ing) several copies of today's morning paper to assure himself that what it said was true." (Wittgenstein, 2009) As a result, the farther away the joint signal is from $\tilde{s}^{PP}$, the more informative the joint signal is.

Clemen and Winkler (1985) assumes the true states follow a normal distribution and each signal is a sum of the true state and a normal error. The conditional covariance matrices of errors do not vary with states. If all the variances of errors are the same and all the covariances of errors are the same, then the more negatively correlated the errors are, the more Blackwell informative the joint signal is. Since the degrees of correlation of signals are the same in all states, it corresponds to $x = y$ in our model. So their result echoes Lemma 3.1 for $S^{C(0,0)}$, that is, the line connecting $\tilde{s}^{PP}$ and $\tilde{s}^{NN}$.

Although Lemma 3.1 implies more negative correlations increase Blackwell informativeness (also for the case where $x \neq y$, which is beyond the scope of Clemen and Winkler (1985)), it relies on the condition that one end of the line is $\tilde{s}^{PP}$. For example, let $\alpha = \frac{2}{3}$. Consider $\tilde{s}$ corresponding to $\left(\frac{1}{9}, \frac{1}{9}\right)$ (two marginal signals are conditionally independent) and $\tilde{s}'$ corresponding to $\left(\frac{1}{3}, \frac{1}{6}\right)$. Since $\frac{1}{9} < \frac{1}{3}$ and $\frac{1}{9} < \frac{1}{6}$, the marginal signals of $\tilde{s}$ are more negatively correlated than the marginal signals of $\tilde{s}'$ in both states. However, $\tilde{s}'$ strictly Blackwell dominates $\tilde{s}$ (The proof is in the appendix.). These two joint signals are shown in Figure 3.4, and we can see the line segment connecting them does not pass through $\tilde{s}^{PP}$.

Example 3.1. *More negative correlations in both states do not guarantee strict Blackwell dominance.*

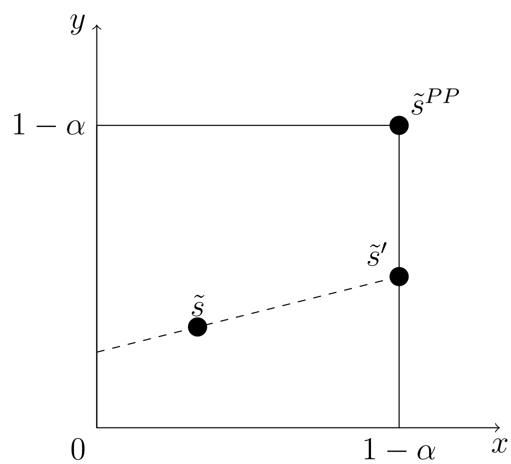


Figure 3.4: $\tilde{s}'$ strictly Blackwell dominates $\tilde{s}$ even though the marginal signals of $\tilde{s}$ are more negatively correlated than the marginal signals of $\tilde{s}'$ in both states

Denote by S^L the set of joint signals corresponding to elements in L. Lemma 3.1 implies that a necessary condition for a joint signal in $J(s^\alpha, s^\alpha)$ not being strictly Blackwell dominated by any joint signal in $J(s^\alpha, s^\alpha)$ is that it belongs to S^L, as is shown in Figure 3.5.

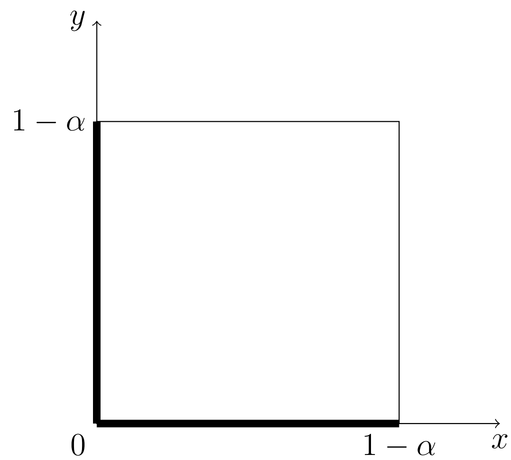


Figure 3.5: The representation of S^L

Then we further refine this condition to get a necessary and sufficient condition.

Proposition 3.1. *When $\alpha \in \left(\frac{1}{2}, \frac{3}{4}\right]$, only $\tilde{s}^{PN}$ and $\tilde{s}^{NP}$ are not strictly Blackwell dominated by any joint signal in $J(s^\alpha, s^\alpha)$.*

When $\alpha \in \left(\frac{3}{4}, 1\right)$, only $\tilde{s}^{PN}$, $\tilde{s}^{NP}$, and joint signals with $xy = 0$, $x < x^*$, and $y < x^*$ are not strictly Blackwell dominated by any joint signal in $J(s^\alpha, s^\alpha)$, where $x^* = \frac{(1-\alpha)(4\alpha-3)}{2\alpha-1}$.

This result is illustrated in Figure 3.6 and Figure 3.7.

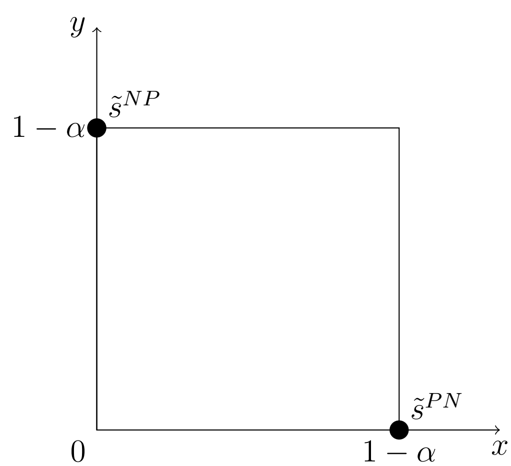


Figure 3.6: The joint signals in $J(s^\alpha, s^\alpha)$ that are not strictly Blackwell dominated by any joint signal in $J(s^\alpha, s^\alpha)$ when $\alpha \in \left(\frac{1}{2}, \frac{3}{4}\right]$

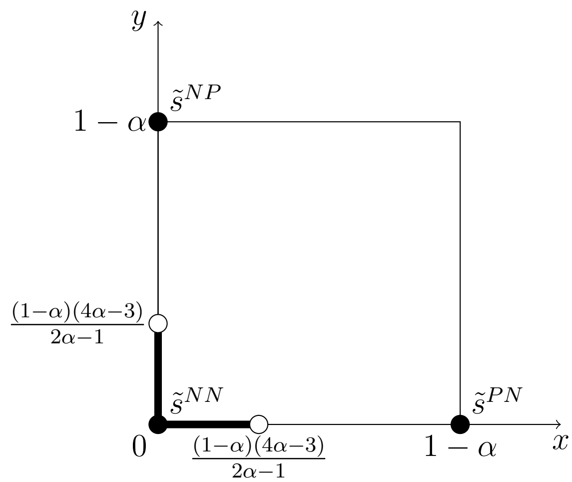


Figure 3.7: The joint signals in $J(s^\alpha, s^\alpha)$ that are not strictly Blackwell dominated by any joint signal in $J(s^\alpha, s^\alpha)$ when $\alpha \in \left(\frac{3}{4}, 1\right)$

Sketch of the proof. According to Lemma 3.1, for any joint signals in S^L, if it is strictly Blackwell dominated by a joint signal in $J(s^\alpha, s^\alpha)$, then it must be strictly Blackwell dominated by a joint signal in S^L. In other words, if a joint signal in S^L cannot be strictly Blackwell dominated by any joint signal in S^L, it cannot be strictly Blackwell dominated by any joint distributions in $J(s^\alpha, s^\alpha)$. Specifically, we prove that there are no distributions of posteriors implied by joint signals in S^L that are mean preserving spreads of the distributions of posteriors implied by the joint signals mentioned in the result. The details are in the appendix. ■

Note that the maximum of x (resp. y) is $1-\alpha$, $\frac{x^*}{1-\alpha}$ is increasing in α, and $\lim\limits_{\alpha \to 1^-} \frac{x^*}{1-\alpha} = 1$.

3.6 Blackwell Dominance by a Convex Combination of Joint Signals: the General Model

In this section, we derive a necessary condition for a joint signal in $J(s^1, s^2)$ not being strictly Blackwell dominated by any convex combination of joint signals in $J(s^1, s^2)$.

Proposition 3.2. *Suppose there exist states $\bar{\omega}, \hat{\omega} \in \Omega$ where $\bar{\omega} \neq \hat{\omega}$, messages $\bar{m}^1, \hat{m}^1 \in M^1$ where $\bar{m}^1 \neq \hat{m}^1$, and messages $\bar{m}^2, \hat{m}^2 \in M^2$ where $\bar{m}^2 \neq \hat{m}^2$, such that $\tilde{s}_{\bar{\omega}}$ assigns positive probability to all elements of $\{\bar{m}^1, \hat{m}^1\} \times \{\bar{m}^2, \hat{m}^2\}$ and $\tilde{s}_{\hat{\omega}}$ assigns positive probability to at least one element of $\{\bar{m}^1, \hat{m}^1\} \times \{\bar{m}^2, \hat{m}^2\}$. Then $\tilde{s}$ is strictly Blackwell dominated by a convex combination of some joint signals in $J(s^1, s^2)$.*

Proof. We first characterize the joint signals associated with $\tilde{s}$ which we will use later to construct a convex combination to strictly Blackwell dominate $\tilde{s}$. We modify $\tilde{s}$ as follows: we change only the conditional probabilities of signal realizations in $\{\bar{m}^1, \hat{m}^1\} \times \{\bar{m}^2, \hat{m}^2\}$ conditional on $\bar{\omega}$, leaving all other conditional probabilities unchanged. The modification is shown in Table 3.5. Let $z \in [\underline{z}, \bar{z}]$ where

$$\underline{z} = \max\left\{-\tilde{s}_{\bar{\omega}}(\bar{m}^1, \bar{m}^2), -\tilde{s}_{\bar{\omega}}(\hat{m}^1, \hat{m}^2)\right\}$$
$$\bar{z} = \min\left\{\tilde{s}_{\bar{\omega}}(\bar{m}^1, \hat{m}^2), \tilde{s}_{\bar{\omega}}(\hat{m}^1, \bar{m}^2)\right\}$$

so that no entry in Table 3.5 is smaller than 0 or larger than 1.

	$\bar{m}^2$	$\hat{m}^2$
$\bar{m}^1$	$\tilde{s}_{\bar{\omega}}(\bar{m}^1, \bar{m}^2) + z$	$\tilde{s}_{\bar{\omega}}(\bar{m}^1, \hat{m}^2) - z$
$\hat{m}^1$	$\tilde{s}_{\bar{\omega}}(\hat{m}^1, \bar{m}^2) - z$	$\tilde{s}_{\bar{\omega}}(\hat{m}^1, \hat{m}^2) + z$

Table 3.5: Modification of $\tilde{s}$

Denote the joint distribution corresponding to z by $\tilde{s}^z$. Clearly $\tilde{s}$ is $\tilde{s}^0$.

Consider a convex combination of $\tilde{s}^{z_L}$ and $\tilde{s}^{z_H}$ with $\underline{z} \leq z_L < 0 < z_H \leq \bar{z}$ and the weight for $\tilde{s}^{z_L}$ is $\frac{|z_H|}{|z_H|+|z_L|}$ and the weight for $\tilde{s}^{z_H}$ is $\frac{|z_L|}{|z_H|+|z_L|}$.

We first prove that for any decision problem, the convex combination of joint signals leads to a higher or equal payoff. Consider any decision problem D. Let us denote by $\vartheta(z)$ those terms in the decision maker's expected utility that depends on z. We have:

$$\begin{aligned}
&\vartheta(z) \\
=& \max_{a \in A}\left[\mu(\bar{\omega})\left(\tilde{s}_{\bar{\omega}}(\bar{m}^1, \bar{m}^2) + z\right)u(a, \bar{\omega}) + \sum_{\omega \in \Omega \setminus \{\bar{\omega}\}} \mu(\omega)\tilde{s}_{\omega}(\bar{m}^1, \bar{m}^2)u(a, \omega)\right] \\
&+ \max_{a \in A}\left[\mu(\bar{\omega})\left(\tilde{s}_{\bar{\omega}}(\bar{m}^1, \hat{m}^2) - z\right)u(a, \bar{\omega}) + \sum_{\omega \in \Omega \setminus \{\bar{\omega}\}} \mu(\omega)\tilde{s}_{\omega}(\bar{m}^1, \hat{m}^2)u(a, \omega)\right] \\
&+ \max_{a \in A}\left[\mu(\bar{\omega})\left(\tilde{s}_{\bar{\omega}}(\hat{m}^1, \bar{m}^2) - z\right)u(a, \bar{\omega}) + \sum_{\omega \in \Omega \setminus \{\bar{\omega}\}} \mu(\omega)\tilde{s}_{\omega}(\hat{m}^1, \bar{m}^2)u(a, \omega)\right] \\
&+ \max_{a \in A}\left[\mu(\bar{\omega})\left(\tilde{s}_{\bar{\omega}}(\hat{m}^1, \hat{m}^2) + z\right)u(a, \bar{\omega}) + \sum_{\omega \in \Omega \setminus \{\bar{\omega}\}} \mu(\omega)\tilde{s}_{\omega}(\hat{m}^1, \hat{m}^2)u(a, \omega)\right]
\end{aligned}$$

Observe that ϑ is the sum of four functions, where each of these four functions in turn is the maximum of a finite set of functions that are linear in z. This implies that ϑ is convex in z. As a result,

$$\vartheta(0) \leq \frac{|z_H|}{|z_H| + |z_L|}\vartheta(z_L) + \frac{|z_L|}{|z_H| + |z_L|}\vartheta(z_H)$$

Then we prove that there is a decision problem for which the convex combination of joint signals leads to a higher payoff. Without loss of generality, suppose $\tilde{s}_{\hat{\omega}}$ assigns positive probability to $(\bar{m}^1, \bar{m}^2)$. Consider the following decision problem. $A = \{a_1, a_2\}$. $u(a_1, \omega) = u(a_2, \omega)$ for all $\omega \in \Omega \backslash \{\bar{\omega}, \hat{\omega}\}$. $u(a_1, \bar{\omega}) \neq u(a_2, \bar{\omega})$, $u(a_1, \hat{\omega}) \neq u(a_2, \hat{\omega})$, and

$$\mu(\bar{\omega})\,\tilde{s}_{\bar{\omega}}\left(\bar{m}^1, \bar{m}^2\right)\left[u(a_1, \bar{\omega}) - u(a_2, \bar{\omega})\right] = -\mu(\hat{\omega})\,\tilde{s}_{\hat{\omega}}\left(\bar{m}^1, \bar{m}^2\right)\left[u(a_1, \hat{\omega}) - u(a_2, \hat{\omega})\right]$$

which implies

$$\begin{aligned} &\mu(\bar{\omega})\tilde{s}_{\bar{\omega}}(\bar{m}^1, \bar{m}^2)u(a_1, \bar{\omega}) + \mu(\hat{\omega})\tilde{s}_{\hat{\omega}}(\bar{m}^1, \bar{m}^2)u(a_1, \hat{\omega}) \\ = &\mu(\bar{\omega})\tilde{s}_{\bar{\omega}}(\bar{m}^1, \bar{m}^2)u(a_2, \bar{\omega}) + \mu(\hat{\omega})\tilde{s}_{\hat{\omega}}(\bar{m}^1, \bar{m}^2)u(a_2, \hat{\omega}) \end{aligned}$$

The expected utility of choosing a_1 observing $(\bar{m}^1, \bar{m}^2)$ is

$$\mu(\bar{\omega})\left(\tilde{s}_{\bar{\omega}}(\bar{m}^1, \bar{m}^2) + z\right)u(a_1, \bar{\omega}) + \mu(\hat{\omega})\tilde{s}_{\hat{\omega}}(\bar{m}^1, \bar{m}^2)u(a_1, \hat{\omega}) + \sum_{\omega \in \Omega \backslash \{\bar{\omega}, \hat{\omega}\}} \mu(\omega)\tilde{s}_{\omega}(\bar{m}^1, \bar{m}^2)u(a_1, \omega)$$

which is linear in z and the slope is $\mu(\bar{\omega})u(a_1, \bar{\omega})$. The expected utility of choosing a_2 observing $(\bar{m}^1, \bar{m}^2)$ is

$$\mu(\bar{\omega})\left(\tilde{s}_{\bar{\omega}}(\bar{m}^1, \bar{m}^2) + z\right)u(a_2, \bar{\omega}) + \mu(\hat{\omega})\tilde{s}_{\hat{\omega}}(\bar{m}^1, \bar{m}^2)u(a_2, \hat{\omega}) + \sum_{\omega \in \Omega \backslash \{\bar{\omega}, \hat{\omega}\}} \mu(\omega)\tilde{s}_{\omega}(\bar{m}^1, \bar{m}^2)u(a_2, \omega)$$

which is linear in z and the slope is $\mu(\bar{\omega})u(a_2, \bar{\omega})$. So the payoffs of choosing a_1 and a_2 observing $(\bar{m}^1, \bar{m}^2)$ are the same when $z = 0$ and their payoffs have different slopes.

Previously, we show that those terms in ϑ relevant for $(\bar{m}^1, \hat{m}^2)$, $(\hat{m}^1, \bar{m}^2)$, and $(\hat{m}^1, \hat{m}^2)$ are all convex in z.

In sum, for this decision problem,

$$\vartheta(0) < \frac{|z_H|}{|z_H| + |z_L|}\vartheta(z_L) + \frac{|z_L|}{|z_H| + |z_L|}\vartheta(z_H)$$

So $\tilde{s}$ is strictly Blackwell dominated by a convex combination of $\tilde{s}^{z_L}$ and $\tilde{s}^{z_H}$. □

Note that if a joint signal has full support conditional on each state, it satisfies the condition in Proposition 3.2. Specifically, we have

Corollary 3.2. *The conditionally independent joint signal is strictly Blackwell dominated by a convex combination of some joint signals in $J(s^1, s^2)$.*

3.7 Blackwell Dominance by a Convex Combination of Joint Signals: the Two-State-Two-Realization Model

In this section, we present a characterization of the joint signals in $J(s^\alpha, s^\alpha)$ that are not strictly Blackwell dominated by any convex combination of joint signals in $J(s^\alpha, s^\alpha)$.

When $\alpha \in \left(\frac{1}{2}, \frac{3}{4}\right]$, in the proof of Proposition 3.1, we show that any joint signals in $J(s^\alpha, s^\alpha)$ other than $\tilde{s}^{PN}$ and $\tilde{s}^{NP}$ is strictly Blackwell dominated by $\tilde{s}^{PN}$ or $\tilde{s}^{NP}$. As a result, for $\tilde{s}^{PN}$ (resp. $\tilde{s}^{NP}$), if it is strictly Blackwell dominated by a convex combination of joint signals in $J(s^\alpha, s^\alpha)$, then it must be strictly Blackwell dominated by $\tilde{s}^{NP}$ (resp. $\tilde{s}^{PN}$). But we show in the proof of Proposition 3.1 that $\tilde{s}^{PN}$ does not strictly Blackwell dominate $\tilde{s}^{NP}$ and vice versa. So both of them are not strictly Blackwell dominated by any convex combination of joint signals in $J(s^\alpha, s^\alpha)$.

When $\alpha \in \left(\frac{3}{4}, 1\right)$, according to Proposition 3.1, joint signals with $x \in \left(0, \frac{(1-\alpha)(4\alpha-3)}{2\alpha-1}\right)$ and $y = 0$ or $x = 0$ and $y \in \left(0, \frac{(1-\alpha)(4\alpha-3)}{2\alpha-1}\right)$ are not strictly Blackwell dominated by any single joint signal in $J(s^\alpha, s^\alpha)$. But according to the proof of Proposition 3.2, they are strictly Blackwell dominated by a convex combination of $\tilde{s}^{PN}$ and $\tilde{s}^{NN}$ or $\tilde{s}^{NP}$ and $\tilde{s}^{NN}$. For example, in Figure 3.8, $\tilde{s}$ is strictly Blackwell dominated by the convex combination of $\tilde{s}^{PN}$ and $\tilde{s}^{NN}$.

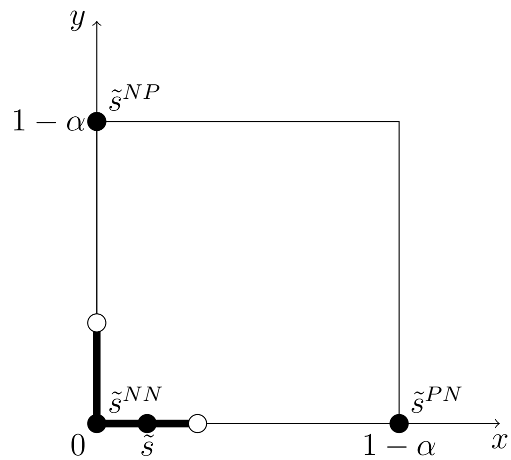


Figure 3.8: The joint signals that are not strictly Blackwell dominated by any single joint signal in $J(s^\alpha, s^\alpha)$ may be strictly Blackwell dominated by a convex combination of joint signals in $J(s^\alpha, s^\alpha)$

So we get a necessary condition for a joint signal not being strictly Blackwell dominated by any convex combination of joint signals in $J(s^\alpha, s^\alpha)$ when $\alpha \in \left(\frac{3}{4}, 1\right)$, that is, the joint distribution must be $\tilde{s}^{PN}$, $\tilde{s}^{NP}$, or $\tilde{s}^{NN}$, as is shown in Figure 3.9.

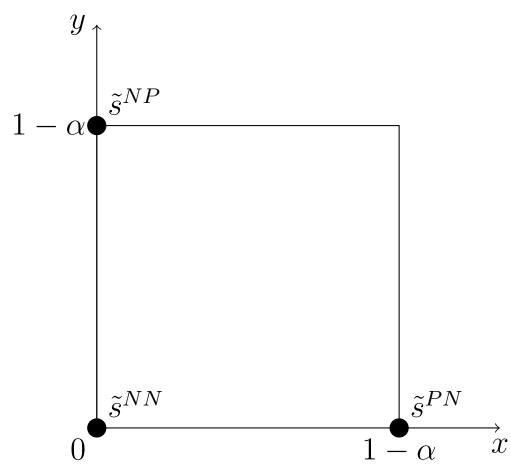


Figure 3.9: A necessary condition for a joint signal not being strictly Blackwell dominated by any convex combination of joint signals in $J(s^\alpha, s^\alpha)$ when $\alpha \in \left(\frac{3}{4}, 1\right)$

As a result, for any of $\tilde{s}^{NN}$, $\tilde{s}^{NP}$ and $\tilde{s}^{PN}$, if it is strictly Blackwell dominated by a convex combination of joint signals in $J(s^\alpha, s^\alpha)$, then it must be strictly Blackwell dominated by a convex combination of the other two joint signals. In other words, if any of them cannot be strictly Blackwell dominated by any convex combination of the other two, it cannot be strictly Blackwell dominated by any convex combination of any joint signals in $J(s^\alpha, s^\alpha)$.

Since there are only three candidates, by checking them one by one, we get a characterization of the set of joint signals that are not strictly Blackwell dominated by any convex combination of joint signals in $J(s^{\alpha}, s^{\alpha})$.

Proposition 3.3. *When $\alpha \in \left(\frac{1}{2}, \frac{4}{5}\right]$, only $\tilde{s}^{PN}$ and $\tilde{s}^{NP}$ are not strictly Blackwell dominated by any convex combination of joint signals in $J(s^{\alpha}, s^{\alpha})$.*

When $\alpha \in \left(\frac{4}{5}, 1\right)$, only $\tilde{s}^{PN}$, $\tilde{s}^{NP}$ and $\tilde{s}^{NN}$ are not strictly Blackwell dominated by any convex combination of joint signals in $J(s^{\alpha}, s^{\alpha})$.

Sketch of the proof. We prove the following statements, which are enough for the claim in this proposition to hold.

(1) For $\alpha \in \left(\frac{1}{2}, \frac{2}{3}\right) \cup \left(\frac{2}{3}, 1\right)$, there is a decision problem in which $\tilde{s}^{PN}$ leads to a higher expected payoff than $\tilde{s}^{NP}$ and $\tilde{s}^{NN}$. For $\alpha = \frac{2}{3}$, $\tilde{s}^{PN}$ and $\tilde{s}^{NP}$ have the same distribution over posteriors, and there is a decision problem in which $\tilde{s}^{PN}$ leads to a higher expected payoff than $\tilde{s}^{NN}$.

(2) For $\alpha \in \left(\frac{1}{2}, \frac{2}{3}\right) \cup \left(\frac{2}{3}, 1\right)$, there is a decision problem in which $\tilde{s}^{NP}$ leads to a higher expected payoff than $\tilde{s}^{PN}$ and $\tilde{s}^{NN}$. For $\alpha = \frac{2}{3}$, $\tilde{s}^{NP}$ and $\tilde{s}^{PN}$ have the same distribution over posteriors, and there is a decision problem in which $\tilde{s}^{NP}$ leads to a higher expected payoff than $\tilde{s}^{NN}$.

(3) For $\alpha \in \left(\frac{1}{2}, \frac{4}{5}\right]$, $\tilde{s}^{NN}$ is strictly Blackwell dominated by a convex combination of $\tilde{s}^{PN}$ and $\tilde{s}^{NP}$.

(4) For $\alpha \in \left(\frac{4}{5}, 1\right)$, there is a decision problem in which $\tilde{s}^{NN}$ leads to a higher expected payoff than $\tilde{s}^{PN}$ and $\tilde{s}^{NP}$.

The details are in the appendix. ■

Note when $\alpha \in \left(\frac{3}{4}, \frac{4}{5}\right]$, $\tilde{s}^{NN}$ is not strictly Blackwell dominated by either of $\tilde{s}^{PN}$ and $\tilde{s}^{NP}$, but is strictly dominated by the convex combination of them.

The analysis of this proposition is as follows. First, consider why $\tilde{s}^{PN}$, $\tilde{s}^{NP}$, and $\tilde{s}^{NN}$ are informative signals. $\tilde{s}^{PN}$ and $\tilde{s}^{NP}$ being informative is intuitive. In one state, the realizations are as positively correlated as possible. In the other state, the realizations are as negatively correlated as possible. So whether or not the two realizations are the same can help to distinguish between two states. But this makes the fact that $\tilde{s}^{NN}$ is not strictly Blackwell dominated surprising. The realizations are as negatively correlated as possible in both states. So whether or not the two realizations are the same cannot help to distinguish between two states. As a result, we need a more careful examination, which is given below.

Note when the true state is ω_1 (resp. ω_2), it is more likely to observe m (resp. $\hat{m}$). First consider $\tilde{s}^{PN}$. When the true state is ω_1, the realizations are as positively correlated as possible (in fact, perfectly positively correlated). So only mm and $\hat{m}\hat{m}$ are possible. When the true state is ω_2, the realizations are as negatively correlated as possible. Note they cannot be perfectly negatively correlated because that implies the probability of m is the same as $\hat{m}$, but the latter should be higher than the former at ω_2. Within mm and $\hat{m}\hat{m}$, mm becomes impossible since the realization tends to be $\hat{m}$ at ω_2. Therefore, mm indicates the true state must be ω_1, and $m\hat{m}$ or $\hat{m}m$ indicates the true state must be ω_2. The only realization that does not indicate the true state for sure is $\hat{m}\hat{m}$. So the intuition that "the realizations being different indicates ω_2" is correct, while the intuition that "the realizations being the same indicates ω_1" is partially correct (only applies to mm). $\tilde{s}^{NP}$ is similar.

Then consider $\tilde{s}^{NN}$. The above analysis suggests that when the true state is ω_1, the realizations can be mm, $m\hat{m}$, and $\hat{m}m$. When the true state is ω_2, the realizations can be $\hat{m}\hat{m}$, $m\hat{m}$, and $\hat{m}m$. Therefore, mm indicates the true state must be ω_1, and $\hat{m}\hat{m}$ indicates the true state must be ω_2. $m\hat{m}$ or $\hat{m}m$ cannot reveal much information about the true state. In sum, when the realizations are as negatively correlated as possible in both states, the realizations being the same is very informative.

Then consider why as the marginal signals become more informative, $\tilde{s}^{NN}$ stands out. The above analysis suggests that the drawback of $\tilde{s}^{PN}$ is $\hat{m}\hat{m}$ while the drawback of $\tilde{s}^{NN}$ is $m\hat{m}$ and $\hat{m}m$. When comparing them, two aspects matter: the extremeness of the posteriors and the probabilities of these realizations. On the one hand, the posterior of $\hat{m}\hat{m}$ in $\tilde{s}^{PN}$ is more extreme than the posteriors of $m\hat{m}$ and $\hat{m}m$ in $\tilde{s}^{NN}$.[4] On the other hand, when the marginal signals are not very informative (resp. very informative), the probability of $\hat{m}\hat{m}$ in $\tilde{s}^{PN}$ is lower (resp. higher) than the probability of $m\hat{m}$ and $\hat{m}m$ in $\tilde{s}^{NN}$.[5] As a result, when the marginal signals are not very informative, $\tilde{s}^{PN}$ is better than $\tilde{s}^{NN}$ on both aspects. So $\tilde{s}^{NN}$ is strictly Blackwell dominated. When the marginal signals are very informative, $\tilde{s}^{PN}$ and $\tilde{s}^{NN}$ each has its advantage. So both of them are not strictly Blackwell dominated.

The above analysis reveals that when the marginal signals are very informative, $\tilde{s}^{PN}$ is better in some decision problems while $\tilde{s}^{NN}$ is better in other decision problems. We further illustrate the features of decision problems for which they are better respectively.

Figure 3.10 illustrates a decision problem with two actions (a_1 and a_2) where $\tilde{s}^{PN}$ is better than $\tilde{s}^{NN}$.

[4] For $\hat{m}\hat{m}$ in $\tilde{s}^{PN}$, the posterior probability of ω_1 is $\frac{1-\alpha}{\alpha}$. For $m\hat{m}$ or $\hat{m}m$ in $\tilde{s}^{NN}$, the posterior probability of ω_1 is $\frac{1}{2}$.

[5] The probability of $\hat{m}\hat{m}$ in $\tilde{s}^{PN}$ is $\frac{\alpha}{2}$. The probability of $m\hat{m}$ and $\hat{m}m$ in $\tilde{s}^{NN}$ is $2(1-\alpha)$. The cutoff of their relative size is $\frac{4}{5}$, which is exactly the cutoff for $\tilde{s}^{NN}$ not being strictly Blackwell dominated.

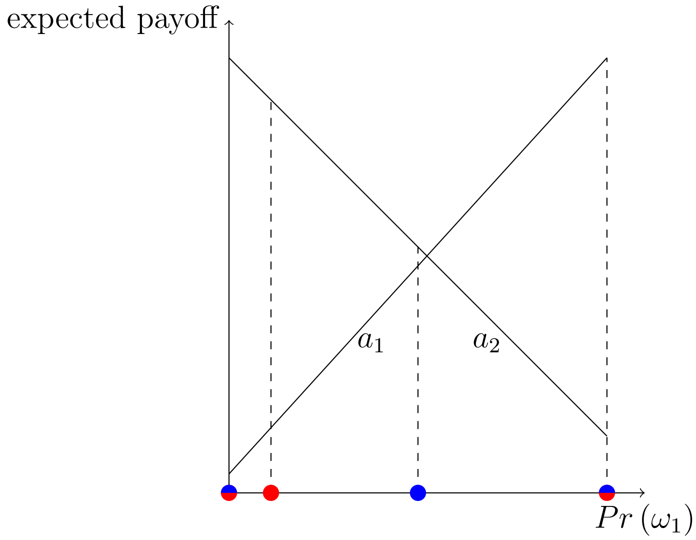


Figure 3.10: A decision problem where $\tilde{s}^{PN}$ is better than $\tilde{s}^{NN}$

The solid lines indicate the expected payoffs of actions in a specific decision problem. The red points indicate the posteriors of $\tilde{s}^{PN}$ while the blue points indicate the posteriors of $\tilde{s}^{NN}$. A half red half blue point means it is a posterior for both $\tilde{s}^{PN}$ and $\tilde{s}^{NN}$.

If we add a safe action a_3 to the above decision problem, we can increase the expected payoff of $\tilde{s}^{NN}$ while keeping the payoff of $\tilde{s}^{PN}$ the same. As a result, $\tilde{s}^{NN}$ becomes better instead, which is shown in Figure 3.11:

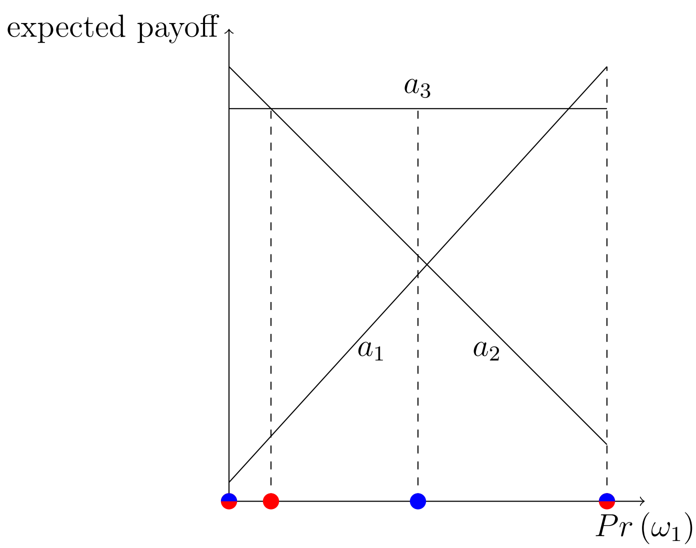


Figure 3.11: A decision problem where $\tilde{s}^{NN}$ is better than $\tilde{s}^{PN}$

3.8 Conclusion

We investigate which joint distributions of signals are not strictly Blackwell dominated among all the joint distributions of two signals with fixed marginal distributions. For a special case with just two states and two signal realizations per signal, we provide a complete characterization of joint distributions that are not Blackwell dominated by any single joint distribution or any convex combination of joint distributions. For the general case, we prove that every joint distribution that has full support conditional on each state is strictly Blackwell dominated by a convex combination of some joint distributions. In all cases, the conditionally independent joint distribution is strictly Blackwell dominated. Nelson Mandela said "I like friends who have *independent* minds because they tend to make you see problems from all angles." (Mandela, 2011) But our paper implies that friends who have *dependent* minds may help you even more.

Our work can be extended in the following ways. First, this paper does not provide any results regarding strict Blackwell dominance by a single joint distribution in the general model, so it is a natural next-step. Second, it would be interesting to make the necessary condition for a joint distribution not being strictly Blackwell dominated by a convex combination of joint distributions in the general model tighter, or even to strengthen it to a sufficient and necessary condition.

APPENDIX A

Appendix for Chapter 1

Proof of Lemma 1.3

Lemma A.1. *If* $I' \in BR(I)$*, then* $C(I, I') < +\infty$*.*

Proof. If $C(I, I') = +\infty$, then

$$U(I') - C(I, I') = -\infty$$

But

$$U(I) - C(I, I) > -\infty$$

Then $I' \notin BR(I)$. So I prove the contrapositive of the lemma. □

Proof. Suppose $(I_\varnothing, I_R, I_S, I_{SR})$ is a distortion.

According to Lemma A.1, $I_R \in BR(I_\varnothing)$ and $I_{SR} \in BR(I_S)$ imply $C(I_\varnothing, I_R) < +\infty$ and $C(I_S, I_{SR}) < +\infty$. Given $C(I_\varnothing, I_S) \geqslant 0$, $C(I_S, I_R) \geqslant 0$, and

$$C(I_\varnothing, I_R) = C(I_\varnothing, I_S) + C(I_S, I_R) \tag{A.1}$$

$C(I_\varnothing, I_R) < +\infty$ implies $C(I_\varnothing, I_S) < +\infty$ and $C(I_S, I_R) < +\infty$. That C satisfies SLP suggests

$$C(I_\varnothing, I_{SR}) \leqslant C(I_\varnothing, I_S) + C(I_S, I_{SR}) \tag{A.2}$$

Then $C(I_\varnothing, I_S) < +\infty$ and $C(I_S, I_{SR}) < +\infty$ suggest $C(I_\varnothing, I_{SR}) < +\infty$. So all the costs involved here are finite, and thus the conventional algebraic properties hold. $I_{SR} \in BR(I_S)$ implies

$$U(I_{SR}) - C(I_S, I_{SR}) \geqslant U(I_R) - C(I_S, I_R) \tag{A.3}$$

It follows that

$$\begin{aligned} & U\left(I_{SR}\right)-C\left(I_{\varnothing},I_{SR}\right) \\ \geqslant\ & U\left(I_{SR}\right)-C\left(I_{S},I_{SR}\right)-C\left(I_{\varnothing},I_{S}\right) \\ \geqslant\ & U\left(I_{R}\right)-C\left(I_{S},I_{R}\right)-C\left(I_{\varnothing},I_{S}\right) \\ =\ & U\left(I_{R}\right)-C\left(I_{\varnothing},I_{R}\right) \end{aligned}$$

where the first inequality is due to (A.2), the second inequality is due to (A.3), and the third equation is due to (A.1). It implies $I_{SR} \in BR\left(I_{\varnothing}\right)$. So $\left(I_{\varnothing}, I_R, I_S, I_{SR}\right)$ is not a distortion.

□

Proof of Theorem 1.2

Proof. According to Matysková and Montes (2021), there exists $0 \leqslant \underline{\gamma_2} \leqslant \overline{\gamma_2} \leqslant 1$ such that the receiver does not learn and takes a_1 when the belief is in $\left[0, \underline{\gamma_2}\right]$, does not learn and takes a_2 when the belief is in $[\overline{\gamma_2}, 1]$, and learns to end up with a distribution over beliefs whose support is $\left\{\underline{\gamma_2}, \overline{\gamma_2}\right\}$ when the belief is in $\left(\underline{\gamma_2}, \overline{\gamma_2}\right)$.

First consider the case where $0 < \underline{\gamma_2} < \overline{\gamma_2} < 1$.

When $\mu \in \left[0, \underline{\gamma_2}\right] \cup [\overline{\gamma_2}, 1]$, it means that $I_R = I_{\varnothing}$. Since $I_S \succ_B I_{\varnothing}$, then $I_{SR} \succsim_B I_S \succ_B I_R$, which violates the definition of diversion.

When $\mu \in \left(\underline{\gamma_2}, \overline{\gamma_2}\right)$, it means that supp $I_R = \left\{\underline{\gamma_2}, \overline{\gamma_2}\right\}$. According the this format of receiver's optimal strategy, it can be seen that whatever I_S the sender chooses, the elements in the support of resulting I_{SR} must be in $\left[0, \underline{\gamma_2}\right] \cup [\overline{\gamma_2}, 1]$. Let supp $I_{SR} = \{\gamma^1, ..., \gamma^s, ..., \gamma^t\}$ where $\gamma^1, ..., \gamma^s \in \left[0, \underline{\gamma_2}\right]$ and others in $[\overline{\gamma_2}, 1]$. Denote the probability of γ^i as r_i. It follows that

$$\begin{aligned} \mu &= \sum_{i=1}^{t} r_i \gamma^i \\ &= \left(\sum_{i=1}^{s} r_i\right) \frac{\sum\limits_{i=1}^{s} r_i \gamma^i}{\sum\limits_{i=1}^{s} r_i} + \left(\sum_{j=s+1}^{t} r_j\right) \frac{\sum\limits_{j=s+1}^{t} r_j \gamma^j}{\sum\limits_{j=s+1}^{t} r_j} \\ &\equiv r_L \gamma_L + r_R \gamma_R \end{aligned}$$

where $r_L \leqslant \underline{\gamma_2}$ and $r_R \geqslant \overline{\gamma_2}$. Consider a distribution over beliefs $\tilde{I}$ whose support is $\{\gamma_L, \gamma_R\}$ and the corresponding probability is r_L and r_R. By construction, $I_{SR} \succsim_B \tilde{I}$.

Lemma A.2. *$\forall I, I' \in \mathcal{B}(I_\varnothing)$ such that supp I is affinely independent, if supp $I' \subseteq$ conv supp I, then $I \succsim_B I'$.*

This result is from Lipnowski et al. (2020). Since $|\text{supp } \tilde{I}| = 2$, supp $\tilde{I}$ is affinely independent. $r_L \leqslant \underline{\gamma_2}$ and $r_R \geqslant \overline{\gamma_2}$ imply supp $I_R \subseteq$ conv supp $\tilde{I}$ and thus $\tilde{I} \succsim_B I_R$ according to Lemma A.2. In sum, $I_{SR} \succsim_B I_R$, which violates the definition of diversion.

It is easy to verify that the theorem holds when any of $0 = \underline{\gamma_2}$, $\underline{\gamma_2} = \overline{\gamma_2}$, and $\overline{\gamma_2} = 1$ holds.

□

Proof of Proposition 1.1

Proof. According to Lemma A.1, $I_R \in BR(I_\varnothing)$ and $I_{SR} \in BR(I_S)$ imply $C(I_\varnothing, I_R) < +\infty$ and $C(I_S, I_{SR}) < +\infty$. Then $C(I_S, I_R) \leqslant C(I_\varnothing, I_R)$ implies $C(I_S, I_R) < +\infty$. So all the costs involved here are finite, and thus the conventional algebraic properties hold. $I_{SR} \in BR(I_S)$ implies

$$U(I_{SR}) - C(I_S, I_{SR}) \geqslant U(I_R) - C(I_S, I_R)$$

Together with $C(I_S, I_R) \leqslant C(I_\varnothing, I_R)$, it implies

$$U(I_{SR}) - C(I_S, I_{SR}) \geqslant U(I_R) - C(I_\varnothing, I_R)$$

So the receiver is weakly better off. □

Proof of Proposition 1.2

According to Lemma A.1, $I_R \in BR(I_\varnothing)$ and $I_{SR} \in BR(I_S)$ imply $C(I_\varnothing, I_R) < +\infty$ and $C(I_S, I_{SR}) < +\infty$. So all the costs involved here are finite, and thus the conventional algebraic properties hold. $I_{SR} \in BR(I_S)$ implies

$$U(I_{SR}) - C(I_S, I_{SR}) \geqslant U(I_S) - C(I_S, I_S) = U(I_S)$$

Together with $U(I_S) \geqslant U(I_R)$ (implied by $I_S \succsim_B I_R$) and $C(I_\varnothing, I_R) \geqslant 0$, it implies

$$U(I_{SR}) - C(I_S, I_{SR}) \geqslant U(I_R) - C(I_\varnothing, I_R)$$

So the receiver is weakly better off.

APPENDIX B

Appendix for Chapter 3

Proof of Lemma 3.1

Proof. Without loss of generality, assume the prior is the uniform distribution. First consider $y = 0$.

β implies the following posterior distribution: the probability of the posterior probability of state 1 being

$$\frac{\beta(1-\alpha)+(1-\beta)x+2\alpha-1}{2\beta(1-\alpha)+(1-\beta)x+2\alpha-1}$$

(observing mm) equals

$$\frac{2\beta(1-\alpha)+(1-\beta)x+2\alpha-1}{2}$$

the probability of the posterior probability of state 1 being

$$\frac{1-\alpha-x}{2(1-\alpha)-x}$$

(observing $m\hat{m}$ or $\hat{m}m$) equals

$$(1-\beta)[2(1-\alpha)-x]$$

and the probability of the posterior probability of state 1 being

$$\frac{\beta(1-\alpha)+(1-\beta)x}{2\beta(1-\alpha)+(1-\beta)x+2\alpha-1}$$

(observing $\hat{m}\hat{m}$) equals

$$\frac{2\beta(1-\alpha)+(1-\beta)x+2\alpha-1}{2}$$

β' implies the following posterior distribution: the probability of the posterior probability of state 1 being

$$\frac{\beta'\left(1-\alpha\right)+\left(1-\beta'\right)x+2\alpha-1}{2\beta'\left(1-\alpha\right)+\left(1-\beta'\right)x+2\alpha-1}$$

(observing mm) equals

$$\frac{2\beta'\left(1-\alpha\right)+\left(1-\beta'\right)x+2\alpha-1}{2}$$

the probability of the posterior probability of state 1 being

$$\frac{1-\alpha-x}{2\left(1-\alpha\right)-x}$$

(observing $m\hat{m}$ or $\hat{m}m$) equals

$$\left(1-\beta'\right)\left[2\left(1-\alpha\right)-x\right]$$

and the probability of the posterior probability of state 1 being

$$\frac{\beta'\left(1-\alpha\right)+\left(1-\beta'\right)x}{2\beta'\left(1-\alpha\right)+\left(1-\beta'\right)x+2\alpha-1}$$

(observing $\hat{m}\hat{m}$) equals

$$\frac{2\beta'\left(1-\alpha\right)+\left(1-\beta'\right)x+2\alpha-1}{2}$$

Suppose that, conditionally on the posterior resulting from β being

$$\frac{\beta\left(1-\alpha\right)+\left(1-\beta\right)x+2\alpha-1}{2\beta\left(1-\alpha\right)+\left(1-\beta\right)x+2\alpha-1},$$

we create a new posterior distribution where the posterior equals

$$\frac{\beta'\left(1-\alpha\right)+\left(1-\beta'\right)x+2\alpha-1}{2\beta'\left(1-\alpha\right)+\left(1-\beta'\right)x+2\alpha-1}$$

with probability

$$\frac{2\beta'\left(1-\alpha\right)+\left(1-\beta'\right)x+2\alpha-1}{2\beta\left(1-\alpha\right)+\left(1-\beta\right)x+2\alpha-1}>0$$

and the posterior equals

$$\frac{1-\alpha-x}{2\left(1-\alpha\right)-x}$$

with probability

$$\frac{\left(\beta-\beta'\right)\left[2\left(1-\alpha\right)-x\right]}{2\beta\left(1-\alpha\right)+\left(1-\beta\right)x+2\alpha-1}>0$$

The expected value of this new posterior distribution is

$$\frac{\beta(1-\alpha)+(1-\beta)x+2\alpha-1}{2\beta(1-\alpha)+(1-\beta)x+2\alpha-1},$$

Similarly, conditionally on the posterior resulting from β being

$$\frac{\beta(1-\alpha)+(1-\beta)x}{2\beta(1-\alpha)+(1-\beta)x+2\alpha-1},$$

we create a new posterior distribution where the posterior equals

$$\frac{\beta'(1-\alpha)+(1-\beta')x}{2\beta'(1-\alpha)+(1-\beta')x+2\alpha-1}$$

with probability

$$\frac{2\beta'(1-\alpha)+(1-\beta')x+2\alpha-1}{2\beta(1-\alpha)+(1-\beta)x+2\alpha-1}>0$$

and the posterior equals

$$\frac{1-\alpha-x}{2(1-\alpha)-x}$$

with probability

$$\frac{(\beta-\beta')[2(1-\alpha)-x]}{2\beta(1-\alpha)+(1-\beta)x+2\alpha-1}>0$$

The expected value of this new posterior distribution is

$$\frac{\beta(1-\alpha)+(1-\beta)x}{2\beta(1-\alpha)+(1-\beta)x+2\alpha-1}$$

The distribution over posteriors after the mean preserving spread is the same as the posterior distribution resulting from β'. So the joint distribution resulting from β' strictly Blackwell dominates the joint distribution resulting from β.

The case for $x=0$ is similar. □

Proof of Example 3.1

Proof. Without loss of generality, assume the prior is the uniform distribution.

$x=\frac{1}{9}$ and $y=\frac{1}{9}$ imply the following posterior distribution: the probability of the posterior probability of state 1 being $\frac{4}{5}$ (observing mm) equals $\frac{5}{18}$, the probability of the posterior probability of state 1 being $\frac{1}{2}$ (observing $m\hat{m}$ or $\hat{m}m$) equals $\frac{4}{9}$, and the probability of the posterior probability of state 1 being $\frac{1}{5}$ (observing $\hat{m}\hat{m}$) equals $\frac{5}{18}$.

$x=\frac{1}{3}$ and $y=\frac{1}{6}$ imply the following posterior distribution: the probability of the posterior probability of state 1 being $\frac{4}{5}$ (observing mm) equals $\frac{5}{12}$, the probability of the posterior probability of state 1 being 0 (observing $m\hat{m}$ or $\hat{m}m$) equals $\frac{1}{6}$, and the probability of the posterior probability of state 1 being $\frac{2}{5}$ (observing $\hat{m}\hat{m}$) equals $\frac{5}{12}$.

Suppose that, conditionally on the posterior resulting from $\left(\frac{1}{9}, \frac{1}{9}\right)$ being $\frac{1}{2}$, we create a new posterior distribution where the posterior equals $\frac{4}{5}$ with probability $\frac{5}{16}$, the posterior equals $\frac{2}{5}$ with probability $\frac{10}{16}$, and the posterior equals 0 with probability $\frac{1}{16}$. The expected value of this new posterior distribution is $\frac{1}{2}$. Similarly, conditionally on the posterior resulting from $\left(\frac{1}{9}, \frac{1}{9}\right)$ being $\frac{1}{5}$, we create a new posterior distribution where the posterior equals $\frac{2}{5}$ with probability $\frac{1}{2}$ and the posterior equals 0 with probability $\frac{1}{2}$. The expected value of this new posterior distribution is $\frac{1}{5}$. The distribution over posteriors after the mean preserving spread is the same as the posterior distribution resulting from $\left(\frac{1}{3}, \frac{1}{6}\right)$. So the joint distribution resulting from $\left(\frac{1}{3}, \frac{1}{6}\right)$ strictly Blackwell dominates the joint distribution resulting from $\left(\frac{1}{9}, \frac{1}{9}\right)$. $\square$

Proof of Proposition 3.1

(1) We prove that there are no distributions of posteriors implied by joint signals in S^L that are mean preserving spreads of the distributions of posteriors implied by $\tilde{s}^{PN}$.

In the posterior distribution implied by $\tilde{s}^{PN}$, the probability of the posterior probability of state 1 being 1 (observing mm) equals $\frac{\alpha}{2}$ and the probability of the posterior probability of state 1 being 0 (observing $m\hat{m}$ or $\hat{m}m$) equals $1-\alpha$. To be a mean preserving spread of this distribution, the probability of the posterior probability of state 1 being 1 must be at least $\frac{\alpha}{2}$ and the probability of the posterior probability of state 1 being 0 must be at least $1-\alpha$. When $x \in [0, 1-\alpha)$ and $y = 0$, the posterior probability of state 1 being 1 (observing mm) equals $\frac{x+2\alpha-1+y}{2} < \frac{\alpha}{2}$. When $x = 0$ and $y \in (0, 1-\alpha)$, the posterior probability of state 1 being 1 equals $0 < \frac{\alpha}{2}$. When $x = 0$ and $y = 1-\alpha$, the probability of the posterior probability of state 1 being 1 (observing $m\hat{m}$ or $\hat{m}m$) equals $1-\alpha$ and the probability of the posterior probability of state 1 being 0 (observing $\hat{m}\hat{m}$) equals $\frac{\alpha}{2}$. So if $\alpha \in \left(\frac{1}{2}, \frac{2}{3}\right) \cup \left(\frac{2}{3}, 1\right)$, either $1-\alpha < \frac{\alpha}{2}$ or $\frac{\alpha}{2} < 1-\alpha$. If $\alpha = \frac{2}{3}$, $\tilde{s}^{PN}$ and $\tilde{s}^{NP}$ have the same distribution over posteriors. In sum, there are no distributions of posteriors implied by joint signals in S^L that are mean preserving spreads of the distributions of posteriors implied by $\tilde{s}^{PN}$.

(2) The proof that there are no distributions of posteriors implied by joint signals in S^L that are mean preserving spreads of the distributions of posteriors implied by $\tilde{s}^{NP}$ is similar to (1).

(3) We prove that, for $\alpha \in \left(\frac{1}{2}, \frac{3}{4}\right]$, the joint signals corresponding to $x \in [0, 1-\alpha)$ and $y = 0$ are strictly Blackwell dominated by $\tilde{s}^{PN}$ and the joint signals corresponding to $x = 0$ and $y \in [0, 1-\alpha)$ are strictly Blackwell dominated by $\tilde{s}^{NP}$.

First consider joint signals with $x \in [0, 1-\alpha)$ and $y = 0$. In the implied posterior distribution, the probability of the posterior probability of state 1 being 1 (observing mm) equals $\frac{x+2\alpha-1}{2}$, the probability of the posterior probability of state 1 being $\frac{1-\alpha-x}{2(1-\alpha)-x}$ (observing $m\hat{m}$ or $\hat{m}m$) equals $2(1-\alpha)-x$, and the probability of the posterior probability of state 1 being $\frac{x}{x+(2\alpha-1)}$ (observing $\hat{m}\hat{m}$) equals $\frac{x+2\alpha-1}{2}$.

Note in the posterior distribution implied by $\tilde{s}^{PN}$, the probability of the posterior probability of state 1 being 1 (observing mm) equals $\frac{\alpha}{2}$, the probability of the posterior probability of state 1 being 0 (observing $m\hat{m}$ or $\hat{m}m$) equals $1-\alpha$, and the probability of the posterior probability of state 1 being $\frac{1-\alpha}{\alpha}$ (observing $\hat{m}\hat{m}$) equals $\frac{\alpha}{2}$.

The posterior distribution implied by $\tilde{s}^{PN}$ is a mean preserving spread of the posterior distribution implied by a joint signals with $x \in [0, 1-\alpha)$ and $y = 0$, if and only if there exists $\beta_i \in \mathbb{R}_+, i = 1, 2, ..., 6$ such that

$$
\begin{aligned}
\beta_1 + \beta_2 + \beta_3 &= 1 \\
\beta_4 + \beta_5 + \beta_6 &= 1 \\
\frac{1-\alpha-x}{2(1-\alpha)-x} &= \beta_1 \cdot 1 + \beta_2 \frac{1-\alpha}{\alpha} + \beta_3 \cdot 0 \\
\frac{x}{x+2\alpha-1} &= \beta_4 \cdot 1 + \beta_5 \frac{1-\alpha}{\alpha} + \beta_6 \cdot 0 \\
\frac{\alpha}{2} &= \frac{x+2\alpha-1}{2} + \beta_1 [2(1-\alpha)-x] + \beta_4 \frac{x+2\alpha-1}{2} \\
\frac{\alpha}{2} &= \beta_2 [2(1-\alpha)-x] + \beta_5 \frac{x+2\alpha-1}{2} \\
1-\alpha &= \beta_3 [2(1-\alpha)-x] + \beta_6 \frac{x+2\alpha-1}{2}
\end{aligned}
$$

It is equivalent to

$$
\begin{aligned}
\beta_1 &= \frac{1-\alpha-x}{2(1-\alpha)-x} - \frac{1-\alpha}{\alpha}\beta_2 \\
\beta_3 &= \frac{1-\alpha}{2(1-\alpha)-x} - \frac{2\alpha-1}{\alpha}\beta_2 \\
\beta_4 &= \frac{2(1-\alpha)[2(1-\alpha)-x]\beta_2 - \alpha(1-\alpha-x)}{\alpha(x+2\alpha-1)} \\
\beta_5 &= \frac{\alpha - 2[2(1-\alpha)-x]\beta_2}{x+2\alpha-1} \\
\beta_6 &= \frac{2(2\alpha-1)[2(1-\alpha)-x]}{\alpha(x+2\alpha-1)}\beta_2
\end{aligned}
$$

where β_2 is the free variable.

$\beta_1 \geq 0$ is equivalent to

$$\beta_2 \leq \frac{\alpha(1-\alpha-x)}{(1-\alpha)[2(1-\alpha)-x]}$$

$\beta_3 \geq 0$ is equivalent to

$$\beta_2 \leq \frac{\alpha(1-\alpha)}{(2\alpha-1)[2(1-\alpha)-x]}$$

$\beta_4 \geq 0$ is equivalent to

$$\beta_2 \geq \frac{\alpha(1-\alpha-x)}{2(1-\alpha)[2(1-\alpha)-x]}$$

$\beta_5 \geq 0$ is equivalent to

$$\beta_2 \leq \frac{\alpha}{2[2(1-\alpha)-x]}$$

$\beta_6 \geq 0$ is equivalent to

$$\beta_2 \geq 0$$

Note

$$\frac{\alpha(1-\alpha-x)}{(1-\alpha)[2(1-\alpha)-x]} \geq \frac{\alpha(1-\alpha-x)}{2(1-\alpha)[2(1-\alpha)-x]}$$

and

$$\frac{\alpha}{2[2(1-\alpha)-x]} \geq \frac{\alpha(1-\alpha-x)}{2(1-\alpha)[2(1-\alpha)-x]}$$

So β_2 exists if and only if

$$\frac{\alpha(1-\alpha-x)}{2(1-\alpha)[2(1-\alpha)-x]} \leq \frac{\alpha(1-\alpha)}{(2\alpha-1)[2(1-\alpha)-x]}$$

which is equivalent to

$$x \geq \frac{(1-\alpha)(4\alpha-3)}{2\alpha-1} \equiv x^*$$

For $\alpha \in \left(\frac{1}{2}, \frac{3}{4}\right]$, $x^* \leq 0$, so the joint signals corresponding to $x \in [0, 1-\alpha)$ and $y = 0$ are strictly Blackwell dominated by $\tilde{s}^{PN}$.

The proof that the joint signals corresponding to $x = 0$ and $y \in [0, 1-\alpha)$ are strictly Blackwell dominated by $\tilde{s}^{NP}$ is similar.

(4) We prove that, for $\alpha \in \left(\frac{3}{4}, 1\right)$, there are no distributions of posteriors implied by joint signals in S^L that are mean preserving spreads of the distributions of posteriors implied by joint signals with $xy = 0$, $x < x^*$, and $y < x^*$.

First consider any joint signal with $x \in [0, x^*)$ and $y = 0$.

(a) The proof in (3) illustrates that it is not strictly Blackwell dominated by $\tilde{s}^{PN}$.

(b) It is also not strictly Blackwell dominated by joint signals with $x' \in [x^*, 1-\alpha)$ and $y' = 0$, otherwise it is strictly Blackwell dominated by $\tilde{s}^{PN}$.

(c) We prove that the posterior distribution implied by $x' \in (x, x^*)$ and $y' = 0$ is not a mean preserving spread of the posterior distribution implied by (x, y).

In the posterior distribution implied by (x, y), the probability of the posterior probability of state 1 being 1 (observing mm) equals $\frac{x+2\alpha-1}{2}$, the probability of the posterior probability of state 1 being $\frac{1-\alpha-x}{2(1-\alpha)-x}$ (observing $m\hat{m}$ or $\hat{m}m$) equals $2(1-\alpha)-x$, and the probability of the posterior probability of state 1 being $\frac{x}{x+(2\alpha-1)}$ (observing $\hat{m}\hat{m}$) equals $\frac{x+2\alpha-1}{2}$.

In the posterior distribution implied by (x', y'), the probability of the posterior probability of state 1 being 1 (observing mm) equals $\frac{x'+2\alpha-1}{2}$, the probability of the posterior probability of state 1 being $\frac{1-\alpha-x'}{2(1-\alpha)-x'}$ (observing $m\hat{m}$ or $\hat{m}m$) equals $2(1-\alpha)-x'$, and the probability of the posterior probability of state 1 being $\frac{x'}{x'+(2\alpha-1)}$ (observing $\hat{m}\hat{m}$) equals $\frac{x'+2\alpha-1}{2}$.

Since

$$x^* < \frac{(1-\alpha)(2\alpha-1)}{\alpha}$$

we have

$$\frac{x}{x+(2\alpha-1)} < \frac{1-\alpha-x}{2(1-\alpha)-x}$$

and

$$\frac{x'}{x'+(2\alpha-1)} < \frac{1-\alpha-x'}{2(1-\alpha)-x'}$$

So the lowest posterior probability of state 1 in the posterior distribution implied by (x, y) is $\frac{x}{x+(2\alpha-1)}$, while the lowest posterior probability of state 1 in the posterior distribution implied by (x', y') is $\frac{x'}{x'+(2\alpha-1)}$. Note

$$\frac{x'}{x'+(2\alpha-1)} > \frac{x}{x+(2\alpha-1)}$$

As a result, the posterior distribution implied by (x', y') is not a mean preserving spread of the posterior distribution implied by $x \in [0, x^*)$ and $y = 0$ because a mean preserving spread does not increase the lowest posterior probability of state 1.

(d) We prove that the posterior distribution implied by $x' \in [0, x)$ and $y' = 0$ is not a mean preserving spread of the posterior distribution implied by (x, y). (This case is trivial when $x = 0$)

In the posterior distribution implied by (x, y), the probability of the posterior probability of state 1 being 1 (observing mm) equals $\frac{x+2\alpha-1}{2}$.

In the posterior distribution implied by (x', y'), the probability of the posterior probability of state 1 being 1 (observing mm) equals $\frac{x'+2\alpha-1}{2}$. Note

$$\frac{x'+2\alpha-1}{2} < \frac{x+2\alpha-1}{2}$$

As a result, the posterior distribution implied by (x', y') is not a mean preserving spread of the posterior distribution implied by (x, y) because a mean preserving spread does not decrease the probability of the posterior probability of state 1 being 1.

(e) We prove that the posterior distribution implied by $x' = 0$ and $y \in (0, 1-\alpha)$ is not a mean preserving spread of the posterior distribution implied by (x, y).

In the posterior distribution implied by (x, y), the probability of the posterior probability of state 1 being 1 (observing mm) equals $\frac{x+2\alpha-1}{2}$.

In the posterior distribution implied by (x', y'), the probability of the posterior probability of state 1 being 1 equals 0. Note

$$0 < \frac{x + 2\alpha - 1}{2}$$

As a result, the posterior distribution implied by (x', y') is not a mean preserving spread of the posterior distribution implied by (x, y).

(f) We prove that the posterior distribution implied by $\tilde{s}^{NP}$ is not a mean preserving spread of the posterior distribution implied by (x, y).

In the posterior distribution implied by (x, y), the probability of the posterior probability of state 1 being 1 (observing mm) equals $\frac{x+2\alpha-1}{2}$.

In the posterior distribution implied by $\tilde{s}^{NP}$, the probability of the posterior probability of state 1 being 1 equals $1 - \alpha$. Note

$$1 - \alpha < \frac{x + 2\alpha - 1}{2}$$

when $\alpha \in \left(\frac{3}{4}, 1\right)$. As a result, the posterior distribution implied by $\tilde{s}^{NP}$ is not a mean preserving spread of the posterior distribution implied by (x, y).

In sum, there are no distributions of posteriors implied by joint signals in S^L that are mean preserving spreads of the distributions of posteriors implied by joint signals with $x \in [0, x^*)$ and $y = 0$. So the joint signals with $x \in [0, x^*)$ and $y = 0$ are not strictly Blackwell dominated by any joint signal in S^L.

The proof that the joint signals with $x = 0$ and $y \in (0, x^*)$ are not strictly Blackwell dominated by any joint signal in S^L is similar.

Proof of Proposition 3.3

Proof. We show the statements in the sketch of the proof one by one.

(1) Consider the following decision problem, denoted by D_δ, where the set of actions $A = \{a_1, a_2\}$ and the utility function u is as follows.

	ω_1	ω_2
a_1	0	$-(1+\delta)$
a_2	-1	0

Table B.1: A class of decision problems: D_δ

According to (3.1), the payoffs of $\tilde{s}^{PN}$, $\tilde{s}^{NP}$, and $\tilde{s}^{NN}$ are as follows.

$$V(D_\delta, \tilde{s}^{PN}) = \max\left\{-\frac{1}{2}(2\alpha-1)(1+\delta), -\frac{1}{2}(1-\alpha)\right\}$$

$$V(D_\delta, \tilde{s}^{NP}) = \max\left\{-\frac{1}{2}(1-\alpha)(1+\delta), -\frac{1}{2}(2\alpha-1)\right\}$$

$$\begin{aligned} &V(D_\delta, \tilde{s}^{NN}) \\ =\ &2\max\left\{-\frac{1}{2}(1-\alpha)(1+\delta), -\frac{1}{2}(1-\alpha)\right\} \\ &+\max\left\{-\frac{1}{2}(2\alpha-1)(1+\delta), 0\right\} \end{aligned}$$

For $\alpha \in \left(\frac{1}{2}, \frac{2}{3}\right)$, let $\delta \in \left(\frac{3\alpha-2}{1-\alpha}, 0\right)$, then

$$\begin{aligned} V(D_\delta, \tilde{s}^{PN}) &= -\frac{1}{2}(2\alpha-1)(1+\delta) \\ V(D_\delta, \tilde{s}^{NP}) &= -\frac{1}{2}(2\alpha-1) \\ V(D_\delta, \tilde{s}^{NN}) &= -(1-\alpha)(1+\delta) \end{aligned}$$

$\alpha > \frac{1}{2}$ implies $\delta > \frac{3\alpha-2}{1-\alpha} > -1$. So $1+\delta \in (0,1)$. As a result, $V\left(D_\delta, \tilde{s}^{PN}\right) > V\left(D_\delta, \tilde{s}^{NP}\right)$. $\alpha < \frac{2}{3}$ implies $-\frac{1}{2}(2\alpha-1) > -(1-\alpha)$. As a result, $V\left(D_\delta, \tilde{s}^{PN}\right) > V\left(D_\delta, \tilde{s}^{NN}\right)$.

For $\alpha = \frac{2}{3}$, $\tilde{s}^{PN}$ and $\tilde{s}^{NP}$ have the same distribution over posteriors. Let $\delta = 0$, then

$$\begin{aligned} V(D_\delta, \tilde{s}^{PN}) &= -\frac{1}{6} \\ V(D_\delta, \tilde{s}^{NN}) &= -\frac{1}{3} \end{aligned}$$

As a result, $V\left(D_\delta, \tilde{s}^{PN}\right) > V\left(D_\delta, \tilde{s}^{NN}\right)$.

For $\alpha \in \left(\frac{2}{3}, 1\right)$, let $\delta \in \left(0, \frac{3\alpha-2}{1-\alpha}\right)$, then

$$V(D_\delta, \tilde{s}^{PN}) = -\frac{1}{2}(1-\alpha)$$
$$V(D_\delta, \tilde{s}^{NP}) = -\frac{1}{2}(1-\alpha)(1+\delta)$$
$$V(D_\delta, \tilde{s}^{NN}) = -(1-\alpha)$$

It is easy to see $V\left(D_\delta, \tilde{s}^{PN}\right) > V\left(D_\delta, \tilde{s}^{NP}\right)$ and $V\left(D_\delta, \tilde{s}^{PN}\right) > V\left(D_\delta, \tilde{s}^{NN}\right)$.

(2) Again we consider D_δ. For $\alpha \in \left(\frac{1}{2}, \frac{2}{3}\right)$, let $\delta \in \left(0, \frac{2-3\alpha}{2\alpha-1}\right)$, then

$$V(D_\delta, \tilde{s}^{PN}) = -\frac{1}{2}(2\alpha-1)(1+\delta)$$
$$V(D_\delta, \tilde{s}^{NP}) = -\frac{1}{2}(2\alpha-1)$$
$$V(D_\delta, \tilde{s}^{NN}) = -(1-\alpha)$$

$\delta > 0$ implies $V\left(D_\delta, \tilde{s}^{NP}\right) > V\left(D_\delta, \tilde{s}^{PN}\right)$. $\alpha < \frac{2}{3}$ implies $-\frac{1}{2}(2\alpha-1) > -(1-\alpha)$. As a result, $V\left(D_\delta, \tilde{s}^{NP}\right) > V\left(D_\delta, \tilde{s}^{NN}\right)$.

The case of $\delta = \frac{2}{3}$ is the same as (1).

For $\alpha \in \left(\frac{2}{3}, 1\right)$, let $\delta \in \left(\frac{2-3\alpha}{2\alpha-1}, 0\right)$, then

$$V(D_\delta, \tilde{s}^{PN}) = -\frac{1}{2}(1-\alpha)$$
$$V(D_\delta, \tilde{s}^{NP}) = -\frac{1}{2}(1-\alpha)(1+\delta)$$
$$V(D_\delta, \tilde{s}^{NN}) = -(1-\alpha)(1+\delta)$$

$\alpha < 1$ implies $\delta > \frac{2-3\alpha}{2\alpha-1} > -1$. So $1+\delta \in (0,1)$. As a result, $V\left(D_\delta, \tilde{s}^{NP}\right) > V\left(D_\delta, \tilde{s}^{PN}\right)$. It is easy to see $V\left(D_\delta, \tilde{s}^{NP}\right) > V\left(D_\delta, \tilde{s}^{NN}\right)$.

(3) $\tilde{s}^{PN}$ implies the following posterior distribution: the probability of the posterior probability of ω_1 being 1 (observing mm) equals $\alpha/2$; the probability of the posterior probability of ω_1 being 0 (observing $m\hat{m}$ equals $1-\alpha$ or $\hat{m}m$), and the probability of the posterior probability of ω_1 being $(1-\alpha)/\alpha$ (observing $\hat{m}\hat{m}$) equals $\alpha/2$.

$\tilde{s}^{NP}$ implies the following posterior distribution: the probability of the posterior probability of ω_1 being $(2\alpha-1)/\alpha$ (observing mm) equals $\alpha/2$; the probability of the posterior probability of ω_1 being 1 (observing $m\hat{m}$ or $\hat{m}m$) equals $1-\alpha$, and the probability of the posterior probability of ω_1 being 0 (observing $\hat{m}\hat{m}$) equals $\alpha/2$.

$\tilde{s}^{NN}$ implies the following posterior distribution: the probability of the posterior probability of ω_1 being 1 (observing mm) equals $\alpha-(1/2)$; the probability of the posterior probability of ω_1 being 1/2 (observing $m\hat{m}$ or $\hat{m}m$) equals $2(1-\alpha)$, and the probability of the posterior probability of ω_1 being 0 (observing $\hat{m}\hat{m}$) equals $\alpha-(1/2)$.

Suppose that, conditionally on the posterior resulting from $\tilde{s}_{NN}$ being 1/2, we create a new posterior distribution according to which the posterior is 0 with probability $(4-4\alpha)/(8-8\alpha)$, and equal to $(2\alpha-1)/\alpha$ with probability $\alpha/(8-8\alpha)$, equal to $(1-\alpha)/\alpha$ with probability $\alpha/(8-8\alpha)$, and equal to 1 with probability $(4-5\alpha)/(8-8\alpha)$. One can verify that the expected value of this new posterior distribution is 1/2. If s_{NN} results in posteriors 0 or 1, we do not change the posterior distribution. The new posterior distribution that we have created is exactly the posterior distribution of the 50-50 convex combination of $\tilde{s}^{PN}$ and $\tilde{s}-NP$. Note that $\alpha \leq \frac{4}{5}$ ensures $(4-5\alpha)/(8-8\alpha) \geq 0$. So the joint distribution resulting from the 50-50 convex combination of $\tilde{s}^{PN}$ and $\tilde{s}^{NP}$ strictly Blackwell dominates the joint distribution resulting from $\tilde{s}^{NN}$.

(4) Consider the following decision problem, denoted by D_α:

	ω_1	ω_2
a_1	0	-1
a_2	-1	0
a_3	$-\frac{1-\alpha}{\alpha}$	$-\frac{1-\alpha}{\alpha}$

Table B.2: A class of decision problems: D_α

The expected payoff of each action in the above decision problem is illustrated in Figure B.1:

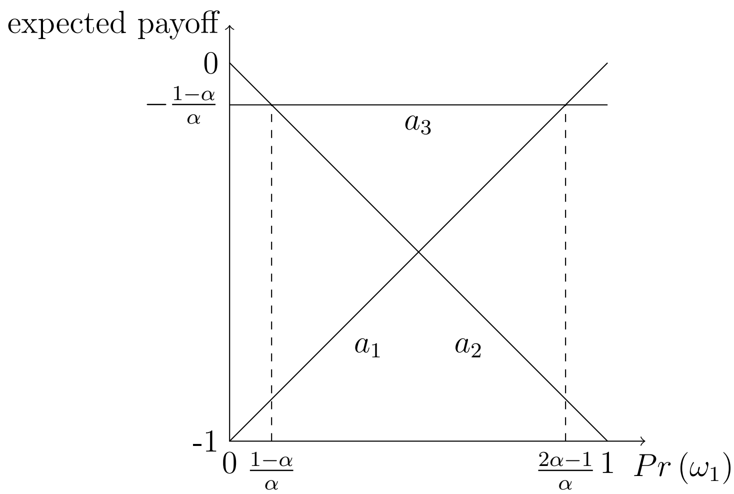


Figure B.1: The expected payoff for each action in D_α

So

$$V(D_\alpha, \tilde{s}^{PN}) = V(D_\alpha, \tilde{s}^{NP}) = \frac{\alpha}{2} \cdot 0 + (1-\alpha) \cdot 0 + \frac{\alpha}{2} \cdot \left(-\frac{1-\alpha}{\alpha}\right) = -\frac{1-\alpha}{2}$$

while

$$V(D_\alpha, \tilde{s}^{NN}) = \left(\alpha - \frac{1}{2}\right) \cdot 0 + (2-2\alpha) \cdot \left(-\frac{1-\alpha}{\alpha}\right) + \left(\alpha - \frac{1}{2}\right) \cdot 0$$

Note that $\alpha > \frac{4}{5}$ implies $2 - 2\alpha < \frac{\alpha}{2}$. So $V\left(D_\alpha, \tilde{s}^{NN}\right) > V\left(D_\alpha, \tilde{s}^{PN}\right)$ and $V\left(D_\alpha, \tilde{s}^{NN}\right) > V\left(D_\alpha, \tilde{s}^{NP}\right)$. □

BIBLIOGRAPHY

Athey, S. and J. Levin (2018). The value of information in monotone decision problems. *Research in Economics 72*(1), 101–116.

Bizzotto, J., J. Rüdiger, and A. Vigier (2020). Testing, disclosure and approval. *Journal of Economic Theory 187*, 105002.

Blackwell, D. (1951). Comparison of experiments. In J. Neyman (Ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 93–102. Berkeley and Los Angeles: University of California Press.

Blackwell, D. (1953). Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 265–272.

Bloedel, A. W. and I. Segal (2021). Persuading a rationally inattentive agent. Working Paper.

Bloedel, A. W. and W. Zhong (2021). The cost of optimally acquired information. Working Paper.

Blom, G. (1975). On raoblackwellization. *Scandinavian Journal of Statistics 2*(3), 145–148.

Caplin, A. and M. Dean (2013). Behavioral implications of rational inattention with shannon entropy. Working Paper.

Caplin, A., M. Dean, and J. Leahy (2019a). Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies 86*(3), 1061–1094.

Caplin, A., M. Dean, and J. Leahy (2019b). Rationally inattentive behavior: characterizing and generalizing shannon entropy. Working Paper.

Clemen, R. T. and R. L. Winkler (1985). Limits for the precision and value of information from dependent sources. *Operations Research 33*(2), 427–442.

Elga, A. (2010). How to disagree about how to disagree. In R. Feldman and T. A. Warfield (Eds.), *Disagreement*, pp. 175–186. Oxford: Oxford University Press.

Fishburn, P. C. (1975). Separation theorems and expected utilities. *Journal of Economic Theory 11*(1), 16–34.

Gentzkow, M. and E. Kamenica (2014). Costly persuasion. *American Economic Review 104*(5), 457–462.

Hébert, B. and M. Woodford (2021). Rational inattention when decisions take time. Working Paper.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Kim, Y. (2022). Comparing information in general monotone decision problems. Working Paper.

Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li (2017). Persuasion of a privately informed receiver. *Econometrica 85*(6), 1949–1964.

Lehmann, E. L. (1988). Comparing location experiments. *The Annals of Statistics 16*(2), 521–533.

Lipnowski, E., L. Mathevet, and D. Wei (2020). Attention management. *American Economic Review: Insights 2*(1), 17–32.

Lipnowski, E., L. Mathevet, and D. Wei (2022). Optimal attention management: a tractable framework. Working Paper.

Lohr, S. L. (2021). *Sampling: design and analysis*. Boca Raton: Chapman and Hall/CRC.

Mandela, N. (2011). *Nelson Mandela by himself: The authorised book of quotations*. Johannesburg: Pan Macmillan.

Martin, D. (2017). Strategic pricing with rational inattention to quality. *Games and Economic Behavior 104*, 131–145.

Mas-Colell, A., M. D. Winston, and J. Green (1995). *Microeconomic Theory*. New York: Oxford University Press.

Matějka, F. and A. McKay (2015). Rational inattention to discrete choices: a new foundation for the multinomial logit model. *American Economic Review 105*(1), 272–298.

Matysková, L. and A. Montes (2021). Bayesian persuasion with costly information acquisition. Working Paper.

Morris, S. and P. Strack (2019). The wald problem and the equivalence of sequential sampling and ex-ante information costs. Working Paper.

O'Connor, R. (2014). Practical tools for teaching news literacy. *The New York Times*.

Pearce, D. G. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica: Journal of the Econometric Society*, 1029–1050.

Quirk, J. P. and R. Saposnik (1962). Admissibility and measurable utility functions. *The Review of Economic Studies 29*(2), 140–146.

Ravid, D. (2019). Bargaining with rational inattention. Working Paper.

Rockafellar, T. (1970). *Convex Analysis*. Princeton: Princeton University Press.

Rothschild, M. and J. E. Stiglitz (1970). Increasing risk: I. a definition. *Journal of Economic theory 2*(3), 225–243.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Vohra, R. (2005). *Advanced Mathematical Economics*. Abingdon: Routledge.

Wei, D. (2021). Persuasion under costly learning. *Journal of Mathematical Economics 94*, 102451.

Weinstein, J. (2020). Best-reply sets. *Economic Theory Bulletin 8*, 105–112.

Wittgenstein, L. (2009). *Philosophical investigations* (4th ed.). John Wiley & Sons.

Wu, W. (2018). Sequential bayesian persuasion. Working paper.

Yang, M. (2012). Coordination with rational inattention. Working Paper.