Dominance and Optimality[*]

by Xienan Cheng[†] ⓡ Tilman Börgers[‡]

March 9, 2023

Abstract

This paper 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.

[*]We are grateful to Yaron Azrieli, Andrew Ellis, Ian Jewitt, Jianpei Li, Shuo Liu, and Balázs Szentes for comments.

[†]Guanghua School of Management, Peking University, xienancheng@gsm.pku.edu.cn.

[‡]Department of Economics, University of Michigan, tborgers@umich.edu.

1. Introduction

In many economic contexts a robust prediction implied by the rationality hypothesis is that agents will not choose "dominated" options. For example, rational agents won't choose "dominated insurance plans," and rational players won't choose "dominated strategies" in games. In all these contexts one option "dominates" another one if the former is better than the latter regardless of some specifics. What these specifics are depends on the context. For example, the relevant specifics may be an agent's risk attitude, or they may be a player's beliefs about other players' behavior. The prediction that rational agents won't choose dominated strategies is *robust* because it is true regardless of these specifics.

But can these predictions be further narrowed down without giving up robustness? The answer is "no" if for every undominated option there are some specifics in which these options are the best choice. In other words, if "not being dominated" and "being potentially optimal" are equivalent, then the best robust prediction is indeed that agents don't choose dominated strategies. In this paper we prove several general theorems that assert such equivalences. We then apply our general theorems to various economic settings. We obtain some new results and re-prove some known results. Our general theory also illuminates when the equivalence of "not being dominated" and "being potentially optimal" may fail.

One of the new results that we obtain is that under some assumptions a signal is not Blackwell dominated (Blackwell, 1951, 1953) by other signals drawn from a given set of signals if and only if there is a decision problem such that, in this decision problem, the signal is the optimal choice from the given set. Examples of dominance notions that are covered by our theorems and for which some versions of our results are known from previous literature apply to strict and weak dominance in games Pearce (1984) and to first-order (Quirk and Saposnik, 1962) and second-order stochastic dominance (Rothschild and Stiglitz, 1970) among monetary lotteries.

Our proofs of our general results are based on separating hyperplane theorems for finite-dimensional Euclidean spaces. Separating hyperplane theorems in finite-dimensional Euclidean space can be illustrated in simple graphs, and therefore we hope that our proofs are easily accessible.

An important detail in our theorems is that these results hold only if one considers the possibility that one option is dominated not by another option by itself alone, but by a convex combination of other options. Indeed, the typical game theory textbook will provide examples that show that a strategy may not be strictly dominated by any other pure strategy, yet not be a best response for any belief that a player may hold about other players' behavior (see Figure 61.1 in Osborne and Rubinstein (1994)). This is a general theme that also matters in the other applications that we consider, and that we shall emphasize throughout.

It is a familiar fact from Pearce (1984) that the details of the notion of dominance among actions and also of the notion of optimality matter for the results that we are aiming for. That is also true in our context. We introduce different notions of being "not dominated" and relate them to a variety of notions of "being optimal." Some of our results will be based on assumptions that hold in some, but not in all applications. The most general result that will hold in all applications characterizes 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 most general result establishes an equivalence between an alternative not being dominated by convex combinations of other alternatives, and an alternative being a uniquely optimal choice.

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 provide sufficient conditions such that, when there are multiple optimal choices in a decision problem, then at least one of those choices is, in some other decision problem, the uniquely optimal choice. 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. Thus, it is a robust prediction that the decision maker will make a choice from this set provided that the decision maker has some very small attention cost.

The general results on which we build our analysis are presented in Section 2. We motivate a focus on uniquely optimal actions in Section 3. In the subsequent sections we apply our general analysis first to dominance relations in games (Section 4), then to dominance relations among experiments (Section 5), and finally to dominance relations among monetary lotteries (Section 6). We postpone a discussion of the previous results concerning these applications to Sections 4, 5 and 6.

2. General Results

Let X and Y be two non-empty sets. Here, we interpret X as the set of actions x that a decision maker can choose from, while we interpret Y as the set of "situations" y that this decision maker might find herself in. In the applications that we shall consider, X is, depending on the context, a player's set of strategies in a game, or a set of experiments, or a set of monetary lotteries. Y is, depending on the context, the set of beliefs the decision maker might hold, the set of value functions corresponding to the decision problems the decision maker might face, or the set of Bernoulli utility functions the decision maker might have. Throughout this paper, unless otherwise stated, we shall make the following assumption:

Assumption 1. *X is finite. Y is a subset of a nontrivial real vector space $\mathcal{Y}$.*

The assumption that X is finite greatly simplifies the analysis below. In the applications listed above it is, on the other hand, convenient to allow Y to be infinite. Moreover, in these examples Y is naturally embedded into a real vector space.

We also assume that a function $u : X \times \mathcal{Y} \to \mathbb{R}$ is given. When $y \in Y$ then $u(x, y)$ is the decision maker's utility if she chooses action x in situation y. We define the domain of the utility function u to be $X \times \mathcal{Y}$ rather than $X \times Y$ because this definition makes it slightly more convenient to express some technical arguments below.

The decision maker first observes the situation $y \in Y$ and then chooses an action $x \in X$.

Definition 1. *An action $x \in X$ is* optimal[1] *if there exists a $y \in Y$ such that:*

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

We denote by X_O the set of optimal actions.

Definition 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)*[2] *such that:*

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

[1] A more accurate phrase is: "potentially optimal," but, to shorten the terminology, we drop "potentially."

[2] Here, we mean by "relative interior" the algebraic relative interior; see Ok (2007, p. 438)

We denote by X_{IO} the set of interior optimal actions. It may not be intuitively obvious why it is relevant whether the situation y in which an action x is a best choice is interior or not. Consider the example where Y represents a set of all beliefs, i.e. probability distributions over, say, a finite set. Then the relative interior of Y is the set of full support beliefs. Having beliefs with full support has been interpreted in the game theoretic literature as a sign of caution by the decision maker. This is why interior situations y will receive special attention.

Definition 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 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 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 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 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 Theorems 1 and 2, our main results in this section. These results generalize some results familiar from game theory.

Theorem 1.

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

(ii) If u is linear on $\mathcal{Y}$, then an action that is interior optimal is not weakly dominated: $X_{IO} \subseteq X_{NWD}$.

(iii) If an action is uniquely optimal then it is not weakly dominated and not redundant: $X_{UO} \subseteq X_{NWD} \cap X_{NR}$.

Proof. We first prove (i). The proof is indirect. Suppose x were optimal for some $\bar{y} \in Y$, but that x were strictly dominated. Let $\{x_1, x_2, \ldots, x_n\} \subseteq X \setminus \{x\}$ 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 (ii). 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 (iii). 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. $\square$

Theorem 2. *Assume Y is convex and u is linear on $\mathcal{Y}$.*

(i) If $\mathcal{Y}$ can be endowed with a topology such that u is continuous on Y and Y is compact, then an action that is not strictly dominated is optimal: $X_{NSD} \subseteq X_O$.
(ii) An action that is not weakly dominated is optimal: $X_{NWD} \subseteq X_O$.
(iii) If $\mathcal{Y}$ is finite-dimensional, then an action that is not weakly dominated is interior optimal: $X_{NWD} \subseteq X_{IO}$.
(iv) An action that is not weakly dominated and not redundant is uniquely optimal: $X_{NWD} \cap X_{NR} \subseteq X_{UO}$.

The assumptions that Y is convex and that u is linear on $\mathcal{Y}$ set the stage for the application of the separating hyperplane theorems which will be used to prove Theorem 2. Note that results (i) and (iii) are based on additional assumptions. Of course, the theorem does not claim that these additional assumptions are necessary. However, if these additional assumptions do not hold, the conclusions of parts (i) and (iii) may fail. The appendix contains an example that demonstrates this for part (i). In Section 6 we illustrate how the conclusion of (iii) may fail if $\mathcal{Y}$ is not finite-dimensional.

The separating hyperplane theorems that we use are elementary in that they apply in finite-dimensional Euclidean space. The idea on which these proofs are based is borrowed from the proofs of Theorems 1 and 2 in Fishburn (1975).[3] We illustrate the idea of the proofs geometrically after presenting the formal proof. We also explain there how Fishburn's approach differs in an important way from the textbook approach to similar results that is also based on the separating hyperplane theorem, and that is exemplified by the proof of Theorem 2.2 in Fudenberg and Tirole (1991).

Corollary 1 in Battigalli et al. (2016) is a more general version of those parts of our Theorems 1 and 2 that apply to strict dominance. Their Corollary 1 is more general than our results because it allows the action set X to be infinite. An earlier, closely related result is the main proposition in Zimper (2005). Like Battigalli et al. (2016) and Zimper (2005) also Shah (2022) allows for infinite action set X. Shah (2022) proves versions of our results on strict and weak dominance.[4] None of these papers has a result relating the property of being not dominated to being uniquely optimal. All three papers use proof strategies that are different from our own. Zimper's proof is based on a separating hyperplane argument that is similar to the textbook approach we mentioned in the previous paragraph. Battigalli et. al.'s proofs are based on the minimax theorem, and Shah's proof is based on Browder's coincidence theorem.

Proof. We first prove (i). The proof is indirect. Let x be an action that is 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 elements of the set $X \setminus \{x\}$. If x is not optimal in any situation $y \in Y$, then:

$$C \cap \mathbb{R}^n_- = \emptyset.^{5}$$

Observe that C is non-empty, 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 separating hyperplane theorem which, to make this paper self-contained, we briefly prove:

[3]We explain in the last paragraph of Section 6 the relation between our Theorem 2 and Theorems 1 and 2 in Fishburn (1975).

[4]The first version of our paper and Shah (2022) were circulated simultaneously.

[5]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.

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

Proof. The strict separating hyperplane theorem for finite-dimensional Euclidean spaces (for example Proposition 5 in Ok (2007, p. 485)) implies that there are $\lambda \in \mathbb{R}^n$ with $\lambda \neq 0$ and $\varepsilon \in \mathbb{R}$ such that $\lambda \cdot c < \varepsilon$ for all $c \in \mathbb{R}^n_-$ and $\lambda \cdot c > \varepsilon$ for all $c \in C$. It easily follows that $\lambda \in \mathbb{R}^n_+$ and that $\varepsilon > 0$. Therefore, $\lambda \cdot c > \varepsilon$ for all $c \in C$ implies $\lambda \cdot c > 0$ for all $c \in C$. □

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:

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

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.

We next prove (ii). The proof is indirect. Suppose x were not optimal for any $y \in Y$. We prove that then a convex combination of actions weakly dominates x. Let C be defined as before. If x is not optimal in any situation $y \in Y$, then:

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

Observe that C is non-empty and convex (because of the linearity of u and the convexity of Y). We can then apply the following separating hyperplane theorem:

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

This is Lemma 5 in Fishburn (1975).

We normalize again 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:

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

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 next prove (iii). Because $\mathcal{Y}$ is a nontrivial finite-dimensional vector space, it is isomorphic to $\mathbb{R}^m$ for some $m \in \mathbb{N}$.[6] It is then without loss of generality to assume that $\mathcal{Y} = \mathbb{R}^m$ for some $m \in \mathbb{N}$. Define the set C as before. Because Y is a convex subset of $\mathbb{R}^m$, by Exercise 21 in Ok (2007, p. 444), the algebraic relative interior of Y is the same as the topological relative interior of Y. Because C is the image of Y in $\mathbb{R}^n$ under a linear mapping, it is convex, and therefore, again, the algebraic relative interior of C is the same as the topological relative interior of C. Moreover, by Theorem 6.6 in Rockafellar (1970), the topological relative interior of C is the image under the linear mapping of the topological 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 3: Suppose $C \subseteq \mathbb{R}^n$ is non-empty and convex. If $ri(C) \cap \mathbb{R}^n_- = \emptyset$ then there exists $\lambda \in \mathbb{R}^n_+$ with $\lambda \neq 0$ such that $\lambda \cdot c \geq 0$ for all $c \in C$ and $\lambda \cdot c > 0$ for at least one $c \in C$.

Proof. By Theorem 6.2 of Rockafellar (1970), $ri(C)$ is non-empty and convex. By Separating Hyperplane Theorem 2, there exists a $\lambda \in \mathbb{R}^n_+$ with $\lambda \neq 0$ such that $\lambda \cdot c \geq 0$ for all $c \in ri(C)$ and $\lambda \cdot c > 0$ for at least one $c \in ri(C)$. By continuity, $\lambda \cdot c \geq 0$ for all c 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, $\lambda \cdot c \geq 0$ for all $c \in C$. Finally, because $ri(C) \subseteq C$, we have $\lambda \cdot c > 0$ for at least one $c \in C$. $\square$

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.

[6] Corollary 3 in Ok (2007, p. 391).

We finally prove (iv). The proof is indirect. Suppose x were not uniquely optimal for any $y \in Y$. Defining the set C as before, this means that:

$$C \cap \mathbb{R}^n_{--} = \emptyset.$$ [7]

We now apply the following separating hyperplane theorem:

Separating Hyperplane Theorem 4: Suppose $C \subseteq \mathbb{R}^n$ is non-empty and convex. If $C \cap \mathbb{R}^n_{--} = \emptyset$ then there exists $\lambda \in \mathbb{R}^n_+$ with $\lambda \neq 0$ such that $\lambda \cdot c \geq 0$ for all $c \in C$.

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

Proof. Minkowski's separating hyperplane theorem (Ok (2007, p. 483)) implies that there are $\lambda \in \mathbb{R}^n$ with $\lambda \neq 0$ and $\varepsilon \in \mathbb{R}$ such that $\lambda \cdot c \leq \varepsilon$ for all $c \in \mathbb{R}^n_{--}$ and $\lambda \cdot c \geq \varepsilon$ for all $c \in C$. It easily follows that $\lambda \in \mathbb{R}^n_+$ and $\varepsilon \geq 0$. Therefore, $\lambda \cdot c \geq \varepsilon$ for all $c \in C$ implies $\lambda \cdot c \geq 0$ for all $c \in C$. □

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 actions $x_1, \ldots, x_n$ 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. □

The key idea of the proof of Theorem 2 is to construct for any action that is not optimal a convex combination of all other actions that dominates it. An alternative approach that we refer to as the "classical approach" is to prove for every action that is not dominated the existence of a situation in which it is optimal. The approach that we take here is borrowed from Fishburn (1975). The alternative, classical approach is exemplified by the proof of Theorem 2.2 in Fudenberg and Tirole (1991). We now illustrate both approaches, and explain why it is crucial that we rely in the proof of Theorem 2 on Fishburn's approach rather than the classical approach.

For concreteness, we shall focus in this discussion on the proof of the claim that in a finite two-player game every pure strategy of player 1 that is not strictly dominated

[7] 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 $x_i < 0$ for all i.

is a best response to some belief about player 2's strategies. This is a special case of part (i) of Theorem 2. Adopting Fishburn's approach, one proves this result by showing the contrapositive: every pure strategy that is not a best response to some belief about player 2's strategy is strictly dominated by a mixed strategy of player 1. Consider the game in Table 1. In this game, player 1's strategy M is not a best response to any belief about player 2's choices.

	L	R
T	0	1
M	0.4	0.4
B	1	0

Table 1. Player 1's utility in a two-player game

Figure 1 illustrates the idea that underlies the proof of part (i) of Theorem 2.

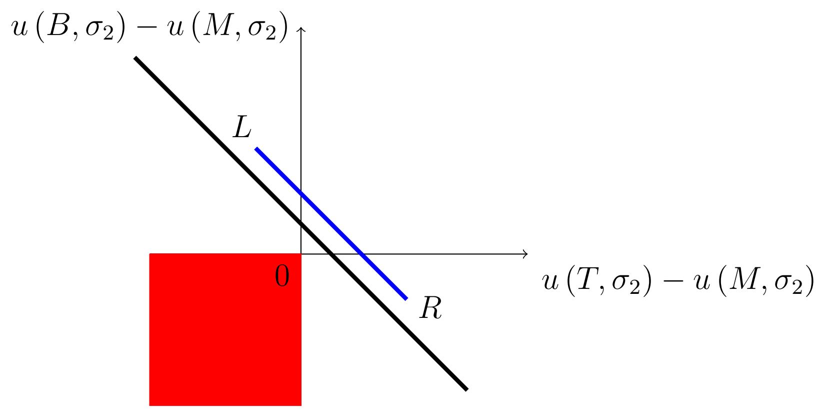

Figure 1. Illustrating the proof of Theorem 2

Letting σ_2 denote any element of the set $\Delta(\{L, R\})$, the probability measures that reflect player 1's beliefs about player 2's pure strategies, and denoting by $u(T, \sigma_2)$, $u(M, \sigma_2)$, and $u(B, \sigma_2)$ the expected utility of player 1 if he chooses any of his three pure strategies, the set C defined in the proof of Theorem 2 can be written as:

$$C \equiv \{(u(T, \sigma_2) - u(M, \sigma_2), u(B, \sigma_2) - u(M, \sigma_2)) \,|\sigma_2 \in \Delta(\{L, R\})\}.$$

In the example, the set C corresponds to the blue line segment in Figure 1. We have colored the non-positive orthant in red in Figure 1. That M is never a best response is reflected by the fact that the blue set and the red set in Figure 1 don't intersect. The black line in Figure 1 is one example of a hyperplane that separates the two sets. The normalized orthogonal vector of this hyperplane corresponds to the convex combination of T and B that places weight 0.5 on both. That this convex combination strictly dominates M is reflected by the fact that the blue line is strictly above the separating hyperplane and the hyperplane itself is above the origin.

An alternative approach, the "classical approach," is to prove for every action that is not dominated the existence of a situation in which it is optimal. Consider the example in Table 2. In this game, strategy M is not strictly dominated.

	L	R
T	0	1
M	0.7	0.7
B	1	0

TABLE 2. Player 1's utility in a two-player game

Figure 2 illustrates the approach taken in the "classical proof."

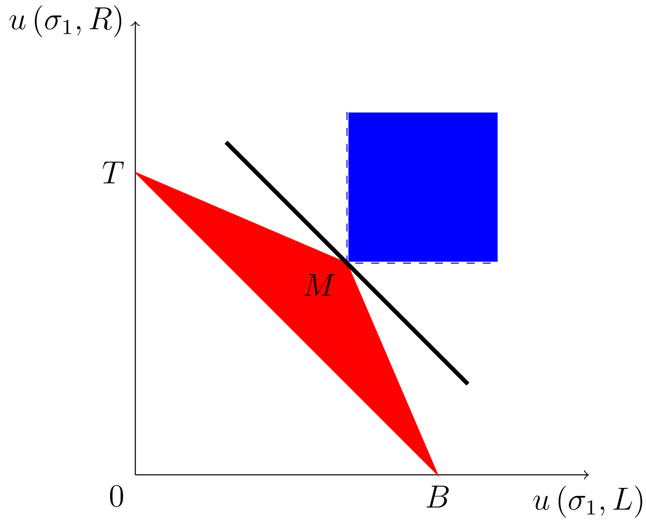

FIGURE 2. Illustrating the classical proof

Denoting by σ_1 the mixed strategies of player 1, the red area in Figure 2 corresponds to the set:

$$\{(u(\sigma_1, L), u(\sigma_1, R)) \,|\sigma_1 \in \Delta(\{T, M, B\})\}.$$

The blue area corresponds to the set of all two-dimensional utility vectors that are greater in both components than the vector $(0.7, 0.7)$ that represents the utility combination that player 1 obtains when choosing M. Because M is not strictly dominated, the blue and the red areas don't intersect. A hyperplane that separates the two areas is indicated in Figure 2 in black. Its normalized orthogonal vector corresponds to a belief about player 2's actions that renders M expected utility maximizing. That M is expected utility maximizing is reflected by the fact that the red area is below the hyperplane.

Observe that in Figure 2 we focus on strict dominance with respect to the set of pure strategies of player 2 only. But in this application Y is in fact the set of all probability distributions over this set. We can focus on the pure strategies because Y is the convex hull of its extreme points, and the pure strategies are the extreme points. Strict dominance with respect to these extreme points implies strict dominance for the whole set Y. The classical approach can therefore use finite-dimensional separating hyperplane theorems as long as the set Y is the convex hull of a finite number of its elements. However, in some applications that we consider below the set Y does not have the property that it can be written as the convex hull of a finite number of its elements. To extend the classical approach we would therefore have to use separating hyperplane theorems in infinite dimensions. By contrast, as long as the set of actions is finite, Fishburn's approach allows us to work with finite-dimensional separating hyperplane theorems. This is what we do in the proof of Theorem 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. Specifically, we state conditions under which the set of uniquely optimal actions is the smallest set of actions 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 $\hat{X}$ 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 is of second-order importance: the decision maker's first priority is to take an optimal action in every situation.[8]

[8]One may think of the set $\widehat{X}$ as a "consideration set" in the sense of the recent decision literature. For a model in which attention cost leads to consideration sets that are so small that material payoff may be lost see Caplin et al. (2019).

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 define is "minimal sufficiency."

Definition 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.[9]

It is obvious that every sufficient set of actions must include the uniquely optimal actions. 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. These conditions imply in other words that, whenever there are multiple optimal actions in a decision problem, at least one of these will also, in some other decision problem, be the unique optimal action.

We assume for this result that no two actions x and x' are identical. Without this assumption, one can obviously not expect the minimally sufficient set to be unique.

Theorem 3. *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.

[9]The literature on statistical decision theory refers to the concepts of "complete classes" and "minimal complete classes" of decision rules and contains characterizations of such classes. That literature also defines "essentially complete classes" and "minimal essentially complete classes" of decision rules. These concepts are related to our concepts of sufficiency and minimal sufficiency, but different. In particular, they are defined using notions of dominance, whereas our concepts are defined using the notion of optimality.

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, Theorems 1 and 2 show assumptions for the primitives of our model that imply $X_{UO} = X_{NWD} \cap X_{NR}$.

The proof of Theorem 3 describes an algorithm for finding the set $X_{NWD} \cap X_{NR}$ that may be of independent interest.

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 $X^1 = X$. For $k = 2, 3, \ldots, |X|$ (where $|X|$ is the number of elements of X), if no action $x \in X^{k-1}$ is either weakly dominated by, or equivalent to, a convex combination of the actions in $X^{k-1} \setminus \{x\}$, then set $X^k = X^{k-1}$. Otherwise, pick arbitrarily some such $x^{k-1} \in X^{k-1}$, and set $X^k = X^{k-1} \setminus \{x^{k-1}\}$. Note that after at most $|X|$ steps no further actions are eliminated. We claim that the final set is $X^{|X|} = X_{NWD} \cap X_{NR}$. It is clear that $X^{|X|}$ 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 $X^{k-1} \setminus \{x\}$, and if $x \in X^k$, then x it is also weakly dominated by, or equivalent to, a convex combination of the actions in $X^k \setminus \{x\}$. Let the elements of $X^{k-1} \setminus \{x\}$ be $\{x_1, x_2, \ldots, x_n\}$, and let the weights of the convex combination be $\lambda_1, \lambda_2, \ldots, \lambda_n$. Without loss of generality assume that x_1 is eliminated in step k, that is: $x^{k-1} = x_1$. This means that x_1 is either weakly dominated or equivalent to, a convex combination of the elements of the set $\{x_2, \ldots, x_n\} \cup \{x\}$. Let the weights of the convex combination be: $\hat{\lambda}_2, \ldots, \hat{\lambda}_n, \hat{\lambda}_x$. It is then obvious that x is also weakly dominated, or equivalent to, a convex combination of $\{x_2, \ldots, x_n\} \cup \{x\}$ with weights: $\lambda_2 + \lambda_1\hat{\lambda}_2, \ldots, \lambda_n + \lambda_1\hat{\lambda}_n, \lambda_1\hat{\lambda}_x$.

We next 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 3.

Now observe that, if x is weakly dominated by, or equivalent to, a convex combination of $\{x_2, \ldots, x_n\} \cup \{x\}$ with weights:

$$\lambda_2 + \lambda_1\hat{\lambda}_2, \ldots, \lambda_n + \lambda_1\hat{\lambda}_n, \lambda_1\hat{\lambda}_x$$

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

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

The division by $1 - \lambda_1\hat{\lambda}_x$ is well-defined because, as we showed in the previous paragraph, $\lambda_1\hat{\lambda}_x < 1$. We have now concluded the proof that the algorithm that we have described terminates with the set $X_{NWD} \cap X_{NR}$.

Now consider any situation $y \in Y$ and suppose $x \in X$ is optimal in situation y. Obviously, $x \in X^1$. Also, either $x \in X^2$, or x is weakly dominated by, or equivalent to, a convex combination of the actions in $X^1 \setminus \{x\} = X^2$. Then one of the actions in X^2 must also be a best response to y. Iterating this argument leads to the conclusion that one of the actions in $X^{|X|} = X_{NWD} \cap X_{NR}$ is optimal in situation y. □

We now provide an example in which $X_{UO} \neq X_{NWD} \cap X_{NR}$, and in which X_{UO} is not the unique minimally sufficient subset of X. In this example $X = \{x_1, x_2, x_3, x_4\}$ and $Y = \{y_1, y_2\}$. The utility function is given in the following table:

	y_1	y_2
x_1	0	2
x_2	1	2
x_3	2	0
x_4	2	1

Table 3. An example where X_{UO} is not minimally sufficient

In this example, $X_{UO} = \emptyset$ and $X_{NWD} \cap X_{NR} = \{x_2, x_4\}$. Obviously, X_{UO} is not sufficient. The minimally sufficient sets of actions are all sets with two elements such that one of the elements is x_1 or x_2, and the other element is x_3 or x_4.

4. Dominance and Optimality of Strategies in Games

We now explain how to apply the results of Section 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} which we also assume to be finite. 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 $\Delta(S_{-i})$.[10] 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 in this application the set X of actions is the set S_i of strategies, the set Y of situations is the set $\Delta(S_{-i})$ of beliefs, and the vector space $\mathcal{Y}$ is $\mathbb{R}^{|S_{-i}|}$. Note that the set of beliefs is a convex subset of $\mathbb{R}^{|S_{-i}|}$, and that the expected utility function u_i can be extended to $\mathbb{R}^{|S_{-i}|}$ as a linear function. We endow $\mathcal{Y}$ with the Euclidean metric. Then u_i is continuous and $\Delta(S_{-i})$ is compact.

The notions of strict and weak dominance introduced in Section 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 $\Delta(S_{-i})$ is 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 4, 5, and 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 Theorems 1 and 2 apply in this setting. In particular, note that part (i) of Theorem 2 applies because the set $\Delta(S_{-i})$ is compact, and part (iii) of Theorem 2 applies because $\mathbb{R}^{|S_{-i}|}$ is finite-dimensional. Part (i) in Theorems 1 and 2, 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 part (ii) of Theorem 1 and part (iii) of Theorem 2 note that an element of the relative interior of $\Delta(S_{-i})$ is a full support

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

belief, and therefore part (ii) of Theorem 1 and part (iii) of Theorem 2 correspond to Lemma 4 in Pearce (1984). Again, the proof in Pearce (1984) is different from ours. Finally, part (iii) of Theorem 1 and part (iv) of Theorem 2, if applied to strategic games, are 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.[11]

There are several other applications of Theorems 1 and 2 in which the set Y of situations is the set of all probability distributions on some finite set Y^*, and u is expected utility. As above, Definitions 4, 5, and 6 would then not change if one replaced the expression "for all $y \in Y$" by the expression "for all $y \in Y^*$."

If Y^* is some abstract set of states of the world, Theorems 1 and 2 can be interpreted as providing necessary and sufficient conditions for the existence of a subjective probability distribution over Y^* that makes a particular act x an expected utility maximizing choice.

Another re-interpretation of the framework of this section considers a decision maker who faces some unknown state of the world which is contained in a finite set of possible states of the world. The decision maker observes a signal that has a known conditional distribution for each state of the world and has finitely many possible realizations. The decision maker has to choose an action from a given finite set, and has a utility function that depends on the action and the state of the world.

Let us call a mapping that assigns to each signal realization an action a "statistical decision rule." We can then interpret the set X as the set of statistical decision rules, and the set Y^* as the set of possible states of the world.[12] Theorems 1 and 2 can be used to derive results that relate statistical decision rules that are not dominated to statistical decision rules that are optimal for some prior over the set of states of the world. Decision rules that are not weakly dominated are called "admissible" in the literature on statistical decision theory, and decision rules that are optimal for some prior are called "Bayesian" in the literature on statistical decision theory.

Theorem 2.3.2 in Ferguson (1967) stating that a Bayesian decision rule with respect to a full support belief is admissible corresponds to part (ii) of our Theorem 1 and

[11] We are grateful to Duarte Gonçalves for bringing Weinstein (2020) to our attention. Amanda Friedenberg pointed out to us that Weinstein's Proposition 3 is a straightforward implication of Lemmas D.2 and D.4 in Brandenburger et al. (2008).

[12] Kuzmics (2017) observes an analogy between statistical decision rules and Anscombe-Aumann acts in decision theory.

Theorem 2.10.1 in Ferguson (1967) stating that every admissible decision rule is Bayesian with respect to some belief is part (ii) of our Theorem 2.[13] [14]

As a final example, one may interpret the set X as the set of alternatives chosen by a social planner, and the set Y^* as the finite set of individuals in society, so that $u(x, y)$ is the (cardinal) utility of some agent $y \in Y^*$ if social alternative x is chosen. Suppose that the social planner is utilitarian, and seeks to maximize a weighted utilitarian welfare function: $\sum_{y \in Y^*} w(y)u(x, y)$. Then the weights $w(y)$ play the same role as the probability measure μ_i in our game theoretical application. Theorems 1 and 2 can then be applied to obtain several results about the connection between utilitarian optimal social alternatives and social alternatives that are not Pareto dominated by any lottery over the other social alternatives. We omit the details.[15]

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 which has full support. 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: $s : \Omega \to \Delta(M_s)$, 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.[16]

The decision maker faces a decision problem $(A, \mathfrak{u})$. Here, A is a finite set of actions and $\mathfrak{u} : A \times \Omega \to [0, 1]$ is a Bernoulli utility function. The assumption that the co-domain is the interval $[0, 1]$ is a normalization. We denote by $\mathcal{D}$ 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. We denote the support of μ_s by Γ_s. Define $\Gamma_{\mathcal{S}} = \cup_{s \in \mathcal{S}} \Gamma_s$ to be the

[13] It does not seem to have been noted in the statistics literature (see, for example, Duanmu and Roy (2021)) that the belief can be taken to have a full support, which is part (iii) of our Theorem 2.

[14] The "complete class theorems" in statistical decision theory (Wald (1947) and Ferguson (1967)) are also closely related to our results in this paper, but, given the differences between our framework and that of the complete class literature which we pointed out in footnote 9, a precise description of the connections is complicated and does not seem to add much insight.

[15] We are grateful to Yaron Azrieli for suggesting this application to us.

[16] We could allow signals to be costly. But then the dominance order that we consider would no longer correspond to the classic Blackwell dominance.

union of all supports. For every decision problem $(A, \mathfrak{u})$ we denote by $v_{A,\mathfrak{u}} : \Gamma_{\mathcal{S}} \to [0, 1]$ the value function:

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

Here, γ stands for an arbitrary posterior belief in $\Gamma_{\mathcal{S}}$. 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:

$$\sum_{\gamma \in \Gamma_s} v_{A,\mathfrak{u}}(\gamma)\, \mu_s(\gamma)$$

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

$$\sum_{\gamma \in \Gamma_s} v_{A,\mathfrak{u}}(\gamma)\, \mu_s(\gamma) \geq \sum_{\gamma \in \Gamma_{\hat{s}}} v_{A,\mathfrak{u}}(\gamma)\, \mu_{\hat{s}}(\gamma)$$

We now explain how to fit Blackwell dominance into our framework. The set X is $\mathcal{S}$. The set Y is the set of all value functions that correspond to a decision problem in $\mathcal{D}$. In fact, Y is the set of all convex functions with domain $\Gamma_{\mathcal{S}}$ and co-domain $[0, 1]$. To clarify our terminology, we define convexity explicitly.

Definition 9. *A value function $v : \Gamma_{\mathcal{S}} \to \mathbb{R}$ is* convex *if for all $\{\gamma_1, \gamma_2, \ldots, \gamma_n\} \subseteq \Gamma_{\mathcal{S}}$ and $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n_+$ with $\sum_{i=1}^n \lambda_i = 1$, if*

$$\gamma = \sum_{i=1}^{n} \lambda_i \gamma_i \in \Gamma_{\mathcal{S}},$$

we have:

$$v(\gamma) \leq \sum_{i=1}^{n} \lambda_i v(\gamma_i)$$

When the inequality is strict for all $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n_+$ with $\sum_{i=1}^n \lambda_i = 1$ and $\lambda_i < 1$ for all $i \in \{1, 2, \ldots, n\}$, then we call the value function v strictly convex.

We set $\mathcal{Y}$ to be $\mathbb{R}^{|\Gamma_S|}$. The utility function $u(x,y)$ from Section 2 is in the setting of this section the expected utility $\sum_{\gamma\in\Gamma_s} v_{A,\mathfrak{u}}(\gamma)\,\mu_s(\gamma)$. Note that Y is convex,[17] and that u can be extended to $\mathbb{R}^{|\Gamma_S|}$ as a linear function. We endow $\mathcal{Y}$ with the Euclidean metric. Then u is continuous and Y is compact.

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 10. *Suppose* $\{s_1, s_2, \ldots, s_n\} \subseteq \mathcal{S}$ *and assume that the vector* $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n_+$ *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 λ_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.

A signal is then not Blackwell dominated if and only if it is neither weakly dominated by a convex combination of the other signals, nor redundant in the sense introduced in Section 2.

All parts of Theorems 1 and 2 apply. However, parts (i) of Theorems 1 and 2 are vacuously true. The reason for this is that we have not ruled out from consideration those decision problems $(A, \mathfrak{u}) \in \mathcal{D}$ in which the utility function u does not depend on the state ω. For such a decision problem, the value function is constant, and all

[17] The convex combination of two value functions with given weights corresponds to the decision problem in which the decision maker chooses a contingent plan, i.e. actions for each of the two corresponding decision problems, and the utility is the expected utility from the two decision problems, where each decision problem is relevant with probability equal to the given weight.

signals are optimal. Therefore, no signal, nor any convex combination of signals, is ever *strictly better* than another signal for *all* value functions. As regards part (ii) of Theorem 1 and part (iii) of Theorem 2 note that the relative interior of Y is the set of all strictly convex value functions v with domain $\Gamma_{\mathcal{S}}$ and co-domain $[0,1]$.

Part (iii) of Theorem 1 and part (iv) of Theorem 2 exactly correspond to Blackwell dominance. We seek to state explicitly the bottom line of these two parts. We first re-state the definition of unique optimality in our setting:

Definition 11. *A signal $s \in \mathcal{S}$ is a uniquely optimal choice in decision problem $(A, \mathfrak{u})$ if:*

$$\sum_{\gamma \in \Gamma_s} v_{A,\mathfrak{u}}(\gamma)\, \mu_s(\gamma) > \sum_{\gamma \in \Gamma_{\hat{s}}} v_{A,\mathfrak{u}}(\gamma)\, \mu_{\hat{s}}(\gamma) \text{ for all } \hat{s} \in \mathcal{S} \text{ with } \hat{s} \neq s.$$

Part (iii) of Theorem 1 and part (iv) of Theorem 2 say in our setting:

Proposition 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 $\mathcal{S} \setminus \{s\}$.*

This result describes a property of Blackwell dominance that, to our knowledge, has not been noticed before. We now show by means of an example that the result would not be true if we 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 4 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 4. Conditional distributions of s_1, s_2, and s_3

Table 5 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 5. 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, and 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 s_4 generates with probability 1/12 a posterior belief that is a Dirac measure on this state. Also, for each pair of states in Ω, signal s_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 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, the convex combination of signals s_1, s_2, and s_3 generates with probability 2/9 ($< 1/4$) a posterior belief that places probability 1/2 on two of the three states. We can thus obtain the poster distribution for the convex combination of signals s_1, s_2, and s_3 from the posterior distribution for s_4 by replacing, for each pair of states, probability mass 1/36 placed on the belief that assigns probability 0.5 on each of the two states by probability mass 1/72 placed on the Dirac measures for each of the two states. Thus 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. Therefore s_4 will never be uniquely optimal. Yet s_4 is not Blackwell dominated by any single of the signals s_1, s_2 and s_3. This shows that Proposition 1 would not be true had we not allowed the Blackwell dominating signal to be a convex combination of the other signals.

We conclude this section by briefly considering some alternatives to the Blackwell order and the applicability of the results in Theorems 1 and 2 to these orders. Suppose that the set of states Ω is a finite 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 12. *Signal s* monotonically dominates *signal $\hat{s}$ if for every monotone decision problem $(A, \mathfrak{u}) \in \mathcal{M}$:*

$$\sum_{\gamma \in \Gamma_s} v_{A,\mathfrak{u}}(\gamma)\, \mu_s(\gamma) \geq \sum_{\gamma \in \Gamma_{\hat{s}}} v_{A,\mathfrak{u}}(\gamma)\, \mu_{\hat{s}}(\gamma)$$

Suppose again that the decision maker has to choose one from a finite set $\mathcal{S}$ of signals. One can again embed this setting in the model of Section 2. The set Y is convex for the same reason as it was convex in the case of the Blackwell order. We can again use part (iii) of Theorem 1 and part (iv) of Theorem 2 to conclude:.

Proposition 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 12 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 Y of value functions is convex no longer applies.[18] This is because that

[18]This is because a contingent plan for two decision problems, as referred to in footnote 17, is multi-dimensional, not single dimensional.

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. Quah and Strulovici (2007) and Kim (2022) allow multi-dimensional action sets but imposes joint conditions on signals and decision problems. A detailed consideration of their orders is outside of the scope of this paper.

One might also modify the Blackwell order by considering only strictly convex 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.

6. Dominance and Optimality of Monetary Lotteries

Consider an expected utility maximizer who chooses one lottery ℓ from a finite set $\mathcal{L}$ of lotteries. We shall assume that there is some $\zeta > 0$ such that the support of all lotteries in $\mathcal{L}$ is a subset of the interval $[0, \zeta]$. The elements of $[0, \zeta]$ are interpreted as money payments. One lottery "first-order stochastically dominates" another lottery if the former lottery yields at least as high expected utility as the latter for all possible Bernoulli utility functions provided that the decision maker's utility function 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 (weakly) concave in money.[19]

We consider two cases. In the first case each lottery $\ell \in \mathcal{L}$ has finite support Q_ℓ. Define $Q_\mathcal{L} = \cup_{\ell \in \mathcal{L}} Q_\ell$ to be the union of all supports. To apply the results of Section 2 we let the set X of actions be $\mathcal{L}$. 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) Bernoulli utility functions $\mathfrak{u}$ with domain $Q_\mathcal{L}$ and co-domain $[0, 1]$. The assumption that the co-domain is the interval $[0, 1]$ is a normalization. To clarify our terminology, we define concavity explicitly.

[19]Our definitions of the stochastic dominance orders are taken from Definition 6.13 in Kreps (2013).

Definition 13. *A Bernoulli utility function* $\mathfrak{u} : Q_{\mathcal{L}} \to [0,1]$ *is* concave *if for all* $\{q_1, q_2, \ldots, q_n\} \subseteq Q_{\mathcal{L}}$ *and* $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n_+$ *with* $\sum_{i=1}^n \lambda_i = 1$, *if*

$$q = \sum_{i=1}^n \lambda_i q_i \in Q_{\mathcal{L}},$$

we have:

$$\mathfrak{u}(q) \geq \sum_{i=1}^n \lambda_i \mathfrak{u}(q_i)$$

When the inequality is strict for all $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n_+$ *with* $\sum_{i=1}^n \lambda_i = 1$ *and* $\lambda_i < 1$ *for all* $i \in \{1, 2, \ldots, n\}$, *then we call the utility function* $\mathfrak{u}$ strictly concave*.*

We set $\mathcal{Y}$ to be $\mathbb{R}^{|Q_{\mathcal{L}}|}$. The utility function u from Section 2 is then the expected utility, that is, if x is the lottery ℓ and y is the utility function $\mathfrak{u}$, then

$$u(x,y) = \sum_{q \in Q_\ell} \mathfrak{u}(q)\, \ell(q)$$

where $\ell(q)$ is the probability of the realization being q in lottery ℓ. In this specification the space $\mathcal{Y}$ is finite-dimensional, and the expected utility function u can be extended to $\mathcal{Y}$ as a linear function. We endow $\mathcal{Y}$ with the Euclidean metric. The set Y is convex and compact, and u is continuous on Y, for both the case of first-order stochastic dominance and the case of second-order stochastic dominance.

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.

All parts of Theorems 1 and 2 apply. However, parts (i) of Theorems 1 and 2 are vacuously true. The reason for this is that we have not ruled out from consideration the constant Bernoulli utility function. For such a utility function, all lotteries are optimal, and also no lottery, nor any convex combination of lotteries, is ever *strictly better* than another lottery for *all* Bernoulli utility functions. As regards part (ii) of Theorem 1 and part (iii) of Theorem 2 note that the relative interior of Y is the set of all strictly increasing or strictly increasing and strictly concave Bernoulli utility functions $\mathfrak{u}$ with domain $Q_{\mathcal{L}}$ and co-domain $[0,1]$. Part (iii) of Theorem 1 and part (iv) of Theorem 2 exactly correspond to first- and second-order stochastic dominance.

We provide in Table 6 an example that demonstrates that part (iv) of Theorem 2 would not hold had we not considered convex combinations. We focus on first-order stochastic dominance. 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.

	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 6. Payoffs of three lotteries

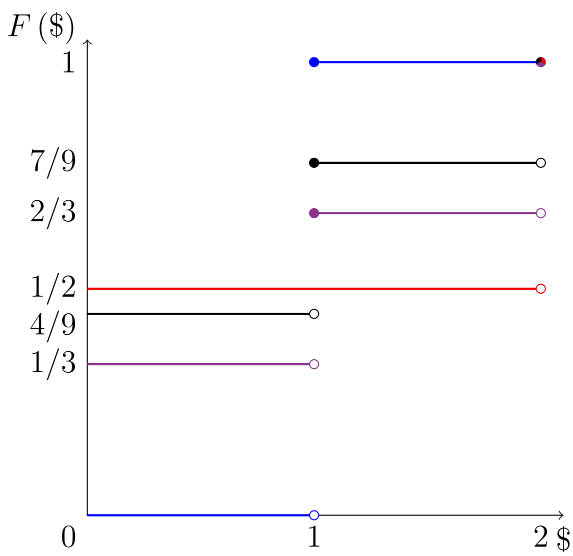

Figure 3. Cumulative distribution functions of lotteries

The cumulative distributions functions of the lotteries are illustrated in Figure 3 where the red line indicates ℓ_1, the blue line indicates ℓ_2, the black line indicates ℓ_3, and the purple line indicates the convex combination of ℓ_1 and ℓ_2 that attaches weight $2/3$ to ℓ_1 and weight $1/3$ to ℓ_2. Note that neither ℓ_1 nor ℓ_2 first-order stochastically dominates ℓ_3.[20] However, the convex combination of ℓ_1 and ℓ_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. For every utility function that is non-decreasing, one of lotteries ℓ_1 or ℓ_2 yields at least as high expected utility as lottery ℓ_3. Yet lottery ℓ_3 is not first-order stochastically dominated by either of the two other lotteries. Thus part (iv) of Theorem 2 would not hold had we not considered convex combinations (nor would, as is easily seen from this example, part (iii) of Theorem 2).

After the analysis for the case of finite support lotteries, we next turn to the case where all the lotteries the decision maker can choose have full supports on $[0, \xi]$. To apply the results of Section 2 we let the set X of actions be $\mathcal{L}$, and we let the set Y be the set of continuous and non-decreasing (in the case of first-order stochastic dominance), or the set of continuous, non-decreasing, and concave (in the case of second-order stochastic dominance) Bernoulli utility functions $\mathfrak{u}$ with domain $[0, \xi]$ and co-domain $[0, 1]$. Moreover, we set $\mathcal{Y}$ equal to the set of Lebesgue-measurable and bounded functions $\tilde{\mathfrak{u}}$ with domain $[0, \xi]$. 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 $\mathfrak{u}$, then:

$$u(x, y) = \int_0^\xi \mathfrak{u}(q) dF.$$

In this specification, the set Y is a convex set, and the utility function u is linear on $\mathcal{Y}$. Next, we list some facts that we use in the following and whose proofs are in the appendix: The relative interior of Y is empty. If we endow $\mathcal{Y}$ with the supremum norm (inducing the topology of uniform convergence), then u is continuous on Y. However, the set Y is not compact.

We can now apply Theorem 1 in this setting. Part (i) of Theorem 1 is vacuously true for the same reason as in the case of lotteries with finite support. The assumption of part (ii) of Theorem 1 holds, but part (ii) of Theorem 1 is vacuous. This is because the relative interior of Y is empty. Thus, the only part of Theorem 1 that is not vacuous in our setting is part (iii): if a lottery is uniquely optimal, then it is not first/second-order stochastic dominated.

[20]In the graph, one lottery first-order stochastically dominates another lottery if the former's cumulative distribution function is everywhere below the latter's cumulative distribution function.

We now turn to Theorem 2. Note that the background assumptions of the first sentence of Theorem 2 hold. Although the assumption of part (i) is not satisfied, the conclusion of part (i) of Theorem 2 is trivially true for the same reason as in the case of lotteries with finite support. Part (ii) of Theorem 2 is trivially true for the same reason. Because $\mathcal{Y}$ is not finite-dimensional, part (iii) of Theorem 2 does not apply. Its conclusion is false because the relative interior of Y is empty. The only part of Theorem 2 that is non-trivial and that applies to this setting is thus part (iv) which says that a lottery that is not first/second-order stochastic dominated is uniquely optimal.

Results that are closely related to some of the results described in this section have been obtained by Fishburn (1975). Under more general assumptions about the lotteries considered and the utility functions allowed Fishburn (1975) showed in his Theorem 1 a result that corresponds to part (iii) of our Theorem 1 and part (iv) of our Theorem 2.[21] Making the additional assumption that $X_{NSD} = X_{NWD}$, Fishburn showed in his Theorem 2 a result that corresponds to part (i) of our Theorem 1 and part (ii) of our Theorem 2. Fishburn (1975) did not consider what we call "interior optimality."

7. Conclusion

We have investigated for various dominance orders in economics the relation between an action not being dominated and an action being optimal in some situation. Our investigation has uncovered shared properties of, as well as differences among, dominance orders in economics. Our paper has emphasized the importance of considering convex combinations of actions in dominance relations.

Our investigation has sometimes suggested modifications of the definition of dominance relations. For example, for dominance relations among signals, we have suggested a modification of the definition of monotone dominance. It remains an open problem whether one can provide simple characterizations of these modified dominance relations.

In some contexts, previous literature has perhaps not sufficiently emphasized the importance of convex combinations. In the theory of the optimal choice of investment portfolios it is of interest to investigate "efficient" portfolios where such portfolios are defined as those that are not dominated by convex combinations of other portfolios.

[21] Assuming that all lotteries q_i (using Fishburn's (1975) notation) are the same.

This point was made by Post (2017) who gave a real world example of a stock portfolio that is second-order stochastically dominated by a convex combination of other portfolios, but not by any single one of those portfolios. In the theory of information acquisition, when discussing the choice of signals of an expected utility maximizing decision maker, it seems worthwhile to investigate which signals are undominated within particular sets of signals. In a forthcoming companion paper we shall 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. This seems of particular interest in the context of theoretical statistics. Our results for two-dimensional signal distributions, for example, have implications for the theory of efficient sampling methods.

References

Athey, S. and J. Levin (2018): "The Value of Information in Monotone Decision Problems," *Research in Economics*, 72, 101–116.

Battigalli, P., S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci (2016): "A Note on Comparative Ambiguity Aversion and Justifiability," *Econometrica*, 84, 1903–1916.

Blackwell, D. (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, Berkeley and Los Angeles: University of California Press, 93–102.

——— (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24, 265–272.

Brandenburger, A., A. Friedenberg, and H. J. Keisler (2008): "Admissibility in Games," *Econometrica*, 76, 307–352.

Caplin, A., M. Dean, and J. Leahy (2019): "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *The Review of Economic Studies*, 86, 1061–1094.

Duanmu, H. and D. M. Roy (2021): "On Extended Admissible Procedures and Their Nonstandard Bayes Risk," *The Annals of Statistics*, 49, 2053–2078.

Ferguson, T. S. (1967): *Mathematical Statistics: A Decision Theoretic Approach*, New York: Academic Press.

Fishburn, P. C. (1975): "Separation Theorems and Expected Utilities," *Journal of Economic Theory*, 11, 16–34.

Fudenberg, D. and J. Tirole (1991): *Game Theory*, MIT Press.

Kim, Y. (2022): "Comparing Information in General Monotone Decision Problems," Working Paper, Department of Economics, Duke University.

Kreps, D. M. (2013): *Microeconomic Foundations I: Choice and Competitive Markets*, Princeton University Press.

Kuzmics, C. (2017): "Abraham Wald's Complete Class Theorem and Knightian Uncertainty," *Games and Economic Behavior*, 104, 666–673.

Lehmann, E. L. (1988): "Comparing Location Experiments," *The Annals of Statistics*, 16, 521–533.

Munkres, J. (2014): *Topology*, Harlow: Pearson Education Limited, 2nd ed.

Niculescu, C. P. and L.-E. Persson (2018): *Convex Functions and Their Applications: A Contemporary Approach*, Springer, 2nd ed.

Ok, E. (2007): *Real Analysis with Economic Applications*, Princeton University Press.

Osborne, M. and A. Rubinstein (1994): *A Course in Game Theory*, MIT Press.

Pearce, D. G. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–1050.

Post, T. (2017): "Empirical Tests for Stochastic Dominance," *Review of Finance*, 21, 793–810.

Quah, J. K.-H. and B. Strulovici (2007): "Comparative Statics with the Interval Dominance Order: Some Extensions," Working Paper, Northwestern University.

Quirk, J. P. and R. Saposnik (1962): "Admissibility and Measurable Utility Functions," *The Review of Economic Studies*, 29, 140–146.

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, 225–243.

Shah, S. A. (2022): "Either/Or: Best Reply Versus Dominance," Working Paper, Delhi School of Economics.

Wald, A. (1947): "An Essentially Complete Class of Admissible Decision Functions," *Annals of Mathematical Statistics*, 18, 549–555.

Weinstein, J. (2020): "Best-Reply Sets," *Economic Theory Bulletin*, 8, 105–112.

Zimper, A. (2005): "Equivalence between Best Respones and Undominated Strategies: A Generalization from Finite to Compact Strategy Sets," *Economics Bulletin*, 3, 1–6.

Appendix A. Counterexample for part (i) of Theorem 2 when Y is not compact

Let $X = \{x_1, x_2, x_3\}$ and $Y = \{(y_1, y_2) \in \mathbb{R} | y_1 \leq 1, y_2 \leq 1, y_1 + y_2 > 0\} \cup \{(y_1, y_2) \in \mathbb{R} | -1 \leq y_1 < 0, y_2 = -y_1\}$. We set for all $y \in Y$:

$$u(x_1, y) = y_1, u(x_2, y) = y_2, u(x_3, y) = 0.$$

The set Y is the blue area in Figure 4 where the dashed edge is excluded. Observe that Y is not closed, hence not compact. All other assumptions of part (i) of Theorem 2 hold: Y is convex, and, if we set $\mathcal{Y} = \mathbb{R}^2$, then the above definition of u can obviously be extended to all of $\mathcal{Y}$ such that u is linear on $\mathcal{Y}$. Obviously u is then also continuous on Y.

We claim that x_3 is not strictly dominated by any convex combination of x_1 and x_2 and yet x_3 is not optimal for any $y \in Y$. To see that x_3 is not strictly dominated note that any convex combination that places weight greater than 0.5 on x_1 yields negative payoffs for $y \in Y$ that are sufficiently close to $(-1, +1)$, and any convex combination that places weight less than 0.5 on x_1 yields negative payoffs for $y \in Y$ that are sufficiently close to $(+1, -1)$. Finally, the convex combination that places weight 0.5 on x_1 and x_2 yields payoff 0 whenever $y_1 + y_2 = 0$. Therefore, no convex combination of x_1 and x_2 strictly dominates x_3. On the other hand, x_3 is never optimal. This is because for all $y \in Y$ we have that at least one of y_1 and y_2 is strictly positive, and therefore at least one of x_1 and x_2 is better than x_3.

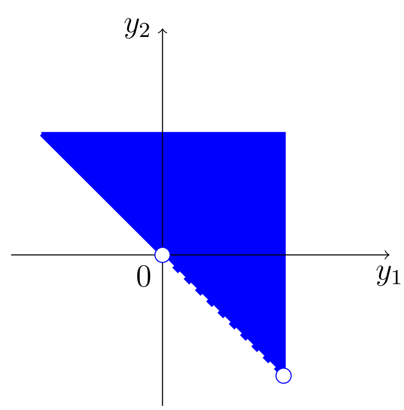

Figure 4. The set Y in the counterexample for part (i) of Theorem 2 when Y is not compact

Appendix B. Proofs of the claims in Section 6

In the second half of Section 6 we let X be a finite subset of the set of all lotteries with full support $[0, \xi]$ for some $\xi > 0$. We obtain first-order stochastic dominance if we let Y in the framework of our paper be the set of continuous and non-decreasing Bernoulli utility functions with domain $[0, \xi]$ and co-domain $[0, 1]$. We embedded Y in Section 6 in the vector space $\mathcal{Y}$ of all Lebesgue-measurable and bounded functions with domain $[0, \xi]$. We claimed in the main text:

Claim 1. *$ri\,(Y) = \emptyset$.*

Proof. Pick any $\mathfrak{u} \in Y$. By Lebesgue's Theorem (Ok (2007, p. 212)) $\mathfrak{u}$ is differentiable almost everywhere. Suppose it is differentiable at $\tau \in (0, \xi)$. Consider the function $\bar{\mathfrak{u}}_\tau$ given by:

$$\bar{\mathfrak{u}}_\tau(q) = (q - \tau)^{\frac{1}{3}} \text{ for all } q \in [0, \xi].$$

There is obviously a positive affine transformation that we can apply to $\bar{\mathfrak{u}}_\tau$ to obtain a function $\mathfrak{u}_\tau$ which has co-domain $[0, 1]$. For any $\beta > 0$ now define $\tilde{\mathfrak{u}}$ by setting: $\tilde{\mathfrak{u}}(q) = (1+\beta)\mathfrak{u}(q) - \beta\mathfrak{u}_\tau(q)$ for all $q \in [0, \xi]$. Because $\mathfrak{u}'_\tau(\tau) = \infty$, there exists Δ such that

$$\frac{\tilde{\mathfrak{u}}\,(\tau + \Delta) - \tilde{\mathfrak{u}}\,(\tau)}{\Delta} < 0.$$

which implies that $\tilde{\mathfrak{u}}$ is not nondecreasing and therefore $\tilde{\mathfrak{u}} \notin Y$. Because this holds for arbitrarily small β it follows that $\mathfrak{u} \notin ri\,(Y)$. □

In Section 6 we endowed $\mathcal{Y}$ with the topology of uniform convergence, and then claimed:

Claim 2. *u is continuous on Y.*

Proof. Consider any sequence $(\mathfrak{u}_n)_{n\in\mathbb{N}}$ of functions in Y that uniformly converges to $\mathfrak{u} \in Y$. Because for any $n \in \mathbb{N}$ and any $q \in [0, \xi]$, $|\mathfrak{u}_n\,(q)| \leq |\hat{\mathfrak{u}}\,(q)|$, where $\hat{\mathfrak{u}}\,(q) \equiv 1$ for all $q \in [0, \xi]$, we can apply the dominated convergence theorem and conclude:

$$\lim_{n\to\infty} \int_0^\zeta \mathfrak{u}_n(q)dF = \int_0^\zeta \mathfrak{u}(q)dF.$$

With the supremum norm, Y is a metric space, so the above equation implies that u is continuous on Y (Theorem 21.3 in Munkres (2014)). □

Claim 3. *Y is not compact.*

Proof. We show that Y is not compact in the topology of pointwise convergence. Because the topology of uniform convergence is finer than the topology of pointwise convergence, it follows that Y is also not compact in the topology of uniform convergence.

To facilitate the proof that Y is not compact with the topology of pointwise convergence, we first establish that $\mathcal{Y}$, if endowed with the topology of pointwise convergence, is a Hausdorff space. Consider any two distinct functions $\mathfrak{u}_1$ and $\mathfrak{u}_2$. So there exists $q \in [0, \xi]$ such that $\mathfrak{u}_1(q) \neq \mathfrak{u}_2(q)$. And there exists ε such that $(\mathfrak{u}_1(q) - \varepsilon, \mathfrak{u}_1(q) + \varepsilon)$ and $(\mathfrak{u}_2(q) - \varepsilon, \mathfrak{u}_2(q) + \varepsilon)$ are disjoint. Consider the following two sets:

$$U_1 = \{\mathfrak{u} | \mathfrak{u} \in \mathcal{Y} \text{ and } \mathfrak{u}(q) \in (\mathfrak{u}_1(q) - \varepsilon, \mathfrak{u}_1(q) + \varepsilon)\}$$

and

$$U_2 = \{\mathfrak{u} | \mathfrak{u} \in \mathcal{Y} \text{ and } \mathfrak{u}(q) \in (\mathfrak{u}_2(q) - \varepsilon, \mathfrak{u}_2(q) + \varepsilon)\}$$

By Munkres (2014, p. 281) these two sets belong to the subbasis of the topology of pointwise convergence, and therefore are open sets in the topology of pointwise convergence. Note that $\mathfrak{u}_1 \in U_1$, $\mathfrak{u}_2 \in U_2$, and U_1 and U_2 are disjoint. Therefore $\mathcal{Y}$, if endowed with the topology of pointwise convergence, is a Hausdorff space.

We now prove that Y is not compact in $\mathcal{Y}$ if we endow $\mathcal{Y}$ with the topology of pointwise convergence. For every $n \in \mathbb{N}$ let:

$$\{\mathfrak{u}_n(q)\} \equiv \left\{ \left(\frac{q}{\xi}\right)^{\frac{1}{n}} \right\}$$

for all $q \in [0, \xi]$. The sequence $(\mathfrak{u}_n)_{n \in \mathbb{N}}$ converges pointwise to $\mathfrak{u}$ where $\mathfrak{u}(q) = 0$ for $q = 0$ and $\mathfrak{u}(q) = 1$ for $q \in (0, \xi]$. $\mathfrak{u} \in \mathcal{Y}$, but it is not continuous, so $\mathfrak{u} \notin Y$, and thus Y is not sequentially closed. But in a Hausdorff space a compact set must be closed, and a closed set must be sequentially closed. Therefore, in a Hausdorff space, a compact set must be sequentially closed. Recall that $\mathcal{Y}$ is a Hausdorff space with the topology of pointwise convergence. So Y is not sequentially closed implies that Y is not compact. $\square$

Next, we turn to second-order stochastic dominance. In this case we defined in Section 6 the set Y to be the set of all continuous, non-decreasing, and concave Bernoulli utility functions with domain $[0, \xi]$ and co-domain $[0, 1]$. We embedded Y again in the vector space $\mathcal{Y}$ of all Lebesgue-measurable and bounded functions with domain $[0, \xi]$.

Claim 4. *$ri(Y) = \emptyset$.*

Proof. By Alexandrov's Theorem (Theorem 3.12.3 in Niculescu and Persson (2018)) every $\mathfrak{u} \in Y$ is twice differentiable almost everywhere. Suppose it is twice differentiable at $\tau \in (0, \xi)$. Consider the function $\bar{\mathfrak{u}}_\tau$ that satisfies for all $q \in [0, \xi]$:

$$\bar{\mathfrak{u}}_\tau(q) = q(\xi - \tau)^{\frac{1}{3}} - \frac{3}{4}(q - \tau)^{\frac{4}{3}}.$$

We have:

$$\bar{\mathfrak{u}}'_\tau(q) = (\xi - \tau)^{\frac{1}{3}} - (q - \tau)^{\frac{1}{3}}.$$

Note that $\bar{\mathfrak{u}}'_\tau$ is decreasing and non-negative on $[0, \xi]$. There is obviously a positive affine transformation that we can apply to $\bar{\mathfrak{u}}_\tau$ to obtain a function $\mathfrak{u}_\tau$ which has co-domain $[0, 1]$. For any $\beta > 0$ now define $\tilde{\mathfrak{u}}$ by setting: $\tilde{\mathfrak{u}}(q) = (1 + \beta)\mathfrak{u}(q) - \beta \mathfrak{u}_\tau(q)$ for all $q \in [0, \xi]$. Because $\mathfrak{u}''_\tau(\tau) = -\infty$, there exists Δ such that

$$\frac{\tilde{\mathfrak{u}}(\tau + \Delta) - 2\tilde{\mathfrak{u}}(\tau) + \tilde{\mathfrak{u}}(\tau - \Delta)}{\Delta^2} > 0$$

which implies $\tilde{\mathfrak{u}}$ is not concave and thus $\tilde{\mathfrak{u}} \notin Y$.Because this holds for arbitrarily small β it follows that $\mathfrak{u} \notin ri\,(Y)$. □

We also claimed in Section 6 that u is continuous on Y and that Y is not compact. The proofs are the same as those of Claims 2 and 3, and are therefore omitted.