

Fixing Stochastic Dominance

Jeffrey Sanford Russell

Abstract

Decision theorists widely accept a stochastic dominance principle: roughly, if a risky prospect A is at least as probable as another prospect B to result in something at least as good, then A is at least as good as B . Recently, philosophers have applied this principle even in contexts where the values of possible outcomes do not have the structure of the real numbers: this includes cases of incommensurable values and cases of infinite values. But in these contexts the usual formulation of stochastic dominance is wrong. We show this with several counterexamples. Still, the motivating idea behind stochastic dominance is a good one: it is supposed to provide a way of applying dominance reasoning in the stochastic context of probability distributions. We give two new formulations of stochastic dominance that are more faithful to this guiding idea, and prove that they are equivalent.

1. *The Idea of Stochastic Dominance*

2. *Counterexamples*

2.1 *Incomparability*

2.2 *Non-separability*

3. *Setwise Stochastic Dominance*

4. *Stochastic Dominance*

5. *Conclusion*

Appendix

1 The Idea of Stochastic Dominance

One of the central ideas in decision theory is dominance. Consider two risky prospects A and B , each of which may turn out several different ways. If it is certain that A will turn out at least as well as B , then we say A (weakly) dominates B . It is widely accepted that if A weakly dominates B , then A is

at least as good a prospect as B .^{1 2}

(Weak) Dominance. For prospects A and B , if A is certain to turn out at least as well as B , then A is at least as good as B .

While Dominance is not utterly uncontroversial, it is one of the least controversial principles there is in normative decision theory. Evaluating a prospect means weighing up the various ways that it might turn out, trading off its chances of turning out well or badly in various ways against one another. If a prospect A is sure to turn out as well as B , then all of the various ways B might turn out are ways A will turn out as well or better. Every point in favour of B is at least as strong a point in favour of A ; and every point against A is at least as strong a point against B . So however we trade off these possible outcomes against each other, A provides at least as good a trade-off as B does.³

A second central idea in decision theory is that the value of a risky prospect is determined by the probabilities of it turning out one way or another.⁴ For each risky option A , there are various ways it might turn out if chosen; and there is some probability of it turning out each of these ways. So a prospect A has a corresponding probability distribution over outcomes.⁵

Stochasticism. Two prospects that have the same probability distribution over outcomes are equally good.

The idea is that probabilities are adequate for capturing all of the information about risk that is relevant to how the value of different outcomes should be weighed. Stochasticism is not without challenges, but it is widely accepted even so.⁶ Call two prospects with the same probability distributions over

¹This notion of ‘good’ might be understood in terms of what it is rational to choose or prefer, or in terms of moral or prudential betterness with respect to incomplete evidence. This essay concerns structural issues that do not depend on which interpretation we apply.

²For now we focus on weak dominance and betterness. We’ll say something about strict versions in section 4.

³Compare similar arguments in Broome ([1991], pp. 95–6, in defence of the sure thing principle); Tarsney ([2020], p. 8, in defence of stochastic dominance).

⁴Again, what kind of probability is at issue—whether it is evidential, or credal, or chancy, or something else—is not important for this essay. I will take for granted that probability distributions satisfy the standard Kolmogorov axioms, including countable additivity. But I anticipate that many of the lessons of this essay will apply beyond that framework—for example, to finitely additive or infinitesimal probabilities—with minimal modification.

⁵Outcomes are understood as ways a risky option might turn out, in whatever respects matter for goodness. These might be modelled as equivalence classes of possible worlds under the relation of being equally good overall. Prospects are understood as possible objects of choice under risk, without commitment to how exactly these should be modelled: prospects are not identified with probability distributions, nor with functions from states to outcomes, though these may be used to represent prospects. See section 4.

⁶Schoenfeld ([2014]) and Bales *et al.* ([2014]) argue against Stochasticism in cases involving incomparability; Seidenfeld *et al.* ([2009]) and Lauwers and Vallyntyne ([2017]) argue against

outcomes **stochastically equivalent**.

It may not be immediately clear how the two ideas of Dominance and Stochasticism should be combined: for Dominance does not apply directly to probability distributions. Suppose Lottie has a choice between two gambles A and B with the following probability distributions.

- A Probability 1/2 of winning a new scooter, and probability 1/2 of winning nothing.
- B Probability 1/4 of winning a new scooter, and probability 3/4 of winning nothing.

Does option A dominate option B ? The probability distributions on their own don't answer this question. Here is one way the game might be played. Say the outcome is determined by drawing a card from a well-shuffled deck.

- A If a red card is drawn, win the scooter, and otherwise win nothing.
- B If a diamond is drawn, win the scooter, and otherwise win nothing.

If these are the rules, then option A dominates option B . But the game could be played a different way.

- B' If a club is drawn, win the scooter, and otherwise win nothing.

In this case, A does not dominate B' — A is not certain to turn out as well as B' , since B' will have a better result than A if a club is drawn. (See table 1.)

Table 1: Lottie's gambles. A dominates B but not B' , while B and B' are stochastically equivalent.

Gambles	♦	♥	♣	♠
A	🎰	🎰	✗	✗
B	🎰	✗	✗	✗
B'	✗	✗	🎰	✗

Nonetheless, even though A does not dominate B' , we do have enough information to compare them. For A does dominate the option B , and B is stochastically equivalent to B' . So even though Dominance on its own does not tell us about how A and B' compare, Dominance and Stochasticism together tell us that A is at least as good as B' .

it in cases involving infinity. For defences, see Nover and Hájek ([2004]); Easwaran ([2014]); Bader ([2018]).

Still, this situation is not entirely satisfactory. Applying Dominance requires a framework for representing risky prospects that includes more information than just their probability distributions over outcomes: we have to reason about distinct stochastically equivalent prospects like B and B' . This seems out of the spirit of Stochasticism. There is a standard way of improving this theoretical situation: we can define a single relationship between probability distributions that simultaneously captures the upshot of both Stochasticism and Dominance—a relation called ‘stochastic dominance’. Here is how the usual definition goes.

Definition 1. Let P and Q be probability distributions over an ordered set of outcomes \mathcal{X} .⁷ We say P naïvely (first-order, weakly) stochastically dominates Q if, for every possible outcome $x \in \mathcal{X}$, P has at least as high a probability of turning out at least as well as x as Q does. Using the shorthand $(\geq x)$ for the set of outcomes y which are at least as good as x :

$$P(\geq x) \geq Q(\geq x) \quad \text{for every outcome } x.$$

We will similarly say that a prospect A naïvely stochastically dominates a prospect B when A 's probability distribution naïvely stochastically dominates B 's probability distribution. Many decision theorists have been attracted to the following normative principle.⁸

Naïve Stochastic Dominance. For any prospects A and B , if A naïvely stochastically dominates B , then A is at least as good as B .

This sounds intuitively very plausible: if A is at least as likely as B to give you something good, for any threshold for what counts as ‘good’, then A seems to be doing as well as B in every respect that matters for comparing prospects.

Furthermore, it is widely believed that Naïve Stochastic Dominance is simply a way of capturing the combination of Stochasticism and Dominance. A basis for this would be the following technical claim:

Naïve Realization. Suppose there are prospects with distributions P and Q , respectively. If P naïvely stochastically dominates Q , then there are some prospects A and B that have the probability distributions P and Q , respectively, such that A dominates B .

Stochasticism, Dominance, and Naïve Realization together imply Naïve Stochastic Dominance. Suppose A naïvely stochastically dominates B . Then Naïve Realization tells us that there is a pair of prospects A' and B' with

⁷We assume that \mathcal{X} is also equipped with a σ -algebra; more on this in section 3 and appendix A.1. But we will keep these technical details in the background as much as possible.

⁸Fine ([2008], pp. 627–8, though without endorsement); Buchak ([2013]); Easwaran ([2014]); Bader ([2018]); Tarsney ([2020]); Wilkinson ([2022a]).

the same distributions, where A' dominates B' . Stochasticism tells us that A is just as good as A' and B' is just as good as B ; and Dominance tells us that A' is at least as good as B' . Putting this together, by transitivity, A is at least as good as B .⁹

But Naïve Realization and Naïve Stochastic Dominance are both wrong. In what follows I will explain why, and consider how to repair stochastic dominance.

Naïve Realization is closely related to the following well-known statistical folklore.¹⁰

Theorem 1. For any probability measures P and Q on the real numbers,¹¹ the following are equivalent:

- (a) For every real number x , $P(\geq x) \geq Q(\geq x)$.
- (b) There exist real-valued random variables X, Y on some probability space such that X has the distribution P , Y has the distribution Q , and $X \geq Y$ pointwise. (That is, $X(s) \geq Y(s)$ for every state s in the probability space.)

It is commonplace to represent prospects by real-valued random variables—that is, measurable real-valued functions defined on some probability space. The underlying probability space represents states of nature, and the real numbers represent outcomes; a prospect is then represented by a function from states to outcomes. If this representation is faithful, then Naïve Realization follows.

The trouble is that not all prospects can be represented by real-valued random variables—because some possible outcomes cannot be represented by real numbers.

⁹Suppose we represent each prospect A by a function f_A from states to outcomes. Define a permutation of a prospect to be a function f' such that, for some probability-preserving bijection π from states to states, $f'(\pi(s)) = f_A(s)$ for each state s . A prospect is stochastically equivalent to any of its permutations. So in this setting, Naïve Realization follows from the following claim:

Naïve Permutation. If A stochastically dominates B , then some permutation of A dominates B .

Easwaran ([2014], p. 20) makes this claim (at least for discrete prospects on a ‘homogeneous’ state space), and it is repeated by Bader ([2018], p. 500) and Tarsney ([2020]). Easwaran does prove a different important thing: for discrete prospects A and A' on a homogeneous state space, if A is stochastically equivalent to A' , then A' is a permutation of A . He asserts that the analogous claim for (strict) stochastic dominance follows. As we will see, this is not true in general.

¹⁰This can be derived as a corollary to theorem 3 in section 4, though it also has a more elementary proof.

¹¹Equipped with the Borel algebra; see section 3 and appendix A.3.

The ordering of the real numbers is total and separable. ‘Total’ means that for any pair of real numbers, one is greater than the other, or else they are equal. ‘Separable’ means that there is a countable set of real numbers—for instance, the rational numbers—which is dense in the reals: for any two distinct real numbers, there is some rational number between them.¹²

It is far from clear that all the values we care about have these features. Many theorists hold that some outcomes are incomparable (or ‘incommensurable’ or ‘on a par’): neither is better than the other, but they are not equally good (see Chang [2002]; Hare [2010]). And many theorists hold that there are infinite values, the ordering of which need not be separable. For example, consider possible outcomes in which there are infinitely many future generations. If the interests of people in the arbitrarily distant future all have the same weight, and a Pareto principle holds—namely, that an outcome where some are better off and none are worse off is better overall—then the ordering of distributions of goods to future generations is not separable (see Basu and Mitra [2003]). I will argue that without totality and separability, the technical claim Naïve Realization can fail—and the normative claim Naïve Stochastic Dominance fails along with it.

Recently quite a few theorists have taken interest in applying stochastic dominance concepts in settings like these, where real-valued representations of outcomes are unavailable (for example, Bader [2018]; Russell [2021]; Wilkinson [2022b]).¹³ So it has become pressing to understand how dominance concepts ought to be deployed in these more exotic settings. Outside the familiar realm of real numbers, we must tread carefully. Stochastic dominance must be fixed.

2 Counterexamples

Incomparability and non-separability each give rise to counterexamples to Naïve Stochastic Dominance. Let’s consider each in turn.

2.1 Incomparability

Coffee and Crowds. Inga is planning to do some writing at a local coffee shop: either the Percolator or Quixote’s. She is hoping to get good espresso and chill vibes. These two things are incommensurable: neither is better than the other, but they are also not equally good. But

¹²In fact, several subtly different countability-related properties are relevant, as will be discussed in section 3.

¹³Fishburn ([1978]) considers generalizations of stochastic dominance in a setting without transitivity; this is not something I will take up here.

getting both good espresso and chill vibes is better than either one alone.

Quixote’s is unreliable on both counts. $1/3$ of the time they have good coffee and chill vibes, but $1/3$ of the time they burn the coffee beans, and $1/3$ of the time they are overtaken by a local church group. (These two calamities never both happen at the same time, because the less competent barista happens to be part of the church group.)

The Percolator always has good espresso, and it is always a great place to work—when it is open. But the owner often closes the shop at unpredictable times—and if Inga goes there when it is closed, she will get neither of the good things she desires, and lose most of her writing day. This outcome has chance $1/3$, and it is much worse than dealing with either sub-par coffee or a crowded cafe.

We can represent Inga’s four possible outcomes as a diamond, with edges pointing upward from worse outcomes to better outcomes, as in figure 1. We can represent Inga’s two options with two probability distributions over these four nodes: P for the Percolator and Q for Quixote’s.

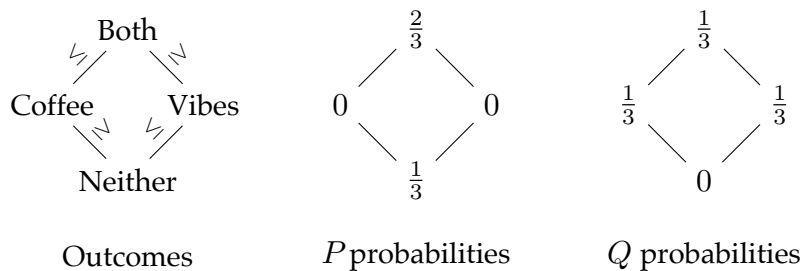


Figure 1: Coffee and Crowds

It seems clear that, as long as the outcome Neither is sufficiently bad, the Percolator is a worse option than Quixote’s for Inga. Taking the $1/3$ -probability risk of getting the worst outcome (Neither) can outweigh the extra probability that the Percolator gives to the best outcome (Both), if the worst outcome is much worse than either just Coffee or just Vibes, while the best outcome is only a little better than these.

Yet P naïvely stochastically dominates Q . We can check this by considering each outcome. The probability of the best outcome (Both) is higher for P than Q : $2/3$ versus $1/3$. The probability of an outcome at least as good as Coffee alone is $2/3$ for either one of P or Q . The same goes for the probability of an outcome at least as good as Vibes alone. And the probability of an outcome at least as good as Neither is 1 in each case. So for every possible outcome x ,

$$P(\geq x) \geq Q(\geq x).$$

But if prospects A and B have distributions P and Q , respectively, then A *can't* dominate B . This is easy to see: A has probability $1/3$ of obtaining the worst outcome, Neither, while B has probability 0 of Neither. So there is at least probability $1/3$ that A turns out strictly worse than B . This tells us that Naïve Realization fails.

Likewise, if the Percolator is not as good as Quixote's, Coffee and Crowds is also a counterexample to the normative principle of Naïve Stochastic Dominance.

2.2 Non-separability

The second example is a bit more technical.

The Wrong Circles. Dante was wrong: there are not just nine circles of Hell, but uncountably many—in fact, one for each of the countable ordinals,

$$1, 2, 3, \dots, \omega, \omega + 1, \dots, 2\omega, \dots, \omega^2, \dots$$

Circle 1 is worst, followed by circle 2, and so on, with earlier ordinals corresponding to worse circles.

There are also uncountably many circles of Heaven—again, one for each countable ordinal. In this case the earliest ordinals correspond to the best circles—so heavenly circle 1 is best, followed by circle 2, and so on.

Sepehr is considering whether to eat of the fruit of the tree of the knowledge of good and evil. If he does, he is doomed to Hell with certainty—but it is uncertain which circle it will be. In fact, the probability distribution over circles of Hell has the interesting property that for every countable ordinal i , the probability that he will be in a circle worse than i is zero. (Even though the union of these probability zero sets has probability one, this does not violate countable additivity, because there are uncountably many of them.)

If Sepehr does not eat of the fruit, then he will certainly be rewarded in Heaven—but again, it is uncertain which of the beatific circles will be his. For each countable ordinal i , the probability of going to a circle of Heaven *better* than i is zero.

It seems clear that eating the fruit is worse than refraining: the consequence is Hell rather than Heaven, with certainty. But, perhaps surprisingly, eating the fruit (weakly) naïvely stochastically dominates refraining. Consider each possible outcome. For any circle of Hell, if Sepehr eats the fruit, the

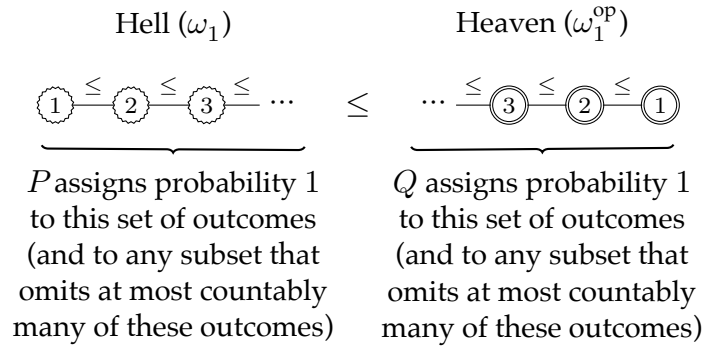


Figure 2: The Wrong Circles

probability of ending up in that circle or better is *one*. And for any circle of Heaven, if Sepehr eats the fruit, the probability of ending up in that circle or better is *zero*. But we get the exact same probabilities if Sepehr refrains from eating the fruit: again, Sepehr is sure to do better than any circle of Hell, but the probability of doing at least as well as any particular circle of Heaven is again zero. In short, if P is the probability distribution Sepehr obtains from eating the fruit, and Q the probability distribution that arises from refraining, then for every possible outcome x , $P(\geq x) = Q(\geq x)$, and thus $P(\geq x) \geq Q(\geq x)$.¹⁴

But it is clear that if a prospect A has the ‘downstairs’ distribution P , while B has the ‘upstairs’ distribution Q , then A can’t possibly dominate B . Since A sends Sepehr to Hell with probability one, and B sends Sepehr to Heaven with probability one, the probability that A turns out worse than B is also one.

Furthermore, since it seems clear that eating the fruit is a worse prospect for Sepehr than refraining, The Wrong Circles is also a counterexample to the normative principle of Naïve Stochastic Dominance.¹⁵

¹⁴The probability measure P assigns probability zero to each countable set of circles of Hell, and probability one to each co-countable set of circles of Hell. This is all we need to say, if we are only concerned with defining a measure on the weak σ -algebra containing just the countable and co-countable sets of countable ordinals. There is also a richer σ -algebra which is natural for the set of countable ordinals: namely, the Borel algebra generated by the order topology. The existence of such a measure P on this algebra is less straightforward: see Folland ([1999], sec. 7.2, ex. 15).

¹⁵The Wrong Circles is a contrived example, but there may be more realistic cases that are structurally similar. I mentioned in section 1 that a non-separable ordering of outcomes arises from the Pareto order on welfare distributions for people in the arbitrarily distant future. If the set of individuals is uncountable, there are sets of such distributions for which the Pareto order is isomorphic to the ordering in the Wrong Circles ($\omega_1 \oplus \omega_1^{\text{op}}$). It is not obvious whether we could find ourselves with probabilities like Sepehr’s over such outcomes, but it seems rash to assume otherwise.

3 Setwise Stochastic Dominance

There is a pretty straightforward modification of Naïve Stochastic Dominance that correctly handles the examples of Coffee and Crowds and The Wrong Circles. This fix stays true to the gloss we originally gave for naïve stochastic dominance: if A is at least as likely as B to give you something good, for any threshold for goodness, then A is at least as good as B . Where we went wrong was in identifying a threshold with an outcome. Intuitively, some thresholds don't line up with particular outcomes, but rather fall in the gaps between outcomes. We can capture this by thinking of a 'threshold of goodness' not as an outcome, but as a set of outcomes.

Definition 2. A set of outcomes U is called upward-closed (or an upper set) iff for each outcome x in U , for any outcome y which is at least as good as x , y is in U as well. For probability distributions P and Q , we'll say P setwise stochastically dominates Q when

$$P(U) \geq Q(U) \quad \text{for every (measurable) upward-closed set of outcomes } U.$$

A set of the form $(\geq x)$, the set of all outcomes at least as good as a given outcome x , is called a principal upper set. So naïve stochastic dominance amounts to restricting setwise stochastic dominance to just the principal upper sets, rather than considering all of them.

For the ordering of the real numbers, countable additivity ensures that the setwise definition of stochastic dominance and the naïve definition coincide.¹⁶ But consider the two counterexamples. In Coffee and Crowds, the set of outcomes U containing Coffee, Vibes, and Both—but not Neither—is a non-principal upper set. Furthermore, in that example, $P(U) = 2/3$ for the Percolator, while $Q(U) = 1$ for Quixote's. In other words, while P is at least as likely as Q to turn out at least as well as any particular outcome, it is less likely than Q to meet the standard of providing something at least as good as either Coffee or Vibes. So while P naïvely stochastically dominates Q , it does not setwise stochastically dominate Q .

In The Wrong Circles, consider the set of outcomes U consisting of all of the heavenly outcomes. This, again, is a non-principal upper set in X . (There is no worst circle of Heaven, no best circle of Hell, and no outcome between

¹⁶ More generally, if a total order is countably generated, in the sense defined in appendix A.3, then every upward-closed set is a countable union of a chain of principal upper sets, which implies that naïve and setwise stochastic dominance coincide. *Proof.* Let U be upward-closed. We can recursively define a transfinite sequence $x_0 \geq x_1 \geq \dots$ of elements of U . For each countable ordinal i , if U is the countable union of principal upper sets $(\geq x_j)$ for $j < i$, we are done. Otherwise, there is some $x_i \in U$ such that $x_i < x_j$ for any $j < i$. If \mathcal{U} generates the order \leq , then for each countable ordinal i there is some set $V_i \in \mathcal{U}$ such that $x_i \in V_i$, but $x_{i+1} \notin V_i$. So these are all distinct, and thus \mathcal{U} is uncountable. \square

Heaven and Hell.) Furthermore, eating the fruit has probability $P(U) = 0$, while refraining has probability $Q(U) = 1$. So once again P does not setwise stochastically dominate Q .

It is tempting, then, to replace the normative principle of Naïve Stochastic Dominance with a principle involving this setwise notion:

Setwise Stochastic Dominance. If prospect A has a distribution that weakly setwise dominates the distribution of prospect B , then A is at least as good as B .

This principle does indeed hold in a broad class of situations—much broader than the naïve principle. Unfortunately, it is still not quite right.

Love and Money. Lexy cares about just two things: love and money. The amount of love in Lexy's life can come in continuous degrees between 0 and 1. It is also better for her to have more money rather than less. But love is infinitely more important to Lexy than money: a larger amount of love is always better for her than a smaller amount, no matter how much money she has in either situation.

Unfortunately, love is utterly unpredictable and uncontrollable: whatever Lexy does, the amount of love she will get is given by a uniform probability distribution over the interval $[0, 1]$. But she *can* decide whether to pick up a \$100 bill that happens to be lying on the ground in front of her.

We can represent Lexy's possible outcomes with two line segments (figure 3): one representing the different possible amounts of love together with the \$100, and the other representing the different possible amounts of love without the \$100. The horizontal axis—the amount of love—is lexicographically more important than the vertical axis—the amount of money. (To be explicit, the outcome space is $[0, 1] \times \{\$0, \$100\}$, consisting of two copies of the line segment.)

If Lexy leaves the \$100 lying on the ground, then the probability of getting any particular outcome is given by a uniform distribution over the bottom segment—where she doesn't get the \$100. Call this distribution P . If she takes the \$100, then the probability of getting any particular outcome is given by a uniform distribution over the top segment, where she gets \$100. Call this Q .

It seems clear that leaving the money is worse than taking it: Lexy loses \$100, with no compensating benefit at all. But it turns out that P (weakly) *setwise* stochastically dominates Q . For each amount of love $x \in [0, 1]$, there are these three upper sets:

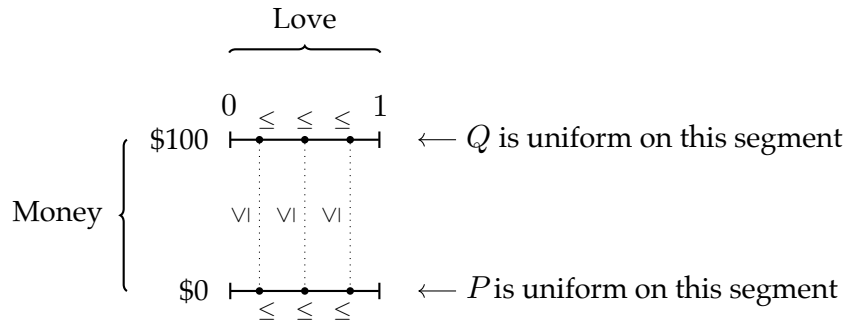


Figure 3: Love and Money

- The set of outcomes at least as good as $(x, \$0)$.
- The set of outcomes at least as good as $(x, \$100)$.
- The set of outcomes strictly better than $(x, \$100)$.

The P -probability and the Q -probability of any of these sets are both exactly $1 - x$. Moreover, every upper set in this space of outcomes is of one of these three kinds. So $P(U) = Q(U)$ for every upper set U . (Note that while the outcome $(x, \$100)$ is possible if Lexy takes the money, and not otherwise, P and Q both assign this outcome probability zero.) Even so, no prospect with distribution P can (weakly) statewise dominate a prospect with distribution Q . (The proof of this is less straightforward than the earlier examples, and can be found in appendix A.5.)

What is going wrong this time? For Lexy's ordering of outcomes, cumulative probabilities—the probabilities of the upward-closed sets—don't provide enough information to pin down a measure. The two measures P and Q agree on every cumulative probability, but they still disagree on other sets of outcomes, such as the upper \$100 segment. When probabilities can float free of cumulative probabilities, setwise stochastic dominance is too crude an instrument to distinguish measures that intuitively should be distinguished.

We can give a more precise diagnosis of how this situation arises by introducing some technical ideas. The lexicographic ordering of outcomes in Love and Money is in fact both total and separable; but it lacks a stronger countability property.¹⁷ In general, partial orders have a natural topology, called the interval topology. (This generalizes the topology that is standardly defined on total orders. Precise definitions and proofs for the topological ideas in this section are in appendix A.3.) A topological space that has a countable base is called 'second-countable'. For example, the standard topology on the real numbers is second-countable, because every open set of real numbers is a union of open intervals with rational endpoints, and there are only

¹⁷Thanks to Kenny Easwaran for prompting this line of thought.

countably many such intervals. In general, every second-countable space is separable, but the interval topology on the lexicographic space $[0, 1] \times \{0, 1\}$ is an example of a space that is separable, but not second-countable.¹⁸

This is connected to a technical issue that we have kept in the background so far. A probability measure is defined on some σ -algebra of sets of outcomes. Partially ordered spaces have a natural σ -algebra inherited from their interval topology, called the Borel algebra; measures defined on this algebra are called Borel measures. When the interval topology is second-countable, then this Borel algebra is precisely the same as the σ -algebra that is generated by the upward-closed Borel sets. Thus any two countably additive probability measures that agree on the probabilities of upward-closed sets also assign the same probability to every set in the Borel algebra.¹⁹

In short, Borel measures on second-countable spaces are characterized by their cumulative probabilities. Moreover, it turns out that Setwise Stochastic Dominance really is correct for such measures, in a sense to be made precise in section 4. But more generally, probability measures on a too-rich algebra of sets can float free from their cumulative probabilities, as in Lexy's case.²⁰ This is a third way that counterexamples can arise, not just to Naïve Stochastic Dominance, but also to Setwise Stochastic Dominance.

4 Stochastic Dominance

Neither naïve stochastic dominance nor setwise stochastic dominance answers to our purpose, in general. What can replace them?

Let us return to the original idea: stochastic dominance is supposed to extend the idea of *dominance* to the stochastic setting of *probability distributions* over outcomes, which need not include all of the information about how these outcomes are associated with particular states. When the values of outcomes can be represented by real numbers, theorem 1 told us that naïve stochastic dominance was good enough, because in that setting it is equivalent to another relation between probability distributions P and Q : namely, that there exist random variables with those distributions such that one dominates the

¹⁸*Separable.* The countable set of outcomes $(q, 0)$ for each rational $0 \leq q \leq 1$ intersects every open interval (x, y) in $[0, 1] \times \{0, 1\}$ with the lexicographic order.

Not second-countable. In the interval topology on the lexicographic order on $[0, 1] \times \{0, 1\}$, for each real number $r \in [0, 1]$, the set of outcomes $I_r = \{x \mid x > (r, 0)\} = \{x \mid x \geq (r, 1)\}$ is open. So any base would have to include an open set U_r such that $(r, 1) \in U_r \subseteq I_r$. These must all be distinct, so the base is uncountable.

¹⁹This is a consequence of Carathéodory's extension theorem (the uniqueness part. See Folland [1999], sec. 1.4).

²⁰As it happens, Lexy's P and Q are not even Borel measures (see appendix A.5). But we can slightly modify Love and Money to construct a counterexample using Borel measures on the lexicographically ordered square $[0, 1] \times [0, 1]$, which is also not second-countable.

other. In more general contexts this equivalence breaks down. For those contexts, it makes sense to simply focus on this other relation directly, taking it as our *definition* of stochastic dominance.

Definition 3. For probability distributions P and Q on an ordered space of outcomes \mathcal{X} , P (weakly) stochastically dominates Q if and only if there exists a probability space (Ω, μ) and random variables $X, Y : \Omega \rightarrow \mathcal{X}$ with corresponding probability distributions $\mu_X = P$ and $\mu_Y = Q$, where $X \geq Y$ pointwise.²¹

In a setting where prospects straightforwardly correspond to outcome-valued random variables (rather than real-valued random variables), this just amounts to combining dominance and stochastic equivalence—which was the original idea all along. Suppose that prospects A and B can be represented by random variables X and Y on a probability space (Ω, μ) . Then A 's distribution is μ_X and B 's distribution is μ_Y . The definition tells us that μ_X stochastically dominates μ_Y if and only if there exist random variables X' and Y' which are stochastically equivalent to X and Y , respectively (that is, $\mu_X = \mu_{X'}$ and $\mu_Y = \mu_{Y'}$), such that X' dominates Y' (that is, $X' \geq Y'$ pointwise).

There is also a corresponding normative principle:

Stochastic Dominance. For prospects A and B , if A has a probability distribution that stochastically dominates the probability distribution of B (in the sense of definition 3), then A is at least as good as B .

This principle really is equivalent to the combination of Stochasticism and Dominance, assuming a correspondence between prospects and random variables.

Still, some reasons for dissatisfaction with definition 3 remain. The idea of stochasticism is that probabilities of outcomes carry all the information we need for comparing prospects. But definition 3 takes us outside that framework: while stochastic dominance is a relation between probability distributions, it is 'extrinsic', in that it appeals to other probability spaces and random variables on those spaces. So it is nice to find that there is another formulation, equivalent to definition 3, which is more at home in the framework of probability distributions over outcomes.

Here is the idea. In the stochastic framework, what we care about is just the probability of prospects turning out one way or another. But the relation of dominance isn't just a matter of the ways one prospect might turn out, and

²¹ For a measure μ on a set \mathcal{X} , a set \mathcal{Y} equipped with a σ -algebra, and a measurable function $f : \mathcal{X} \rightarrow \mathcal{Y}$, the induced measure μ_f on \mathcal{Y} is given by $\mu_f(E) = \mu(f^{-1}(E))$ for each measurable set $E \subseteq \mathcal{Y}$.

the ways another prospect might turn out. It is about whether one prospect would turn out as well as another would—which is crucially a matter of how both prospects would turn out, taken together.

This is something that it makes sense to model directly. Consider again the example of Lottie’s gambles from section 1.

- A* If a red card is drawn, win the scooter, and otherwise win nothing.
- B* If a diamond is drawn, win the scooter, and otherwise win nothing.
- B'* If a club is drawn, win the scooter, and otherwise win nothing.

Here *A* dominates *B*, but *A* does not dominate the stochastically equivalent prospect *B'*. But it isn’t essential to talk about the cards in order to capture this contrast. We can instead consider the joint probability distributions, which tell us the probabilities of the various ways each of two prospects might turn out. For example, the probability is 1/4 that choosing gamble *A* would result in winning the scooter and choosing gamble *B* would result in winning nothing. In general, the joint distribution for two prospects assigns probabilities to pairs of outcomes. The two joint distributions are shown in table 2.

Table 2: Joint distributions for Lottie’s gambles.

Gambles	(\vec{d}_s, \vec{d}_s)	(\vec{d}_s, \times)	(\times, \vec{d}_s)	(\times, \times)
<i>A</i> and <i>B</i>	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
<i>A</i> and <i>B'</i>	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

The only case where the first gamble turns out worse than the second is the pair of outcomes (nothing, scooter). The crucial difference between the two joint distributions is that in the case of dominance (*A* and *B*), this pair gets probability zero; meanwhile, in the case of non-dominance (*A* and *B'*) the probability of this pair is non-zero.

In short, dominance is not a property of a pair of probability distributions over outcomes, but it is a property of a joint probability distribution over pairs of outcomes. If *A* and *B* are prospects with joint distribution μ , then for *A* to dominate *B* is for their joint distribution to put all of its probability into pairs of outcomes (x, y) such that x is at least as good as y .²² To give

²²In general it may be important to distinguish between it being certain that *A* turns out as well as *B*, and it being merely almost certain that *A* turns out as well as *B*—that is, with probability one. But in the stochastic framework, this distinction between certainty and almost certainty disappears: the picture is that the probabilities of the outcomes are all that matters, and certainty, as opposed to almost certainty, is not merely a matter of probability. This may be an important philosophical challenge to Stochasticism, or to the

that a name, we say their joint distribution is supported by the set of pairs

$$\{(x, y) \in \mathcal{X} \times \mathcal{X} \mid x \geq y\}.$$

(The precise definition of support is in appendix A.1.)

A joint distribution for two prospects includes all of the information about each of those prospect's individual distributions: these are marginal distributions. A distribution μ on pairs of outcomes $\mathcal{X} \times \mathcal{X}$ has a marginal distribution μ_1 on single outcomes in \mathcal{X} , where the μ_1 -probability of an outcome x is the μ -probability of getting any pair of outcomes with x as its first coordinate. Likewise, μ has a marginal distribution μ_2 for the second coordinate. If A and B have the joint distribution μ , then A has the distribution μ_1 and B has the distribution μ_2 .²³

This gives us yet another way of defining stochastic dominance.

Definition 4. For probability distributions P and Q on an ordered space of outcomes \mathcal{X} , P (weakly) stochastically dominates Q if and only if there exists a joint probability distribution μ on $\mathcal{X} \times \mathcal{X}$ such that $\mu_1 = P$, $\mu_2 = Q$, and μ is supported by the set $\{(x, y) \in \mathcal{X} \times \mathcal{X} \mid x \geq y\}$.

The idea of definition 4 is that there is some way of 'lining up' the different outcomes of P and Q such that one dominates the other. This definition, unlike definition 3, does not appeal to random variables on some extraneous state space. In the spirit of Stochasticism, it is all about the probabilities of prospects turning out one way or another—now including joint probabilities for ways of pairing up the outcomes two different prospects each would have.

We have defined stochastic dominance twice over: once in terms of random variables, and once in terms of joint distributions. But there is no conflict between the two definitions.

Theorem 2. Definitions 3 and 4 are equivalent.

Furthermore, we can now say precisely how setwise stochastic dominance (definition 2) comes close to the correct definition, as we alluded in section 3.

Theorem 3. Definition 2 is equivalent to definitions 3 and 4 for Borel probability measures on a partially ordered space of outcomes whose interval topology is second-countable.

standard probability axioms. But this is not the place to fuss over it; here we are working out the implications of these commitments.

²³The marginal distributions μ_1 and μ_2 are simply the induced measures μ_{π_1} and μ_{π_2} (in the sense of footnote 21), where π_1 and π_2 are the projection functions that take pairs in $\mathcal{X} \times \mathcal{X}$ to their first and second coordinates, respectively.

The proofs of these theorems are in appendices A.2 and A.4.

Since the topology of the real numbers is second-countable, and naïve stochastic dominance and setwise stochastic dominance come to the same thing for measures on the reals (footnote 16), theorem 3 subsumes the original ‘folklore’ theorem 1 discussed in section 1. Another consequence is that all three definitions are equivalent for discrete probability measures on any partial order at all. (In that case we can think of the measures as ‘living on’ a countable set of outcomes that they both assign probability one.) So setwise stochastic dominance is a pretty good approximation for what I consider to be the correct notion—quite a bit better than the naïve definition, though imperfect in general.

Representing prospects as random variables naturally goes with a Savage-style framework, which is commonly understood as follows. A ‘state of nature’ is supposed to specify a way that everything independent of an agent’s choice might be. An ‘outcome’ is supposed to capture everything relevant to how good or bad it is for things to turn out a certain way. A prospect and a state of nature together are supposed to determine an outcome; thus a prospect determines a function from states to outcomes. Moreover, if every (measurable) function from states to outcomes is to represent some evaluable prospect, then states and outcomes must be ‘freely recombinable’: any state of nature is consistent with any possible outcome. While they may be a reasonable idealization for modelling certain simple decision problems, in general these assumptions are dubious.²⁴

The joint distribution approach suggests a different picture. That is not to say that the state-outcome formalism is wrong—after all, definitions 3 and 4 are equivalent, not rivals. But thinking in terms of joint distributions suggests a more lightweight interpretation of the formalism. What plays the role of a ‘state’ is simply a pair of outcomes. One natural interpretation of the joint distribution is that the pair (x, y) stands in for the conjunction of two counterfactuals: if A were chosen, then x would result, and if B were chosen, then y would result. A joint distribution represents a probability distribution over these conjunctions. Perhaps the counterfactuals are ultimately grounded in some deeper, more sweeping ‘state of nature’, but our decision-theoretic formalism need not entangle us in any such metaphysics. Furthermore, there is no temptation to try to make sense of recombining these lightweight ‘states’ with arbitrary outcomes.

Thus far I have streamlined discussion by focusing on weak dominance principles, about what is at least as good, rather than what is strictly better. It is not difficult to extend these ideas to a relation of strict stochastic dominance as well.

²⁴For critical discussion, see Joyce ([1999], p. 107).

Definition 5. A probability distribution P strongly stochastically dominates Q if and only if P weakly stochastically dominates Q , while Q does not weakly stochastically dominate P .

There is a corresponding normative principle:

Strong Stochastic Dominance. If a prospect A has a distribution that strongly stochastically dominates the distribution of a prospect B , then A is strictly better than B .

By the same kind of reasoning as before (and given the same assumptions), the combination of Weak and Strong Stochastic Dominance is equivalent to the combination of Stochasticism, (Weak) Dominance, and a further principle:

Strong Dominance. If A is certain to turn out at least as well as B , and B is not certain to turn out at least as well as A , then A is strictly better than B .

5 Conclusion

We considered two normative principles, Naïve Stochastic Dominance and Setwise Stochastic Dominance. Each of these faces counterexamples, involving incomparability, failures of countability properties, or measures on especially fine-grained σ -algebras. But this isn't because the idea of stochastic dominance is on the wrong track. Rather, it is because neither of the two relations, naïve stochastic dominance or setwise stochastic dominance, is a good way of capturing this idea, in general—though the setwise version comes much closer. The good news is that we have two better ways of capturing the idea of stochastic dominance: one appealing to random variables, and the other appealing to joint probabilities. And we don't have to choose between these two explications, because they are equivalent.

A Appendix

A.1 Definitions

Let's introduce convenient labels for the various stochastic dominance notions we have considered.

Definition 6. For a finite measure space (\mathcal{X}, μ) and a set $S \subseteq \mathcal{X}$, we say S supports μ if $\mu(E) = \mu(\mathcal{X})$ for every measurable set E containing S .²⁵

²⁵The reason for this slightly roundabout definition is that, in general, the set of pairs of outcomes we will be interested in need not be measurable in the product algebra on $\mathcal{X} \times \mathcal{X}$.

Definition 7. Let P and Q be probability measures on an ordered set \mathcal{X} equipped with a σ -algebra.

- (a) $P \geq_{\text{naive}} Q$ if $P(\geq x) \geq Q(\geq x)$ for every outcome $x \in \mathcal{X}$. (See definition 1.)
- (b) $P \geq_{\text{set}} Q$ if $P(U) \geq Q(U)$ for every measurable upward-closed subset $U \subseteq \mathcal{X}$. (See definition 2.)
- (c) $P \geq_{\text{r.v.}} Q$ if there exists a probability space (Ω, μ) and random variables $X, Y : \Omega \rightarrow \mathcal{X}$ such that $\mu_X = P, \mu_Y = Q$, and $X \geq Y$ pointwise (See definition 3.)
- (d) $P \geq_{\text{joint}} Q$ if there exists a probability measure on $\mathcal{X} \times \mathcal{X}$ (with its product σ -algebra) such that $\mu_1 = P$ and $\mu_2 = Q$, which is supported by the set

$$\Omega = \{(x, y) \in \mathcal{X} \times \mathcal{X} \mid x \geq y\}.$$

(See definition 4.)

A.2 Proof of theorem 2

Recall what theorem 2 says:

For any probability measures P and Q on an ordered space \mathcal{X} (equipped with a σ -algebra), $P \geq_{\text{r.v.}} Q$ if and only if $P \geq_{\text{joint}} Q$.

Proof. Right-to-left. Let μ be a joint distribution on $\mathcal{X} \times \mathcal{X}$ supported by Ω . Then we can build a probability space and appropriate random variables living on Ω . The σ -algebra consists of sets $E \cap \Omega$ for E in the product algebra on $\mathcal{X} \times \mathcal{X}$. Let $\mu^*(E \cap \Omega) = \mu(E)$. The condition that Ω supports μ guarantees that μ^* is a well-defined probability measure. Let X and Y be the restrictions of μ_1 and μ_2 to Ω , respectively. By construction, $X \geq Y$ pointwise, and it is straightforward to check that X and Y are measurable and $\mu_X^* = \mu_1$ and $\mu_Y^* = \mu_2$.

Left-to-right. Suppose X and Y are \mathcal{X} -valued random variables on a probability space (Ω, μ) such that $X \geq Y$ pointwise. Then let (X, Y) be the product function from Ω to $\mathcal{X} \times \mathcal{X}$ such that $(X, Y)(s) = (X(s), Y(s))$ for each $s \in \Omega$. This is measurable, and it is straightforward to check that the induced joint distribution $\mu_{(X, Y)}$ on $\mathcal{X} \times \mathcal{X}$ has the desired properties: its projections are μ_X and μ_Y , and it is supported by the set of $(x, y) \in \mathcal{X} \times \mathcal{X}$ such that $x \geq y$. \square

For example, if \mathcal{X} is the set of countable ordinals equipped with its Borel algebra, then the set of outcomes (x, y) such that $x \geq y$ is not in the product algebra on $\mathcal{X} \times \mathcal{X}$ (see Folland [1999], p. 231, ex. 28 in sec. 7.4).

A.3 Topologies and algebras

Definition 8. Let \mathcal{X} be a partially ordered set.

- (a) The interval topology on a partially ordered set \mathcal{X} is the coarsest topology containing every set $\mathcal{X} \setminus (\geq x)$ or $\mathcal{X} \setminus (\leq x)$ for $x \in \mathcal{X}$.
- (b) The Borel algebra on \mathcal{X} is the smallest σ -algebra that contains every open set of the interval topology.

The interval topology generalizes the standard order topology for totally ordered sets. It sensibly counts each closed interval

$$[x, y] = \{z \in \mathcal{X} \mid x \leq z \leq y\} = (\geq x) \cap (\leq y)$$

as a closed set.

Definition 9. A family \mathcal{U} of subsets of \mathcal{X} is said to generate the order \leq such that, for all $x, y \in \mathcal{X}$, $x \leq y$ if and only if every $U \in \mathcal{U}$ that contains x also contains y . A countably-generated order is one that is generated by some countable family of sets.

Note that in the order that \mathcal{U} generates, every $U \in \mathcal{U}$ automatically counts as upward-closed.

Proposition 1. Let \mathcal{X} be a partially ordered set with a second-countable interval topology.

- (a) The order on \mathcal{X} is countably generated.
- (b) The Borel algebra on \mathcal{X} is generated by the set of all principal upper sets and principal lower sets. (Thus the Borel algebra is also generated by the set of all upward-closed open or closed sets.)

Proof. Part (a). Let \mathcal{B} be a countable base for the interval topology. For any $x, y \in \mathcal{X}$ such that $x \not\leq y$, the set $\mathcal{X} \setminus (\leq x)$ is open, and so there is some $B \in \mathcal{B}$ such that $y \in B$ and $(\leq x) \cap B = \emptyset$; in other words, x is not in the upward-closure of B . So \leq is generated by the countable set of all upward-closures of sets $B \in \mathcal{B}$.

Part (b). Let Σ be the smallest σ -algebra containing every set $(\geq x)$ or $(\leq x)$ for $x \in \mathcal{X}$. Since these are all closed sets in the interval topology, clearly the Borel algebra contains Σ . Let a basic open set be a finite intersection of sets of the form $\mathcal{X} \setminus (\geq x)$ or $\mathcal{X} \setminus (\leq x)$. These sets comprise a base for the interval topology. Clearly every basic open set is in Σ . Moreover, if the interval topology is second-countable, then in particular it has a countable base consisting of basic open sets (see Willard [2004], p. 113, exercise 16B). So every open set is a countable union of basic open sets, and is thus in Σ . So Σ contains the Borel algebra as well. \square

A.4 Proof of theorem 3

We will in fact prove a modest strengthening of theorem 3 (in light of proposition 1 from appendix A.3).

Theorem 4. Let \mathcal{X} be a set equipped with a partial order generated by a countable set \mathcal{U} , as well as a countably generated σ -algebra Σ that contains \mathcal{U} . For probability measures P and Q on \mathcal{X} , the following are equivalent:

- (a) $P \geq_{\text{r.v.}} Q$;
- (b) $P \geq_{\text{joint}} Q$;
- (c) $P(U) \geq Q(U)$ for every set $U \in \mathcal{U}$.

We have already shown that (a) and (b) are equivalent. It is clear that (a) implies (c): if $X \geq Y$ pointwise, then $X^{-1}(U) \supseteq Y^{-1}(U)$ for every measurable upward-closed set U , and thus

$$\mu_X(U) = \mu(X^{-1}(U)) \geq \mu(Y^{-1}(U)) = \mu_Y(U).$$

For the remaining step we will show that (c) implies (b)—and consequently, if $P \geq_{\text{set}} Q$ then $P \geq_{\text{joint}} Q$. This is surprisingly difficult.

The basic idea of the proof is to successively split P and Q into pieces, producing subprobability measures P_F and Q_F for each set F in a certain algebra. Finally, we will assemble all of these little measures into a single big measure on the product space.

We will use an extended notion of setwise stochastic dominance.

Definition 10. Let \mathcal{B} be any Boolean algebra of subsets of a partially ordered set \mathcal{X} , and let P and Q be finitely additive subprobability measures on \mathcal{B} . For any family \mathcal{U} of subsets of \mathcal{X} , say $P \geq_{\mathcal{U}} Q$ if $P(U) \geq Q(U)$ for each $U \in \mathcal{B} \cap \mathcal{U}$.

We can also state this another way, generalizing from measures of sets to integrals of simple functions.

Definition 11. For any family \mathcal{A} of subsets of \mathcal{X} , let $\mathcal{S}(\mathcal{A})$ be the set of non-negative simple \mathcal{A} -measurable functions: that is, functions of the form $\sum_i c_i 1_{A_i}$ for some finitely many non-negative numbers c_1, \dots, c_n and sets $A_1, \dots, A_n \in \mathcal{A}$. (We use the notation 1_A for the characteristic function of A .)

Note that if $\mathcal{A} \subseteq \mathcal{B}$ then $\mathcal{S}(\mathcal{A}) \subseteq \mathcal{S}(\mathcal{B})$. Notice also that the functions in $\mathcal{S}(\mathcal{U})$ are *monotone* with respect to the partial order generated by \mathcal{U} .

Lemma 1. For any Boolean algebra of sets \mathcal{A} , any finitely additive subprobability measures P and Q on \mathcal{A} , and any family of sets \mathcal{U} , we have $P \geq_{\mathcal{U}} Q$ if and only if $\int f \, dP \geq \int f \, dQ$ for every function $f \in \mathcal{S}(\mathcal{A} \cap \mathcal{U})$.

Proof. Straightforward. □

The following lemma is the heart of the proof. In this lemma and its proof, understand ‘measure’ as ‘finitely additive subprobability measure’, and ‘algebra’ as ‘Boolean algebra’.

Lemma 2 (Decomposition Lemma). Let \mathcal{B} be a countable Boolean algebra of subsets of \mathcal{X} , and let \mathcal{U} be a family of subsets of \mathcal{X} . Let P and Q be measures on \mathcal{B} such that $P \geq_{\mathcal{U}} Q$. Let Q_1 and Q_2 be measures such that $Q = Q_1 + Q_2$. Then there exist measures P_1 and P_2 such that

$$P = P_1 + P_2 \quad P_1 \geq_{\mathcal{U}} Q_1 \quad P_2 \geq_{\mathcal{U}} Q_2$$

Proof. The strategy is to start with measures on the trivial subalgebra $\{\emptyset, \mathcal{X}\}$, and inductively extend these measures to richer and richer subalgebras of \mathcal{B} . The induction is a bit tricky, though.

Call a (signed) additive function P_1 on a subalgebra $\mathcal{A} \subseteq \mathcal{B}$ *nice* if it has the following property:

For any $f, g \in \mathcal{S}(\mathcal{U})$, any $h \in \mathcal{S}(\mathcal{B})$, and any $\hat{f}, \hat{g} \in \mathcal{S}(\mathcal{A})$, if $f \leq \hat{f} + h$ and $g \leq \hat{g} + h$, then

$$\int f \, dQ_1 + \int g \, dQ_2 \leq \int (\hat{f} - \hat{g}) \, dP_1 + \int (\hat{g} + h) \, dP.$$

To understand this condition, note that if P_1 is nice, this implies:

- (a) $P_1 \geq 0$. (Let all the functions but \hat{f} be zero.)
- (b) $P_1 \leq P$. (Similarly for \hat{g} .)

Thus if we define $P_2 = P - P_1$ on \mathcal{A} , then P_1 and P_2 are both subprobability measures. Instead of ‘nice function’ we can say ‘nice measure’. Also:

- (c) $P_1 \geq_{\mathcal{U}} Q_1$. (Let $f = \hat{f} \in \mathcal{S}(\mathcal{A} \cap \mathcal{U})$, and the other functions zero.)
- (d) $P_2 \geq_{\mathcal{U}} Q_2$. (Similarly for \hat{g} .)

Furthermore, in the case where $\mathcal{A} = \mathcal{B}$, niceness is in fact equivalent to the combination of these four properties. (In that case, the right-hand side of the inequality can be rewritten $\int (\hat{f} + h) \, dP_1 + \int (\hat{g} + h) \, dP_2$.)

Thus our goal is to find a nice measure P_1 on the countable algebra \mathcal{B} . To do this, it suffices to show two things: first, that there is a nice measure on the trivial subalgebra; second, that each nice measure on a finite subalgebra can be extended to a nice measure on a larger finite subalgebra.

For the base case, let \mathcal{A} be the trivial subalgebra $\{\emptyset, \mathcal{X}\}$, and let P_1 be the measure on \mathcal{A} such that $P_1(\mathcal{X}) = Q_1(\mathcal{X})$. Suppose $f, g, h, \hat{f}, \hat{g}$ are as in the definition of niceness. Since \hat{f} and \hat{g} are in $\mathcal{S}(\{\emptyset, \mathcal{X}\})$, they must each be constant functions, for constants a and b such that $f \leq a + h$ and $g \leq b + h$. Suppose $a \leq b$. Let $f^+ = f + b - a$, and let k be the pointwise maximum of f^+ and g . This function is also in $\mathcal{S}(\mathcal{U})$, and $k \leq b + h$. Thus:

$$\begin{aligned} \int f^+ dQ_1 + \int g dQ_2 &\leq \int k dQ_1 + \int k dQ_2 \quad (\text{since } f^+ \leq k \text{ and } g \leq k) \\ &= \int k dQ \leq \int k dP \quad (\text{since } P \geq_{\mathcal{U}} Q) \\ &\leq \int (b + h) dP \quad (\text{since } k \leq b + h) \end{aligned}$$

Thus

$$\begin{aligned} \int f dQ_1 + \int g dQ_2 &= (a - b)Q_1(\mathcal{X}) + \int f^+ dQ_1 + \int g dQ_2 \\ &\leq (a - b)Q_1(\mathcal{X}) + \int (b + h) dP \\ &= \int (\hat{f} - \hat{g}) dP_1 + \int (\hat{g} + h) dP \end{aligned}$$

(using the fact that $P_1(\mathcal{X}) = Q_1(\mathcal{X})$). So P_1 is nice. The case $a \geq b$ goes similarly (using $k = \max(f, g + a - b)$).

For the inductive step, we will show the following.

Let \mathcal{A} be a finite subalgebra of \mathcal{B} , let E be a set in $\mathcal{B} \setminus \mathcal{A}$, and let \mathcal{A}^* be the algebra generated by $\mathcal{A} \cup \{E\}$. Any nice measure P_1 on \mathcal{A} can be extended to a nice measure P_1^* on \mathcal{A}^* .

Let $\{A_1, A_2, \dots, A_n\}$ be the set of atoms of \mathcal{A} . For any sequence $\bar{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, there is a unique (signed) additive function $P_1^{\bar{x}}$ defined on \mathcal{A}^* such that, for each i ,

$$P_1^{\bar{x}}(A_i \cap E) = x_i \quad P_1^{\bar{x}}(A_i \setminus E) = P_1(A_i) - x_i$$

We will show that there is some $\bar{x} \in \mathbb{R}^n$ such that $P_1^{\bar{x}}$ is nice.

We can split up a simple function in $\mathcal{S}(\mathcal{A}^*)$ into two parts: a ‘flattened out’ function which is simple with respect to the smaller algebra \mathcal{A} , together

with an ‘offset’ function that tell us about the steps the function takes within the atoms of \mathcal{A} —the difference between the function’s value on $A_i \cap E$ and its value on $A_i \setminus E$. That is, any function in $\mathcal{S}(\mathcal{A}^*)$ can be written in the form

$$f + \sum a_i 1_{A_i \cap E} + \sum a'_i 1_{A_i \setminus E}$$

where $f \in \mathcal{S}(\mathcal{A})$ and the two sequences (a_1, \dots, a_n) and (a'_1, \dots, a'_n) are non-negative and orthogonal, in the sense that at least one of a_i or a'_i is equal to zero for each i .

So this is what it means to say that $P_1^{\bar{x}}$ is nice:

Let $f, g \in \mathcal{S}(\mathcal{U})$, $h \in \mathcal{S}(\mathcal{B})$, and $\hat{f}, \hat{g} \in \mathcal{S}(\mathcal{A})$. Let (a_1, \dots, a_n) and (a'_1, \dots, a'_n) be non-negative and orthogonal, and likewise let (b_1, \dots, b_n) and (b'_1, \dots, b'_n) be non-negative and orthogonal. Let

$$\begin{aligned} a &= \sum_i a_i 1_{A_i \cap E} + \sum_i a'_i 1_{A_i \setminus E} \\ b &= \sum_i b_i 1_{A_i \cap E} + \sum_i b'_i 1_{A_i \setminus E} \end{aligned}$$

Suppose $f \leq \hat{f} + a + h$ and $g \leq \hat{g} + b + h$. Then:

$$\int f \, dQ_1 + \int g \, dQ_2 \leq \int ((\hat{f}+a) - (\hat{g}+b)) \, dP_1^{\bar{x}} + \int (\hat{g}+b+h) \, dP$$

The first integral on the right-hand side can be rewritten:

$$\begin{aligned} &\sum (a_i - b_i)x_i + \sum (a'_i - b'_i)(P(A_i) - x_i) + \int (\hat{f} - \hat{g}) \, dP_1^{\bar{x}} \\ &= \sum (a_i - b_i - (a'_i - b'_i))x_i + \sum (a'_i - b'_i)P(A_i) + \int (\hat{f} - \hat{g}) \, dP_1 \end{aligned}$$

In short, what it takes for $P_1^{\bar{x}}$ to be nice is for (x_1, \dots, x_n) to satisfy a family of inequalities of the form

$$\sum c_i x_i \geq d.$$

Moreover, this family of inequalities is closed under positive linear combinations. (The sets of functions $\mathcal{S}(\mathcal{U})$, $\mathcal{S}(\mathcal{B})$, and $\mathcal{S}(\mathcal{A}^*)$ are each closed under positive linear combinations.) We can now simplify things using a result from linear programming: in fact, it is enough to show that the inequalities hold in the case where each coefficient c_i is zero. To state the result precisely:

Lemma 3. Let $L \subseteq \mathbb{R}^{n+1}$ be a set of sequences real numbers closed under positive linear combinations. The following are equivalent:

- (a) There exist $x_1, \dots, x_n \in \mathbb{R}$ such that $\sum_i c_i x_i \geq d$ for every sequence $(c_1, \dots, c_n, d) \in L$;
- (b) $0 \geq d$ for every sequence of the form $(0, \dots, 0, d) \in L$.

This is a variant of Farkas' lemma, and it can be proved in the same way, by eliminating variables one by one. (See Kuhn [1956], Theorem III.)

So we may assume that $a_i - b_i = a'_i - b'_i$ for each i . We know that at least one of a_i and a'_i is zero, at least one of b_i and b'_i is zero, and all four of these terms are non-negative. It follows that $a_i = b_i$ and $a'_i = b'_i$ for each i ; in other words, the functions a and b are equal. In that case, the inequality simplifies to this form:

$$\int f \, dQ_1 + \int g \, dQ_2 \leq \int (\hat{f} - \hat{g}) \, dP_1 + \int (\hat{g} + b + h) \, dP$$

Since $a = b$, we know $f \leq \hat{f} + (b + h)$ and $g \leq \hat{g} + (b + h)$. So the inductive hypothesis that P_1 is nice ensures that each of these inequalities holds. Thus there exists $\bar{x} \in \mathbb{R}^n$ such that $P_1^{\bar{x}}$ is a nice measure that extends P_1 to \mathcal{A}^* .

This completes the induction. It follows that there exists a chain of nice measures defined on finite subalgebras $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \dots$, where $\bigcup_i \mathcal{A}_i$ is the whole countable algebra \mathcal{B} . Taking the union of these measures gives us a nice measure defined on all of \mathcal{B} , completing the proof of the lemma. \square

To complete the proof of the theorem, we need to use a family of finitely additive measures on \mathcal{X} to construct a countably additive measure on $\mathcal{X} \times \mathcal{X}$. We will use the following technical lemma.

Lemma 4 (Radon–Nikodym Theorem). Let \mathcal{B} be Boolean algebra of subsets of \mathcal{X} , and let Σ be the σ -algebra generated by \mathcal{B} . Suppose ν is a finitely additive measure defined on \mathcal{B} , μ is a countably additive finite measure defined on Σ , and $\nu \leq \mu$. Then there exists a Σ -measurable non-negative function $f : \mathcal{X} \rightarrow \mathbb{R}$ such that

$$\nu(A) = \int_A f \, d\mu \quad \text{for each } A \in \mathcal{B}.$$

Proof. See Folland ([1999]), secs. 1.4 and 3.2. \square

Proof of theorem 4. Let \mathcal{U} be a generating set of upward-closed subsets of \mathcal{X} . (Assume without loss of generality that $\mathcal{X} \in \mathcal{U}$.) Let Σ be a countably generated σ -algebra containing \mathcal{U} . Then in particular there is a countable Boolean algebra $\mathcal{B} \supseteq \mathcal{U}$ that generates Σ . Let P and Q be countably additive probability measures on Σ such that $P \geq_{\mathcal{U}} Q$. For $F \in \mathcal{B}$, let $Q_F =$

$Q(- \cap F)$. Using the decomposition lemma (lemma 2), we can iteratively decompose P into finitely additive subprobability measures on finer and finer subalgebras of \mathcal{B} . (Given a measure $P_F \geq_{\mathcal{U}} Q_F$ and a set $E \in \mathcal{B}$, we can decompose P_F into $P_{F \cap E} \geq_{\mathcal{U}} Q_{F \cap E}$ and $P_{F \setminus E} \geq_{\mathcal{U}} Q_{F \setminus E}$.) By a straightforward induction argument, it follows that there exists a family of finitely additive measures P_F on \mathcal{B} for each $F \in \mathcal{B}$ such that

$$\begin{aligned} P_{\mathcal{X}} &= P \\ P_F &\geq_{\mathcal{U}} Q_F \\ P_F + P_{F'} &= P_{F \cup F'} \quad \text{whenever } F \cap F' = \emptyset \end{aligned}$$

Let $R(E, F) = P_F(E)$; so $R(E, -)$ and $R(-, F)$ are finitely additive measures on \mathcal{B} for each $E, F \in \mathcal{B}$. Also,

$$R(-, F) \leq P \quad \text{and} \quad R(E, -) \leq Q \quad \text{for each } E, F \in \mathcal{B}.$$

Thus we can apply the Radon–Nikodym theorem (lemma 4) twice, producing a function $g : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ such that, for each $E, F \in \mathcal{B}$,

$$R(E, F) = \int_E \left(\int_F g \, dQ \right) dP$$

By Fubini's theorem, equivalently,

$$R(E, F) = \int_{E \times F} g \, d(P \otimes Q)$$

where $P \otimes Q$ is the product measure. Now we can at last define the countably additive measure on $\mathcal{X} \times \mathcal{X}$ that we have sought: for any A in the product σ -algebra on $\mathcal{X} \times \mathcal{X}$, let

$$\mu(A) = \int_A g \, d(P \otimes Q)$$

There are a couple of things left to check.

First:

$$\begin{aligned} \mu_1(E) &= R(E, \mathcal{X}) = P_{\mathcal{X}}(E) = P(E); \\ \mu_2(F) &= R(\mathcal{X}, F) = P_F(\mathcal{X}) = Q_F(\mathcal{X}) = Q(F). \end{aligned}$$

So $\mu_1 = P$ and $\mu_2 = Q$ (using the uniqueness part of Carathéodory's extension theorem).

Finally, we show that μ is supported by Ω . Since \mathcal{U} generates the order, this means that for any $x \not\preceq y$ there is some $U \in \mathcal{U}$ such that $y \in U$ and $x \notin U$. Thus

$$(\mathcal{X} \times \mathcal{X}) \setminus \Omega = \{(x, y) \mid x \not\preceq y\} \subseteq \bigcup_{U \in \mathcal{U}} (\mathcal{X} \setminus U) \times U.$$

Furthermore, for each $U \in \mathcal{U}$, by additivity we have

$$P_U(\mathcal{X} \setminus U) + P_U(U) + P_{\mathcal{X} \setminus U}(\mathcal{X}) = P_{\mathcal{X}}(\mathcal{X}) = 1.$$

Also, since $Q_F = Q(- \cap F)$:

$$Q_U(U) + Q_{\mathcal{X} \setminus U}(\mathcal{X}) = Q(U) + Q(\mathcal{X} \setminus U) = 1.$$

Moreover, $P_U \geq_U Q_U$ implies $P_U(U) \geq Q_U(U)$, and $P_{\mathcal{X} \setminus U} \geq_U Q_{\mathcal{X} \setminus U}$ implies $P_{\mathcal{X} \setminus U}(\mathcal{X}) \geq Q_{\mathcal{X} \setminus U}(\mathcal{X})$. Thus

$$\mu((\mathcal{X} \setminus U) \times U) = P_U(\mathcal{X} \setminus U) = 0.$$

So $(\mathcal{X} \times \mathcal{X}) \setminus \Omega$ is covered by countably many measure zero sets, and thus has measure zero. So μ is supported by Ω . \square

A.5 The lexicographic order

Here we show what was claimed in section 3 regarding the lexicographic order counterexample.

Proposition 2. Let $\mathcal{X} = [0, 1] \times \{0, 1\}$ with its lexicographic order, equipped with the σ -algebra generated by the Borel algebras on each copy of the unit interval. Let P be the uniform measure on $[0, 1] \times \{0\}$ and let Q be the uniform measure on $[0, 1] \times \{1\}$. It is not the case that $P \geq_{\text{r.v.}} Q$ or $P \geq_{\text{joint}} Q$.

Proof. Suppose for contradiction that there is a probability measure μ on $\mathcal{X} \times \mathcal{X}$ where $\mu_1 = P$, $\mu_2 = Q$, and μ is supported by $\Omega = \{(x, y) \in \mathcal{X} \times \mathcal{X} \mid x \geq y\}$.

The μ_1 -probability of the set of pairs $(x, 1)$ is zero, and the μ_2 -probability of the set of pairs of the form $(y, 0)$ is also zero. It follows that, letting $B \subset \Omega$ be the set of pairs $((x, 0), (y, 1))$ such that $x > y$, we have $\mu(\Omega \setminus B) = 0$. So μ is supported by B .

For each for rational $p \in [0, 1]$, let

$$I_p = \{(x, 0) \mid x \geq p\} \quad J_p = \{(y, 1) \mid y \leq p\}$$

The rectangles $I_p \times J_p$ are a countable cover of B : if $x > y$ and p is a rational number between x and y , then $(x, 0) \in I_p$ and $(y, 1) \in J_p$.

But also, $\mu(I_p \times J_p) = 0$ for every p . Whenever $x > y$, we have $x \geq p$ or $y < p$, so $(x, 0) \in I_p$ or $(y, 1) \in J_p$. Thus $B \subseteq (I_p \times \mathcal{X}) \cup (\mathcal{X} \times J_p)$. Since μ is supported by B ,

$$1 = \mu((I_p \times \mathcal{X}) \cup (\mathcal{X} \times J_p)) = \mu_1(I_p) + \mu_2(J_p) - \mu(I_p \times J_p).$$

By construction, $\mu_1(I_p) = P(I_p) = 1 - p$, and $\mu_2(J_p) = Q(J_p) = p$. So $\mu(I_p \times J_p) = 0$.

Thus $\mu(\mathcal{X}) = \mu(B) = 0$, contradicting the assumption that μ is a probability measure. \square

Acknowledgments

Thanks to Cian Dorr, Kenny Easwaran, Jeremy Goodman, Zachary Goodsell, John Hawthorne, Yoaav Isaacs, Weng Kin San, Laura Schaposnik, Christian Tarsney, Teru Thomas, Hayden Wilkinson, and two anonymous referees for very helpful discussions and comments. This work was supported by a grant from Longview Philanthropy.

Jeffrey Sanford Russell
School of Philosophy
University of Southern California
Los Angeles, California, United States
jeff.russell@usc.edu

References

- Bader, R. M. [2018]: ‘Stochastic Dominance and Opaque Sweetening’, *Australasian Journal of Philosophy*, **96**, pp. 498–507.
- Bales, A., Cohen, D. and Handfield, T. [2014]: ‘Decision Theory for Agents with Incomplete Preferences’, *Australasian Journal of Philosophy*, **92**, pp. 453–70.
- Basu, K. and Mitra, T. [2003]: ‘Aggregating Infinite Utility Streams with Intergenerational Equity: The Impossibility of Being Paretian’, *Econometrica*, **71**, pp. 1557–63.
- Broome, J. [1991]: ‘Weighing Goods: Equality, Uncertainty and Time’, Wiley-Blackwell.
- Buchak, L. [2013]: ‘Risk and Rationality’, Oxford University Press.
- Chang, R. [2002]: ‘The Possibility of Parity’, *Ethics*, **112**, pp. 659–88.
- Easwaran, K. [2014]: ‘Decision Theory Without Representation Theorems’, *Philosophers’ Imprint*, **14**.
- Fine, T. L. [2008]: ‘Evaluating the Pasadena, Altadena, and St Petersburg Gambles’, *Mind*, **117**, pp. 613–32.
- Fishburn, P. C. [1978]: ‘Stochastic Dominance Without Transitive Preferences’, *Management Science*, **24**, pp. 1268–77.
- Folland, G. B. [1999]: ‘Real Analysis: Modern Techniques and Their Applications’, Vol. 40 John Wiley & Sons.
- Hare, C. [2010]: ‘Take the Sugar’, *Analysis*, **70**, pp. 237–47.

- Joyce, J. M. [1999]: 'The Foundations of Causal Decision Theory', Cambridge University Press.
- Kuhn, H. W. [1956]: 'Solvability and Consistency for Linear Equations and Inequalities', *The American Mathematical Monthly*, **63**, pp. 217–32.
- Lauwers, L. and Vallentyne, P. [2017]: 'A Tree Can Make a Difference', *Journal of Philosophy*, **114**, pp. 33–42.
- Nover, H. and Hájek, A. [2004]: 'Vexing Expectations', *Mind*, **113**, pp. 237–49.
- Russell, J. S. [2021]: 'On Two Arguments for Fanaticism', *GPI Working Paper*,.
- Schoenfield, M. [2014]: 'Decision Making in the Face of Parity', *Philosophical Perspectives*, **28**, pp. 263–77.
- Seidenfeld, T., Schervish, M. J. and Kadane, J. B. [2009]: 'Preference for Equivalent Random Variables: A Price for Unbounded Utilities', *Journal of Mathematical Economics*, **45**, pp. 329–40.
- Tarsney, C. [2020]: 'Exceeding Expectations: Stochastic Dominance as a General Decision Theory', *Global Priorities Institute Working Papers Series*,.
- Wilkinson, H. [2022a]: 'In Defense of Fanaticism', *Ethics*, **132**, pp. 445–77.
- [2022b]: 'Infinite Aggregation and Risk', *Australasian Journal of Philosophy*, pp. 1–20.
- Willard, S. [2004]: 'General Topology', *Dover Publications*,.