Philosophers' Imprint

volume 14, no. 27
august, 2014

DECISION THEORY WITHOUT REPRESENTATION THEOREMS

Kenny Easwaran

Texas A&M University

Naive applications of decision theory often assume that it works by taking a specification of probabilities and utilities and using them to calculate the expected utilities of various acts, with a rational agent being required to take whichever act has the highest (or sufficiently high) expected utility. However, justifications of the formal framework of expected utility theory generally work in the opposite way — they start with an agent's preferences among acts, and use them to calculate an implied probability and utility function. The theory of expected utility is justified by showing that any agent who satisfies some reasonable-seeming constraints on her preferences can be represented in such a way, by postulating appropriate abstract "probability" and "utility" functions.

In section 1 I give some reasons why the view based on representation theorems is unsatisfying, and why some other potential justifications of expected utility theory won't work. In section 2 I describe an alternate way to build decision theory that is more in agreement with the naive view, taking probability and utility to define norms on preference. I show how to prove that the resulting set of preferences is consistent, without just taking that as a brute requirement, as in representation theorems. In section 3 I then develop a version of this theory beginning with a very restricted dominance principle and extending it by means of some indifference relations. I show that with some natural assumptions, the resulting theory generates all of the preferences arising from classical expected utility theory, and more. I will give some minor support for the resulting theory, but I will also show how versions can be developed that use different assumptions. This allows the modular development of decision theory, so that responding to a problem case doesn't require starting over from scratch. I don't claim to have a sufficient argument for rejecting the orthodox view, but I hope to demonstrate that there are several advantages of this alternative view.

1. Traditional Theory

1.1 Representation Theorems

The orthodox view of decision theory endorsed by Savage (1954) and Jeffrey (1965) takes preference over acts with uncertain outcomes to be the fundamental concept of decision theory, and shows that if these preferences satisfy a particular set of axioms, then they can be represented by a probability function and a utility function. This "representation theorem" involves showing that the expected utilities calculated from the probability and utility function correspond to the preferences that the agent actually has. Importantly, on standard interpretations of these results, the probabilities and utilities are not themselves fundamentally real, but are mere mathematical constructs out of the preferences.

This conflicts with a naive reading of the concept of expected utility, which was perhaps the dominant understanding of theories that were popular in the 17th to 19th centuries. One often assumes that utilities and probabilities are prior to preference, and that decision theory says that you *should* prefer an act with a higher expected utility over any act with a lower expected utility. And this is how the theory of expected utility is often applied in practical contexts. However, according to the orthodox theory, there is no normative requirement, but rather a conceptual necessity! *What it is* to have probabilities and utilities *just is* to have preferences that line up with the resulting expected utilities. If an actual agent's preferences don't line up with the calculations, then the preferences and utilities one is calculating with just aren't those of the agent. (A similar point is made by Meacham and Weisberg (2011).)

A natural thought is that one might step back from taking on the full set of axioms as a normative theory for agents. Rather, one might allow that actual agents fall short of full rationality. Perhaps some further sort of representation might still be applied to calculate probabilities and utilities that provide a good (though not perfect) representation of the agent's preferences. If this can be done, then calculating expected utilities according to these probabilities and utilities might be able to provide some normative guidance for an agent to fix her other preferences so that they better approach ideal rationality. But it's far from obvious that it will be possible to divide the agent's preferences in this way, so that a subset is sufficient to define the probabilities and utilities, and even if this can be done, it's not obvious why *these* probabilities and utilities should be normative on the others, rather than vice versa.

Regardless of the prospects for this sort of roundabout use of the representation theorems, there have been many challenges to the axioms on preference that are required for the theory to get off the ground. A particular interpretation of McGee (1999) shows that these assumptions (or, at least, slightly generalized versions of them) mean that utilities cannot be unbounded if there are arbitrarily small positive probabilities. Thus, potential paradoxes like St. Petersburg, Two Envelope, and Pasadena (Nover and Hájek, 2004) just can't arise. But as Nover and Hájek argue, the theory gives no reason for any particular value to be an upper bound on utility — it just requires that some ad hoc bound exist to block the paradoxes. Traditional expected utility theory allows these cases to arise, though it doesn't say what agents should prefer (Fine, 2008). However, representation theorems just rule them out of consideration entirely, because they violate conditions of continuity and the like. I intend to give a theory that can allow for them and also give some guidance on what to prefer, though my theory won't settle all questions involving them.

Another problem for this use of representation theorems is the source of the normativity involved. On the standard representation theorems, it appears that the norms on belief arise out of the role of belief in governing action. However, many philosophers have argued that the fundamental norms on belief must arise out of the aim of seeking the truth (Shah, 2003; Wedgwood, 2002). Although it has been argued that decision-theoretic justifications of probabilism may not be essentially pragmatic, and may use preferences just as an example of a mental attitude in which non-probabilistic agents must be inconsistent (Christensen, 1996), it's still hard to see how an inconsistency of preferences connects to the fundamental aim of closeness to the truth.

Joyce (1998) uses a notion of "accuracy", or closeness to the truth, to give an argument for probabilism in one's degrees of belief, but gives no connection between this probability function and one's actions. Representation theorems give no indication that the probability function constructed out of preference must correspond to one's degrees of belief. (This is the assumption called "Representation Accuracy" criticized by Meacham and Weisberg (2011, p. 655).) Instead, one would like to start with Joyce's totally non-pragmatic probability function, and use it to generate a norm of preferring acts on the basis of their expected utility, which can then explain why one's preferences ought to satisfy the axioms of a representation theorem. This is especially important for "expected accuracy" arguments (exemplified by Greaves and Wallace (2006), Easwaran (2013), and Leitgeb and Pettigrew (2010)), which rely on Joyce's foundations, and then seek to evaluate "epistemic acts" by means of expected accuracy. For these arguments, a justification of expected value calculations is central to the purely epistemic program, and not just (as for Joyce) a connection between it and a theory of pragmatic value. One goal of this paper is to show that these arguments can be carried out while taking accuracy and degree of belief to both be prior to expected value.

Furthermore, there are paradoxes like those of Allais, Ellsberg, and the like to which expected utility theory can't give the intuitively correct answers. If these intuitive answers are correct, then the axioms involved in the representation theorems can't be. There have been some theories (Buchak, 2013) that give representation theorems for decision theories yielding the intuitive preferences in some of these cases, but they involve reconstructing the decision theory from the axioms of preference up. In constructing my theory, I will show how various alternative choices along the way can at least accommodate the unorthodox preferences, and I will hint at extensions of my theory that may even be able to generate these preferences. (Ellsberg is discussed to some extent in section 3.3.2, while Allais is dealt with more thoroughly in section 3.5.5.) I will not aim to fully justify one alternative or another, but hopefully this will show how the theory can be modified without starting over from the beginning.

1.2 *Laws of Large Numbers*

An alternative justification of expected utility theory that takes probability and utility as fundamental is sometimes given in terms of the laws of large numbers. These are theorems of probability theory that show that for a long enough sequence of independent acts, the sum of their actual utilities is very likely to be extremely close to the sum of the expected utilities. Thus, if one makes a series of choices of lower expected utility, then one will very likely end up in a worse situation than if one had made a series of choices of higher expected utility. This long-run behavior is supposed to then justify each individual choice. (For an example of an argument using the weak law of large numbers rather than the strong law of large numbers, see Easwaran (2008).)

However, this justification is not very often endorsed by philosophers, because of some major problems. First, there is the obvious problem that it seeks to justify each individual choice on the basis that it is part of a sequence of choices that is good overall. While it is plausible that a sequence of choices that are individually good must itself be a good sequence of choices, there is no reason to suppose that the converse holds — a sequence of choices that is overall good may contain some individual choices that are themselves bad. Rule utilitarianism and Kantian ethics make parallel arguments — individual acts are said to be good if they are part of an overall rule or policy or maxim that is a good one. But this is a conceptual problem they all share, and many criticisms of these views focus on the odd recommendations they give in individual cases — most philosophers find at least some of these recommendations to be bad, and take this as a reductio of the theory, rather than accepting the principle that an individual act is good if it is part of a sequence that is good.

Second, the laws of large numbers tell us about the limiting behavior of infinitely long sequences of choices, but don't necessarily give guidance about how things turn out when only finitely many choices are actually made. If an act has positive probability both of gaining and of losing utility for the agent, then the same will be true for any large finite sum of independent copies of this act. One can't get certainty out of finitely many independent copies of uncertain choices. And although the repeated sequence of copies of the act with highest expected utility will have the highest probability of giving rise to the best overall outcome, the repeated sequence of copies of the act with the individual best possible outcome will itself have the individual best possible outcome, and the repeated sequence of copies of the act that minimizes potential loss will itself minimize the potential loss. Some principle is still needed to say why high probability of a good outcome is better than a possibility of the best possible outcome, or than guaranteeing avoidance of the worst possible outcome.

One can't merely appeal to a principle that says that one should prefer one series of acts to another if it is sufficiently likely that the first gives a better result than the second, because this principle actually contradicts expected utility in cases where a gamble is extremely likely to yield a minorly good outcome but has a small chance of giving an extremely bad one. One seems to need a principle that weighs the probability of the bad outcome and the relative magnitudes of the goodness or badness of other outcomes, which appears to be tantamount to assuming the expected utility principle itself.

This second worry is most pressing when a choice is essentially nonrepeatable, and when one knows that there are only a limited number of further decisions that one will make in one's life. But it applies even when the sequence is large but known not to be infinite. One can't get certainty out of the law of large numbers, as long as the large numbers are finite. Without certainty, we are back where we started at trying to evaluate a sequence of acts whose outcome could be either good or bad, but with the additional problem of trying to connect our decision about the sequence of acts to the decision about the single act.

1.3 *The Goal of Decision Theory*

Expected utility theory aims to evaluate acts by assigning a real number, known as the "expected utility", to each one. However, at least some of the cases that motivate alternatives to representation theorems involve acts that could have infinite value. Furthermore, there are some situations in which it appears to be impossible to represent ordinal preferences correctly with just the real numbers.[1] It's conceivable that there might be some numerical valuation that uses a set of numbers far larger than the standard real numbers, like Abraham Robinson's "hyperreals", or John Conway's "surreals". But I think this numerical representation is inessential.

Instead, I will just aim to provide an ordering that represents the normative preferences for a set of acts. If probabilities and utilities are well-defined, then they will play a role in determining these norms.

1. For instance, consider the following example:

> One is entering heaven, and knows that each day will be finitely good, but there are infinitely many days. One wants to choose how good each day should be. We might represent this by saying that each day corresponds to a natural number, and the happiness of that day is also represented by a natural number. Thus, the choices can be represented by sequences of natural numbers. It is plausible to say that if one sequence is strictly higher than another on all but finitely many of the terms, then it is better.
>
> However, this preference ordering can't be represented by the standard ordering on the real numbers, no matter how one assigns real numbers to these sequences.

There is clearly no best sequence, so the set of real numbers that represents the values must not have a highest element. But for any non-empty set S of real numbers with no highest element, there is a countable subset C that "exhausts" the set, in the sense that for every member of S, there is a member of C that is higher than it. However, no countable subset of the sequences of natural numbers exhausts this set — one can always construct a sequence of natural numbers that is higher than every member of some countable set by letting the nth member of the sequence be higher than any of the first n members of any of the first n sequences. Thus, no set of real numbers can have the same ordinal structure as these sequences.

But certain alternatives will allow versions of my theory to work even without having well-defined numerical probabilities and utilities.

A project like mine has been worked on by Martin Peterson (2002; 2004). He aims to justify something like standard expected utility theory, starting with probability and utility. However, instead of constructing the whole ordering, he only aims to say which options are best. Part of this is because he aims to work only with sets of acts that could be the available choices for an agent, while I will aim to give preferences over very rich sets that include many options that are not actually available. But even among the available options, I think it is important to figure out which non-ideal options are better or worse than others. Some ways of being irrational are more irrational than others. (This is especially true if some sort of "satisficing" view of choiceworthiness is the right one.) Additionally, I will show how my ordering can be modified to give alternatives to expected utility theory, in addition to showing how it can give rise to that theory.

The ordering I give will not be a total ordering. That is, there will be some pairs of acts such that my theory says neither that the agent should prefer one to the other nor that the agent should be indifferent between them. There are two potential sources of this incompleteness.

One source of incompleteness is that there may in fact be pairs of acts that are incomparable. This could be because they give rise to outcomes that are incomparable; an agent might radically prefer a life in the army in some ways and a life in the priesthood in other ways, but they may be so different that there is no way to make an overall comparison. Or it could be because the outcomes are all measurable in a standard way, but the pattern of probabilities and utilities makes it impossible for there to be such a thing as the overall value, as one might think for Nover and Hájek's "Pasadena game" and other related gambles.

Another source is that this incompleteness is an essential part of the intermediate steps of my construction, and I don't claim to have reached the final step of the construction so far. I start with an ordering that is very incomplete and gradually show how to fill it in, piece by piece. For the intermediate orderings in the construction, much of the incompleteness is just a way to leave things open for further development of the theory. The version of the theory I reach by the end likely has some of the same incompleteness — there's no clear reason to think that it includes all rationally required preferences. Cases where the resulting ordering is incomplete could be cases that will eventually be determined by some further addition to the theory, or they could be cases of essential incomparability. It would be interesting to pursue a development of the theory that could in fact naturally give rise to a complete ordering, but I don't yet see a good way to do so.

However, the important point is that judgments made by some intermediate version of the theory will never be reversed in later extensions. And at each stage, I will check that the resulting theory is consistent and not merely intuitively plausible. Thus, even an intermediate theory can be useful, since its judgments will also be endorsed by the final theory. But I take no stand on whether the incomparabilities that remain at the end are essential, or whether they too can be decided by further extensions of the theory.

2. Structure of a New Theory

In this section, I will outline the way a theory of my sort can be developed without representation theorems. In section 3 I will then give an extended development of a theory of this type, showing that it can generate all the preferences given by expected utility calculations, and more. But the general framework of the theory could also be developed in a different way by theorists who endorse some alternative role for credence and utility other than maximizing expected utility. (Most such theorists, like Buchak (2013), are motivated by considerations of risk-aversion, but there could conceivably be alternative reasons to pursue such a theory.) The subsections of this section give the mathematical formalism behind the characterization given here.

The central observation in the development of my theory is that if one ought to prefer act A to act B, and one ought to be indifferent between acts B and C, then one ought to prefer A to C. Thus, the way a theory of my sort is built up is to start with a basic relation that characterizes cases where a rational agent ought to prefer one act to another, and then find further relations that characterize cases where a rational agent ought to be indifferent between acts. Using these relations, one can then characterize a wider set of cases in which agents ought to prefer one act to another.

For technical reasons, it will be useful to start with a pair of preference relations, $\succ$ and $\succeq$, with the former indicating cases where an agent is rationally required to strictly prefer one act to another, and the latter indicating cases where an agent is rationally required to find one act at least as good as another. There will also be an indifference relation $\approx$, which can be defined by saying that $A \approx B$ iff $A \succeq B$ and $B \succeq A$. Note that I will *not* assume that $A \succeq B$ iff $A \succ B$ or $A \approx B$ — which of the two possibilities holds may be undecided at the beginning of the process and settled only later.

The essential thing to make sure that the resulting theory is consistent is to make sure that if one is required to strictly prefer one act to another, then one is never required at a later stage to be indifferent between the acts or prefer them in the reverse order. This is the condition of "compatibility" that will be formally defined in section 2.2. Compatibility will ensure that the theory is consistent. Axiomatic decision theories with representation theorems just take the consistency and transitivity of preference as a primitive assumption that is necessary in order to be represented by a probability and a utility, while my theory will show how the existence of probability and utility guarantees that the resulting norms on preference are consistent and transitive.

To simplify both the checking of compatibility and the definition of the extended preference relations from the basic preference and indifference relations, I define a notion of "commutativity" in section 2.2. The idea is that if one starts with an arbitrary preference relation $\succ$ and indifference relation $\approx$, then to find out that one is required to prefer act A to A', one might need to note a long sequence of intervening relations. For instance, perhaps $A \succ B \approx B' \succ B'' \approx B''' \succ A'$. It can be quite difficult to make sure that no long chain like this shows that some new preference is required that is incompatible with old preferences. If two relations are commutative, however, any chain like this can be compressed into a single preference followed by a single indifference. Although there's no reason to think that every normatively required preference relation should commute with every normatively required indifference relation, it will turn out that all the relations I consider do commute, at least when a suitably large collection of acts are considered. Thus, because this greatly simplifies the arguments, my mathematical framework will require that the relevant relations always commute with each other. To guarantee this, I need the set of acts under consideration to be suitably rich. However, my arguments require only the *norms* governing an agent's preferences to apply to this large set of acts, and they don't require the agents to actually ever consider such acts, much less to have such acts available.

The remainder of this section develops the mathematical tools needed to give a version of the theory and prove that it is consistent. Some of the details are inessential to understand the general ideas of section 3, but they will be needed if one wants to track the details. In section 2.1, I define the basic features that preference and indifference relations have, as well as some features that other related relations may have. In section 2.2 I then give the formal definition of what it takes for two relations to commute, and show how commutativity enables a simple definition of the preference relation implied by an initial preference relation and an indifference relation. I also give the formal definition of what it takes for an indifference relation to be compatible with a preference relation. In section 2.3 I then show how these features for pairs of relations can be generalized to larger sets, and show how this can be used to build up a full decision theory from a very sparse initial preference relation and a sequence of indifference relations. Section 3 then develops a version of decision theory in this way.

2.1 Orderings and equivalence relations

The symbol I will use for an abstract binary relation is '$\sim$', possibly with a subscript, though I will use other symbols for specific types of relations. For each such relation, there will be some set of objects called its *domain* that are the possible relata. (For the preference and indifference relations, the domain will be acts, but I will discuss some relations on other domains as well.) I won't ask of two objects whether they bear some relation or not unless both objects are members of the domain of that relation.

I will use the following standard definitions (in all cases, the quantifiers range only over the domain of the relevant relation):

- $\sim$ is *reflexive* iff for all A, $A \sim A$.
- $\sim$ is *irreflexive* iff there is *no* A such that $A \sim A$.
- $\sim$ is *symmetric* iff for all A and B, $A \sim B$ iff $B \sim A$.
- $\sim$ is *anti-symmetric* iff there are no A and B such that $A \sim B$ and $B \sim A$.
- $\sim$ is *transitive* iff for all A, B, and C, if $A \sim B$ and $B \sim C$ then $A \sim C$.

A *weak ordering* (which I will generally indicate with the symbol '$\succeq$', often with a subscript) is any relation that is reflexive and transitive.[2] A *strict ordering* (which I will generally indicate with the symbol '$\succ$', often with a subscript) is a relation that is irreflexive and transitive (and is thus also anti-symmetric). I will sometimes emphasize that an ordering is *partial* when there are some pair A, B of elements of the domain such that neither bears the ordering to the other. Conversely, I will say that an ordering is *total* or *linear* if for all $A \neq B$, either $A \succ B$ or $B \succ A$ (and similarly for $\succeq$). An *equivalence relation* (which I will generally indicate with the symbol '$\approx$', often with a subscript) is a relation that is reflexive, transitive, and symmetric.

For two relations $\sim_X$ and $\sim_Y$, I will say that $\sim_Y$ *extends* $\sim_X$ iff for every A and B, if $A \sim_X B$ then $A \sim_Y B$. When one relation in a set of relations extends all relations in that set, I will say that it is the *largest* in that set, and when all relations in a set extend a single one, I will say that the relation is the *smallest* in that set. As an example, if $\succeq$ is a weak ordering, and we define $A \approx B$ iff $A \succeq B$ and $B \succeq A$, then we can see that $\approx$ is the largest equivalence relation that $\succeq$ extends.

2. I don't assume any sort of anti-symmetry for weak orderings. That is, there may be distinct A and B such that $A \succeq B$ and $B \succeq A$. Some mathematicians use the term "pre-order" in this case, and reserve the word "order" for a relation such that $A \succeq B$ and $B \succeq A$ entails that $A = B$, but I will not.

2.2 *Composition, commutativity, and compatibility of pairs of relations*

Consider two relations $\sim_X$ and $\sim_Y$ on the same domain. I will define their *composition* $\sim_{XY}$ by saying that $A_1 \sim_{XY} B_2$ iff there is A_2 such that $A_1 \sim_X A_2$ and $A_2 \sim_Y B_2$. Similarly, their composition in the other direction is defined by saying that $A_1 \sim_{YX} B_2$ iff there is B_1 such that $A_1 \sim_Y B_1$ and $A_2 \sim_X B_2$. (The choice of particular letters and subscripts for these relata and others later is meant to assist in remembering the asserted relations between them, but is not meant to entail anything about them.)

I will say that $\sim_X$ and $\sim_Y$ *commute* with each other iff $\sim_{XY}$ is the same relation as $\sim_{YX}$. That is, they commute iff:

- For any A_1 and B_2, there exists an A_2 such that $A_1 \sim_X A_2$ and $A_2 \sim_Y B_2$ iff there exists a B_1 such that $A_1 \sim_Y B_1$ and $B_1 \sim_X B_2$.

Theorem 1. If $\sim_X$ and $\sim_Y$ commute, and both are transitive, then their composition is transitive too.

Theorem 2. If $\sim_X$ and $\sim_Y$ are both reflexive, then their composition is reflexive too.

Theorem 3. If $\sim_X$ and $\sim_Y$ commute, and both are symmetric, then their composition is symmetric too.

(For these and all further theorems, the proofs are in the appendix.)

Given these results, we can say a lot about the composition of various relations that commute. The composition of two commuting weak orderings is a weak ordering, the composition of two commuting equivalence relations is an equivalence relation, and the composition of a weak ordering with an equivalence relation it commutes with is a weak ordering. It is straightforward to see that the composition of two equivalence relations that commute is the smallest equivalence relation that extends both. Thus, since it is plausible that indifference between acts is required to be an equivalence relation, if we can find two indifference relations that are both normatively required, then their composition will be as well.

For two relations $\sim_X$ and $\sim_Y$, I will say that $\sim_X$ is *transitive over* $\sim_Y$ iff the following two conditions hold:

- If $A \sim_X B$ and $B \sim_Y C$, then $A \sim_X C$.
- If $A \sim_Y B$ and $B \sim_X C$, then $A \sim_X C$.

It is straightforward to see that if a weak ordering and an equivalence relation commute, then their composition is the smallest weak ordering that extends the original weak ordering and is transitive over the equivalence relation. Thus, since it is plausible that weak preference between acts is required to be transitive over indifference, if we can find a weak preference relation and an indifference relation that are both normatively required, then their composition will be as well.

Note that there are further issues with strict orderings. If $\sim_X$ and $\sim_Y$ are strict orderings that commute, their composition $\sim_{XY}$ is not necessarily a strict ordering. It is true that $\sim_{XY}$ is transitive, but it is not clear whether it is irreflexive, as needed for being a strict ordering. For $\sim_{XY}$ to be irreflexive, meaning there is no A with $A \sim_{XY} A$, is just for there not to be A and B with $A \sim_X B$ and $B \sim_Y A$.

Thus, I will say that $\sim_X$ and $\sim_Y$ are *compatible* iff there are no A and B such that $A \sim_X B$ and $B \sim_Y A$, since this is precisely the condition that must hold between two commuting transitive relations for their composition to be a strict ordering. I will normally talk about this condition only in cases where at least one of $\sim_X$ or $\sim_Y$ is itself a strict ordering.

It is straightforward to see that if a strict ordering and an equivalence relation commute, then their composition is the smallest strict ordering that extends the original and is transitive over the equivalence relation. Since it is plausible that strict preference must be transitive over indifference, if we can find normatively required strict preference and indifference relations, then their composition will itself yield a normatively required set of strict preferences.

Note that none of this suggests that individual relations that are normatively required will generally commute with each other. All we can say so far is that discussion of what is normatively required is much simpler when the relations do commute. As it turns out, the relations I will discuss in section 3 do in fact commute, but if the set of acts I consider didn't have a certain richness to it, then they would not. However, as long as the relations we consider are interpreted as saying that the agent *ought* to prefer one act to another, then it won't matter whether the agent *actually does* have preferences over this rich set of acts, as long as there are norms that apply to the agent's preferences among these acts.

2.3 *Interactions of multiple relations*

The theory I build up will be constructed in multiple steps. Thus, we won't always be checking commutativity, compatibility, and composition of pairs of relations, but will sometimes need to check them for larger sets of relations as well.

Theorem 4. If $\sim_X$, $\sim_Y$, and $\sim_Z$ are three relations, and $\sim_{XY}$ is the composition of $\sim_X$ with $\sim_Y$, and $\sim_{YZ}$ is the composition of $\sim_Y$ with $\sim_Z$, then the composition $\sim_{(XY)Z}$ of $\sim_{XY}$ with $\sim_Z$ is the same as the composition $\sim_{X(YZ)}$ of $\sim_X$ with $\sim_{YZ}$.

Thus, if $\sim_X$, $\sim_Y$, and $\sim_Z$ commute with each other pairwise, then the composition of the three relations is uniquely defined, regardless of the order in which they are combined, and this composition can be referred to as $\sim_{XYZ}$, or $\sim_{ZXY}$, or any other permutation of the three. Similar considerations apply to larger sets of relations that all commute with each other pairwise.

If $\succ_X$ is a strict ordering, and $\sim_Y$ and $\sim_Z$ are both transitive, and all three relations commute with each other, then previous results show that their composition is a strict ordering if $\succ_X$ is compatible with $\sim_{YZ}$, or if $\succ_{XY}$ is compatible with $\sim_Z$, or if $\succ_{XZ}$ is compatible with $\sim_Y$. That is, there must be no A, B, C with $A \succ_X B \sim_Y C \sim_Z A$. In such a case, I will say that the relations are *mutually compatible*. Because of commutativity, if there are no such A, B, C bearing these relations, then no triplet bears these three relations in a different order either. Thus, proving the non-existence of such a triplet with one ordering of the relations is sufficient for proving mutual compatibility. The same holds for compositions of larger sets of relations, and I will call such larger sets mutually compatible in the parallel case.

If $\succ_X$ is a strict ordering, and $\sim_Y$ and $\sim_Z$ are both equivalence relations, then the set of all three is mutually compatible only if both $\sim_Y$ and $\sim_Z$ are compatible with $\succ_X$. However, this is not sufficient (even in the presence of commutativity) for the set to be mutually compatible. (See the appendix for a counterexample.) Thus, although commutativity can be checked pairwise for any set of relations, compatibility must be checked for the set as a whole.

3. One Version of the Theory

The remainder of this paper presents a version of decision theory that works according to the framework defined above. I think that the relations presented here are normative for decision theory, but if some of them are not, then they can be switched out for others that might do some of the same work. Thus, this final section is primarily for purposes of giving a worked-out example of the decision theory, showing that it can be used to extend standard expected utility theory. As mentioned earlier, I think this is particularly important in justifying the calculations made by expected-accuracy theorists, who often want to define probability in a way that is prior to decision theory. But I think it is also of interest to see how the resulting theory relates acts that can't be related by expected utility theory itself. I also discuss several weakenings of the assumptions that might allow generalization of the theory.

In section 3.1, I define the notion of "act" over which preferences will be defined. In particular, I assume that the agent has a basic preference relation over possible "outcomes" of acts, and I represent acts by arbitrary functions from "states" to outcomes. However, I don't assume that acts are defined over the same sets of states — different acts may create different amounts of indeterminacy in the world that could play a role in determining their different outcomes.

In section 3.2, I then describe the basic preference relations $\succ$ and $\succeq$ that the theory will be based on. These relations are relations of dominance, in which one act gives a better outcome than another in every state. These relations are defined only for acts defined over the same set of states, so they are quite limited indeed. I contrast the two specific relations $\succ$ and $\succeq$ with a relation $\succ'$ that one might expect to be normative, but will turn out in later subsections to have some problematic interactions with various indifference relations. Clarifying the status of this relation, and its interaction with various proposed indifference relations, will be a major topic for future research.

In section 3.3, I introduce the first basic indifference relation, which says that one should be indifferent between two acts not only if they give the same outcomes in the same states, but if there is an "appropriate" correspondence between the states the two acts are defined on such that they give the same outcome in corresponding states. In section 3.4, I then introduce various amounts of comparative and numerical probability theory, and show how they can be used to sharpen up the notion of appropriateness here. I show that the resulting indifference relation is compatible with dominance under certain plausible assumptions (if either the probability function has no "null" sets other than the empty set, or the probability function is countably additive and the set of outcomes is "separable"). I also show that (assuming a sort of "homogeneity" of the probability function) the resulting indifference relation, when combined with the basic dominance relations, gives all the preferences of "stochastic dominance".

In section 3.5, I introduce a notion of "differences" among outcomes, and use this to define a new indifference relation between acts. In particular, two acts will be equivalent if the set of states in which the first is better has the same probability as the set of states in which the second is better, and if the differences between the outcomes in corresponding states are all the same. The idea is that the better outcomes of one act can serve as a "trade-off" for the worse outcomes that it gives. I also show various results about the interaction of this trade-off equivalence with the other relations mentioned earlier. In particular, I show that trade-off equivalence is compatible with the dominance relations, assuming that outcomes are linearly ordered and that probabilities are linearly ordered and additive (though not necessarily numerical). I also show that if probabilities and outcomes are numerical, then trade-off equivalence and dominance together give all preferences between acts of finite expected value on a single state space. Finally, I show that under these conditions, the resulting relation is compatible with the indifference relation given by permutations of states that respect the probability function, which allows the overall preference relation to account for preferences among acts of finite expected value on different state spaces, and also for many intuitive comparisons of acts whose expected utilities are infinite or undefined. In section 3.5.5, I also suggest that a modified version of my theory may be able to accommodate the phenomenon of risk-sensitivity, which representation theorems must be completely redone to do.

Thus, I show that, starting with a very restricted preference relation, a few intuitively reasonable indifferences can give rise to full expected utility theory under some natural conditions. But I also show that the theory built from these basic preferences and indifferences can function without the full strength of these structural assumptions, unlike the theory built from representation theorems. The set of acts over which these relations are defined is quite rich, and plausibly includes many acts that agents never in fact consider. However, if we interpret these preference and indifference relations as ones that the agent ought to have, then we don't need to assume that agents in fact have these preferences, as we would if we were basing a representation theorem on these preferences.

3.1 Background for Decision Theory

I adopt the basic framework of Savage (1954) rather than Jeffrey (1965) — I assume that an act is a function that assigns some "outcome" to each "state of the world". Additionally, every such function counts as an act. (In section 3.4.1 I make a slight restriction on this full combinatorial space of acts, but it is still quite large.) Most such acts are not available for an agent to perform in any situation, but there is still some sense to be made of them as hypothetical acts. For Savage and Jeffrey, it is important to idealize and assume that agents actually *have* preferences over all these acts. However, for my theory, it is important just that an agent's credences and values for outcomes implicitly define preferences that she *ought* to have over all these acts if she considers them. (See Dreier (1996) for more on hypothetical preferences.)

Unlike standard decision theorists, I do not assume that there is a single set $\mathcal{S}$ of "states of the world" that all acts are defined on. Instead, I allow that different acts might depend on distinct sets of states. This is because states are supposed to represent the features of the world that determine what the outcome of a given act is, and in cases of indeterminism, some of those relevant features may themselves depend on the agent's act. For instance, if one act involves flipping a coin, and another one doesn't, then the former act is going to have states in which the coin comes up heads and other states in which the coin comes up tails. Perhaps the agent thinks that there are objective features of the world that causally determine how the coin *would have* come up, even if the coin isn't actually flipped, but this seems like a strong metaphysical commitment.[3] A better solution is to allow that the acts have different state spaces as their domains.[4]

3. An interesting version of a theory like this has been developed by Stefánsson and Bradley (2015).

4. This allowance also means that I don't automatically give a potentially problematic treatment of the Newcomb problem. If one-boxing and two-boxing are acts defined on the same state space, with a given probability distribution, then there is no way for the states to have different probabilities for the two acts. Jeffrey allows this problem by using possible worlds instead of states, and identifying both acts and outcomes with sets of worlds. Savage does not consider

For a given act A, $\mathcal{S}_A$ is the set of states that A is defined on. I assume that these state spaces for distinct acts are either identical or completely disjoint. That is, for any two acts, either every state that could determine the outcome of one could determine the outcome of the other, or there is no state that could determine the outcome of both.

I assume, however, that there is a single set $\mathcal{O}$ of outcomes. They represent everything that an agent values in the world. That is, it shouldn't matter to the agent which act gives rise to a given outcome, if they each really give rise to the same outcome. If there are two ways to get the same outcome that an agent has a preference between, then the theorist has not made a fine enough distinction among outcomes.

The outcomes come with a strict ordering $>_v$ (the v stands for "value") representing the agent's preference of one outcome over another, and an equivalence relation $\equiv_v$ representing the agent's indifference between two outcomes. I will assume that the relations are compatible, and that $>_v$ is transitive over $\equiv_v$, and define $\geq_v$ as the relation that holds whenever either $>_v$ or $\equiv_v$ does. At some points (which I will mention explicitly), it will be useful to assume that these relations give a *linear* ordering, which means that for every pair of outcomes o_1, o_2, either $o_1 >_v o_2$, $o_1 \equiv_v o_2$, or $o_2 >_v o_1$. I will sometimes replace talk of outcomes with talk of "utilities", which I will just assume are outcomes whose value ordering has a natural representation by the standard ordering on the real numbers. But most of the time I will just work with the general framework of partially ordered outcomes.

the Newcomb problem, and in fact appears to rule it out entirely. He implicitly assumes throughout that acts play no role in causally determining which state of the world is actual. And his remarks on "grand world" and "small world" decision problems in section 5 of Chapter 5 seem to imply that states, and the values of outcomes, must be probabilistically independent of acts as well. But in the Newcomb problem, there is no identification of states that is both probabilistically and causally independent of acts. I want to preserve Savage's conceptual distinctions while not automatically ruling out cases like this. This will be discussed further in section 3.3.2, though I won't attempt to fully address the problem.

3.2 *Dominance*

Since an act is a function from states to outcomes, we can use these relations on outcomes to define relations on acts. The basic idea is that preferences among acts should be governed by the values of the outcomes they can produce. If one act is guaranteed to give a better outcome than another, then one ought to prefer the former. Thus, we can define:

$$A \succ B \text{ iff } \mathcal{S}_A = \mathcal{S}_B \text{ and } (\forall s \in \mathcal{S}_A)(A(s) >_v B(s)),$$

$$A \succeq B \text{ iff } \mathcal{S}_A = \mathcal{S}_B \text{ and } (\forall s \in \mathcal{S}_A)(A(s) \geq_v B(s)),$$

$$A \approx B \text{ iff } \mathcal{S}_A = \mathcal{S}_B \text{ and } (\forall s \in \mathcal{S}_A)(A(s) \equiv_v B(s)).$$

I will use these as my basic relations. It is straightforward to see that $\succ$ is transitive over $\succeq$ and $\approx$, and that $\succeq$ is transitive over $\approx$. Thus, these can form the core of a decision theory if they are supplemented with a variety of equivalence relations that are compatible with each other, and with all of these relations.

One can see that $A \approx B$ iff $A \succeq B$ and $B \succeq A$. However, it is not the case that $A \succ B$ iff $A \succeq B$ and $B \not\succeq A$ — if A gives strictly better outcomes than B in some states and exactly equal outcomes in others, then $A \succeq B$ and $B \not\succeq A$, but $A \not\succ B$. Thus, one might define:

$$A \succ' B \text{ iff } A \succeq B \text{ and } B \not\succeq A,$$

or equivalently

$$A \succ' B \text{ iff } \mathcal{S}_A = \mathcal{S}_B \text{ and } (\forall s \in \mathcal{S}_A)(A(s) \geq_v B(s)) \text{ and } (\exists s \in \mathcal{S}_A)(A(s) >_v B(s)).$$

However, I will show that although $\succ$ is compatible with each indifference relation I will mention, $\succ'$ is compatible with only some of them. Thus, I think we should focus on $\succ$ rather than $\succ'$ as the basis of decision theory. For now, however, I introduce both relations just to help clarify what $\succ$ does and doesn't mean.

These relations are quite restricted, so that they don't let us compare many pairs of acts. For instance, an act that yields \$10 if a Democrat is elected president and \$30 if a Democrat is not elected can't be compared to an act that yields \$20 in either case. However, the most basic preferences can be accounted for. If you prefer apples to oranges, and oranges to nothing, then an act that gives you an apple if a black ball is drawn from an urn and an orange otherwise should be preferred to an act that gives you an orange if a black ball is drawn from an urn and nothing otherwise. These preferences are plausible regardless of whether or not there is any further structure to the values of the outcomes and the probabilities of the states.

3.3 Permutations and Correspondences

These existing relations so far only compare acts that are defined over the same state space. If two acts have different state spaces, then there is so far no way to compare them. An act whose outcome depends on the result of a coin flip can't be compared to an act whose outcome depends on a different indeterministic process. The goal of this subsection is to give a mathematical formalism for connecting different state spaces, and show conditions under which the resulting equivalence relation is compatible with and commutes with all the orderings given so far. But the resulting relation is also able to introduce new comparisons among acts on a single state space.

The basic idea is that even though two acts may be defined on different state spaces, there may be an important sense in which states in one space naturally "correspond" to states in another space. If act A involves flipping a fair coin, and act B involves rolling a six-sided die and seeing whether the face that comes up is odd or even, then there is a sense in which one can say that heads corresponds to odd and tails to even, or vice versa. This can then give a sense in which certain acts are equivalent to others, if they give the same outcomes in corresponding states, even though they depend on different states. This idea is suggested by Schick (2003), though I am generalizing it quite a bit.

Section 3.3.1 gives the formal requirements on any such relation, and shows that any such relation is compatible with the dominance relations. Section 3.3.2 shows that appropriate understanding of the notion of "correspondence" can allow for compatibility with intuitive preferences in some cases that expected utility has a hard time with. But full development of the idea will wait until section 3.4, where I introduce probability theory.

3.3.1 Formal definitions

A *correspondence* of one set to another is a one-to-one, total function from the first set onto the other. The first set is called the *domain* of the correspondence, and the second is called the *range*. That is, a correspondence of S_1 to S_2 is a function f that is defined for all elements of S_1, such that for every s_2 in S_2, there is a unique s_1 in S_1 such that $f(s_1) = s_2$. We can denote this unique element s_1 as $f^{-1}(s_2)$, and it is straightforward to check that in this case, f^{-1} is itself a function, and in fact is a correspondence of S_2 to S_1. f^{-1} is known as the *inverse* of f. If f is a correspondence from S_1 to S_2, and g is a correspondence from S_2 to S_3, then there is a correspondence gf from S_1 to S_3 defined by $gf(s) = g(f(s))$, and gf is known as the *composition* of f and g. Of course, the composition of two correspondences is defined only if the domain of g is the range of f.

A correspondence of a set to itself is called a *permutation*. For any set S there is a trivial permutation id_S called the "identity", which has the definition that for all s in S, $id_S(s) = s$. id_S has the special feature that it is its own inverse. If S has more than one element, then there are of course plenty of other permutations of a set as well, and some of them even have the feature of being their own inverse. If f is any correspondence from S to any set, then the composition $f^{-1}f$ just is

id_S.

When talking about multiple acts that are defined on multiple different sets of states, we will be interested in a whole collection of correspondences between the various state spaces. Consider a collection of sets, and a set G of correspondences among them. We say that G is a *groupoid* iff it satisfies the following three conditions:

- For each set S in the collection, id_S is in G.
- For each correspondence f that is in G, f^{-1} is also in G.
- For any two correspondences f and g in G, if the domain of g is the range of f, then their composition gf is also in G.

(For those that are familiar with the mathematical notion of a *group*, it is just the abstract characterization of a groupoid where all correspondences are in fact permutations of a single set, though many presentations of group theory leave out any mention of this set and focus instead on the compositions of correspondences.)

Let G be any groupoid on the set of state spaces for all acts. We can then say that $A \approx_G B$ iff there is some g in G whose domain is $\mathcal{S}_A$ and whose range is $\mathcal{S}_B$, such that for all s in $\mathcal{S}_A$, $A(s) = B(g(s))$. The idea is that if G is chosen appropriately, one should be indifferent between A and B, because they give the same outcome in corresponding circumstances. (For most choices of G, $\approx_G$ will not be plausible as an actual normative requirement, but for the mathematical development at this point this is not essential. Conditions under which $\approx_G$ is more plausible will be developed in section 3.4.) It is straightforward to see that $\approx_G$ is an equivalence relation.

Theorem 5. If G is any groupoid on the set of state spaces for all acts, then $\approx_G$ commutes with $\succ, \succeq, \approx$, and $\succ'$.

(Again, the proof of this theorem, like all others, is in the appendix.)

To make use of this relation, we will have to find some specific groupoid G and show that $\approx_G$ is compatible with $\succ$, and possibly $\succ'$. So first I will find the most general conditions under which compatibility can fail. In later sections I will discuss specific groupoids, and show certain more specific conditions under which compatibility holds or fails.

Compatibility fails iff there are acts A and B with $A \succ B$ (or $A \succ' B$) and $B \approx_G A$. $B \approx_G A$ means there is some $g \in G$ with $B(s) = A(g(s))$ for all s. But since $A \succ B$, this means that for all s, $A(s) >_v B(s) = A(g(s))$ (or $A(s) \geq_v B(S) = A(g(s))$ for $\succ'$, with at least one such pair having a strict inequality). Similarly, $A(g(s)) >_v B(g(s)) = A(g(g(s)))$, $A(g^{-1}(s)) >_v B(g^{-1}(s)) = A(s)$, and so on. Thus, for $\approx_G$ to be incompatible with $\succ$, there must be a g such that for *every* state s, the states $g^n(s)$ are all distinct, while for $\approx_G$ to be incompatible with $\succ$, there must be a g such that for *some* state s, the states $g^n(s)$ are all distinct.

Conversely, if the states $s, g(s), g(g(s)), \ldots$ are all distinct, then so are the states $g^{-1}(s), g^{-1}(g^{-1}(s)), \ldots$. If there are outcomes $\cdots >_v o_{-2} >_v o_{-1} >_v o_0 >_v o_1 >_v o_2 >_v \ldots$, then there is an act that assigns $A(g^n(s)) = o_n$ for each n. In this case, we can show that $\approx_G$ is incompatible with $\succ'$, and similarly for $\succ$ if every state s has this property. It is straightforward to check that these conditions are necessary and sufficient for incompatibility of the relevant relations.

If each state space $\mathcal{S}_A$ is finite, then it doesn't matter what groupoid G is considered — for every permutation g and every state s, the sequence $s, g(s), g(g(s)), \ldots$ forms a cycle, and it is impossible for any act to assign a strictly decreasing sequence of outcomes to all these elements. Thus, I have shown that for at least some groupoids and sets of acts, the equivalence relation $\approx_G$ is compatible with both $\succ$ and $\succ'$, and we can define a decision theory using its compositions with them, $\succ_G$ and $\succ'_G$. In section 3.4 I will give some more general conditions for compatibility that don't assume that the sets of states or outcomes are finite, as well as developing plausible requirements on G for $\approx_G$ to be normatively significant.

3.3.2 Applications

This general proposal is inspired by some ideas of Frederick Schick. In his (2003), he proposes a consideration of acts as being defined on different state spaces as a way to clarify some of the issues around the Newcomb problem. In this problem, the agent is given the choice between taking just the contents of an opaque box, or taking the contents of that box together with a transparent box that contains $1000. The contents of the opaque box were determined by a very reliable predictor, which has put $1,000,000 in it if it predicted that the agent will take only the one box, and nothing if it predicted that the agent will take both.

A naive approach says that the two acts are defined on the same state space, which can be taken to have one state where the opaque box is full of money and one state where the opaque box is empty. In either state, it is better to take both boxes, so dominance says that two-boxing is better than one-boxing.

However, Schick suggests that the reliability of the predictor means that the state where the money is in the box when the agent takes one box doesn't naturally correspond to the state where the money is in the box when the agent takes two boxes. Thus, we should take the two acts to be defined over separate state spaces, which happen to have homophonous descriptions of states that don't correspond.

Some causal decision theorists argue that Schick is wrong, and say that the correspondence is in fact natural. Schick instead proposes a correspondence that connects the states in the two spaces based on whether the predictor has predicted correctly. Under the first correspondence, the composition of dominance and correspondence again supports two-boxing. But under the second correspondence, the composition of dominance and correspondence doesn't yet determine anything, and some further rule (possibly involving probability) must be used.

Thus, this proposal doesn't immediately resolve the Newcomb paradox. It is compatible with either response, provided that one has further means to determine which correspondence(s) are relevant. I won't aim to resolve this question in this paper.

As another example, consider the "proto-Ellsberg" situation where a ball is about to be drawn from an urn containing an unknown number of black and white balls. Consider the acts B_1, B_2, W_1, W_2, where B_1 (resp. W_1) results in a gain of $1 if a black (resp. white) ball is drawn and nothing otherwise, and B_2 (resp. W_2) results in a gain of $2 if a black (resp. white) ball is drawn and nothing otherwise. We can identify the states here as "a black ball is drawn" and "a white ball is drawn". In this case, there seems to be no objection to thinking of the state space as identical for all acts. Assuming that $\$2 >_v \$1 >_v 0$, we thus have $B_2 \succ B_1$ and $W_2 \succ W_1$, but we have no comparisons between one of B_1 or B_2 and one of W_1 or W_2.

There are two natural ways to think about this situation. One might think that the subjective uncertainty about the number of black and white balls makes these other pairs all incomparable, so that dominance tells us all the normative facts there are about preference. Or one might think that the symmetry of the situation makes the states "a black ball is drawn" and "a white ball is drawn" interchangeable for the agent.[5] In the latter case, we can then consider the group G of permutations that includes both the identity id and the interchange g of the two states. Then we can see that since $W_1(s) = B_1(g(s))$ for all s, this means that $W_1 \approx_G B_1$, and similarly $W_2 \approx_G B_2$. Composing this relation with dominance, we see that $W_2 \succ_G B_1$ and $B_2 \succ_G W_1$. This would then settle all comparisons among the four acts.

This illustrates the sorts of considerations that go into determining what G to use — it depends on which states are taken to be interchangeable as far as preference among acts should go.

5. A third intuitive response might be to prefer B_2 to W_1 and W_2 to B_1, on the grounds that they give strictly better benefits, while taking W_1 and B_1 to be incomparable on the grounds of radical uncertainty, as well as W_2 and B_2. This set of preferences can't be achieved by any version of the theory described in this paper, because it includes more preferences than strict dominance but has no indifferences other than identity.

3.4 Probability

As suggested by the last example, probability is a natural concept to use to limit the sorts of correspondences between state spaces that are appropriate to use in an indifference relation. For much of the 20th century, the primary justifications for using probability theory to describe notions of confidence or degree of belief have based the theory of probability on that of preference. These are the "Dutch book argument" used by Ramsey (1926) and de Finetti (1937), and the standard representation theorems of decision theory, which also date back to Ramsey but were later extended by Savage, Jeffrey, and others. Since these arguments seek to derive probabilism from a full theory of preference, there would be a kind of circularity in then basing a theory of preference on probability.

However, there has also been a competing tradition among epistemologists that seeks to justify probabilism on the grounds of some pure epistemic value, without assuming a general theory of preference. One version of this, exemplified by Cox (1946) and Jaynes (2003), makes some technical mathematical assumptions about the structure of uncertainty. (For instance, they assume that an agent's degree of belief in a proposition determines her degree of belief in the negation, and that an agent's degree of belief in two incompatible propositions determines her degree of belief in the disjunction.) From these assumptions, they show that any such set of degrees of belief can be represented by a numerical function satisfying the probability axioms. The assumptions that go into the proof are quite strong ones, and may be seen as question-begging (since they assume features that correspond to all of the probability axioms, even if they are not identical to them), but for at least some researchers, this has been considered a strong enough motivation for probabilism that can be carried out completely prior to and independent of any theory of rational preference.

Another tradition, dating back at least as far as van Fraassen (1983), but perhaps put forward most widely to formal epistemologists by Joyce (1998), has more detailed foundations. Joyce considers (inspired by James (1896)) the idea that belief is a state aimed at the truth, and extends it to a notion of confidence that is aimed at "closeness to the truth". He shows that if this notion of closeness satisfies certain basic conditions, then any credal state that fails to satisfy the probability axioms is dominated by a different credal state that does satisfy them. Thus, he is able to justify the probability axioms as constraints on rational credence without assuming any constraint on rational preference beyond dominance, which I have also taken as basic.

On this picture one can start with dominance as the basic principle governing rational preference. Given a notion of "closeness to the truth", one can then argue that credences ought to obey probabilism. Using the ideas discussed in the rest of sections 3.4 and 3.5, we can then defend the standard role of expected value in determining preference. Finally, the results of Greaves and Wallace (2006), Leitgeb and Pettigrew (2010), and Easwaran (2013) can then use expected closeness to the truth to get further constraints on credence and the way it is updated. Thus, there is a satisfying and non-circular interplay between credence and preference.

In section 3.4.1 I give the basic background for any theory of probability, laying the groundwork for which sets of states even have probabilities. In section 3.4.2, I then set out a basic theory of comparative probability (which doesn't assume as much as the numerical probability theories endorsed by Jaynes and Joyce) and use it to give some general conditions under which $\approx_G$ is compatible with $\succ$ and $\succ'$, and some conditions under which it is compatible with $\succ$ but not $\succ'$. In section 3.4.3, I then define full numerical probability, and show that combining $\approx_G$ with dominance can give plausible preferences between acts that can't be compared by expected utility theory (though it does not yet give preferences for all acts that can be compared by expected utility theory). Finally, in section 3.4.4, I show that under some homogeneity assumptions about the spaces of states, the combination of $\approx_G$ with dominance gives a preference relation that is exactly that of "stochastic dominance". That final section is not important for the later parts of the paper and can be skipped by readers who want to avoid the most dense mathematical technicalities.

3.4.1 *Measurability*

Probability is a feature of sets of states. However, not every set of states is eligible to have a probability. For instance, if there are multiple state spaces, then a set that contains states from different spaces doesn't have a probability. But in many cases, if a single state space is infinite, then not every subset of that space has a probability either. Instead, we define some collection of sets of states to be the *measurable* ones, and say that probability applies to all and only the measurable sets.

For each state space, I will assume that the collection of subsets of it that are measurable forms an *algebra*. This means that the empty set is measurable, the complement of any measurable subset of a state space is measurable, and the intersection of two measurable subsets is measurable. This implies that each state space also counts as a measurable set (since it is the complement of the empty set), and that the union of any two measurable subsets of the same state space is itself measurable. I will sometimes further assume that the collection of measurable sets forms a *σ-algebra*, which means that not only is the intersection of two measurable sets measurable, but also the intersection of any countably infinite collection of measurable sets is measurable.

Although probability doesn't apply to sets of outcomes or sets of real numbers directly, it is also useful to refer to some of these sets as "measurable" as well. For any given outcome o, I define the sets $\mathcal{O}_{>o} = \{x \in \mathcal{O} | x >_v o\}$, $\mathcal{O}_{\geq o} = \{x \in \mathcal{O} | x \geq_v o\}$, $\mathcal{O}_{<o} = \{x \in \mathcal{O} | x <_v o\}$, and $\mathcal{O}_{\leq o} = \{x \in \mathcal{O} | o \leq_v x\}$ to be measurable, and the measurable sets of outcomes to be exactly the ones that can be achieved from these by iterating the operations of complement and intersection (finite or countable, depending on which assumption has been made for sets of states).[6] Similarly for real numbers — I define the sets $\mathbb{R}_{>x} = \{y \in \mathbb{R} | y > x\}$ and $\mathbb{R}_{<x} = \{y \in \mathbb{R} | y < x\}$ to be measurable, and define the measurable sets of real numbers to be exactly the ones that can be achieved from these by iterating the operations of complement and intersection. (For those that are familiar with the term, even countable intersection gives rise only to the Borel measurable sets — one could go further and use the Lebesgue measurable sets, but this won't be needed for my purposes.)

Given these notions of measurability for sets, it is also useful to define a notion of measurability for a function. A function is said to be *measurable* iff for every measurable subset S of its range, the set $\{x | f(x) \in S\}$ is itself measurable. Since I have defined measurability for sets of states, sets of outcomes, and sets of real numbers, this gives notions of measurability for functions from states to states (such as permutations and correspondences), from states to outcomes (such as acts), from states to real numbers, and from outcomes to real numbers.

In general, from here on I will restrict attention to measurable functions. In particular, since acts are functions from states to outcomes, this means that I will restrict attention to acts that are measurable functions. This is a change from the earlier discussion that included *all* functions from states to outcomes as acts.

With these notions of measurability, I can now start the discussion of how probability limits the appropriate correspondences between states in various spaces, and eventually use probability to define some further indifference relations between acts.

3.4.2 *Comparative Probability*

I start with what I take to be the most minimal notion of probability. This consists in relations of comparative probability on the measurable sets of states. The basic assumptions are these:

- $>_p$, $\geq_p$, and $\equiv_p$ are relations on the set of measurable sets of states.
- $>_p$ is a strict ordering, $\geq_p$ is a weak ordering, and $\equiv_p$ is an equivalence relation.
- $>_p$ is compatible with and transitive over each of the other two relations, and $\geq_p$ is transitive over $\equiv_p$.

6. If the set of outcomes is separable (as defined in section 3.4.2) then any of these four types of sets are constructible from the others by iteration of complement and countable intersection. However, if the set of outcomes is not separable, and if only finite intersections are allowed, then we may need to include several of these types of sets as basic measurable sets of outcomes.

- For any measurable s_1 and s_2, if $s_1 \subseteq s_2$ then $s_2 \geq_p s_1$.
- If $\mathcal{S}$ is any full state space and $\emptyset$ is the empty set, then $\mathcal{S} >_p \emptyset$.
- If $\mathcal{S}_1$ and $\mathcal{S}_2$ are two full state spaces, then $\mathcal{S}_1 \equiv_p \mathcal{S}_2$.[7]

Note that I don't assume that comparative probability is a linear ordering. In particular, it is natural to think that for at least some measurable sets that are subsets of disjoint state spaces, there might be no comparative probability relation that holds between them. But the fewer sets are comparable, the more restricted the notion of $\approx_G$ governed by comparative probability turns out to be as an indifference relation.

One piece of notation that is useful to have is the application of a function to a set of elements rather than to an individual element. If f is any function and S is any subset of its domain, then I define $f[S] = \{f(x)|x \in S\}$, which is the set we get by applying f to each element of S.

A correspondence g between two state spaces (or a permutation on one state space) is said to be *measure-preserving* iff it is measurable, and for every measurable subset S of the domain of g, $S \equiv_p g[S]$. It is clear that the identity on any state space is measure-preserving, that the composition of two measure-preserving correspondences is measure-preserving, and that the inverse of a measure-preserving correspondence is measure-preserving. Thus, the set of measure-preserving correspondences forms a groupoid.

I suggest that the equivalence relation generated by this groupoid is the one that should be applied as an indifference relation. More generally, it seems plausible that whatever groupoid is used must be a subset of this one. I say that any such groupoid is itself *measure-preserving*.

Given that $\approx_G$ is a natural indifference relation to apply to acts when G is a measure-preserving groupoid, I now give some characterizations of situations in which $\approx_G$ is compatible with $\succ$ or $\succ'$. Each will need some new definitions.

Say that a set S is *null* iff for every n, there are n disjoint subsets of a single state space that are all at least as probable as S. A set S is then non-null iff there is some n such that no n disjoint subsets of a single state space are all at least as probable as S. (Equivalently, any n subsets of the same state space that are at least as probable as it must have some overlap.)

Say that the comparative probability relation is *regular* iff the empty set is the only null set. This is a way of generalizing the feature of finite probability spaces that guarantees that $\approx_G$ is compatible with the preference relations.

Theorem 6. If the probability is regular, and G is measure-preserving, then $\approx_G$ is compatible with $\succ'$ (and therefore with $\succ$).

Regularity in this sense is a very strong condition on probability functions, and is often rejected for reasons given in Easwaran (2014). Thus, it is useful to give another set of more commonly accepted conditions under which $\approx_G$ is compatible with $\succ$, though it turns out that they allow $\approx_G$ to be incompatible with $\succ'$.

Say that a set D of outcomes is *dense* iff the following three conditions hold:

- For every $o \in \mathcal{O}$ there is $d \in D$ with $d <_v o$.
- For every $o \in \mathcal{O}$ there is $d \in D$ with $d >_v o$.
- For every $o_1 <_v o_2 \in \mathcal{O}$ there is $d \in D$ with $o_1 <_v d <_v o_2$.

Say that $\mathcal{O}$ is *separable* iff there is a dense set $D \subseteq \mathcal{O}$ that has only countably many members. Separability is one of the characteristic features of the real numbers (the rational numbers form a countable dense subset of the reals) but it applies to many outcome spaces that aren't

7. Other sets of conditions for comparative probability are discussed in Chapter 11 of Fine (1973) and sections 7 and 8 of Chapter 2 of Halpern (2003), as well as Hawthorne (2015).

straightforwardly numerically measurable, and allows a sort of partial numerical representation.

Theorem 7. If $\mathcal{O}$ is separable, then there is a function R from $\mathcal{O}$ to the real numbers, such that whenever $o_1 <_v o_2$, it is the case that $R(o_1) < R(o_2)$.

A comparative probability is *countably partitive* iff no non-null set is a countable union of null sets. Equivalently, every countable partition of a non-null set has at least one non-null member.[8]

Theorem 8. If the probability is countably partitive, and the outcome space is separable, and G is measure-preserving, then $\approx_G$ is compatible with $\succ$.

This theorem will be particularly relevant in the case where outcomes have real-valued utilities and sets of states have a countably additive real-valued numerical probability function. But I give the result here in its more general form to show that these assumptions aren't necessary.

However, these conditions don't guarantee that $\approx_G$ is compatible with $\succ'$.

To see this, let the state space consist of the set of points on the unit circle, with the comparative probability relation given by the standard countably additive probability function. Let the outcome space be the real numbers. Let G be the group of rotations of the unit circle, which are all measure-preserving correspondences.

Let A be the act that assigns outcome n to every point that is exactly n radians counterclockwise from a given point on the circle (allowing positive and negative integers n), and assigns outcome 0 to all other points. Let g be the rotation by 1 radian counterclockwise. Then $A(s) >_v A(g(s))$ for every s that is an integer number of radians from the given point, and $A(s) \equiv_v A(g(s))$ for all other s. Therefore, $A \succ' Ag$, so $\succ'$ is incompatible with $\approx_G$. In this case it seems more plausible to me that we should hold onto $\approx_G$ as a norm of indifference than $\succ'$ as a norm of preference, but consideration of cases like this will surely be important for future research.

I have given two sets of conditions on the comparative probability relation that suffice to make $\succ$ compatible with $\approx_G$ for every measure-preserving groupoid, but there are surely other sufficient conditions for compatibility that may be more relevant in particular applications of decision theory.

3.4.3 Numerical Probability

A numerical probability P is a function that assigns a real number to every measurable set of states, satisfying the following conditions:

- For any measurable set S, $P(S) \geq 0$.
- If $\mathcal{S}$ is any full state space, then $P(\mathcal{S}) = 1$.
- If S_1 and S_2 are two disjoint subsets of the same state space, then $P(S_1 \cup S_2) = P(S_1) + P(S_2)$.

As I mentioned above, Joyce (1998) has given a justification of these axioms that depends only on dominance and some ideas of "closeness to truth".

Given a numerical probability function, one might naturally define a comparative probability by saying that for any two measurable sets of states S_1 and S_2:

- $S_1 >_p S_2$ iff $P(S_1) > P(S_2)$.
- $S_1 \geq_p S_2$ iff $P(S_1) \geq P(S_2)$.
- $S_1 \equiv_p S_2$ iff $P(S_1) = P(S_2)$.

However, I will not assume these full connections, and instead assume only that for any two measurable sets of states S_1 and S_2:

- If $P(S_1) > P(S_2)$, then $S_1 >_p S_2$.

8. In the presence of a homogeneity assumption of the sort used in section 3.4.4, it is sufficient to assume that no state space is a union of countably many null sets: if some countable union of null sets were non-null, then a union of finitely many copies of that non-null set would cover the whole space, and this would induce a covering by n copies of the original countable union, which would itself be a countable cover of the whole space by null sets.

- If $S_1 \geq_p S_2$, then $P(S_1) \geq P(S_2)$.
- If $S_1 \equiv_p S_2$, then $P(S_1) = P(S_2)$.

The stronger definition clearly satisfies these constraints, but if there are reasons to accept a more refined notion of comparative probability than the numerical probability function, then one will have to use something more like these weaker axioms (Easwaran, 2014).

Even given these weak axioms, it is straightforward to prove that if g is a measure-preserving correspondence for the comparative probability, then for any measurable set of states S, $P(S) = P(g[S])$. Similarly, one can prove that if S is null, then $P(S) = 0$. If we assume that the probability function P is *countably additive* (meaning that the probability of a disjoint union of countably many subsets of a single state space is equal to the sum of their probabilities), then we also get the consequence that the comparative probability is countably partitive. Additionally, there are many probability comparisons that are possible across state spaces given this connection — the only cases where sets of states might not bear a comparative probability relation to each other is when they have the same numerical probability.

If there is a numerical probability function, and we let $\approx_G$ be the groupoid of all measure-preserving correspondences between state spaces for the associated comparative probability relation, then the combined preference relation $\succ_G$ makes many judgments that standard expected utility theory can't make. For instance, consider the St. Petersburg game, where a fair coin is repeatedly flipped until it comes up heads, and the agent gets 2^n units of utility if it takes n flips. The theory under consideration can tell us to be indifferent between this St. Petersburg game and one where the coin is repeatedly flipped until it comes up tails, even though standard expected utility theory merely tells us that neither of these acts has a finite expected utility. Similarly, if we consider the Leningrad game defined by Colyvan (2008), where the agent gets $2^n + 1$ units of utility if the coin takes n flips, then the theory here can tell us to prefer a Leningrad game, played for either heads or tails, to any St. Petersburg game.

Of course, this is all still very weak. For instance, we still can't decide between a coin flip for 3 units of utility or nothing, and an act that guarantees 1 unit of utility. In section 3.4.4 I will demonstrate how far this theory does go (assuming a technical notion of "homogeneity" on the sets of states), and in section 3.5 I will add another indifference relation that allows us to recover all of classical expected utility theory as well. Putting these relations together, we will have a decision theory that makes all the comparisons that classical expected utility theory does, and more, and it will be built in a modular way that allows for easier modifications.

3.4.4 Stochastic Dominance

This section is more mathematically technical and is unnecessary for the general results. Readers can thus skip it without much loss. However, it demonstrates some very strong consequences of some structural conditions on comparative probability.

Say that a probability function on a set of state spaces is *homogeneous* iff whenever S_1 and S_2 are measurable sets with $S_1 \leq_P S_2$, there is a measure-preserving correspondence g whose domain is the state space S_1 is in and whose range is the state space S_2 is in, with $g[S_1] \subset S_2$.

Homogeneity holds for the sorts of state spaces that usually arise in consideration of continuous probability. For instance, consider a set of state spaces consisting of dartboards whose coordinates are given by real numbers, or of infinite sequences of coin flips, with the standard numerical probability functions on them. Say that $S_1 <_P S_2$ iff $P(S_1) < P(S_2)$ or $P(S_1) = P(S_2)$ and S_1 has a strictly smaller cardinality than S_2 or $P(S_1) = P(S_2)$ and the complement of S_1 has a strictly larger cardinality than the complement of S_2. Say that $S_1 \equiv_P S_2$ iff $P(S_1) = P(S_2)$, S_1 and S_2 have the same cardinality, and their complements also have the same cardinality. For each of these probability spaces, the comparative probability relation given by this definition is homogeneous.

The following two theorems show that when probabilities are homogeneous, any two sets of the same probability can be interchanged, and this interchange can be done without disturbing the rest of the space. The second theorem uses the following definition: comparative probability is *additive* iff for every set of states A, and every two sets of states B and B' both disjoint from A, $A \cup B \equiv_P A \cup B'$ iff $B \equiv_P B'$.

Theorem 9. If $S_1 \equiv_P S_2$ and the comparative probability relation is homogeneous, then there is a measure-preserving correspondence g with $g[S_1] = S_2$.

Theorem 10. If S_1 and S_2 are in the same state space, and the comparative probability relation is homogeneous and additive, then there is a measure-preserving permutation h such that $h[S_1] = S_2$, $h[S_2] = S_1$, and h is the identity on all states outside of $S_1 \cup S_2$.

The next theorem shows that if probabilities are homogeneous, then two acts with the same probabilities of various outcomes can also be interchanged. The theorem states that this holds if the act has only countably many distinct outcomes, and I conjecture that it holds in general.

Say that an act A is *discrete* iff it is measurable, and for any o, either there is no s with $A(s) = o$, or $P(A(s) = o) > 0$. (Here, and in what follows, I use "$A(s) = o$" to refer to the set $\{s \in \mathcal{S}_A | A(s) = o\}$, and similarly with other conditions on $A(s)$.)

Theorem 11. Let P be a homogeneous probability function. Let A_1 and A_2 be any discrete acts such that for every o, $(A_1(s) = o) \equiv_P (A_2(s) = o)$. Then there is a measure-preserving correspondence g such that for all s, $A_1(s) = A_2(g(s))$.

Conjecture. Let P be a homogeneous probability function. Let A_1 and A_2 be acts such that for every o, $(A_1(s) > o) \equiv_P (A_2(s) > o)$. Then there is a measure-preserving correspondence g such that for all s, $A_1(s) = A_2(g(s))$. (We may need to assume that $\mathcal{O}$ is separable for this.)

If true, this conjecture means that we can represent an act by its cumulative distribution function. This is a function CDF_A defined on $\mathcal{O}$ such that $CDF_A(o) = \{s \in \mathcal{S}_A | A(s) <_v o\}$. If for every o, $CDF_{A_1}(o) \equiv_P CDF_{A_2}(o)$, then these results say that $A_1 \approx_G A_2$.

Importantly, this also means that if for every o, $CDF_{A_1}(o) <_P CDF_{A_2}(o)$, then $A_1 \succ_G A_2$. This condition is called *stochastic dominance* and is a standard generalization of dominance. Thus, if the probability is homogeneous, and the conjecture is true, then we have shown that $\succ_G$ gives all decisions that stochastic dominance does, and even without the conjecture, this holds for discrete acts. Stochastic dominance is still not enough for expected utility, but it is a very strong condition that is often used in applications of decision theory.

3.5 *Trade-Offs*

The idea of using permutations of the state space to find acts that one should be indifferent between is that if sets of states S_1 and S_2 are interchangeable (meaning that some permutation in the relevant group interchanges them while leaving other states fixed), then it shouldn't matter if an act gives a good outcome in S_1 and a bad outcome in S_2, or vice versa. The idea of using measure-preserving permutations is that probability alone should determine whether sets of states are interchangeable. But this suggests two further ways that the idea can be generalized.

First, if probability isn't homogeneous (for instance, consider a state space with just three states, one of which has probability $1/2$ and the other two of which have probability $1/4$ — the singleton of the former has the same probability as the set of the other two, but no permutation interchanges them), permutations alone won't let us get all the invariances we want. Second, it's natural to consider interchanges where we don't entirely interchange the goodness of the outcomes associated with two states but somehow only exchange some of this goodness. To do this, we will need to be able to compare not just ordinal goodness among outcomes, but also differences of goodness between outcomes.

In section 3.5.1 I give the mathematical background for a theory of differences in value of outcomes. In section 3.5.2 I use this to define the notion of "trade-off" equivalence formalizing the above reasoning. In section 3.5.3 I then show that this equivalence relation commutes with all the earlier relations, and give various sets of conditions under which it is compatible with them. (This section can be skipped by readers who are willing to assume compatibility and wish to skip some of the most technical sections.) In section 3.5.4, I then show that combining trade-off equivalence with dominance gives all judgments that are given by traditional expected utility theory on individual state spaces, and that combining both with permutation equivalence gives all of traditional expected utility theory and more. In section 3.5.5 I finally consider prospects for a slight modification of the notion of trade-off equivalence that might allow for a version of the theory that captures modifications of expected utility that take risk sensitivity into account.

3.5.1 Differences of Outcomes

I assume that for certain outcomes o_1, o_2 there is a "difference" $o_1 - o_2$ between them, and I denote the set of differences by '$\mathcal{D}$'. I require that differences satisfy the following assumptions:

- For any difference d and outcome o, there is an outcome o' such that $o' - o = d$.
- There is an operation $+_d$ of addition on differences such that for any differences d_1, d_2, d_3, $d_1 +_d d_2 = d_2 +_d d_1$, and $d_1 +_d (d_2 +_d d_3) = (d_1 +_d d_2) +_d d_3$.
- For any outcomes o_1, o_2, o_3, if two of their pairwise differences exist, then so does the third, and $(o_1 - o_2) +_d (o_2 - o_3) = (o_1 - o_3)$.
- There is a difference 0_d such that for any difference d, $d +_d 0_d = d$.
- For any outcomes o_1, o_2, $o_1 =_v o_2$ iff $o_1 - o_2 = 0_d$.
- For any outcomes o_1, o_2, if $o_1 - o_2$ exists, then $o_2 - o_1$ does too.
- For any outcomes o_1, o_2, if $o_1 >_v o_2$, then their difference exists, and $o_1 - o_2 >_d 0_d$.

These assumptions about differences of outcomes are consistent, and can be satisfied by a variety of different structures. One way is if outcomes and differences are both represented by real numbers. But there are also other structures satisfying these features, while giving up some of the other characteristic features of the real numbers, including linearity, density, and the Archimedean principle. The specific features of differences that I have defined above may themselves not be essential to my project — perhaps a different way of setting things up would work as well.

As an example of a non-linear ordering satisfying these features, consider the following: an idealized agent values money (and debt) linearly, and finds a career in the military incomparable to a career in the priesthood, and values nothing else. This agent could have outcomes with the structure of two separate, incomparable copies of the real numbers, with differences defined only among outcomes that involve the same career, while differences have the structure of the real numbers.

As another example, an agent could value having friends, disvalue having enemies, and value money, but nothing else. Say that the value of one more friend is always exactly equal to the value of having one fewer enemy, while being incomparable to the value of any amount of money. Then every pair of outcomes will have a difference, which can be thought of as an ordered pair of an integer (the number of additional friends minus the number of additional enemies) and a real number (the amount of additional money), but there will be differences that can't be compared (when one difference involves adding more friends, while the other involves adding more money). Additionally, there will be pairs of outcomes in which one is strictly better than the other, but there is no outcome that is strictly between them — just consider two outcomes that have exactly the same amount of money, but one more friend or one fewer enemy.

Furthermore, we could have a different agent with the same sort of preferences, except that an additional friend is strictly more valuable than any amount of money, rather than being incomparable with it. In that case, the differences have a total ordering, but this ordering is non-Archimedean — for two outcomes whose difference is purely monetary, no number of copies of that difference is enough to make up for the loss of a friend.

Exactly what feature of the agent's psychology makes it the case that they value outcomes in a way that has this sort of difference structure is beyond the scope of this paper. On a standard representation-theorem framework, one can derive the difference structure from the preference ordering. But I can't do this, because I want to reverse the order of explanation. I will ignore the question of what determines the facts about differences, and just show what follows if these facts can be argued to hold. It is possible that this sort of consideration will eventually lead one back to a representation-theorem-based account of decision theory, but perhaps some other technique discussed by Krantz et al. (1971) can give us a foothold here.

3.5.2 *Trade-Offs*

For two acts A and B on the same state space, I will define their *trade-off* $T(A, B)$ to be (S_1, S_2, d) iff the following conditions hold:

- S_1 and S_2 are disjoint sets of states, and d is a difference.
- For all $s \in S_1$, $A(s) - B(s) = d$.
- For all $s \in S_2$, $B(s) - A(s) = d$.
- For all s outside $S_1 \cup S_2$, $A(s) \equiv_v B(s)$.

The idea is that the triple (S_1, S_2, d) tells us the two sets of states on which the acts differ in opposite directions, and the amount by which they differ.

Now, we can define $A \sim_v B$ iff $T(A, B) = (S_1, S_2, d)$ and $S_1 \equiv_P S_2$ — that is, there is a trade-off between the two acts such that the states where A is better than B are equally probable as the states where B is better than A, and the differences in values are the same, in opposite directions. It is straightforward to see that $\sim_v$ commutes with $\succ, \succeq, \succ'$, and also with $\approx_G$ for any measure-preserving groupoid G. (It is important that for every d and o, there is an o' with $o' - o = d$, to make sure that if $A_1 \sim_v A_2$ and $A_2 \succ B_2$, then there is a B_1 with $T(A_1, A_2) = T(B_1, B_2)$. Given that B_1 exists, it is not hard to see that $A_1 \succ B_1$, and similarly for the other relations.)

Note that $\sim_v$ is reflexive (since we can take S_1 and S_2 to be the empty set) and symmetric (since we can reverse the role of S_2 and S_1). However, it is not transitive. So I define the full trade-off equivalence relation $\approx_v$ to be the transitive closure of $\sim_v$. That is, $A \approx_v B$ iff there exists a finite sequence $A_0, \ldots, A_n$ such that $A = A_0 \sim_v A_1 \sim_v A_2 \ldots A_{n-1} \sim_v A_n = B$. This is an equivalence relation. Since $\sim_v$ commutes with $\succ, \succeq, \succ', \approx_G$, it is straightforward to see that $\approx_v$ does too. (Note that $\approx_v$ and $\sim_v$ are relations on acts, while $\equiv_v$ and $>_v$ are relations on outcomes.)

3.5.3 *Commutativity, Compatibility, and Composition of Trade-Off Equivalence*

Theorem 12. Assume the outcomes are linearly ordered. Assume that probabilities are linearly ordered, additive, and homogeneous. Then $\approx_v$ is compatible with $\succ$. (This result needs only comparative probability, not numerical probability.)

Thus, $\approx_v$ is compatible and commutes with $\succ$ and $\succeq$, so we can take their compositions to define relations $\succ_v$ and $\succeq_v$. (Note that $\succ_v$ and $\succeq_v$ are relations on acts, while $>_v$ and $\geq_v$ are relations on outcomes.)

Note that $\approx_v$ is *not* in general compatible with $\succ'$. This is trivial if it is possible for a non-empty set to be equiprobable with the empty set. In that case, two acts A and B can bear the relation $\sim_v$ to each other when one is better than the other on the empty set, while the other is better than the first on a non-empty set equiprobable with the empty set. In that case, the latter bears $\succ'$ to the former, so the relations are not compatible.

Even a restriction of the definition of $\sim_v$ to not allow it to directly deal with the empty set doesn't solve things, as long as it is possible for $S_1 \subsetneq S_2$ with $S_1 \equiv_P S_2$. For instance, consider three bets on the throw of an infinitely thin dart at a line that is 1 unit long. The state space can be considered to be the unit interval $[0, 1]$. Let A_0 be the act that gives value 0 at every point. Let A_1 be the act that gives value $+1$ for all states in $[0, 1/2]$ and -1 for all states in $(1/2, 1]$ — that is, we increase the value of the outcome by 1 for all states in the closed left half of the interval, and decrease the value of the outcome by 1 for all states in the open right half of the interval. If $[0,1/2] \equiv_P (1/2,1]$, then $A_0 \sim_v A_1$. Let A_2 be the act that gives the value $+2$ for the state $1/2$ and gives the value 0 for all other states. Then $A_2 \succ' A_0$, because there is one state in which it is strictly better, and no state in which it is not at least as good. But if it's also the case that $[0,1/2) \equiv_P [1/2,1]$, then $A_1 \sim_v A_2$, because A_2 gives an outcome that is one unit better in all states in the closed right half of the interval, and one unit worse in all states in the open left half of the interval. The only way to block this contradiction with $\succ'$ is to say that some sets of the same numerical probability are nevertheless not equally probable. This may be reasonable, but one needs to make sure that one doesn't rule out too many of these equiprobability relations for $\sim_v$ to say anything.

I conjecture that $\succ'$ is compatible with $\approx_v$ if the probability satisfies what I call *strong regularity*: If $S_1 \subseteq S_2$ and $S_1 \equiv_P S_2$, then $S_1 = S_2$. At any rate, there are certainly some state spaces and probability functions that make $\approx_v$ compatible with $\succ'$. For instance, if the state space has finitely many states, all of equal (and therefore positive) probability, then the argument for compatibility with $\succ$ will work for $\succ'$ as well.

But this general lack of compatibility between the two is the reason why I prefer to take $\succ$ as the base relation rather than $\succ'$. Further research to investigate what conditions on the probability, or modifications of $\sim_v$ and $\approx_v$, allow us to use $\succ'$ instead is certainly in order.

3.5.4 Expected Utility

Finally, we can put together all the pieces. As Theorem 13 below shows, under some standard conditions, the relations of dominance, correspondence under a measure-preserving groupoid, and trade-off equivalence are all mutually compatible. Theorem 14 shows that the composition of trade-off equivalence with dominance includes all decisions made by standard expected utility theory, assuming a divisibility condition on the state spaces. Thus, the decision theory made by combining all of the relations described so far is consistent, and gives all the decisions of expected utility theory, plus substantially more.

Theorem 13. If outcomes and differences are real numbers, and probabilities are numerical and countably additive, and G is any measure-preserving groupoid, then $\approx_v$ is compatible with $\succ_G$.

Say that a probability function is *divisible* if whenever $P(A) > 0$, there is $A' \subset A$ with $P(A) > P(A') > 0$.

Theorem 14. If probabilities, values of outcomes, and differences of outcomes are all real numbers that satisfy the standard relations, and if probability is divisible, then if A and B are acts on the same space with finite upper and lower bounds on the utility of their outcomes, and if the expected utility of A is strictly greater than the expected utility of B, then $A \succ_v B$.

Thus, $\succ_v$ gives a decision theory that extends classical expected utility theory. However, it also gives us decisions between acts that classical expected utility theory can't handle. For instance, if gamble A is the St. Petersburg game, and gamble B is the St. Petersburg game together with an independent coin flip that gives \$3 if heads and costs \$1 if tails, then $B \succ_v A$ (just move \$1.50 from the heads states to the tails states), but both have expected value that is infinite and so can't be compared by traditional expected utility theory.

The relation $\succ_v$ by itself doesn't directly allow us to compare acts on different state spaces, but when it can be combined with $\approx_G$ (as theorem 13 shows it often can) to yield the relation $\succ_{vG}$, we can further extend expected utility theory.

For instance, consider act A, which yields the outcome of a St. Petersburg gamble played for heads, together with the result of an independent coin flip that gives \$3 if heads and costs \$1 if tails. Let act B yield the outcome of a St. Petersburg gamble played for tails on a different coin. We can see that $A \succ_{vG} B$, even though these two acts can't be compared by classical expected utility theory (since both have infinite expectation), or dominance (since they are on different state spaces), or $\succ_G$ (since even with a useful identification of state spaces, the additional coin flip in A gives possibilities of either winning or losing additional money). Thus, $\succ_{vG}$ is a strict extension of all of the decision theories mentioned so far. There are still acts that it doesn't tell us how to compare (for instance, it can't compare Nover and Hájek's "Pasadena game" to any constant act), but it allows us to make other decisions involving these acts, and leaves room for further extension, perhaps if some method like that of Easwaran (2008) can be combined with the relations described here.

3.5.5 Risk-Weighted Expected Utility

One prominent challenge to classical expected utility theory, raised prominently by Allais (1953), is the phenomenon of risk aversion. Although Savage famously claimed that Allais' example is just an example of the irrationality of actual human decision-making, many others have thought that decision theory must be modified to allow for sensitivity to risk.

One such theory is that of Buchak (2013). On this theory, in addition to a probability function and a utility function, an agent has a risk-sensitivity function. Instead of each state contributing value to an act based on its utility times its probability, each state is counted in proportion to its position in the gamble. That is, states that give rise to the best possible outcomes may be counted differently from states that give rise to middling outcomes, which in turn may be counted differently from states that give rise to the worst outcomes, even if these states all have the same probability. On this theory, a risk-averse individual may care very strongly about low-probability outcomes if they are bad, but not if they are neutral or good, while a risk-seeking individual may care much more strongly about low-probability outcomes if they are good, but not if they are neutral or bad.

Thus, if one wants to compare bets involving the flip of a fair coin, improving the outcomes where one already wins could affect the value of the bet differently from improving the outcomes where one loses. A risk-averse individual could prefer the latter, while a risk-seeking individual could prefer the former.

This theory is obviously in conflict with the relation $\sim_v$ that I proposed above. However, in comparing her theory to classical expected utility theory, Buchak shows that the two theories can be axiomatized in a way on which the axiom that distinguishes the two corresponds to this trade-off relation. She shows that by replacing "Trade-Off Consistency" with "Co-Monotone Trade-Off Consistency", one can turn a particular set of axioms for the classical theory into a set of axioms for her theory.

I suspect that a suitable modification of the trade-off equivalence relation that I discuss should be able to similarly turn my relations $\succ_v$ and $\succ_{vG}$ into relations that bear the same relation to Buchak's risk-weighted theory as my relations do to classical expected utility theory. They would be able to compare acts for which risk-weighted expected utility theory is either undefined or infinite, in a way that is compatible with (and extends) the decisions made by that theory in simpler cases.

4. Conclusion

I have shown in section 3 (and in particular in section 3.5.4) that a decision theory can be built along the lines suggested in section 2 in such a way that it yields all of the decisions of standard expected utility theory, and more. Additionally, the assumptions that are required to achieve this result (apart from the divisibility condition on the sets of states) are all required for standard expected utility theory to even make sense. When these conditions aren't satisfied, the defender of standard expected utility theory must do something totally different to give any sort of decision theory. However, in such cases, my earlier results show that some of these relations are still compatible, and so their composition still gives some requirements on preference, though they may be weaker than those of expected utility theory. Finally, I've suggested a way to modify the trade-off part of the theory to accommodate risk sensitivity while retaining some of the other advantages of my general approach.

A defender of standard expected utility theory can use my results to extend her theory to the one I describe in section 3.5.4. She can use a standard representation theorem to establish the general numerical conditions on probability and utility, then argue for the normative force of the relations I describe, and then use the fact that $\succ_{vG}$ is the smallest relation that extends dominance and is transitive over the indifference relations to suggest that agents should follow the recommendations of this theory as well. Thus, even if there is a problem with isolating differences of value between outcomes separately from representation theorems, there can still be value to this overall project.

But I think it may be more interesting to suggest a different foundation for the theory. One can take credence and utility to be features of an agent's psychology that are not metaphysically dependent on facts about preference among acts. From this starting point, one can use Joyce's arguments to show that credences ought to be probabilistic, and there may be analogous arguments one can use to show that utilities ought to have the features described in section 3.5.1. One can then argue that the various relations I describe are normative, and thus come to the conclusion that rational agents ought to have preferences that line up with $\succ_{vG}$. Thus, they ought to obey the prescriptions of standard expected utility theory (and more). Or, if there are problems in Joyce's argument, or reasons to doubt the parallel theory of value for outcomes, or reasons to allow for risk-sensitivity, then one of the weaker results can be used. This picture is more in line with the naive view of decision theory, and it avoids the worries raised by Meacham and Weisberg (2011). I recommend this final interpretation, though the formal results don't require it.[9]

9. This paper has gone through several versions before reaching this final form. I would like to thank audiences for helpful comments in 2006 at the Berkeley-Stanford logic meeting, the Australian National University, the Australasian Association of Philosophy, and the Texas Decision Theory Workshop, as well as in 2009 at the Formal Epistemology Festival in Michigan, and in 2013 at the Workshop in Epistemic Utility Theory at Bristol. There are also many individuals that have given very helpful feedback along the way, including Rachael Briggs, Lara Buchak, Branden Fitelson, Alan Hájek, James Joyce, Richard Pettigrew, Jake Ross, Sahotra Sarkar, Michael Titelbaum, Matt Weiner as well as referees for this journal, and probably many other people I have unintentionally left out.

Appendix: Proofs of Theorems

Theorem 1. If $\sim_X$ and $\sim_Y$ commute, and both are transitive, then their composition is transitive too.

Proof. (See the following diagram.) Assume that $A_1 \sim_{XY} B_2$ and $B_2 \sim_{XY} C_3$. This means that there is a B_1 such that $A_1 \sim_X B_1$ and $B_1 \sim_Y B_2$ and there is a C_2 such that $B_2 \sim_X C_2$ and $C_2 \sim_Y C_3$. Thus $B_1 \sim_{YX} C_2$. But, by commutativity, this means that $B_1 \sim_{XY} C_2$, so there is a C_1 such that $B_1 \sim_X C_1$ and $C_1 \sim_Y C_2$. By transitivity of $\sim_X$, we have $A_1 \sim_X C_1$, and by transitivity of $\sim_Y$, we have $C_1 \sim_Y C_3$. Thus, we have $A_1 \sim_{XY} C_3$.

$$\begin{tikzcd}
A_1 \arrow[d, "X"] \arrow[dr, "XY"] & & \\
B_1 \arrow[r, "Y"] \arrow[d, "X"] & B_2 \arrow[d, "X"] \arrow[dr, "XY"] & \\
C_1 \arrow[r, "Y"] & C_2 \arrow[r, "Y"] & C_3
\end{tikzcd}$$

□

Theorem 2. If $\sim_X$ and $\sim_Y$ are both reflexive, then their composition is reflexive too.

Proof. Since $\sim_X$ and $\sim_Y$ are reflexive, we have that $A \sim_X A$ and $A \sim_Y A$. Thus, $A \sim_{XY} A$. □

Theorem 3. If $\sim_X$ and $\sim_Y$ commute, and both are symmetric, then their composition is symmetric too.

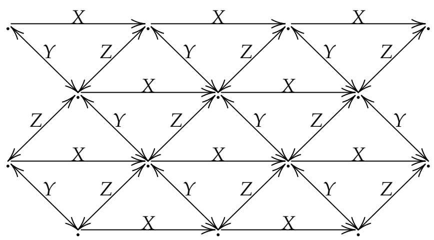
Proof. Assume that $A_1 \sim_{XY} B_2$. This means that there is an A_2 such that $A_1 \sim_X A_2$ and $A_2 \sim_Y B_2$. By symmetry of both relations, this means that $B_2 \sim_Y A_2$ and $A_2 \sim_X A_1$. Thus, $B_2 \sim_{YX} A_1$, and by commutativity, $B_2 \sim_{XY} A_1$. □

Theorem 4. If $\sim_X$, $\sim_Y$, and $\sim_Z$ are three relations, and $\sim_{XY}$ is the composition of $\sim_X$ with $\sim_Y$, and $\sim_{YZ}$ is the composition of $\sim_Y$ with $\sim_Z$, then the composition $\sim_{(XY)Z}$ of $\sim_{XY}$ with $\sim_Z$ is the same as the composition $\sim_{X(YZ)}$ of $\sim_X$ with $\sim_{YZ}$.

Proof. Based on the definitions, it is straightforward to see that $A \sim_{(XY)Z} D$ iff there is C with $C \sim_Z D$ and $A \sim_{XY} C$, which means that there is B with $A \sim_X B$ and $B \sim_Y C$. But by definition, this holds iff $B \sim_{YZ} D$, and so $A \sim_{X(YZ)} D$, so the two relations are the same. $\square$

Example. A strict ordering and two equivalence relations compatible with it, that all commute, but which are not mutually compatible:

Consider relations $\succ_X$, $\sim_Y$, and $\sim_Z$, defined as the transitive closure of the infinite continuation of the following diagram. One can check that $\succ_X$ is a strict ordering, $\sim_Y$ and $\sim_Z$ are equivalence relations, $\succ_X$ is compatible with each of the others individually, and any two of them commute, but they are not mutually compatible. (Every triangle of any size in the diagram provides a counterexample.)



Theorem 5. If G is any groupoid on the set of state spaces for all acts, then $\approx_G$ commutes with $\succ, \succeq, \approx, \succ'$.

Proof. I will show this for $\succ$, but the other three proofs are exactly parallel. If $A_1 \succ A_2$ and $A_2 \approx_G B_2$, then we must have $\mathcal{S}_{A_1} = \mathcal{S}_{A_2}$, so we can call this set $\mathcal{S}_A$. Additionally, there is some g in G such that for all s in $\mathcal{S}_A$, $A_2(s) = B_2(g(s))$.

Let $\mathcal{S}_B$ be the set of states on which B_2 is defined. Then g^{-1} is also in G, and so we can define, for s in $\mathcal{S}_B$, $B_1(s) = A_1(g^{-1}(s))$, and see that $A_1 \approx_G B_1$. Since $g^{-1}(g(s)) = s$, for all s in $\mathcal{S}_A$, we can see that $B_2(s) = A_2(g^{-1}(s))$. Since $A_1 \succ A_2$, we see that for all s in $\mathcal{S}_B$, $A_1(g^{-1}(s)) >_v A_2(g^{-1}(s))$. But this is just what it means to say that $B_1 \succ B_2$. Thus, by assuming the existence of A_2 with $A_1 \succ A_2 \approx_G B_2$, we have shown the existence of B_1 such that $A_1 \approx_G B_1 \succ A_2$.

The converse direction, and the versions replacing $\succ$ with $\succeq, \approx, \succ'$ all work exactly the same way. $\square$

Theorem 6. If the probability is regular, and G is measure-preserving, then $\approx_G$ is compatible with $\succ'$ (and therefore with $\succ$).

Proof. Assume that $\succ'$ is incompatible with $\approx_G$. This means there are acts A and B with $A \succ' B$ and $B \approx_G A$. $B \approx_G A$ means that there is a measure-preserving permutation g with $B(s) = A(g(s))$ for all s. $A \succ' B$ means that there is some s with $A(s) >_v A(g(s))$, and that for every s, $A(s) \geq_v A(g(s))$.

Let s_0 be some state with $A(s_0) >_v A(g(s_0))$, and let S_0 be the set of all s' such that $A(s') \geq_v A(s_0)$ and $A(g(s')) <_v A(s_0)$. Since A and g are measurable, S_0 is too.

By the definitions of S_0 and g, it is clear that S_0 is disjoint from $g[S_0], g[g[S_0]]$, etc., since $A(s) \geq_v A(s_0)$ for any s in S_0, and $A(s) <_v A(s_0)$ for any s in any of the latter sets. Since g is a correspondence, an equal number of applications of g to both sides will preserve disjointness. Thus, all of these sets are disjoint. Since g is measure-preserving, these are all measurable and equiprobable. But since there are infinitely many of them, they must be null. This contradicts the fact that the probability is regular, so $\succ'$ must be compatible with $\approx_G$. $\square$

Theorem 7. If $\mathcal{O}$ is separable, then there is a function R from $\mathcal{O}$ to the real numbers, such that whenever $o_1 <_v o_2$, it is the case that $R(o_1) < R(o_2)$.

Proof. Using the Axiom of Choice, we can extend $<_v$ to a linear ordering $<'_v$ on $\mathcal{O}$. The rationals contain an isomorphic copy of every countable linearly ordered set, so we can use the isomorphic copy of D to define R on the members of D, in such a way that whenever $d_1 <'_v d_2$, we have $R(d_1) < R(d_2)$, where both numbers are rational. (This also means that whenever $d_1 <_v d_2$, we have $R(d_1) < R(d_2)$, though the converse doesn't necessarily hold.)

Then for any $o \in \mathcal{O}$, we can consider the sets $D_{<o}$ and $D_{>o}$ which consist of all members of D that are (respectively) below or above o in $<'_v$. These sets are both non-empty, and by transitivity, every member of $R[D_{<o}]$ is less than every member of $R[D_{>o}]$. Thus, we can assign $R(o)$ to be any real number in this gap. Because D is dense, there can be no contradiction between these assignments, so they can all be made independently. The resulting function R has the desired property. □

Theorem 8. If the probability is countably partitive, and the outcome space is separable, and G is measure-preserving, then $\approx_G$ is compatible with $\succ$.

Proof. Let R be a function given by Theorem 7.

Now assume that there is some A and some g such that for all s, $A(s) >_v A(g(s))$. We can define a function f on states such that $f(s) = R(A(s)) - R(A(g(s)))$. This function is measurable. (The set of s with $f(s) > \epsilon$ is the union of the sets of s such that $A(s) >_v R^{-1}(x)$ and $A(g(s)) < R^{-1}(x - \epsilon)$ [which is always measurable, since A and g are measurable] where x ranges over all rationals. This is a countable union of measurable sets.)

The whole space is then the countable union of the sets with $f(s) > \epsilon$ where ϵ is any positive rational number. Thus, by countable partitivity, at least one of these sets is non-null. Let S be this set.

Now consider $S_k = \{s \in S | R(f(s)) < k\epsilon\}$, where k is any integer. Let $T_k = S_k - S_{k-1}$. Then $T_k, g[T_k], g[g[T_k]], \ldots$ are disjoint sets (T_k is disjoint from the following sets because they are all contained within S_{k-1} by definition of S, and each other set is disjoint from all sets following it because g is a permutation) that are all equally probable (because g is measure-preserving). Thus, T_k must be null. But S is non-null, and it is the union of the countably many T_k, which is a contradiction.

Therefore, the assumption that there were such an A and g must be false, so we see that $\succ$ and $\approx_G$ are compatible. □

Theorem 9. If $S_1 \equiv_P S_2$ and the comparative probability relation is homogeneous, then there is a measure-preserving correspondence g with $g[S_1] = S_2$.

Proof sketch:. The definition of homogeneity means that there is a measure-preserving correspondence g_1 with $g_1[S_1] \subseteq S_2$ and a measure-preserving correspondence g_2 with $g_2[S_2] \subseteq S_1$. By following the standard proof of the Schröder-Bernstein theorem in set theory, one can use these two measure-preserving correspondences to create g. □

Theorem 10. If S_1 and S_2 are in the same state space, and the comparative probability relation is homogeneous and additive, then there is a measure-preserving permutation h such that $h[S_1] = S_2$, $h[S_2] = S_1$, and h is the identity on all states outside of $S_1 \cup S_2$.

Proof. Let $S'_1 = S_1 \setminus S_2$ and $S'_2 = S_2 \setminus S_1$ be the sets of members of one set but not the other. By additivity, $S'_1 \equiv_P S'_2$. Let g be a measure-preserving permutation with $g[S_1] = S_2$ as given by Theorem 9. Then define $h(s) = g(s)$ if $s \in S'_1$, $h(s) = g^{-1}(s)$ if $s \in S'_2$, and $h(s) = s$ otherwise. This function satisfies the stated conditions. □

Theorem 11. Let P be a homogeneous probability function. Let A_1 and A_2 be any discrete acts such that for every o, $(A_1(s) = o) \equiv_P (A_2(s) = o)$. Then there is a measure-preserving correspondence g such that for all s, $A_1(s) = A_2(g(s))$.

Proof. Because A_1 and A_2 are discrete, there are only countably many o that are the outcome of either act on any state. Number these as $o_1, o_2, \ldots$. By homogeneity, there is a correspondence g_0 between $\mathcal{S}_{A_1}$ and $\mathcal{S}_{A_2}$. By induction, I will construct a sequence of measure-preserving correspondences g_n such that $g_n[A_1(s) = o_i] = (A_2(s) = o_i)$ for any $i \leq n$. For each state s, there is some i such that $A_1(s) = o_i$, and I will then define $g(s) = g_i(s)$. It will be straightforward to see that g is the desired correspondence.

Now assume that g_n has already been constructed. We can construct g_{n+1} by noting that $(g_n[A_1(s) = o_{n+1}]) \equiv_P (A_2(s) = o_{n+1})$. Thus, by

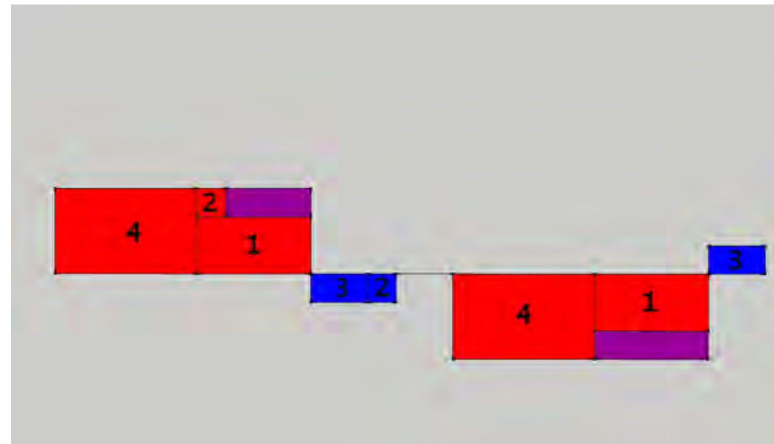
Theorem 10, there is a measure-preserving permutation of $\mathcal{S}_{A_2}$ that interchanges these two sets and leaves everything else fixed. Composing this permutation with g_n gives g_{n+1}. □

Theorem 12. Assume the outcomes are linearly ordered. Assume that probabilities are linearly ordered, additive, and homogeneous. Then $\approx_v$ is compatible with $\succ$. (This result needs only comparative probability, not numerical probability.)

Proof. I will prove a lemma showing that if $A_0 \sim_v A_1 \sim_v A_2 \ldots A_{n-1} \sim_v A_n$, then there exists a sequence that starts and ends with the same A_0 and A_n, which is "monotonic", in the sense that for each state s, either $A_0(s) \leq_v A_1(s) \leq_v \cdots \leq_v A_{n-1}(s) \leq_v A_n(s)$ or $A_0(s) \geq_v A_1(s) \geq_v \cdots \geq_v A_{n-1}(s) \geq_v A_n(s)$.

Given this lemma, the proof can then be completed as follows: assume there are some A_0 and A_n such that $A_0 \succ A_n$ and $A_0 \approx_v A_n$. Applying the lemma, we can find a sequence $A_0 \sim_v A_1 \sim_v \cdots \sim_v A_{n-1} \sim_v A_n$ that is monotonic in the sense just mentioned. Let S_i be the set of states s in which $A_i(s) >_v A_{i+1}(s)$. By monotonicity, and since $A_0 \succ A_n$, the union of the S_i must be the whole state space. By finite additivity, there must be some i such that S_i is strictly more probable than the empty set. But then the set of states for which $A_i(s) <_v A_{i+1}(s)$ must also be strictly more probable than the empty set. Thus, by monotonicity of the sequence, these states must be ones for which $A_0(s) <_v A_n(s)$, which contradicts the claim that $A_0 \succ A_n$. Thus, there can be no $A_0 \succ A_n$ with $A_0 \approx_v A_n$, as required.

Now it just remains to prove the lemma. I will prove it for the case when the sequence has just two steps, but the argument should generalize to any number of steps. The idea is to consider $T(A_0, A_1) = (S_{1+}, S_{1-}, d_1)$ and $T(A_1, A_2) = (S_{2+}, S_{2-}, d_2)$. The goal will be to replace these two trade-offs with four trade-offs such that the states in which each trade-off increases the values of outcomes is completely disjoint from the sets of states in which each trade-off decreases the values of outcomes. The following diagram may be helpful. The red and purple areas on the left represent S_{1+} and d_1, and the ones on the right represent S_{1-} and d_1; the blue and purple areas on the right represent S_{2+} and d_2, and the ones on the left represent S_{2-} and d_2.



Without loss of generality, we can assume that $d_1 >_d d_2 >_d 0_d$ (d_1 is the height of the red regions, and d_2 is the height of the blue regions), and that $S_{1+} \cap S_{2-} >_P S_{1-} \cap S_{2+}$ (these are respectively the widths of the right and left purple regions). Then we can define the four trade-offs shown in the diagram, so that no state has its outcome increased by one trade-off and decreased by another.

The set-theoretic definitions (which guarantee that these are in fact trade-offs of the relevant sort that add up to the same result as the initial two trade-offs) are as follows: the first trade-off (indicated by the two regions with "1" on the diagram) is $(S^*, S_{2+} \cap S_{1-}, d_1 - d_2)$, where $S_{1+} \cap S_{2-} \subset S^* \subset S_{1+}$ and $S^* \equiv_P S_{2+} \cap S_{1-}$. The second trade-off (indicated by the two regions with "2" on the diagram) is $(S^* \setminus S_{2-}, S_*, d_2)$, where $S_* \subset S_{2-} \setminus S_{1+}$ and $S_* \equiv_P S^* \setminus S_{2-}$. The third trade-off (indicated by the two regions with "3" on the diagram) is $(S_{2+} \setminus S_{1-}, S_{2-} \setminus (S_{1+} \cup S_*), d_2)$. The fourth tradeoff (indicated by the two regions with "4" on the diagram) is $(S_{1+} \setminus S^*, S_{1-} \setminus S_{2+}, d_1)$.

This proves the lemma for the case of two consecutive trade-offs, and the construction should generalize to any finite number of consecutive trade-offs, which completes the proof of the theorem. □

Theorem 13. If outcomes and differences are real numbers, and probabilities are numerical and countably additive, and G is any measure-preserving groupoid, then $\approx_v$ is compatible with $\succ_G$.

Proof. As before, for any act A, we can define $CDF_A(x) = P(A(s) \leq x)$ for any real-number outcome x. Now we can define $CDF^-(A,B) = \int (CDF_B(x) - CDF_A(x))dx$, where the integral is taken over all real numbers. By Theorem 11, we can see that if $A \succ_G B$, then A stochastically dominates B, so that for all x, $CDF_A(x) < CDF_B(x)$. Then $CDF^-(A,B)$ is positive.

To complete the proof, I will show that if $A \approx_v B$, then $CDF^-(A,B) = 0$. To prove this, it will suffice to prove that if $A \sim_v B$, then $CDF^-(A,B) = 0$, and that $CDF^-(A,B) + CDF^-(B,C) = CDF^-(A,C)$. The second part holds because of linearity of subtraction, addition, and integration. So it just remains to show that if $A \sim_v B$, then $CDF^-(A,B) = 0$.

To do this, I will show that if $T(A,B) = (S,\emptyset,k)$ (meaning that $A(s) - B(s) = k$ whenever $s \in S$ and $A(s) = B(s)$ otherwise), then $CDF^-(A,B) = kP(S)$. Then the fact that $A \sim_v B$ iff there is C with $T(A,C) = (S_1,\emptyset,k)$ and $T(C,B) = (S_2,\emptyset,-k)$ with $P(S_1) = P(S_2)$ completes the proof.

So assume that $T(A,B) = (S,\emptyset,k)$. Then $CDF_B(x) - CDF_A(x) = P(S \cap (x-k < A(s) \leq x))$, because the states where $B(s) > x$ are exactly the states in S with $A(s) > x-k$. Integrating this over all x, we should get $kP(S)$. (This is because, for any $s \in S$, it will show up in $CDF^-(A,B)$ for all x between $A(s)$ and $B(s)$.) □

Theorem 14. If probabilities, values of outcomes, and differences of outcomes are all real numbers that satisfy the standard relations, and if probability is divisible, then if A and B are acts on the same space with finite upper and lower bounds on the utility of their outcomes, and if the expected utility of A is strictly greater than the expected utility of B, then $A \succ_v B$.

Proof. Say that an act is *simple* iff it has only finitely many distinct outcomes. It is a standard result that for any act whose outcomes have a finite upper and lower bound, there are simple acts dominating it whose expected utilities are arbitrarily close to its expected utility, and there are simple acts it dominates whose expected utilities are arbitrarily close to its expected utility. These simple acts can also be arranged to have outcomes with rational utility and to have rational probability of achieving each outcome (because of divisibility).

Thus, we can find simple acts A' and B' with $A \succ A'$ and $B' \succ B$, with A' and B' both simple and rational, and with the expected utility of A' greater than that of B'. But any simple and rational act is trade-off equivalent to the act whose outcome in every state is equal to the expected utility of that act. (This can be seen by breaking the space of states into subsets whose probabilities are all equal to the greatest common divisor of the probabilities of the outcomes of the simple and rational act, and shifting utility around in units equal to the greatest common divisor of the utilities of the outcomes of the act and the expected utility of the act.)

Thus, A' is trade-off equivalent to an act that dominates one that is trade-off equivalent to B', so A and B are related by a sequence of dominance and trade-off-equivalence relations, and so, by commutativity of these relations, we must have $A \succ_v B$. □

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axioms de l'école Americaine. *Econometrica*, 21(4):503–546.

Buchak, L. (2013). *Risk and Rationality*. Oxford University Press.

Christensen, D. (1996). Dutch-book arguments depragmatized: Epistemic consistency for partial believers. *The Journal of Philosophy*, 93(9):450–479.

Colyvan, M. (2008). Relative expectation theory. *The Journal of Philosophy*, 105(1):37–44.

Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7(1):1–68.

Dreier, J. (1996). Rational preference: Decision theory as a theory of practical rationality. *Theory and Decision*, 40(3):249–276.

Easwaran, K. (2008). Strong and weak expectations. *Mind*, 117(467):633–641.

Easwaran, K. (2013). Expected accuracy supports conditionalization — and conglomerability and reflection. *Philosophy of Science*, 80(1):119–142.

Easwaran, K. (2014). Regularity and hyperreal credences. *The Philosophical Review*, 123(1):1-41.

Fine, T. L. (1973). *Theories of Probability*. Academic Press Inc.

Fine, T. L. (2008). Evaluating the Pasadena, Altadena, and St. Petersburg gambles. *Mind*, 117(467):613–632.

Greaves, H. and Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115(459):607–632.

Halpern, J. Y. (2003). *Reasoning about Uncertainty*. MIT Press.

Hawthorne, J. (2015). A logic of comparative support: Qualitative conditional probability relations representable by Popper functions. In Hájek, A. and Hitchcock, C., editors, *Oxford Handbook of Probabilities and Philosophy*. Oxford University Press.

James, W. (1896). The will to believe. *The New World*, 5:327–347.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jeffrey, R. C. (1965). *The Logic of Decision*. McGraw-Hill.

Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4):575–603.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971). *Foundations of Measurement, Volume 1: Additive and Polynomial Representations*. Academic Press.

Leitgeb, H. and Pettigrew, R. (2010). An objective justification of Bayesianism I: Measuring inaccuracy. *Philosophy of Science*, 77:201–235.

McGee, V. (1999). An airtight Dutch book. *Analysis*, 59(264):257–265.

Meacham, C. J. G. and Weisberg, J. (2011). Representation theorems and the foundations of decision theory. *Australasian Journal of Philosophy*, 89(4):641–663.

Nover, H. and Hájek, A. (2004). Vexing expectations. *Mind*, 113(450):237-249.

Peterson, M. (2002). An argument for the principle of maximizing expected utility. *Theoria*, 68(2):112–128.

Peterson, M. (2004). From outcomes to acts: a non-standard axiomatization of the expected utility principle. *Journal of Philosophical Logic*, 33(4):361–378.

Ramsey, F. P. (1926). Truth and probability. In Braithwaite, R. B., editor, *The Foundations of Mathematics and other Logical Essays* (1931). Harcourt, Brace and Company.

Savage, L. J. (1954). *The Foundations of Statistics*. Dover.

Schick, F. (2003). *Ambiguity and Logic*. Cambridge University Press.

Shah, N. (2003). How truth governs belief. *The Philosophical Review*, 112(4):447–482.

Stefánsson, H. O. and Bradley, R. (2015). Counterfactual desirability. *British Journal for the Philosophy of Science*.

van Fraassen, B. C. (1983). Calibration: A frequency justification for personal probability. In Cohen, R. S. and Laudan, L., editors, *Physics, Philosophy, and Psychoanalysis*, pages 295–319. D. Reidel.

Wedgwood, R. (2002). The aim of belief. *Noûs*, 36(s16):267–297.